Effects of dual grammatical gender systems on conceptual representations in simultaneous bilinguals: Evidence from Ukrainian-Russian speakers across behavioural and neural measures

Oleksandra Osypenko

Supervisors: Professor Silke Brandt and Dr Aina Casaponsa

Department of Linguistics and English Language
Lancaster University

Thesis submitted for the degree of
Doctor of Philosophy

May 2025

Oleksandra Osypenko

Effects of dual grammatical gender systems on conceptual representations in simultaneous bilinguals: Evidence from Ukrainian-Russian speakers across behavioural and neural measures

Abstract

This PhD thesis investigates how grammatical gender influences cognitive processes in simultaneous bilinguals of Ukrainian and Russian. Specifically, it examines whether simultaneous bilinguals of two three-gendered languages (neuter, feminine, and masculine genders) exhibit grammatical gender effects from their first languages (L1s) on conceptual representations when tested in genderless English.

Four experiments were conducted using behavioural and/or electrophysiological (EEG) measures across three empirical studies: participants engaged in similarity ratings (Chapter 3), memory recall (Chapter 4), and non-verbal categorisation (Chapter 5) tasks. Each experiment included stimuli with both matching and mismatching genders across two L1s, allowing assessment of overall gender effects when genders overlap, and the influence of participants' more proficient L1 when genders mismatch.

Findings revealed that grammatical gender modulated cognition in simultaneous bilinguals, but only under specific conditions. Specifically, when nouns with neuter gender were excluded from the stimuli (Chapter 3), or when the task had moderate to high gender salience (Chapters 3 and

4), participants showed stronger effects of grammatical gender on their responses. These effects were particularly pronounced when grammatical gender was matching across both languages (emerged in both similarity rating and memory recall tasks), compared to the nouns matching in participants' more proficient L1 (only in similarity ratings). However, in a low-salience EEG task (Chapter 5), no significant effects of grammatical gender were observed, suggesting limited automaticity of grammatical gender effects at early perceptual stages.

The thesis offers novel theoretical insights into the structural-feedback hypothesis by demonstrating the context-sensitive nature of the observed effects. It also reopens a previously underexplored question: whether grammatical gender effects extend to speakers of three-gendered languages, and not just those with two-gendered systems (cf. Sera et al., 2002). Overall, this work represents the most comprehensive investigation to date of grammatical gender effects in simultaneous bilinguals of two gendered languages.

Declaration

The contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other university. This thesis does not exceed the permitted maximum of 80,000 words including footnotes and appendices. All Lancaster University ethical requirements were strictly upheld throughout this project.

Oleksandra Osypenko

Authorship statement

This thesis contains published work conducted in collaboration with my supervisors Professor Silke Brandt and Dr Aina Casaponsa. The thesis is being submitted under the Alternative Format framework to enable me to build a publication portfolio prior to graduation. Each of the empirical studies included in this thesis has been either accepted for publication, is currently under review, or will be submitted before the viva to peer-reviewed journals (see *Published papers*). I conceived, conducted, and wrote up each of the studies presented. Professor Silke Brandt and Dr Aina Casaponsa contributed to the design and analysis of the studies presented in chapters three, four, and five, and provided comments on the text for all chapters.

Oleksandra Osypenko

Acknowledgements

During the years I have spent doing my PhD, I have had the privilege of meeting many people who, directly or indirectly, have supported, influenced, and inspired me. Without them, this journey would not have been the same. While I may not be able to name everyone here, please know that I am deeply grateful to all who have been part of this experience.

First and foremost, I want to express my deepest gratitude to my supervisors, **Silke Brandt and Aina Casaponsa**, for their unwavering support, guidance, and patience. Their feedback and expertise have been invaluable, and I truly appreciate the opportunities they provided for me to learn, grow, and refine my skills. I am especially thankful for their patience as I worked to improve my academic writing and tackled the challenges of learning EEG and MATLAB (and basically everything else there is to learn about academic work). Their encouragement and willingness to support me in this learning process made all the difference, and I could not have completed this PhD without them.

I would also like to express my sincere gratitude to **Panos Athanasopoulos**, who took a chance on an inexperienced but very enthusiastic candidate at the interview. Panos' insightful feedback and guidance throughout my PhD shaped this experience into something far richer and more rewarding than I could have imagined. I am especially grateful for all our productive collaboration, the countless rounds of revisions, and the invaluable time spent refining ideas - particularly during my visiting stay in Lund, which made this journey all the more special.

Of course, my deepest love and gratitude go to my **family**. I know it was not always easy to stand by me during this journey, which was both exciting and exhausting in equal measure. Your unwavering support, encouragement, and patience helped me navigate the challenges that came my way. I am especially grateful to my **mom, Denys, and Valeriia** for being by my side in Lancaster, particularly in the final year of my PhD. Just having you there meant more to me than I can express. Seeing how hard you worked to adjust and succeed in the UK was a constant source of inspiration, pushing me to work harder and keep going. I am also deeply grateful for the opportunity to stay in the UK and pursue this project - something that would not have been possible without your support and belief in me.

Published and submitted for review papers

Chapter Three

Osypenko, O., Brandt, S., & Athanasopoulos, P. (2025). The influence of three-gendered grammatical systems on simultaneous bilingual cognition: The case of Ukrainian-Russian bilinguals. *Language and Cognition*, *17*, e25. doi:10.1017/langcog.2024.73

Chapter Four

Osypenko, O., Brandt, S., & Athanasopoulos, P. (revised and resubmitted). Between Two Grammatical Gender Systems: Exploring the Impact of Grammatical Gender on Memory Recall in Ukrainian-Russian Simultaneous Bilinguals. *Cognitive Science*.

Chapter Five

Osypenko, O., Brandt, S., & Casaponsa, A. (under review). Early Perceptual Effects (or lack of thereof) of Conflicting Grammatical Genders: ERP Evidence from Simultaneous Bilinguals. *Language, Cognition and Neuroscience*.

Contents

List of Figures xiv
List of Tables xviii
Chapter 1. General Introduction 1
Chapter 2. Literature review 12
2.1. Historical overview of linguistic relativity research 12
2.1.1. Formulation of linguistic relativity 16
2.1.2. Decline of Whorfianism and nativist criticism of linguistic relativity 18
2.1.3. Revival of linguistic relativity (1980s-1990s) and emergence of Neo-Whorfian research 20
2.2. The contemporary period and the latest trends in Neo-Whorfian research 22
2.2.1. Bilingualism and impact of cognitive restructuring and individual differences on linguistic relativity effects 24
2.2.2. Combining behavioural and neural measures to examine Whorfian effects 31
2.2.3. Mechanisms underpinning language effects 34
2.3. Grammatical gender in linguistic relativity research 41
2.3.1. Why grammatical gender is an effective testbed for linguistic relativity? 43
2.3.2. Early studies on grammatical gender effects in monolingual speakers 45
2.3.3 The advantages of incorporating bilingualism in grammatical gender research in linguistic relativity and the role of L1-L2 pairing 50
2.3.4. Key considerations for bilingual research on grammatical gender 58

2.4. Study design issues in grammatical gender research 64
2.4.1. Salience of gender/sex and language in the task and the transparency of grammatical gender manipulation 65
2.4.2. Language of the testing.. 70
2.4.3. Alternative explanations for the found grammatical gender effects 72
2.5. The current study 76
2.5.1. The relevance of studying simultaneous bilinguals...................... 76
2.5.2. Focus on Ukrainian-Russian simultaneous bilinguals................. 78
2.5.3. Research questions and selected stimuli...................................... 79
2.5.4. Justification for the selected tasks 80
Chapter 3. The Influence of Three-Gendered Grammatical Systems on Simultaneous Bilingual Cognition: The case of Ukrainian-Russian Bilinguals 85
3.1. Abstract 85
3.2. Introduction.. 86
3.2.1. Grammatical gender in language and mind................................. 89
3.2.2. Online vs offline nature of the grammatical gender effects 93
3.2.3. Case of Ukrainian simultaneous bilingualism and typological differences in Ukrainian and Russian languages.................................... 95
3.3. Aims and the scope of the current study 98
3.4. Method 101
3.4.1. Experiment 1 101
3.4.1.1. Participants 101
3.4.1.2. Materials 104
3.4.1.3. Procedure and design... 105
3.4.1.4. Analysis 107
3.4.1.5. Results 108

3.4.2. Experiment 2 113
3.4.2.1. Participants 114
3.4.2.2. Materials 115
3.4.2.3. Procedure and design 116
3.4.2.4. Results 118
3.5. Discussion 122
3.6. References 129
Chapter 4. Between Two Grammatical Gender Systems: Exploring the Impact of Grammatical Gender on Memory Recall in Ukrainian-Russian Simultaneous Bilinguals 138
4.1. Abstract 138
4.2. Introduction 139
4.2.1. Effects of Grammatical Gender in Linguistic Relativity Research in Bilinguals 141
4.2.2. Grammatical Gender Effects in Two- vs. Three-Gendered Languages 145
4.2.3. Memory recall effects in LR research 149
4.2.4. The current study 155
4.3. Methods 157
4.3.1. Pre-test 157
4.3.2. Stimuli 158
4.3.3. Participants 162
4.3.4. Procedure 164
4.3.5. Analysis 166
4.4. Results 168
4.4.1. Comparison of English Monolingual and Ukrainian-Russian Bilingual Participants 168

4.4.2. Comparison of Ukrainian-Russian Bilingual Participants based on Language Proficiency 170
4.5. Discussion and conclusion 170
4.6. References 178
Chapter 5. Early Perceptual Effects (or lack thereof) of Conflicting Grammatical Genders: ERP Evidence from Simultaneous Bilinguals 185
5.1. Abstract 186
5.2. Introduction 186
5.2.1. Grammatical gender effects on perception: behavioural evidence across bilinguals 187
5.2.2. Limitations of behavioural findings 191
5.2.3. Grammatical gender effects on perception: ERP evidence and ERP components related to grammatical gender processing 192
5.2.4. Aims and scope of the current study 197
5.3. Methods 200
5.3.1. Participants 200
5.3.2. Materials 203
5.3.2.1. Pre-test 203
5.3.2.2. Main testing 203
5.3.3. Procedure 206
5.3.3.1. Electroencephalography recording 208
5.3.4. Data analysis 210
5.3.4.1. Behavioural analysis 210
5.3.4.2. EEG analysis 213
5.4. Results 215
5.4.1 Analysis for Research Question 1 (stimuli with matching grammatical gender; conditions CR-CU and IR-IU) 215
5.4.1.1. Behavioural analysis 215

5.4.1.2. ERP analysis 219
5.4.2. Results for Research Question 2 (stimuli with matching grammatical gender; conditions CR-IU and IR-CU) 222
5.4.1.1. Behavioural analysis 223
5.4.1.2. ERP analysis 227
5.5. Discussion and conclusion 229
5.6. References 237
Chapter 6. General Discussion 245
6.1. Summary of major contributions, by chapter 245
6.2. Directions for future research in linguistic relativity 253
6.2.1. Future directions of grammatical gender research and taking it outside of the laboratory setting 253
6.2.2. Future directions of bilingual research in linguistic relativity 256
Chapter 7. Consolidated Bibliography 260
Appendix A. Supplementary Materials for Chapter 3 291
A.1. The role of participants' gender on the similarity ratings 291
A.2. Reanalysis of Experiment 2 results using stimuli from Experiment 1 (excluding neuter-gender stimuli) 300
A.3. Distribution of Likert scores in Experiments 1 and 2 303
A.3.1. Comparing Ukrainian-Russian bilinguals and English monolinguals (using stimuli with matching grammatical gender in both L1s) 303
A.3.2. Results of Ukrainian-Russian bilinguals (using stimuli with mismatching grammatical gender in both L1s) 306
Appendix B. Supplementary Materials for Chapter 4 308
Appendix C. Supplementary Materials for Chapter 5 311
C.1. Examining the role of cognates on the findings 311

C.2. Analysis of the variable "Accuracy" as reported in the pre-registration 315
C.2.1. Response Accuracy for Research Question 1 (CR-CU vs IR-IU) 316
C.2.2. Response Accuracy for Research Question 2 (CR-IU vs IR-CU) 318

List of Figures

Figure 2.1: Distribution of gender systems in languages classified by number of grammatical genders (adapted from Corbett, 2013) 42

Figure 3.1: Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent, incongruent, and neuter stimuli pairs in Experiment 1 109

Figure 3.2: Mean Likert Scale Responses from Experiment 1 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent in Russian, (b) congruent in Russian and incongruent in Ukrainian 112

Figure 3.3: Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent and incongruent stimuli pairs in Experiment 2. 119

Figure 3.4: Mean Likert Scale Responses from Experiment 2 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent in Russian, (b) congruent in Russian and incongruent in Ukrainian 122

Figure 4.1: Example of the stimuli used in the learning phase................. 165

Figure 4.2: Example of the stimuli used in the testing phase 166

Figure 4.3: Accuracy of responses (%) based on condition (Congruent in both L1s and Incongruent in both L1s) by group (English monolinguals and Ukrainian-Russian bilinguals). Error bars indicate 95% confidence intervals. 169

Figure 5.1: Sequential stimulus presentation in each trial..... 209

Figure 5.2: Density plots for (a) L1 Proficiency and (b) L1 Dominance coefficients..... 212

Figure 5.3: Distribution of reaction times by Condition (CR-CU vs IR-IU) 216

Figure 5.4: Proportion and Count of Yes and No responses by Condition (CR-CU vs IR-IU) 217

Figure 5.5: Effects of (a) English Proficiency and (b) English Dominance on Response Types (Yes vs No) across Conditions (CR-CU vs IR-IU) .. 218

Figure 5.6: Grand average ERP waveforms of the early discrimination effects (35-75 ms), N1, P2/VPP, and N300 components (circled with a grey line) showing the effects of grammatical gender congruency for Ukrainian-Russian bilinguals across conditions 220

Figure 5.7: Distribution of reaction times by Condition (CR-IU vs IR-CU) 224

Figure 5.8: Proportion and Count of Yes and No responses by Condition (CR-IU vs IR-CU) 225

Figure 5.9: Response types (Yes vs No) as a function of (a) L1 proficiency coefficient and (b) L1 dominance coefficient across CR-IU and IR-CU conditions.. 226

Figure A.1: Distribution of Likert scores in Experiments 1 and 2, by participant group and using stimuli with matching grammatical gender in 2L1s 303

Figure A.2: Distribution of Likert scores of Ukrainian-Russian bilinguals in Experiments 1 and 2, using stimuli with mismatching grammatical gender in 2L1s 307

Figure B.1: Reaction Times by Condition (Congruent in both L1s vs Incongruent in both L1s) and Participant Group (English Monolinguals vs. Ukrainian-Russian Bilinguals) 309

Figure B.2: Reaction Times by Condition (Congruent in Ukrainian & Incongruent in Russian vs Congruent in Russian & Incongruent in Ukrainian) and L1 Proficiency (-100; +100)... 310

Figure C.1: Grand average ERP waveforms of the N1, P2/VPP, and N400 components (circled with a grey line) showing the processing of primes by Ukrainian-Russian bilinguals 314

Figure C.2: Accuracy of responses per condition (CR-CU vs IR-IU)...... 317

Figure C.3: Effects of (a) English Proficiency and (b) English Dominance on Accuracy Across Conditions (CR-CU vs IR-IU) 318

Figure C.4: Interaction of (a) L1 Dominance and (b) L1 Proficiency Coefficients with Accuracy by Condition (CR-IU vs IR-CU) 319

Figure C.5: Interaction of (a) English Proficiency-L1 Proficiency-Accuracy and (b) English Dominance - L1 Dominance - Accuracy by Condition (CR-IU vs IR-CU) 320

List of Tables

Table 2.1: Mechanisms of language influence on thought, based on Wolff and Holmes (2011) 36

Table 2.2: Methods of assessing L1-L2 language proficiency and dominance across key studies on grammatical gender 60

Table 3.1: Proficiency Scores and Distribution of Ukrainian-Russian Bilingual Participants in Experiment 1 103

Table 3.2: Example of stimuli used for both Experiment 1 and 2 105

Table 3.3: Proficiency Scores of Ukrainian-Russian bilinguals in Experiment 2 115

Table 4.1: Examples of the selected stimuli types 160

Table 4.2: Names used in the study by Schmidt and Boroditsky (2000) . 161

Table 4.3: Proficiency scores and distribution of Ukrainian-Russian bilingual participants 163

Table 5.1: Participants' language profiles 202

Table 5.2: Examples of stimuli based on grammatical gender congruency and incongruency in Ukrainian and Russian 206

Table 5.3: L1 Proficiency and L1 Dominance coefficients across participants 211

Table A.1: Number of participants in Experiments 1 and 2, based on their gender 291

Table A.2: Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 1, comparing Ukrainian-Russian bilinguals and English Monolinguals, including participant's gender as a predictor 292

Table A.3: Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 1, comparing Ukrainian-Russian bilingual group based on Language Proficiency, including participant's gender as a predictor 295

Table A.4: Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 2, comparing Ukrainian-Russian bilinguals and English Monolinguals, including participant's gender as a predictor 297

Table A.5: Results of the the cumulative link mixed model (CLMM) for similarity ratings from Experiment 2, comparing Ukrainian-Russian bilingual group based on Language Proficiency, including participant's gender as a predictor 299

Table A.6: Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 2, comparing Ukrainian-Russian bilinguals and English monolinguals, using pairs of stimuli from Experiment 1 (without neuter gender) 299

Table A.7: Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 2, comparing Ukrainian-Russian bilinguals based on Language Proficiency, using pairs of stimuli from Experiment 1 300

Table A.8: Descriptive statistics for Likert scores across Experiments 1 and 2 by participant group 305

Table A.9: Descriptive statistics for Likert scores across Experiments 1 and 2 for Ukrainian-Russian bilingual group 307

Table C.1: Distribution of primes based on spelling and grammatical gender matching (Match vs Mismatch) 312

Chapter 1. General Introduction

The concept of linguistic relativity, often associated with the works of Benjamin Lee Whorf and his mentor Edward Sapir, proposes that the grammatical and lexical properties of a language influence its speakers' cognition and perception of the world. This idea, traced back to ancient debates from Homer's era, gained formal recognition through the Sapir-Whorf hypothesis, suggesting that the lexical inventories and grammatical structures inherent in a language can shape habitual thought patterns (Athanasopoulos et al., 2016; Lucy, 1997; Whorf, 1956). For much of the 20th century, this notion was seen as overly deterministic, particularly after critiques by scholars like Steven Pinker in the 1990s, who dismissed it as implausible or trivial (see Thierry et al., 2024). However, the hypothesis has seen a resurgence in recent decades, with the "neo-Whorfian" movement bringing new empirical evidence to the fore, resulting in what Athanasopoulos and Bylund (2020, p. 1) describe as "Whorf is back. Big time." The revival of interest has been driven by advances in cognitive neuroscience and experimental psychology, which have allowed researchers to investigate the more subtle, unconscious effects of grammar and lexical inventories available in one's language(s) on thought and perception (Athanasopoulos & Bylund, 2020).

One area where this renewed interest is evident is in the study of grammatical gender. This grammatical feature is present in 40% of the world's languages (Corbett, 2013) and assigns nouns to categories such as masculine, feminine, or neuter, often in ways that are arbitrary and unrelated to natural gender distinctions (Sato & Athanasopoulos, 2018).

The key question in this line of research is whether these grammatical gender categories "rub off" on the way speakers perceive and conceptualise objects (Samuel et al., 2019, p. 1767). Specifically, do speakers of gendered languages attribute masculine or feminine characteristics to inanimate objects based on their grammatical gender? The present PhD research addresses this question in the context of simultaneous bilinguals of two partially contrasting three-gendered languages. More precisely, it investigates *whether such bilinguals exhibit grammatical gender effects from their first language(s), even when tested in a genderless second language (L2).*

To explore the extent of grammatical gender's influence on cognition, this thesis draws on a variety of experimental tasks - ranging from categorisation-based studies on perceived gender associations (Chapters 3 and 5) to memory recall studies (Chapter 4). These studies incorporate both linguistic and non-linguistic paradigms to capture how grammatical gender may shape thought beyond language itself. Collectively, all experiments in the current thesis were designed to address the following open issues, which together converge on the central question outlined above.

First, this research is focused on speakers of two languages with three-gendered grammatical systems. It is motivated by the understudied distinction between *two-gendered and three-gendered grammatical gender systems*, a topic explored in detail in throughout this thesis. In Chapters 2, 3 and 4, I engage with the ongoing debate surrounding the contradictory findings providing evidence both in favour (Haertlé, 2017; Lambelet, 2016; Maciuszek et al., 2019; Phillips & Boroditsky, 2003) and against (Kousta et al., 2008; Pavlidou & Alvanoudi, 2019; Sera et al., 2002; Vigliocco et al., 2005) grammatical gender effects in speakers of three-gendered languages, particularly when comparing with speakers of two-gendered languages (e.g., Spanish and Italian).

To advance this discussion, I propose examining speakers of Ukrainian and Russian, two three-gendered languages that have received little empirical attention. Ukrainian and Russian exhibit a typologically distinct system of gender marking: rather than using articles, as in Italian, Spanish, or German, these languages encode gender morphologically through inflections on nouns, adjectives, and, in some cases, verbs. This contrasts with the article gender marking, where grammatical gender does not always align with the biological sex of the referent. For example, the typically used example from German, the word "*das Mädchen*" ("girl") uses the neuter article "das", despite referring to a female person. Similarly, in Italian, *"la guida"* ("the guide") always takes the feminine article "la", regardless of whether the guide is male or female. Therefore, the absence of this grammatical feature offers a unique framework for studying grammatical gender effects without the confounding influence of article-related cues.

Extending the discussion on how effects of grammatical gender may differ between two- and three-gendered systems, this thesis contributes to the field by addressing whether the internal structure of three-gendered systems – particularly the role of the neuter category - attenuates gender salience and reduces gender effects (Sera et al., 2002; Vigliocco et al., 2005).

Second, this PhD thesis examines participants who are not only speakers of three-gendered languages but also *simultaneous bilinguals* whose linguistic profiles include two partially contrasting three-gendered grammatical systems. Most studies examining the effects of grammatical gender rely on sequential bilinguals, who are either compared with another group of sequential bilinguals (whose first language has a different grammatical gender system) or with monolingual controls from a genderless language (typically English). However, as detailed throughout Chapters 2-4, this line of research could greatly benefit from the inclusion of simultaneous bilinguals due to their unique linguistic profile with important theoretical implications. Specifically, unlike sequential bilinguals, who typically acquire their L2 through formal instruction later in life, simultaneous bilinguals acquire two grammatical gender systems naturally and implicitly during early development. This distinction provides an opportunity to explore how two deeply ingrained grammatical systems interact and influence cognition, free from the explicit learning biases often associated with sequential bilinguals. Furthermore, to enable a comprehensive comparison, our empirical studies (Chapters 3 and 4) also included a traditional control group of English monolinguals, representing a genderless language background.

Third, an additional aim of this PhD research, arising from the unique linguistic profile of our participant group, was to explore how stimuli with matching and mismatching grammatical gender in participants' two L1s might influence speakers' conceptual representations. Specifically, this thesis investigates whether grammatical gender effects are strengthened when gender categories match across both languages, or weakened when they mismatch by introducing competing cues that may increase cognitive load. This conflict provides an opportunity to explore how bilinguals resolve competing linguistic cues and whether the more proficient or dominant L1 exerts greater influence on cognition in such cases. This investigation is central to understanding how two grammatical systems interact within a bilingual mind. Moreover, given the scarcity of research on the cognitive consequences of mismatching grammatical gender in two L1s (see Bassetti, 2007 and discussion in Chapter 2), a novel analytical approach was developed to assess how language-specific grammatical structures influence perception and memory recall in adult bilingual speakers.

To achieve this, we went beyond the self-reported measures commonly employed in earlier studies (e.g., in Phillips & Boroditsky, 2003; Boroditsky & Schmidt, 2000 adapted for Chapters 3 and 4 respectivelly) and incorporated formal assessments of language proficiency (Cambridge University Press, 2024; Oxford University Press, 2001; Ukrainian Center for Educational Quality Assessment, 2020) across all studies. Crucially, when analysing responses of Ukrainian-Russian bilingual participants, we treated simultaneous bilingualism as a continuum, comparing to earlier study (i.e., Bassetti, 2007) who approached it as a binary variable. For that, language proficiency was calculated as a coefficient. This approach allowed for a more detailed examination of the effects of grammatical gender, particularly when investigating the effects of mismatching grammatical gender between Ukrainian and Russian.

Fourth, this study aims to address whether the effects of grammatical gender are *consciously engaged or occur unconsciously during cognitive processing*. While much of the existing research has relied on behavioural tasks in which participants may consciously or strategically engage with gender information, it remains difficult to entirely rule out the influence of conscious processing. To address this limitation, our final study was designed to investigate whether grammatical gender effects also manifest at an unconscious level. To this end, we employed event-related potentials (ERPs), focusing on both early pre-linguistic (i.e., N1, P2/VPP) and late (i.e., N300) components, as well as an even earlier gender discrimination effect that occurs as early as at 45-85 ms after the trigger onset. The use of electroencephalography (EEG) when looking at the grammatical gender effects is particularly critical because it allows researchers to move beyond subjective or overtly strategic responses and tap into implicit, pre-attentive processes that are less susceptible to conscious modulation. Unlike behavioural measures, which often reflect higher-order cognitive strategies, EEG provides a window into the neural correlates of perception and cognition, offering insights into the automatic and unconscious engagement of grammatical gender (Athanasopoulos & Casaponsa, 2020; Thierry et al., 2024).

Additionally, in all of our studies (using both behavioural measures and ERPs), we incorporated a thorough debriefing session to verify whether participants were aware of and strategically used grammatical gender information during the experiment. This approach allows us to minimise this central concern in grammatical gender research.

Furthermore, the current PhD study builds upon the *structural-feedback hypothesis* (Sato & Athanasopoulos, 2018), an extension of the label-feedback hypothesis (Lupyan, 2012), which posits that grammatical gender influences perception not merely through explicit linguistic labels but through deeper, habitual patterns embedded within the grammatical system itself. Unlike label-feedback hypothesis (Lupyan, 2012), which focuses on real-time modulations by internal or external labels, structural-feedback hypothesis highlights how grammatical categories unconsciously shape perception by reinforcing specific distinctions, such as those between masculine and feminine genders, in the moment of task completion (Sato & Athanasopoulos, 2018). No prior study to our knowledge has addressed this question using the same task while introducing controlled variations, such as the inclusion or exclusion of the neuter gender. By adopting this manipulation, we aim to investigate whether the effects of grammatical gender are consistent across two experiments or are context-dependent (see Chapter 3). If similar effects are observed across the two experiments - regardless of whether the neuter gender is included - this would support the offline nature of grammatical gender effects. Such a finding would suggest that these effects are not modulated by the immediate linguistic context but are instead driven by long-term, habitual patterns ingrained through extensive exposure to the grammatical system. Conversely, if the effects differ between the two experiments, this would lend support to the hypothesis of online grammatical gender effects. Specifically, it would indicate that these effects are sensitive to immediate task conditions and context (see Lupyan et al., 2020 and Section 2.2.3 in Chapter 2 of this thesis for the in-depth discussion of the online vs offline effects). By designing two analogous experiments in Chapter 3 to systematically manipulate the presence of the neuter gender, this study seeks to disentangle the online and offline influences of grammatical gender, contributing to our understanding of how linguistic structures interact with cognitive processing.

Finally, an important objective of this thesis is to contribute to the replication of two seminal studies in linguistic relativity that have been highly influential and widely cited, serving as a catalyst for subsequent research in this field: the similarity judgment task (Phillips & Boroditsky, 2003; see Chapter 3, Experiment 1) and the memory recall task (Boroditsky & Schmidt, 2000; see Chapter 4). These foundational studies played a critical role in advancing the investigation of linguistic relativity and grammatical gender effects. However, attempts to replicate these findings have often been unsuccessful (e.g., Elpers et al., 2022; Pavlidou & Alvanoudi, 2013, 2019). Chapter 2 provides further discussion of this account. Moreover, we argue that the difficulty in replicating these results may not necessarily reflect the absence of grammatical gender effects on perception and memory, but could instead be attributed to methodological limitations, including the lack of openly shared materials and analysis protocols in the early studies. In recognition of the replication crisis in psychology and linguistics, this thesis adheres to the principles of open science to enhance transparency and reproducibility. All data, materials, and analysis scripts are made accessible via the Open Science Framework, with links provided within each chapter. Additionally, the final study (Chapter 5) follows a pre-registered protocol.

In the course of this PhD project, I designed and conducted four experiments: three behavioural studies and one that combined both behavioural and neural measures (i.e., ERPs) organised into three chapters (Chapter 3 presents Experiments 1 and 2, Chapter 4 presents Experiment 3, and Chapter 5 – Experiment 4). The rationale provided above serves as a unifying theme that connects the three studies presented in this thesis. Nevertheless, each study has distinct objectives, employs its own methodology, and seeks to offer a unique contribution to the field. The studies are presented as follows.

Chapter 3, "The Influence of Three-Gendered Grammatical Systems on Simultaneous Bilingual Cognition: The Case of Ukrainian-Russian Bilinguals", features the first two empirical studies. The first experiment adapted the similarity judgment task from Phillips and Boroditsky (2003) where we compared two groups of participants – Ukrainian-Russian simultaneous bilinguals and English monolinguals. As in all subsequent experiments, we used two types of stimuli: nouns with matching and mismatching grammatical gender in Ukrainian and Russian, with all experiments being conducted entirely in English. In the first experiment, participants were shown pairs comprising of a conceptually neutral depicted noun (with grammatical genders in their two L1s being either matching or mismatching, and belonging to masculine, feminine, or neuter genders) and a personified character with an implied biological sex (e.g., a ballerina). Participants were then asked to rate the similarity of the two pictures on a scale from 1 to 9. The second experiment in this chapter involved analogous experimental design but excluded neuter-gendered nouns. This manipulation allowed us to explore the discussion about nature of the grammatical gender effects (i.e., online vs offline effects), as well as assess how different experimental paradigms might yield nuanced results rather than merely indicating the presence or absence of effects.

In chapter 4, **"Between Two Grammatical Gender Systems: Exploring the Impact of Grammatical Gender on Memory Recall in Ukrainian-Russian Simultaneous Bilinguals"**, we present the findings of the third behavioural experiment, adapting the memory recall task by Schmidt and Boroditsky (2000). In this task, participants were shown pairs of a depicted noun presented alongside a gendered name and they were instructed to memorise as many pairs as possible. Following a brief distraction task, they were asked to recall the names. The manipulation mirrored that of Chapter 3, focusing on the grammatical gender of the object and the biological sex of the name, with two types of stimuli as discussed before, with the two key differences. First, this experiment extended the investigations from Chapter 3 by employing a task with less explicit gender salience (i.e., time-limited responses with clear right or wrong answers), thereby reducing the likelihood of strategic gender use. Second, whereas Chapter 3 focused on such cognitive process as categorisation, the current experiment assessed the effects of two three-gendered languages on bilinguals' memory recall accuracy, offering a more indirect measure of grammatical gender effects.

Chapter 5, **"Investigating the Neural Basis of Partially Conflicting Grammatical Gender Systems on Categorisation Mechanisms of Ukrainian-Russian Bilinguals"** presents the final experiment, which combines behavioural and ERP measures to examine whether grammatical gender effects occur unconsciously in Ukrainian-Russian simultaneous bilinguals. Using pairs of conceptually neutral objects and faces with a distinct biological sex, in the adapted experimental framework by Sato et al. (2020), this study explored the effect of grammatical gender congruence, with stimuli either matching or mismatching in gender in Ukrainian and Russian. ERP components (N1, P2/VPP, and N300) and early gender discrimination effects (occurring at 45-85 ms) were analysed to capture early perceptual and categorisation processes. This design allowed us to investigate whether matching and/or mismatching grammatical gender modulate neural responses and if participants' more dominant/proficient L1 influences cognitive processing. While earlier studies used grammatical gender as a covert manipulation, focusing overtly on tasks related to conceptual gender or semantic relatedness - and demonstrated that grammatical gender activates unconsciously even when not explicitly required - the contribution of this study is to investigate whether grammatical gender alone (i.e., in the absence of semantic or conceptual cues) would directly affect participants' perception.

Finally, chapter 6, **"General Discussion and Future Directions"**, revisits the key issues outlined at the beginning of the current thesis. Here, I integrate the findings from all experiments, addressing the identified gaps in the literature and offering a comprehensive discussion of the implications for linguistic relativity and grammatical gender research. This chapter also outlines potential avenues for future research, aiming to build on the current findings and continue deepening the understanding of the language effects on cognition.

Overall, this PhD project offers insights into several important questions in linguistic relativity, contributes to addressing the replication crisis, and proposes directions for future research. I believe that the findings presented here will be of interest not only within the field of cognitive linguistics but also across disciplines such as psychology, cognitive semiotics, phenomenology, and beyond, where the debate on linguistic relativity remains a significant focus of interest.

Chapter 2. Literature review

2.1. Historical overview of linguistic relativity research

Long before Edward Sapir and Benjamin Lee Whorf, the relationship between language and worldview had already been theorised and examined, particularly during the 18th and early 19th centuries in the context of the Romantic movements in France and Germany. Thinkers such as Jean-Jacques Rousseau (1712–1778) in France, and Johann Georg Hamann (1730–1788) and Johann Gottfried von Herder (1744–1803) in Germany, associated language with the essence of people and their worldview (Pavlenko, 2014). Hamann and von Herder, for example, proposed that language is the "organ of thought" and argued that each "*Volk*" (nation) possesses its own "*Volksgeist*" (spirit) and a distinct way of thinking, which is expressed through its language (Miller, 1968). Von Herder specifically contended that if people are unable to think without thoughts and the way they learn to think is by using words, then language defines the boundaries and structure of our knowledge, making thought nearly indistinguishable from speaking in daily life (see interpretations in Leavitt, 2011; Pavlenko, 2014).

These foundational ideas were further developed by Wilhelm von Humboldt, who argued that "every language draws about the people that possess it a circle whence it is possible to exit only by stepping over at once into the circle of another one" (Humboldt, 1836/1988, p. 60). Humboldt (1836/1988) also distinguished between the concepts of "*Weltansicht*" - the mind's ability to perceive and organise the world into concepts through language - and Weltanschauung, a subjective interpretive framework that exists independently of language. Yet, Humboldt's (1836/1988) linguistic philosophy was based on "Weltansicht", asserting that speakers of different languages navigate and experience daily reality through the patterns inherent in their languages (Elffers, 2012; Underhill, 2009).

The philosophical perspective on language and thought was further advanced by Ludwig Wittgenstein, whose work, while not directly addressing linguistic relativity, echoes Humboldt's ideas. According to Chatterjee (1985), Wittgenstein's views later aligned with some of Whorf's perspectives on language and thought. In his early work, Wittgenstein (1922, 5.61) explored how language structures thought, suggesting that "the limits of my language mean the limits of my world", implying that our conceptual understanding of the world is bounded by linguistic structures. Later, however, Wittgenstein (1953) shifted his perspective, emphasising that the meaning of words is determined by their use in specific cultural and social contexts, which he termed "language games". He argued that language is embedded in forms of life ("*Lebensformen*"), underscoring the inseparability of language and culture (O'Grady, 2004).

Despite the later shift in Wittgenstein's philosophy, these and other contributions laid the groundwork for the linguistic relativity principle. Yet, the "foundation period" of linguistic relativity, as described by Lucy (2011), emerged primarily in the early 20th century with the rise of linguistic anthropology in America. This period was significantly shaped by Franz Boas, often referred to as the father of American anthropology (Lucy, 2011). Firstly, Boas (1920/1940) rejected the notion that Native American languages were inferior to the better-known European languages and attributed this misconception to researchers misinterpreting these languages through their own linguistic frameworks (see further discussion in Boas, 1920/1940; Lucy, 2016). He additionally that "the categories of language compel us to see the world arranged in certain definite conceptual groups which, on account of our lack of knowledge of linguistic processes, are taken as objective categories and which, therefore, impose themselves upon the form of our thoughts" (Boas, 1920/1940, p. 289). To illustrate how languages differ in categorising experience, Boas famously compared the English word "snow" to multiple Inuit terms that distinguish types of snow. While English uses one term, Inuit languages, according to (Boas, 1911), has four distinct lexical items, requiring English speakers to modify their term to capture similar nuances (e.g., "snow on the ground" in English vs "*aput*" in Inuit; "snow falling from the sky" in English vs "*qanik*" in Inuit). Boas argued that this illustrates that different languages use unique systems of classification to achieve comparable referential outcomes (Lucy, 2016). Although this example was later exaggerated and criticised in popular discussions (e.g., Pinker, 1994; see Kodish, 2003 for a full discussion), leading to critiques that sometimes overshadowed its original validity, it underscores Boas' central point: diverse linguistic systems can organise meaning in distinct ways while achieving similar denotational goals (Lucy, 1992, 2011, 2016).

Building on the foundational work of Franz Boas, the development of linguistic relativity (LR) is closely associated with the work of Edward Sapir, one of Boas' most prominent students, and Benjamin Lee Whorf, who was in turn deeply influenced by Sapir's ideas (Lucy, 1992).

Edward Sapir echoed Boas' (1920/1940) ideas in the following ways. While Boas had shown that languages classify experience differently, Sapir (1929/1949) emphasised that these classifications are not merely arbitrary but are part of formally complete systems. According to Sapir (1921/1949), each language constitutes a coherent symbolic system that enables speakers to fully articulate their experiences. This systemic nature of language not only reflects experience but actively organises and channels it, shaping how individuals conceptualise their world. In this way, he argued that language acts as a creative symbolic tool, imposing its structure on thought and experience.

Additionally, Sapir (1921/1949) supported his arguments with detailed empirical examples, further advancing the comparative approach initiated by Boas. For instance, by examining how seemingly simple sentences in English encode multiple concepts and compared this structure with equivalent expressions in German, Yana, Chinese, and Kwakiutl. For example, English explicitly marks plurality with a bound morpheme attached directly to the noun (e.g., book vs. books), where the "-s" functions as an overt grammatical marker indicating number. In contrast, in Kwakiutl (a Wakashan language), the expression of number is structurally different. As Sapir (1921/1949) explains, the noun itself does not necessarily carry the plural marking; instead, plurality is often conveyed elsewhere in the sentence through classificatory particles or inflectional elements applied to associated words, such as demonstratives or verbs, rather than the noun. Therefore, the number in Kwakiutl is not obligatorily marked on the noun but may be distributed across the sentence structure depending on what is communicatively necessary. Similarly, where English requires explicit subject and tense marking (e.g., "it is raining"), Wakashan speakers, as described by Sapir (1921/1949), express similar ideas differently, often using expressions like "it stones down". Such constructions convey the event without assigning a grammatical subject or explicitly marking tense, illustrating that languages differ not just in the elements they encode but also in where and how grammatical meaning is distributed across the sentence. These contrasts underscored his assertion that the structure of a language significantly influences how its speakers conceptualise and articulate their experiences. For a further in-depth analysis of Sapir's contributions, see seminal work by Lucy (1992). Yet, while Edward Sapir played a crucial role in shaping the linguistic relativity hypothesis, the formal articulation of the hypothesis itself was carried out by his student - Benjamin Lee Whorf.

2.1.1. Formulation of linguistic relativity

Whorf (1956) advanced and formally presented the linguistic relativity principle in his writing. Similarly to Boas and Sapir, Whorf (1956) shared the view about the classificatory nature of language, viewing language not as a haphazard collection of categories but a generative, internally coherent framework that systematically organises and shapes the experiences of its speakers. However, he extended Sapir's approach by examining less apparent morphological and grammatical structures, uncovering the intricate classificatory structure of language, and revealing the full extent of how linguistic classifications can interact with and shape thought (Lucy, 1992).

To explain different ways in which languages classify experience, Whorf (1956) discussed the distinction between overt and covert categories. Overt categories are marked by consistent, visible features, such as the English plural suffix "-s" or vowel changes (e.g., "mouse" becoming "mice"). These markers are explicitly noticeable and systematically applied. In contrast, covert categories lack consistent explicit markers and are identifiable only in specific contexts. For instance, in English, intransitive verbs form a covert category, as they cannot be used in passive constructions (e.g., "sleep" or "arrive" cannot be used in sentences like "it was slept"). Whorf (1956) suggested that these categories work together in a subtle but cohesive way to influence how people think and interpret the world. For instance, while English marks plural overtly, in Southern Paiute, plurality is not marked directly on the noun but is instead indicated by a morpheme attached to the first word in a sentence. Whorf (1956) suggested that these cross-linguistic structural differences mean that speakers of different languages are subtly guided to pay attention to different aspects of experience when forming and interpreting sentences. For example, because plurality is grammatically obligatory in English, speakers may be more attuned to whether something is singular or plural in daily interactions. In contrast, speakers of Southern Paiute, a language where plurality is inferred from sentence structure, may conceptualise number in a more context-dependent way rather than as a fixed grammatical category (Whorf, 1956). These distinctions have provided the foundation for modern explorations of neo-Whorfian effects, including investigations of grammatical gender.

Another major contribution of Whorf was his focus on how language remains largely out of speakers' conscious awareness due to its backgrounded nature. He argued that linguistic patterns are often invisible to those who use them because they are intrinsic to the cognitive frameworks shared by the speech community and speakers of a specific language do not recognise linguistic categories or pay conscious attention to them because of the lack of contrasting examples. Whorf (1956) also claimed that speakers are so bound by this background knowledge that even when acquiring a second language they tend to analyse it in terms of their own linguistic categories. Continuing Sapir's idea, Whorf argued that if language classifications influence thought and there are cross-linguistic differences in linguistic classifications, then we can assume the difference of thought between speakers of different languages, calling it the *linguistic relativity principle*:

"These automatic, involuntary patterns of language are not the same for all men but are specific for each language and constitute the formalised side of the language, or its "grammar" ... From this fact proceeds what I have called the "linguistic relativity principle," which means, in informal terms, that users of markedly different grammars are pointed by the grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world." (Whorf, 1956, p. 221)

2.1.2. Decline of Whorfianism and nativist criticism of linguistic relativity

By the early 1990s, linguistic relativity had largely fallen out of favour, with Wolff and Holmes (2011, p. 253) describing it as "all but given up for dead". Whorf's work, initially met with enthusiasm, gradually waned with the rise of nativist theories and Chomskyan perspectives (Everett, 2013). Nativism, with its emphasis on universal grammar, questioned how linguistic differences could produce significant cognitive disparities. If grammatical differences are merely surface-level phenomena and language operates independently of other cognitive processes, as nativism suggests, the mechanisms by which language shapes thought become difficult to explain (Everett, 2013). Also, if language is modular and it is detached from other cognitive processes (Fodor, 1975, 1983; Piaget, 2005), what available mechanisms could allow language to influence non-linguistic thought? Given the dominance of nativism during this period, claims of linguistic relativity faced significant theoretical challenges at that time and had often been dismissed by many in cognitive science (e.g., Devitt &

Sterelny, 1999). Interestingly, while universalist perspectives were considered anti-relativistic at the time, even Whorf (1956, p. 239) wrote that language was "in some sense a superficial embroidery upon deeper processes of consciousness", suggesting that relativity effects do not imply the rejection of universal aspects of human cognition.

While linguistic relativity to this day has its detractors, some scholars reject it outright. For example, Pinker (1994, p. 57) declaring that "it is wrong, all wrong", when mistakenly equating Whorfian claims with the Orwellian question (see Casasanto, 2008). Pinker (1994, p. 67) accused Whorf of participating in a "hoax" related to the famous Innuit snow example, claiming that while Boas reported Innuit speakers as having four words for snow, Whorf exaggerated this count to seven and implied there were even more. However, Whorf in fact wrote in his article that Innuit languages had three words for snow while English has one (Whorf, 1956, p. 210). The exaggeration of this example arose from second- and third-hand discussions, which eventually led to the popularised but incorrect claim that Innuit languages have one hundred words for snow (Pullum, 1991). This distortion and misinterpretation of Whorf's claims also contributed to further criticism of his ideas. As Monaghan (2011, p. 227) observes, "the ongoing distance between popular notions of Whorf and his legacy as seen from within linguistic anthropology is something members of the field continually attempt to correct".

Overall, the hypothesis was particularly susceptible to shifts in theoretical trends, as little compelling evidence was provided to support it - either by Whorf or by other proponents of the idea - except for the initial evidence presented in Whorf's work (see Lucy, 1992; 1997; 2016; Everett, 2013 for detailed analyses of Whorf's work). Stronger evidence for linguistic relativity, however, began to emerge only in the final decades of the 20th century.

2.1.3. Revival of linguistic relativity (1980s-1990s) and emergence of Neo-Whorfian research

Empirical research on the linguistic relativity hypothesis gained back its momentum in the early to mid-1990s, most prominently associated with the writings of Lucy (1992; 1997) and Gumperz and Levinson (1996), although related research had been conducted even earlier (Bloom, 1981; Kay & Kempton, 1984).

One of the seminal studies associated with the revival of Whorfianism was the research on colour discrimination by Kay and Kempton (1984). To examine the Whorfian hypothesis, they conducted an experimental study with speakers of English and Tarahumara, a Mexican indigenous language. In Tarahumara, the term "*siyóname*" encompasses the colour spectrum that English speakers categorise as "green" and "blue". The authors hypothesised that if language influences colour perception, English speakers will distinguish more strongly between green and blue colour chips than between samples belonging to the same colour category. Tarahumara speakers, lacking this linguistic distinction, were expected to base their judgements solely on physical differences. Participants were presented with triads of colour samples and were asked to decide which was the most different. Kay and Kempton (1984) found that Tarahumara speakers judged the differences between colour stimuli based on the physical discrimination distances between the samples, whereas English speakers exhibited category-based distinctions influenced by their lexical labels. However, it was suggested that English speakers might have relied on a "name strategy", labelling the stimuli prior to making their judgments.

This strategy was unavailable to Tarahumara speakers, who lacked separate terms for “green” and “blue”.

To test this explanation and to investigate whether the Whorfian effect exhibited by English speakers could be eliminated, Kay and Kempton (1984) designed a second experiment, where stimuli were presented sequentially, two chips were shown at a time (e.g., A and B, then B and C), with the critical chip (B) presented in both pairs. Participants were explicitly instructed to judge which pair, A-B or B-C, contained the greater difference in colour. This design forced participants to make a relative judgment about perceptual distance rather than relying on overall similarity or categorical labels such as “green” or ‘blue”, thus preventing participants from using linguistic categorisation. By doing so, the researchers aimed to determine whether English speakers could rely purely on physical discrimination distances, as the Tarahumara speakers did in the first experiment. The findings showed that when the “name strategy” was unavailable, English speakers’ judgments mirrored those of Tarahumara participants in the first experiment. The exaggerated perception of differences across the blue-green boundary disappeared, and there was consistency in judgments across the two groups. These findings suggested that the effect observed in the first experiment was not a permanent feature of cognition but rather a context-sensitive strategy. When linguistic labels were accessible, English speakers unconsciously relied on them to make judgments. However, when the experimental design prevented the use of linguistic categorisation, their judgments were based solely on physical properties.

This study remains influential in LR research, particularly in the domain of colour, which continues to be a prominent area of study in the field (e.g., Athanasopoulos et al., 2010; Drivonikou et al., 2007; Gilbert et al., 2006; Roberson et al., 2005; Thierry et al., 2009; Winawer et al., 2007). Moreover, the second experiment presented findings that resonate with key themes in neo-Whorfian research, including whether language effects on perception are context-sensitive and under what conditions such effects emerge or disappear (see further discussion in Athanasopoulos & Casaponsa, 2020).

2.2. The contemporary period and the latest trends in Neo-Whorfian research

Modern approaches to LR have departed from Whorf's original focus on grammatical structures as the primary drivers of cognitive differences. While Whorf (1956) emphasised that speakers of different grammars are guided towards distinct observations and worldviews (see full quote in Section 2.1.1), contemporary research extends beyond grammatical categories to include lexical distinctions.

Empirical studies now provide evidence in favour of linguistic relativity effects across a wide range of conceptual domains, including colour perception (Athanasopoulos, 2009; Roberson et al., 2005), spatial and temporal cognition (Boroditsky, 2001; Bylund & Athanasopoulos, 2017), and motion perception (Athanasopoulos & Bylund, 2013). More recent body of work has expanded into sensory modalities less commonly studied, such as touch (Miller et al., 2018), olfaction (Speed & Majid, 2019; Vanek et al., 2021), and taste (Bylund et al., 2024). Collectively, these findings illustrate how both grammatical and lexical features of language shape non-linguistic cognitive processes. As for grammatical gender, the primary grammatical property examined in this PhD thesis, these studies investigate whether the presence of grammatical gender distinctions influences the conceptual representation and perception of gendered entities. For instance, in Russian and Ukrainian, inanimate nouns are assigned grammatical gender: "*карандаш*" and "*олівець*" respectively ("pencil") is masculine, while "fork", "*вилка*" in Russian or "*виделка*" in Ukrainian, is feminine. A detailed discussion of this topic is presented in Section 2.3, as well as outlined throughout Chapters 3-5.

The latest trends in LR also include investigating whether language effects emerge as a result of conscious language use or occur unconsciously during the task by combining both behavioural and neural measures (e.g., EEG), as well as looking at the mechanisms that underpin the language effects. Furthermore, examining bilingual and multilingual populations has allowed researchers to delve into the effects of acquiring an additional language (i.e., cognitive restructuring, Athanasopoulos & Bylund, 2023), as well as the effects of participants' individual differences on previously observed language effects of monolingual speakers of their respective L1s (Athanasopoulos et al., 2010). This section will cover the aforementioned emerging trends in linguistic relativity, underscoring their relevance for the current PhD research.

The empirical shift in linguistic relativity research is further highlighted in the definitions of LR hypothesis (Everett, 2013). For example, Hunt and Agnoli (1991) frame LR hypothesis as the influence of language on thought, while more recent studies interpret it as a hypothesis that examines whether cross-linguistic differences have any demonstrable effects on thought and non-linguistic cognition (Lucy, 1997; Samuel et al., 2019; Wolff & Holmes, 2011). Casasanto (2016) describes LR as a concept in cognitive science that is interpreted differently by nearly every researcher studying it. In the current PhD thesis, when describing Whorfian effects, linguistic relativity hypothesis, linguistic relativity principle, or

Sapir-Whorf hypothesis, I am referring to an investigation of language effects on cognitive processes and perception.

Finally, LR has been investigated across both monolingual (e.g., Franklin et al., 2005; Iwasaki et al., 2010; Papafragou et al., 2008; Vernich, 2017) and bi-/multilingual (e.g., Athanasopoulos, 2009; Bassetti & Filipović, 2022) populations, with research on bi-/multilingual speakers highlighting how second language acquisition (SLA) alters the language effects of speakers' L1 (see Bylund & Athanasopoulos, 2014). These findings underscore the importance of including bilinguals in investigations of linguistic relativity. The next section explores this crucial dimension of LR research.

2.2.1. Bilingualism and impact of cognitive restructuring and individual differences on linguistic relativity effects

It has been estimated that more than half of the world's population uses two or more languages, including dialects in their day-to-day life (Grosjean, 2021). In Europe, over half the population is at least bilingual, while in North America, approximately 35% of Canadians and 23% of Americans are bilingual, the latter representing around 70 million individuals (Grosjean, 2021).

Early studies on language and thought began with comparison of monolingual speakers of different languages to make inferences about language processing and representation, with an objective to determine whether monolingual speakers of various languages differ not only in language-related but also in non-linguistic tasks (for grammatical gender research see Section 2.3). Bilingual research was not originally considered within this context, except for a few rare cases, such as Ervin (1961), who investigated how bilingual Italian-English speakers (divided into Italian-dominant or English-dominant groups determined by fastest reaction times in a picture-naming task) recall learned material when the language of learning and the language of recall differ. Participants were asked to name pictorial stimuli in either English or Italian and later recall them in a designated language. The findings revealed that recall was strongest when both learning and recall occurred in the bilingual's dominant language, while switching between languages led to a significant decrease in recall performance. Ervin's (1961) study is particularly relevant to this PhD research because it provides early empirical evidence on how bilinguals recall learned material depending on the language used during encoding and retrieval. These findings offer key insights into bilingual memory processes and are particularly relevant for Chapter 4 of this thesis.

Although bilingualism was not initially considered in early LR research, it follows that if speakers of different languages exhibit distinct cognitive patterns, bilinguals - who are influenced by the linguistic structures of all their languages - would also think differently from monolinguals of each of their languages (Bassetti & Filipovic, 2022). The introduction of bilingual participants into linguistic relativity research opened up a range of new questions, for example, do bilinguals transfer concepts from one language to the other? What are the overall cognitive consequences of bilingualism? Does the language in which a participant is tested have an effect (Athanasopoulos & Aveledo, 2012)? Hunt and Agnoli (1991), while not talking about Whorfian theory per se, suggested that bilingual speakers may possess distinct representations of the world due to alternating between two different reality-conceptualising frameworks. Green (1998) drew from this idea, using Levelt's (1989) framework for language production, and theorised that a “conceptualiser” or “message generator” stage must be language-specific, assuming that a message is built based on lexical concepts. He further distinguished between general (e.g., metalinguistic awareness or selective attention) and language-specific (i.e., rooted in lexico-grammatical features) effects. Language-specific effects, according to Green (1998), concern higher level cognitive processing, namely categorisation and reasoning, and are also central to linguistic relativity.

Building on these theoretical propositions, the importance of studying the relationship between bilingualism and thought was emphasised in multiple calls for research (Bassetti & Filipović, 2022; Cook, 2002; Pavlenko, 1999) and gained traction about a decade ago with the publication of foundational papers on the topic (Cook & Bassetti, 2011; Pavlenko, 2011). Since then, this field has grown significantly, exploring the broader topic of LR hypothesis (Bylund & Athanasopoulos, 2014) and specific areas, such as how cross-linguistic differences shape bilinguals’ perception and cognitive processes in domains of time (e.g., Boroditsky et al., 2011; Miles et al., 2011), space (e.g., Haun et al., 2011), motion (e.g., Bylund & Athanasopoulos, 2014), and grammatical gender (Bassetti & Nicoladis, 2016; Kousta et al., 2008; Sato et al., 2020).

Current LR research on bilingual/multilingual speakers focuses on two main questions: (1) how bilinguals or multilinguals experience cognitive effects from their multiple languages, and (2) how to best identify and explain these effects (Bassetti & Filipovic, 2022). Addressing these questions requires both greater depth of research, such as using various experimental techniques (e.g., Athanasopoulos & Casaponsa, 2020; Thierry et al., 2024) and a broader range of research areas (Bassetti & Filipovic, 2022).

This evolving perspective naturally leads to the next area of investigation: how bilingualism facilitates cognitive restructuring. As research has shown, bilingual individuals are not merely adapting to linguistic norms but are also reconfiguring their cognitive frameworks to accommodate multiple linguistic systems, resulting in conceptual or cognitive restructuring. The latter encompasses a gradual process of conceptual changes that bilinguals undergo when acquiring a new language and is thought to occur in both verbal and non-verbal behavioural of bilinguals (Athanasopoulos & Bylund, 2023; Pavlenko, 2014; Wang & Wei, 2019). The process of cognitive restructuring spans across various stages, from reliance on L1-based conceptual patterns to internalisation of L2-based concepts, co-existence, and convergence of linguistic categories, and, in some cases, shifts to L2 dominance or attrition of L1 distinctions (see Wang, 2020 for further detail).

The central question in such investigation is the extent to which bilinguals can flexibly alternate between language-specific thought patterns, the degree to which their cognitive processes align with those of monolingual speakers of each respective language, and the factors that influence this alignment or lack thereof (Athanasopoulos & Bylund, 2023). Existing evidence found across various domains (e.g., motion, colour, time perception) on cognitive flexibility suggests that bilingual individuals can adapt their cognitive behaviours based on the linguistic context of the testing. For example, Athanasopoulos et al. (2015) showed that German-English bilinguals categorised motion based on the specific patterns available in their L1 or L2, depending on the language of the instruction and experimental conditions (i.e., presence of verbal interference). Similar findings were also presented by Kersten et al. (2010), where Spanish-English bilinguals categorised motion events in ways dependent on the experiment's linguistic context. In English, a satellite-framed language, manner of motion is often encoded directly in the main verb (e.g., "run", "skip", "slide"), which makes this aspect highly salient in event descriptions, and additional information about the path (e.g., "into the house") is expressed through prepositional phrases or particles (Talmy, 1985). In contrast, Spanish is a verb-framed language, where the path of motion is typically encoded in the main verb (e.g., "*entrar*" - to enter, "*salir*" – to exit), while manner, if specified, is optional and often expressed peripherally using adverbs or subordinate clauses. In their study, Kersten and colleagues (2010) found that English monolinguals and Spanish-English bilinguals tested in English classified events based on the manner of motion - a feature prominently marked in English. Meanwhile, monolingual Spanish speakers and Spanish-English bilinguals tested in Spanish focused more on path information, which is more salient in Spanish.

Furthermore, the extent to which acquiring a foreign language or simply learning new categories can restructure one's cognitive processes can be assessed with training paradigms. Focusing on colour perception, Özgen and Davies (2002) found evidence that English speakers trained to categorise hues along a Turkish blue continuum (distinguishing "*mavi*" and "*lacivert*" for light and dark blues) developed heightened perceptual sensitivity at the boundary between these categories. Zhou et al. (2010) extended these findings by training Mandarin speakers to assign new lexical labels to colours within a blue-green continuum, specifically distinguishing between shades that previously fell into the same category. Using lateralised visual field tasks, the study found that after training, participants showed enhanced categorical perception for trained categories in the right visual field, that is processed by the left hemisphere, where language processing typically occurs. This suggests that the acquisition of new lexical categories can influence perceptual discrimination, highlighting the flexibility of cognitive processes driven by language. However, both studies underline that these effects may depend on experimental design and context in which the task is performed. For example, in Zhou et al.'s (2010) study, the effects of linguistic training were more pronounced in the right visual field, consistent with left-hemisphere language dominance.

In contrast, some studies challenge the notion of malleability of language effects depending on the language context. For example, Filipović (2011) observed that English-Spanish bilinguals exhibited consistent behaviour in motion categorisation and memory tasks across both English and Spanish contexts. Similarly, Athanasopoulos (2007) reported no differences in object categorisation preferences among Japanese-English bilinguals, regardless of the linguistic context in which the task was performed. In both of these studies, participants consistently performed the tasks relying on their L1s.

Such variability in findings highlights the importance of considering individual background variables when comparing the cognitive patterns of monolinguals and bilinguals. Factors like L1 and L2 proficiency, length of residence, and frequency of language use may significantly influence the degree of cognitive flexibility. For example, being more proficient in L2 has been shown to result in the patterns more comparable with monolingual speakers of that language, compared to less proficient speakers (Athanasopoulos, 2007). Meanwhile, other findings also evidenced effects of length of residence. Athanasopoulos et al. (2010) found that early, automatic perceptual processes could be reshaped by cultural and linguistic immersion. Athanasopoulos et al. (2010) focused on the domain of colour perception, specifically, the distinction in Greek between light blue (“*ghalazio*”) and dark blue (“*ble*”), compared to English that does not have separate labels for these shades. Their findings revealed that responses of short-stay Greek bilinguals (average 7.2 months in the UK) were consistent with Greek monolinguals. Long-stay bilinguals (average 42.6 months in the UK), however, showed no such effect, mirroring English monolinguals. These findings suggest that extended immersion in an L2-speaking environment, such as living in the UK, can reshape both linguistic and perceptual categorisation, reducing L1-specific effects. Interestingly, Athanasopoulos et al. (2010) also reported no significant effects of language proficiency (cf. Athanasopoulos & Bylund, 2023; Boroditsky, 2001). This could be attributed to lack of comparability between used proficiency tests. Several studies have shown that the frequency of language use significantly influences cognitive restructuring in bilinguals. Specifically, the more frequently the language is used, the more closely bilinguals’ cognitive patterns align with those of monolingual speakers of that language (Bylund et al., 2013; Park & Ziegler, 2014).

In conclusion, bilingual cognitive restructuring represents a dynamic and multifaceted process, influenced by a range of individual variables and allowing bilinguals to flexibly adapt their cognitive frameworks to include multiple languages. However, understanding the degree of such effects on bilinguals’ cognitive processes, as well as the nature of language effects overall (i.e., unconscious language effects vs strategic language use), requires a comprehensive approach that considers not only behavioural outcomes but also the underlying neural mechanisms. For example, with the use of electroencephalography (EEG). The following subsection will discuss the findings from these studies and the advantages of using EEG in investigating LR.

2.2.2. Combining behavioural and neural measures to examine Whorfian effects

Mostly researchers tend to use behavioural measures to investigate Whorfian effects, focusing on such variables as reaction times (e.g., whether participants respond faster when the condition is congruent in their L1 and slower if incongruent in their L1) or accuracy (e.g., categorising stimuli according to the labels available in their L1 or not). The extensive empirical research, as shown in the previous subsections, has enabled nuanced examinations of language effects on cognition (e.g., through verbal interference, task complexity manipulations, or visual hemifield studies; Athanasopoulos & Casaponsa, 2020). However, the common criticism arising from such studies is whether participants in fact show unconscious language effects in such studies (see Chapter 5) or whether it is an overt manifestation of metalinguistic knowledge (Athanasopoulos & Casaponsa, 2020; Samuel et al., 2019). Therefore, various studies have been aiming to investigate the LR hypothesis within more biologically grounded frameworks, using neural tools such as EEG (Boutonnet et al., 2012; Casaponsa et al., 2024; Flecken et al., 2015; Thierry et al., 2009).

To discuss the key findings of studies using EEG tool, I will discuss two types of language properties, lexical and grammatical, with the key focus on the latter, as they are most relevant to the current thesis. The use of neural measures in linguistic relativity is critical for understanding how language shapes cognitive and perceptual processes beyond subjective evaluation of experience (Thierry et al., 2024). Thierry and colleagues (2024) argue that traditional methods relying on overt linguistic judgments or meta-cognitive evaluations are often limited by their reliance on conscious awareness and strategic responses. Neural methods, such as event-related potentials (ERPs) and neuroimaging, offer direct and implicit means to capture unconscious and pre-attentive cognitive processes. The key advantage of introducing EEG when investigating linguistic relativity effects is that it captures the rapid temporal dynamics of language-driven perceptual processes, making it more suitable to track the rapid temporal unfolding of automatic, unconscious cognitive processes after stimuli onset (Athanasopoulos & Casaponsa, 2020; Thierry et al., 2024). See Chapter 5 for further discussion of the advantages of using ERPs in LR hypothesis testing, specifically in grammatical gender research.

In the lexical properties of language, one of the most studied areas in LR research is categorical perception (CP), particularly using a visual oddball paradigm (Athanasopoulos & Casaponsa, 2020). This paradigm has been applied to investigate visual detection in such areas of research as colour (Flecken et al., 2015; Xia et al., 2019), shapes of everyday objects (Boutonnet et al., 2013), and objects themselves (Casaponsa et al., 2024; Maier & Abdel Rahman, 2019). These studies often examine the visual mismatch negativity (vMMN), an ERP component reflecting automatic detection of deviant stimuli, providing evidence that language-specific categories modulate early, pre-attentive perception. However, while informative, such studies mainly target lexical-level distinctions and do not fully address the grammatical structures central to the linguistic relativity hypothesis.

When examining effects of grammatical properties of language, the use of ERPs has not been that widely incorporated. As pointed out by Flecken et al. (2015), ERP studies that analysed language-perception interactions have primarily focused on static domains such as colour and object categorisation, leaving grammatical structures under-represented. Specifically, in case of Flecken et al's (2015) study, in the domain of motion events. Flecken et al. (2015) examined the distinction between aspect and non-aspect languages, where cross-linguistic differences between speakers of aspect (English) and non-aspect (German) languages were investigated using ERPs and a visual oddball paradigm. Participants viewed animated motion events (i.e., a dot moving along a trajectory toward an endpoint) followed by target pictures that either matched or mismatched the trajectory or endpoint. Participants were instructed to press a button only when both trajectory and endpoint matched (full match trials), and EEG was used to measure the P3 (or P300) component, which reflects attention allocation and stimulus evaluation (Polich, 2007). The results revealed that German speakers exhibited larger P3 amplitudes for endpoint matches compared to trajectory matches, indicating an attentional bias toward endpoints consistent with their language's emphasis on goal-oriented motion. English speakers, by contrast, showed no such difference, reflecting equal attention to both elements, aligning with their language's focus on event temporal contours. Importantly, these neural effects were not reflected in behavioural measures, as accuracy and reaction times did not differ significantly between groups. The lack of behavioural effects in such paradigms has also been a staple in ERP research on LR, particularly because the selected time windows may be too short to elicit a behavioural response (Flecken et al., 2015; Xue & Williams, 2024).

Overall, behavioural measures are a useful tool and a crucial starting point to approach language effects on cognitive and perceptual processes. However, it is particularly vulnerable to verbal interference, where tasks may implicitly or explicitly rely on verbal information. Consequently, findings from such studies often reflect in-the-moment language activation (see "language-on-language" in the next section) influencing performance rather than showing long-lasting effects of language on perceptual encoding, which stem from neural adaptations over time. Therefore, to fully examine processes at foundational levels - those that are low-level, pre-attentive, and preverbal - and how they are modulated by language, incorporating event-related potentials is vital (see Athanasopoulos & Casaponsa, 2020; Lupyan et al., 2020; Thierry, 2016). This also enables an examination of the underlying mechanisms driving the observed language effects, which will be explored in the next subsection.

2.2.3. Mechanisms underpinning language effects

The discussion regarding the nature and the mechanisms underpinning language effects on thought has been entrenched in LR research, since the proposition of the hypothesis. According to Wolff and Holmes (2011), even Whorf (1956) himself acknowledged varying "intensities" of language effects in his writings. Wolff and Holmes (2011) propose five possible mechanisms that drive language effects on perception and that gathered support from empirical studies (see Table 2.1). These mechanisms are categorised into three classes: thinking before, with, or after language. Thinking before language (or thinking for speaking in Slobin, 1996) suggests that language influences cognition immediately before speech, requiring speakers to attend to specific linguistic features, such as grammatical tense. Thinking with language involves real-time interaction between linguistic and non-linguistic cognition, where language can either meddle by interfering with decision-making or augment cognition by extending representational capacity (Wolff & Holmes, 2011). Thinking after language refers to how habitual language use shapes cognitive processing beyond linguistic contexts, either by acting as a spotlight, making certain distinctions more salient (Boroditsky et al., 2003), or as an inducer, priming a schematic mode of thinking that persists beyond language use, such as relational language (e.g., verbs and prepositions) can induce speakers to focus on abstract relational structures rather than detailed perceptual features (Holmes & Wolff, 2010). Two mechanisms are particularly relevant to grammatical gender: "language as a meddler" and "language as a spotlight" as both account for how grammatical gender categories, though semantically arbitrary, can either subtly intrude on non-linguistic judgments (meddler) or increase the salience of gendered associations (spotlight), influencing perception and reasoning even in tasks unrelated to language.

Table 2.1.
Mechanisms of language influence on thought, based on Wolff and Holmes (2011)

<table>
<tr><th>Types of mechanisms</th><th>Main idea</th><th>Examples of studies supporting the idea</th></tr>
<tr><td>1. Thinking BEFORE language (“thinking FOR speaking”)</td><td>Thinking occurs immediately prior to language production, because speakers are required to attend to certain aspects of experience when using language.</td><td>Papafragou et al. (2008): investigated speakers of path and manner languages, showed effects evoked by differing verb typologies, but only in a linguistic task (when participants are preparing to produce speech), and not in a non-linguistic task (when participants allocate attention without being required to verbally describe the motion).</td></tr>
<tr><td>2. Thinking WITH language</td><td colspan="2">Proposes that processes associated with language are activated alongside non-linguistic processes and can be disrupted by verbal interference.</td></tr>
<tr><td>2.1. “Language as Meddler”</td><td>Language effects emerge from spontaneous recruitment of linguistic codes (LC) and non-linguistic codes (NLC). LC meddle with NLC during decision-making. Therefore, when there is a congruency between LC and NLC, participants’ reaction times and accuracy in a task is boosted, compared to incongruent codes.</td><td>(1) Discussing findings by Paparfagou et al. (2008), W&H suggest that in the non-linguistic task, when participants shifted their attention at the end of the animation (“reverse Whorf” effect), language was in fact acting as meddler. Specifically, the later-arising cross-linguistic differences in gaze allocation found due to the unprompted, involuntary generation of LC that meddled with attention to the presented recordings.
(2) Studies on colour discrimination (Gilbert et al., 2006; Winawer et al., 2007): Faster colour discrimination when linguistic labels align with task. Effect disappears under verbal interference, which could be interpreted as language meddling through the LC and NLC interaction.</td></tr>
<tr><td>2.2. “Language as augmenter”</td><td>Emerge in some cases (e.g., grammatical number, category learning), when both linguistic and non-linguistic representations combine to allow performing the task that could not be completed with one representation alone.</td><td>Studies on category learning (Lupyan et al., 2007; Vanek et al., 2021). For example, Lupyan et al. (2007) trained participants to differentiate between approachable and non-approachable aliens and evidenced that, while learning was possible both with and without assigned verbal labels, category learning occurred faster when it was accompanied by either auditory or written labels.</td></tr>
<tr><td>3. Thinking AFTER language</td><td colspan="2">Described as such where language may direct habitual attention, even in a non-linguistic context, to specific properties in one’s surrounding.</td></tr>
<tr><td>3.1. “Language as a spotlight”</td><td>Habitual use of lexical or grammar patterns highlights specific properties and may result in some aspect of the world appear more salient than others and attract more attention from speakers.</td><td>Studies on grammatical gender (e.g., Boroditsky et al., 2003): grammatical gender can direct attention to objects’ gendered properties, influencing perception and memory.</td></tr>
<tr><td>3.2. “Language as an inducer”</td><td>Language is hypothesised to prime a specific processing mode that is engaged even when language is no longer required or used.</td><td>Simulation studies by Holmes and Wolff (2010): relational language led participants to simulate effects (e.g., when presented with a line drawing of a scene where a pedestal that supported a plant disappeared, participants simulated effects of gravity) in schematic mental representations.</td></tr>
</table>

Alternatively, when discussing the nature of language influences on perception, particularly in visual modality (i.e., recognition, discrimination, and detection), Lupyan et al. (2020) propose dividing these effects into two distinct groups: online and offline effects. Lupyan and colleagues (2020, p. 936) define perception as "a process of predictive inference ... Percepts reflect 'best guesses' of the world and these guesses are informed by prior knowledge, current sensory evidence, and context-varying estimations of their relative reliability". This perspective inherently integrates language into perception, suggesting that language shapes perceptual experiences by influencing the generated "best guesses". However, the precise nature of these effects remains an open debate, with ongoing research, including study presented in Chapter 3, aiming to resolve this question.

The key paradox when discussing the nature of language effects on cognitive and perceptual processes, as emphasised by Lupyan (2012), arises from the need to reconcile how the language effects be both deep (i.e., finding evidence in favour of language effects being found even in basic visual processes), yet vulnerable to down-regulation, such as verbal interference? To address this issue, Lupyan (2012) proposes two points: (1) to reject the distinction between verbal and non-verbal processing and representations, and (2) adopt a framework in which language modulates processing in a flexible and task-dependent manner. These two points are combined in what Lupyan (2012) calls the label-feedback hypothesis. Using an example from the colour domain (Winawer et al., 2007; Thierry et al., 2009), Lupyan (2012, p. 2) argues that cross-linguistic differences found in these studies can be explained with the "gradual perceptual warping caused by learning". According to this hypothesis, long-term experience in categorising colour spectrum using linguistic labels gradually warps perceptual representations of colour. This in turn results in warped representation of colour and separation of the colour spectrum into more or less similar depending on whether they are in the same category. As different languages use different labels, that leads to different patterns of discrimination.

Recent work using training paradigms (e.g., Xue & Williams, 2024) provides further support for the label-feedback hypothesis by showing that learning novel linguistic labels, specifically grammatical morphemes marking transitivity ("*ro*" signalling transitive events, i.e., actions involving two entities; "*gi*" marking intransitive events, i.e., actions involving a single entity), can upregulate attention to previously less salient visual distinctions. Xue and Williams' (2024) findings show that once these labels are learned, they can feed back to modulate both attentional (P300) and pre-attentive (vMMN) visual processing. Xue and Williams (2024) attribute such findings to the label-feedback hypothesis. Specifically, that experimental group learned distinct mappings for two concepts of transitivity and during an oddball task, certain perceptual features covertly activated the associated labels. Then, these labels passed activation back down to the associated concepts and perceptual features, thereby warping subsequent attentional and pre-attentive visual processing.

It is crucial to further note that in the label-feedback hypothesis, Lupyan (2012) discusses primarily the effects of lexical properties on such cognitive processes as categorisation and categorical perception. As Lupyan (2012) himself points out, this hypothesis does not claim to be encompassing of all previously found Whorfian effects. For instance, it does not predict that differences in grammatical properties across languages lead to meaningful differences in thought. Similarly, the studies presented above also focus on the lexical properties rather than grammatical. However, what is the nature of the effects that grammatical properties (if any) have on cognitive and perceptual processes? As Thierry (2016, p. 701) emphasises, “One essential question, much closer to Whorf’s original speculations, is whether effects of language on perception through a mind-shaping effect of grammar could be more entrenched than those originating in terminology”.

To answer this, an extension of label-feedback hypothesis was proposed - *structural-feedback hypothesis* (Sato & Athanasopoulos, 2018). In their study, Sato and Athanasopoulos (2018), using behavioural measures, investigated the effects of grammatical properties (i.e., grammatical gender) on categorisation of French-English bilinguals, and concluded that their results could be attributed to online activation of grammatical gender, leading to the development of perceptual biases (see Chapter 3 for further description). This conclusion is based on the crucial finding from Experiment 2, where French-English bilinguals’ perceptual judgments of genderless faces were influenced by the grammatical gender of previously shown objects, even though the task did not require attention to gender or linking objects to faces. The authors suggest that their findings generally concur to the account put forward by the label-feedback hypothesis - participants unconsciously activated linguistic information in the moment of task completion (i.e., online) and that in turn flexed their perception by highlighting associated features (in case of this study - similarities between stimuli of the same gender in French). So, they propose to extend the notion of label-feedback hypothesis, by including grammatical properties in this prediction, thus formulating structural-feedback hypothesis.

Importantly, the structural-feedback hypothesis somewhat echoes “language as spotlight”, proposing that even though grammatical gender may not hold semantic relevance for lexical or conceptual representations of an entity, it still remains a prominent and mandatory feature of the language. The consistent attention required to encode this information likely shapes a speakers' tendency to adhere to this category. As a result, it influences or biases how perceptual categorisation is applied to incoming information (Sato & Athanasopoulos, 2018). However, the difference is that structural-feedback hypothesis argues that these effects are recruited online (in the moment of task completion) and prompt perceptual biases to emerge: "the unconscious activation of linguistic information occurs during on-line categorical perception, and in so doing, flexes perception by means of emphasising associated features" (Sato & Athanasopoulos, 2018, p. 228). While they view these effects as a result of top-down feedback, it is noteworthy that in a subsequent study (Sato et al., 2020, see Chapter 5 for full description), the effects of grammatical gender were found as early as in the N1 component time window (120-200 ms after stimulus onset). As argued by Thierry (2016), if these effects were purely online in nature, they would typically emerge 200 ms and later.

Finally, one of the most encompassing suggestion can be found in the study by Rączaszek-Leonardi (2010), where the author argued that the interaction between a grammatical feature and cognitive processes can be viewed on at least three timescales: (1) the online influences, (2) ontogenetic timescale, and (3) the timescale of diachronic (Macwhinney, 2005) or glossogenetic (Smith et al., 2003) language change. In this view, Raczaszek-Leonardi (2010) argues against treating grammatical gender solely as an abstract syntactic feature that activates fixed syntactic or semantic representations. Instead, grammatical gender is viewed as a dynamic feature, interacting with cognitive processes in real time (online) and over the course of development (ontogenetic). This combined the idea of language working as a "spotlight" (i.e., as a result of habitual use certain features of concepts become salient, as well as a result of diachronic/historical evolution of grammatical gender forms), and online (i.e., acts as a constraint on real-time cognitive processes, such as during task completion or perception). Overall, this idea suggests that online and offline effects are not independent from each other. Yet, while Chapters 3 proposes a first glance into the nature of grammatical gender effects, future studies could examine this further.

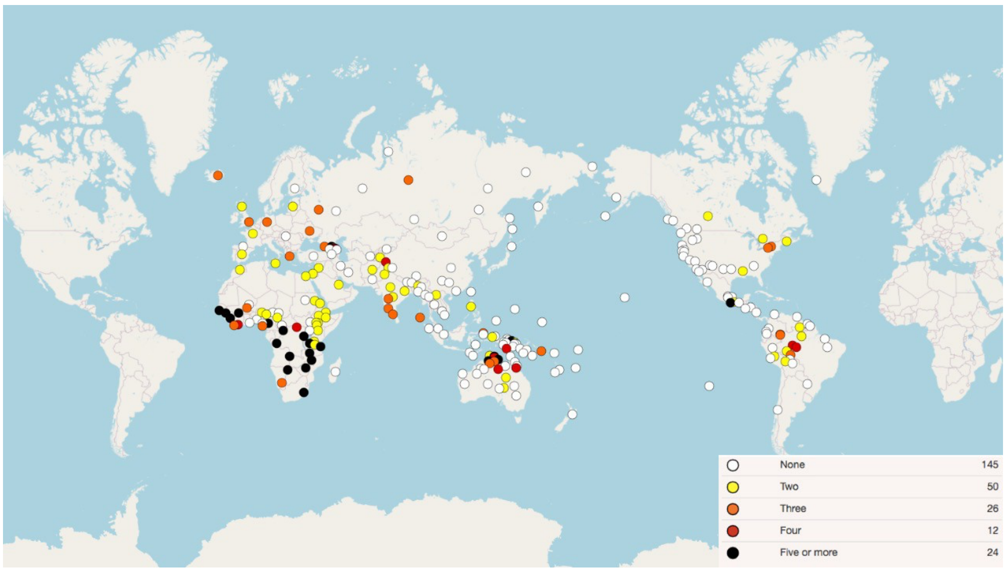
To sum up, so far this literature review presented an overview of historical development of Sapir-Whorf hypothesis and its modern-day iterations. Besides, I outlined the most seminal research in various domains (i.e., motion, colour, grammatical number, etc.), as well as how linguistic relativity research has benefited from the inclusion of bilingual and multilingual speakers. Lastly, I have discussed potential mechanisms that underpin the found language effects on cognition. However, one domain that has yet to be discussed is the domain of grammatical gender. The next section will turn to this area of LR research.

2.3. Grammatical gender in linguistic relativity research

Languages vary significantly in how they express gender. Some languages, considered genderless, exhibit minimal or no grammatical manifestations of gender. Examples include Turkish, Finnish, and Chinese (Mikić Ljubi et al., 2022; Stahlberg et al., 2007). These languages often reflect gender distinctions only through lexical conventions rather than grammatical structures. For example, in Turkish, while the noun "*doktor*" ("doctor") is gender-neutral, speakers may add "*kadın*" (woman) or "*erkek*" (man) to specify gender when necessary (e.g., "*kadın doktor*"). In contrast, other languages, such as English and Swedish, are characterised by conceptual or semantic gender, where a noun's gender is primarily dictated by its biological sex (Samuel et al., 2019). A third category includes languages with grammatical gender, which classify nouns into categories such as masculine, feminine, and, in some cases, neuter, regardless of biological sex. It encompasses many Romance, Germanic, and Slavic languages (Corbett, 1991; Everett, 2013). Grammatical gender is present in approximately 40% of the world's languages (Corbett, 2013; see Figure 2.1.), where distinctions are often marked through noun suffixes, as well as in articles, adjectives, pronouns, and, in specific instances (see discussion on Ukrainian and Russian typology in Chapter 3), verbs (Corbett, 1991; Stahlberg et al., 2007). Furthermore, gendered languages differ in the number of gender categories they employ, ranging from two-gender systems to more complex, as illustrated in Figure 2.1.

Figure 2.1

Distribution of gender systems in languages classified by number of grammatical genders (adapted from Corbett, 2013)



2.3.1. Why grammatical gender is an effective testbed for linguistic relativity?

Yet, out of all the possible grammatical properties, what makes grammatical gender such a compelling testbed for the linguistic relativity hypothesis? Several reasons can be identified:

1. Interplay between grammatical gender and biological sex. Bassetti (2007) argues that grammatical gender provides an ideal context for studying linguistic relativity effects because it provides evidence of pure linguistic effects on cognition, minimising potential interference from non-linguistic cognitive factors. However, other perspectives highlight the complexities inherent in the relationship between grammatical gender (or "*genus*" in Fatemi, 2024) and biological sex. Drawing on Ferdinand de Saussure's semiotic theory, this connection can be understood through the interplay between the signifier (the linguistic form, such as grammatical gender markers) and the signified (the concept or meaning it represents). While grammatical gender markers (e.g., pronouns, declensions, and articles) function as linguistic signifiers, they are often imbued with sociocultural and biological associations, or signifieds (Fatemi, 2024). For example, the masculine grammatical gender of "*der Mond"* (the moon) and the feminine of "*die Sonne"* (the sun) in German reflect not only purely linguistic categorisations, but also cultural meanings, such as linking the sun to nurturing or life-giving qualities rather than just grammatical rules. These cultural and biological connotations influence how grammatical gender is understood and may impact the attributes assigned to nouns, complicating the notion of "pure" linguistic effects. Vigliocco et al.

(2005) propose the "sex and gender" hypothesis as a framework for understanding how grammatical gender can influence cognition through its associations with biological and cultural attributes. According to this hypothesis, the grammatical gender of nouns may evoke male- or female-like qualities, particularly in languages where there is a transparent mapping between grammatical gender and the biological sex of referents.

2. Obligatory syntactic agreement. One of the consequences of grammatical gender system, as highlighted by Samuel et al. (2019, p. 1768) is the "obligatory conformity or agreement with the syntactic rules of that class". This includes various forms of gender marking, such as definite or indefinite articles. For example, "bed" is gender-neutral in English ("a/the bed") but feminine in Spanish ("*la cama"*) and masculine in French ("*le lit*"). This grammatical gender also influences the endings of adjectives describing the noun; for instance, in Spanish, "small bed" becomes "*la cama pequeñ**a***" (feminine form of the adjective), and in French, "small bed" is "*le lit petit*", also reflecting masculine agreement with the noun.

3. Arbitrariness and cross-linguistic variability. Grammatical gender often appears arbitrary (except for natural gender with its male/female distinction), as evidenced by cross-linguistic comparisons (Everett, 2013). For instance, the difference in gender assignment for the word "boat", which is masculine in Portuguese ("*o barco*") and French ("*le bateau*"), feminine in Spanish ("*la barca*"), and neuter in German ("*das Boot*"). Besides, Boutonnet et al. (2012) emphasise, the assigned gender can change over time, for instance in Polish the word for "girl" used to be feminine ("*ta dziewczyna*"), but over time it changed to neuter ("*to dziewcze*").

Therefore, Boutonnet and colleagues (2015, p. 73) claim: “both within a language and cross-linguistically, the relation between grammatical gender and word meaning appears to escape logic”. Grammatical gender organises entities into categories that may not share any inherent similarities in the real world but are grouped together because their corresponding nouns belong to the same morphosyntactic class.

4. Irreplaceable linguistic feature. Unlike classifier systems in grammatical number studies (e.g., Lucy, 1992; Athanasopoulos, 2007), grammatical gender lacks alternative lexicalisation patterns when absent, making it unique and irreplaceable (Boutonnet et al., 2012).

5. Influence of grammatical gender on conceptual gender. Studies both with behavioural (Sato & Athanasopoulos, 2018) and neural (Sato et al., 2020) measures show that grammatical gender may unconsciously flex conceptual representations among speakers of gendered languages. This will be discussed more in the subsequent sections.

2.3.2. Early studies on grammatical gender effects in monolingual speakers

Early studies looking at grammatical gender effects mostly focused on monolingual speakers, examining whether speakers of a gendered language would consistently associate objects with the gender assigned to them in their language. While many studies have found evidence supporting this effect, particularly in Spanish (Flaherty, 2001; Sera et al., 1994), French (Sera et al., 2002), and Italian (Vigliocco et al., 2005), results have been less consistent for German (Sera et al., 2002; Vigliocco et al., 2005) and Hebrew (Beit-Hallahmi et al., 1974; Guiora, 1984). These mixed findings suggest that grammatical gender effects may not be universally observed across all gendered languages (see discussion in Section 2.4 and in Chapters 3-4).

In one of the earliest studies on grammatical gender, Clarke et al. (1981) replicated experiments by the University of Michigan Personality and Language Research Group (Guiora & Acton, 1979) grammatical gender effects, referred to as effects of "*gender loading*" in one's native language. They found that Arabic monolinguals categorised "essentially asexual objects" based on grammatical gender (Clarke et al., 1981). In the original set of experiments, researchers investigated whether gender loading in a speaker's native language affected what characteristics (masculine or feminine) are ascribed to objects. These studies (Beit-Hallahmi et al., 1974; Guiora & Acton, 1979; Guiora & Sagi, 1978) compared children and adult speakers of Hebrew (where all nouns and modifiers are overtly gender-marked, "maximum gender loading"), English ("minimum gender loading") and Finnish ("zero gender loading") using a semantic differential task. Participants rated nouns (e.g., "gun", "apron", "table") on a masculinity-femininity scale, examining consonant words (grammatical and conceptual gender aligned, e.g., skirt - grammatical and conceptually feminine), dissonant words (conflicting, e.g., "pregnancy" - grammatically masculine), and neutral (e.g., "table" - feminine, and "window" - masculine) words. Across all groups, both adults and children classified referents mostly by conceptual gender, rather than grammatical gender, leading researchers to conclude that grammatical gender did not influence classifications.

Clarke et al. (1981) adapted this paradigm to compare Arabic and English speakers and found that Arabic speakers were influenced by grammatical gender when categorising both neutral and dissonant words. Importantly, their participants were Arabic speakers living in the US and being enrolled in the English as a Second Language (ESL) courses, raising the possibility that different effects might be observed in monolinguals. Besides, one of the key limitations was the linguistic nature of this task and reliance on words rather than images may have exhibited language-on-language effects (see Section 2.2.3) and emphasised grammatical gender over the categorisation of the objects they represent (Kousta et al., 2008).

Some of the early studies emerging in the 1990s and early 2000s also were mainly focusing on monolinguals, with one of the most widely studied phenomena in this area is how grammatical gender influences object categorisation. For instance, Sera et al. (1994) showed that Spanish speakers, when asked to classify objects as “masculine” or “feminine”, were influenced by the grammatical gender of the noun, even when linguistic labels were absent. Similarly, Flaherty (2001) found that Spanish-speaking adults and older children (8–10 years old) relied on grammatical gender when assigning gendered attributes to images, whereas younger children (5–7 years old) were more influenced by their own gender (e.g., girls and boy assigning more female and male attributes respectively). These findings suggest that grammatical gender effects develop over time and become more consistent as speakers gain proficiency in their linguistic system. More recently Vernich (2017) investigated grammatical gender effects on monolingual speakers of Estonian, Italian, and Lithuanian monolingual speakers using a voice attribution task (adapted from Sera et al., 1994). The voice attribution task (see detailed review in Samuel et al., 2019) is one of the most commonly used experiments in the study of grammatical gender. In this task, participants are asked to assign a male or a female voice to depicted nouns. While Italian and Lithuanian are two-gendered languages, Estonian was chosen over the traditionally studied English because it is not only genderless but also lacks gendered pronouns such as “he” and “she”. Instead, Estonian uses a third-person pronoun (“*tema*” or “*ta*”) for both male and female referents (Vernich, 2017). The findings revealed significant grammatical gender effects for Italian and Lithuanian speakers, with 76% and 70% of their responses aligning with the grammatical gender of objects, respectively. In contrast, Estonian speakers assigned voices to objects with almost equal distribution between male and female. These results also supported Sera et al.’s (1994) assumption regarding universality of perception of artifacts as more masculine and natural objects as more feminine, at least for Estonian speakers. However, among speakers of gendered languages, grammatical gender often took precedence, particularly when it conflicted with the artefact/natural object distinction.

Yet, when looking at all studies discussed above (Flaherty, 2001; Sera et al., 1994; Vernich, 2017) might have provided evidence in favour of the effects of language-on-language (see also “thinking for speaking”) instead of true Whorfian effects, as participants were tested in their L1s, and nothing was done (or at least reported in such a way) that prevents them from either overt or covert verbalisation. Therefore, one of the key challenges when testing monolingual participants is to determine whether the observed effects truly reflect unconscious grammatical gender influence. Furthermore, the choice of tasks in the discussed studies, such as the voice attribution paradigm in Vernich (2017) or classifying objects as masculine or feminine in Sera et al. (1994), are highly gender-salient and have no objectively correct answer (Samuel et al., 2019; see discussion in

Section 2.4.). This could further prompt participants to use grammatical gender of their L1 in a more conscious manner to "solve" the task. To address these limitations, future research could compare these results with findings from the same populations tested in a genderless L2, such as English or Estonian, as well as look at the experimental design with more covert grammatical gender manipulations.

To address the latter and in attempt to find Whorfian effects beyond "thinking for speaking", here have been some early efforts to provide evidence that grammatical gender also affects perception in non-linguistic contexts. Flaherty (2001, Experiment 2) provided evidence that grammatical gender influences categorisation even when participants were presented with unlabelled visual stimuli. Although speakers were not explicitly instructed to rely on linguistic information, their responses reflected gendered associations tied to their native language's grammatical structure. Examining grammatical gender effects in non-linguistic contexts is crucial; however, this research has been conducted primarily with bilingual populations (see Section 2.3. for examples), with Flaherty (2001) being, to my knowledge, the only study to have investigated this issue in the early years of grammatical gender research.

Overall, early studies on grammatical gender effects among monolingual speakers have shown that the effects can emerge as early as early as 8 years old. However, these effects were found when participants were tested in their respective native languages, which could have resulted in a more conscious use of grammatical gender or verbalising of the stimuli with the grammatical gender indicators. Consequently, a question arising from these studies was whether these effects would pertain when participants are tested in a different language, particularly in a genderless L2? This question motivated the inclusion of bilingual populations, who navigate more than one grammatical system, offering a unique opportunity to investigate whether, and how, the two languages influence cognition. This will be explored in the following section.

2.3.3 The advantages of incorporating bilingualism in grammatical gender research in linguistic relativity and the role of L1-L2 pairing

Incorporating bilingual speakers into grammatical gender research on linguistic relativity provides several advantages compared to a monolingual approach. These advantages become particularly evident when considering specific types of L1-L2 pairings, based on the presence or absence of grammatical gender in each language of a bilingual, namely gendered L1 with genderless L2, genderless L1 with gendered L2, and gendered L1 with gendered L2.

First, bilingualism provides a way to distinguish "language-on-language" effects from those actually affecting conceptual representations, Additionally, it allows to test the stability of grammatical gender effects across languages (Kousta et al., 2008). Monolingual studies have been criticised for capturing only linguistic rather than cognitive influences of grammatical gender on categorisation, attributing findings in support of LR effects to covert verbalisation (Bassetti & Nicoladis, 2016). Testing bilinguals - especially those who speak both gendered and genderless languages - helps determine whether grammatical gender shapes thought beyond language-specific processes. Most studies addressing this matter select participants with a gendered first language (L1) and a genderless second language (L2), aiming to examine whether the grammatical gender effects observed in the L1 persist when tested in the L2 (see section 2.4.2 for further discussion). However, findings from these studies have been somewhat inconsistent. For example, Kousta and colleagues (2008) asked both monolingual Italian speakers and bilinguals with Italian as L1 and English as L2 to complete the same linguistic error-induction task to assess whether grammatical gender affects the semantic substitution errors. Bilinguals were tested in both languages. They predicted that if bilinguals show differences in their cognitive behaviour in this linguistic task depending on the language they are using, this will support the idea that language affects only language-specific processes, not general conceptual or non-linguistic thought. However, if bilinguals show similar pattern to the monolingual speakers of their L1 regardless of the language of the context, this will evidence that grammatical gender has affected conceptual representations. The findings revealed that Italian–English bilinguals adapted their semantic representations based on the language they were using. In English, they performed similarly to monolingual English speakers, while in Italian, their performance mirrored that of monolingual Italian speakers. Kousta et al. (2008) interpreted these results as evidence that grammatical gender effects are confined to language-specific processes and do not extend to non-linguistic cognition. Similarly, Pavlidou & Alvanoudi (2013) found that Greek-English bilinguals showed grammatical gender effects when tested in Greek but not in English, suggesting that grammatical gender effects may be restricted to the linguistic domain rather than conceptual thought (see Chapter 4).

On the other hand, multiple studies (e.g., Forbes et al., 2008; Sato & Athanasopoulos, 2018) found contrasting evidence to that of Kousta et al. (2008). Forbes et al. (2008) expanded this line of research by using a voice attribution task. They used colour drawings of people, animals and inanimate objects and divided conditions into linguistic (pictures presented with auditory labels) and non-linguistic (pictures without labels). This study is particularly interesting as it also examined the order of acquisition of gendered and genderless languages, by investigating French-English and English-French bilinguals, Spanish-English and English-Spanish, and using English monolinguals as controls. Bilingual participants were selected based on their responses to the bilingual history and fluency questionnaire and divided into early or late L2 learning (those who started learning L2 before or after the age of 7, respectively). Items were chosen with opposite grammatical genders in Spanish and French and half of the participants completed the experiment in their L1 and others - in their L2. Forbes et al. (2008) found that French-English and Spanish-English participants were significantly influenced in their classifications by French and Spanish gender respectively, while English-French bilinguals' responses mirrored those by English monolinguals. Besides, neither the language of testing nor the presence of labels had any significant effect on the results. Unexpectedly, Spanish grammatical gender influenced classifications not only by Spanish-English, but also English-Spanish bilinguals and English monolinguals. Both latter groups classified items in line with Spanish grammatical gender. This led Forbes et al. (2008) to speculate whether Spanish grammatical gender may be less arbitrary than often assumed, largely reflecting four observable characteristics ("femininity scores" in Sedlmeier et al. (2016), see Section 2.4.3): items being (1) artifacts or natural objects, (2) angular or curved, (3) typically used by males or females, and (4) being dense or not dense. After examining stimuli based on these characteristics, Forbes et al. (2008) also argued that Spanish grammatical gender and the number of masculine attributes (artifact, angular, typically used by males, dense) are equally reliable in predicting participants' gender classifications, for French speakers as well. This furthered the idea, earlier presented by Sera et al. (2002) that Spanish gender assignment might be universal.

Overall, findings by Forbes et al. (2008) and Sato and Athanasopoulos (2018) suggest that effects of grammatical gender of participants' L1 can be found even when participants are tested in a genderless L2. However, what happens with the increase in language proficiency and/or dominance of a genderless L2? Previous study on colour categorisation (Athanasopoulos et al., 2009) showed that with increased proficiency and exposure to different lexical properties in L2, effects of L1 diminish. Would the same hold true for grammatical properties? Specifically, would increasing proficiency in a genderless L2 reduce effects in a gendered L1 or leave them unchanged?

In an attempt to answer this question, Sato et al. (2013) investigated the effects of grammatical and stereotypical gender information in French-English bilinguals (half with English as L1 and others with French as L1), using sentence evaluation paradigm. Participants judged the acceptability of whether a sentence referring to a group of men or women was a sensible continuation of a preceding sentence that introduced a role noun with either female (e.g., social workers), male (e.g., surgeons), or neutral (e.g., musicians) stereotypical associations. For example, a sentence "The social workers walked through the station" was followed by one of the three options: "At the end of the day, the majority of the [men /women / social workers] seemed to want to go home". Importantly, in the French version, these role nouns were always presented in the masculine plural form, which, while formally generic, can trigger a male-specific interpretation due to the grammatical gender marking. In contrast, English role nouns were grammatically neutral, making gender inferences reliant solely on stereotypes. Each bilingual group was tested in both L1 and L2. The findings revealed that bilinguals construct mental gender representations in alignment with the grammatical and stereotypical cues of the language they are using, but the extent of this alignment was modulated by their proficiency in the second language. Advanced bilinguals showed decreased reliance on their L1 grammatical gender and instead relied more strongly on stereotypical associations, more closely resembling native speakers of their L2. For instance, French-English bilingual adults showed less influence from the French grammatical gender system on their stereotypical gender attitudes as their English proficiency increased. However, no other studies addressing the shift in grammatical gender effects with increased L2 proficiency/dominance have been found, suggesting the need for further investigation of the effects of the increasing/decreasing L2 language proficiency on flexing bilingual's perception.

Second, expanding research to include bilingual speakers that have genderless L1 and a gendered L2 allows researchers to investigate whether grammatical gender effects still emerge when a gendered language is acquired later in life, as well as how early cognitive restructuring occurs during second language acquisition (SLA) and how grammatical gender might influence cognitive processing at various stages of SLA. Kurinski and Sera (2011) conducted a longitudinal study that observed effects of cognitive restructuring, albeit limited, in English speakers learning Spanish. In this study, three groups of participants - college students with English as L1 enrolled in the Spanish course (Beginner level), advanced Spanish L2 learners, and native Spanish speakers - took part in two tasks four times in one academic year. The first task tested their acquisition of Spanish grammatical gender, while the second task (voice attribution) assessed how they categorise inanimate objects. The study revealed two key findings. Firstly, aligning with prior research, native Spanish speakers were exhibiting effects of grammatical gender on categorisation. Secondly, beginner learners of Spanish showed changes in their categorisation as early as 10 weeks into learning, with a gradual increase in grammatical gender effects. The most significant improvement occurred after 20 weeks of instruction, followed by a plateau at 30 weeks, beyond which no further changes emerged. Despite these insights, the study had notable limitations. Although the voice attribution task preceded the language task to minimise priming, repeated testing over the year may have increased participants' awareness of the grammatical gender manipulation. Furthermore, the use of identical stimuli across tasks, with only a small number of control trials, may have influenced participants' categorisation tendencies. The plateau effect after 30 weeks could reflect participants' understanding of the task's purpose rather than a true limit in grammatical gender effects. Employing varied tasks or stimuli in future research might yield different results.

Building on the findings of Kurinski and Sera (2011), Athanasopoulos and Boutonnet (2016) provided additional evidence of grammatical gender influencing cognition in learners with a genderless L1 and acquiring a gendered L2. Their longitudinal study of English speakers learning French found that learners increasingly aligned their categorisation patterns with French grammatical gender. This effect strengthened with the increasing proficiency and exposure to a gendered L2. Importantly, unlike in the study by Kurinski and Sera (2011), these findings reflect not just task-specific effects but effects of cognitive restructuring, as new stimuli were introduced across testing sessions. The latter provides stronger evidence that changes in categorisation were due to internalised learning rather than repeated exposure to identical items.

Integrating the two previously discussed points regarding the two types of L1-L2 pairings based on the order of acquisition of genderless and gendered languages, Chen and Faitaki (2024) recently investigated both speakers of gendered L1 - genderless L2 and genderless L1 - gendered L2. Using a voice assignment task, they examined the effects of French grammatical gender on object categorisation in French monolinguals, English-French and French-English bilinguals, with English monolinguals as controls. Participants rated the gender of objects by assigning them a masculine or feminine voice using a slider, which provided a continuum of responses rather than binary choices as in a traditional voice assignment task. Crucially, L2 proficiency was assessed through standardised testing, but simultaneous and sequential bilinguals were not separated, with 14% (n = 10) of participants being simultaneous bilinguals, 39% (n = 28) began second language acquisition between the ages 4-9, while 47% (n = 32) started L2 learning at age ten or later. The results revealed that French monolinguals and English–French bilinguals aligned their gender ratings with the French grammatical gender system, with no evidence that acquiring English diminished these effects in French–English bilinguals. Furthermore, the effect was not dependent on L2 proficiency or on other self-reported measures, such as age of acquisition of L2 or length of L2 exposure. These findings suggest that learning a gendered L2 influences object categorisation, while learning a genderless L2 does not in fact diminish the effects of a gendered L1, contrasting with findings by Sato et al. (2013). This contradiction may be due to differences in task design and different degree of salience grammatical gender. It is also unclear whether these results would emerge if Chen and Faitaki (2024) used a more covert gender manipulation.

Third, including bilingual speakers that have both gendered L1 and L2 offer additional insights into the interaction of two grammatical gender systems in a bilingual mind and their effects on bilingual perception.

Bilinguals who speak two gendered languages must navigate potentially conflicting gender assignments (see Bassetti, 2007; Phillips & Boroditsky, 2003 in Chapter 3). In a more recent study, Lambelet (2016) aimed to identify connotation in the mental lexicon of French L2 learners that are linked to their L1 (study included speakers of 21 different L1s), as well as how are these connotations modified after learning a gendered L2 (French). Participants with varying French proficiency (self-rated reports from A1 to C1 levels) were asked to complete four tasks, two of which were reported in the study: a voice attribution task and French grammatical gender production task (i.e., assigning appropriate article according to French grammatical system). Importantly, participants had varying grammatical gender systems in their L1 (languages without grammatical gender, with a two- or three- gendered systems) and knew English that was used as the language of the written instructions. Findings showed that grammatical gender of participants' L1s had significant effect on voice attribution task, while speakers of genderless languages assigned voices randomly (51% of male voices). Interestingly, these grammatical gender-linked effects appear to become more pronounced when a second grammatical gender system is learned, particularly for items that share the same grammatical gender in both participants' L1 and French L2. However, when L1's and L2's grammatical genders were contrasting, the L1's grammatical gender effects were diminished, aligning with the earlier findings (Bassetti, 2007; Bassetti & Nicoladis, 2016). Finally, in the voice attribution task, although no direct effects were observed based on the actual grammatical gender of items in French, a significant effect was found with the grammatical gender that participants believed the items had in French.

In their review, Bassetti and Nicoladis (2016, p. 9) also highlight that while it is not yet clear why knowing more than one grammatical gender system may reduce the effects of L1's effects on perception, there are two possible reasons for it. Firstly, bilinguals may come to recognise that gender assignments are semantically arbitrary, or, secondly, they may develop different habitual ways of thinking compared to monolinguals due to the need to refer to the same entity with one gender in one language and a different gender in the other. However, another possibility to explain the discrepancies in findings is that grammatical gender of some languages may affect its speakers more than others, prompting researchers to consider certain aspects of methodological design in studies with bilingual participants. I will turn to these aspects next.

2.3.4. Key considerations for bilingual research on grammatical gender

There are several important factors to consider when including bilingual participants in studies on grammatical gender, particularly regarding methodological design. Careful attention must be given to obtaining detailed information about participants' linguistic profiles to ensure both the validity of findings and comparability across studies. This subsection outlines the following key considerations: (1) how to measure L2 proficiency and dominance, (2) whether to include sequential or simultaneous bilinguals, (3) differences between two- and three-gendered languages, and (4) the importance of replication studies. Each is discussed in detail below.

One of the key matters to discuss when testing bilingual participants is how best to measure L2 proficiency and dominance. As shown in Table 2.2, studies have employed various approaches. Nearly half of the presented studies relied on self-reported measures (i.e., self-reported proficiency in Bassetti, 2007; Boutonnet et al., 2012; Lambelet, 2016;

Phillips & Boroditsky, 2003; Sato et al., 2013), two studies (Boroditsky & Schmidt, 2000; Pavlidou & Alvanoudi, 2013, 2019) did not explain the way L2 proficiency was determined and simply stated the participants knew L2 at an acceptable level to participate. Yet, 4 out of 14 studies presented in the table 2.2 also relied on various standardised tests (Nicoladis & Foursha-Stevenson, 2012), as well as combining them with self-reported questionnaires (Chen & Faitaki, 2024; Sato et al., 2013; Sato & Athanasopoulos, 2018). However, each study relied on a different proficiency test, as well as different self-rated questionnaires.

Table 2.2

Methods of assessing L1-L2 language proficiency and dominance across key studies on grammatical gender

Study	L1-L2 pairing of main participant group	How proficient / dominant language was determined
Bassetti (2007)	Italian-German simultaneous bilingual children	**Self-report** questionnaire assessing: (a) preferred language (L1 or L2); (b) perceived better-spoken language; (c) self-rated proficiency (native-like or not) in each language.
Sato et al. (2013)	French-English and English-French sequential adult bilinguals	C-Test (Grotjahn, 1992; Rahimi & Saadat, 2005); Self-evaluation questionnaire assessing L2 background, Age of Acquisition (AoA), years of L2 study, and self-rated written, listening, writing, and speaking competence in L2.
Nicoladis and Foursha-Stevenson (2012)	French-English simultaneous bilingual children	Peabody vocabulary test (Dunn & Dunn, 1997).
Sato et al. (2020)	French-English sequential adult bilinguals	**Self-rated** questionnaire (AoA, proficiency, current language usage) for both L1 and L2.
Sato and Athanasopoulos, 2018	French-English sequential adult bilinguals	Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007); Oxford Quick Placement Test (Oxford University Press, 2001).
Phillips and Boroditsky (2003)	Spanish-English and German-English sequential adult bilinguals	**Self-reported** L2 proficiency and years of experience with L2.
Boroditsky and Schmidt (2000)	Spanish-English and German-English sequential adult bilinguals	No detailed assessment provided; authors only noted participants' high proficiency in English.
Pavlidou and Alvanoudi (2013)	Greek-English sequential adult bilinguals and German-English sequential adult bilinguals (the results for this group were not reported)	No detailed proficiency assessment provided; participants described as having "very good knowledge" of English (Pavlidou & Alvanoudi, 2013, p. 113).
Boutonnet et al. (2012)	Spanish-English sequential adult bilinguals	**Self-rated** L1 and L2 proficiency and language experience (i.e., daily usage of L1 and L2, AoA of L2, lengths of immersion)
Athanasopoulos and Boutonnet (2016)	English-French sequential adult bilinguals	**Self-rated** background questionnaire in L1 and L2
Chen and Faitaki (2024)	English-French and French-English adult bilinguals (both sequential and simultaneous)	Vocabulary tests for English and French; English Lexical Test for Advanced Learners of English (LexTALE) for French-English bilinguals and French LexTALE for English-French bilinguals (Lemhöfer & Broersma, 2012; Miralpeix & Muñoz, 2018)
Lambelet et al. (2016)	Bilinguals with 21 different L1s, with French as L2	**Self-rated** proficiency in French on the Common European Framework of Reference (CEFR) scale from A1 (beginner) to C1 (effective operational proficiency).
Kurinski and Sera (2011)	English-Spanish sequential adult bilinguals	No direct proficiency assessment: proficiency levels determined by assigned university course level.

Note. The table summarises methods explicitly reported in each study. Additional proficiency/dominance measures may have been utilised but not reported.

A consistent and standardised approach to L2 proficiency and dominance assessment is essential to compare the findings across studies. Importantly, the lack of effects in studies relying only on standardised tests (e.g., Chen & Faitaki, 2024; Kurinski & Sera, 2011) suggest that a combined approach, using both self-reports and standardised testing, may provide a broader understanding of L2 effects. The latter also influenced our choice of standardised proficiency tests and self-rated questionnaire across experiments (see Section 2.5).

Another matter to consider is the inclusion of sequential and simultaneous bilingual speakers. For instance, in the study by Chen and Faitaki (2024) both groups were included in the same analysis. As also mentioned in Chapter 3, there is a scarce body of evidence on the effects of grammatical gender using simultaneous bilinguals. The only two studies specifically focusing on simultaneous bilinguals (Bassetti, 2007; Phillips & Boroditsky, 2003) are discussed in Chapter 2.5. The key objective of including simultaneous bilinguals, particularly those with two grammatical gender systems, is to further address how two distinct grammatical systems are interacting in bilingual's mind. This question is addressed in the current PhD research by incorporating stimuli with both matching and mismatching grammatical gender across the two L1s of simultaneous bilingual participants.

A further area of interest involves comparisons between speakers of two- vs three-gendered languages represents an important (see Chapters 3 and 4 for an extensive discussion). In the previously discussed research, scholars examined speakers of both two-gendered (typically Romance languages) and three-gendered languages (typically German). However, the results for speakers of the latter group were quite contradictory, with findings showing both evidence in favour of language effects in three-gendered languages (e.g., Konishi, 1993; Lambelet, 2016; Phillips & Boroditsky, 2003) and arguing against such effects (e.g., Kousta et al., 2008; Pavlidou & Alvanoudi, 2013, 2019; Sera et al., 2002; Vigliocco et al., 2005), with multiple studies claiming that two-gendered languages have stronger effects on cognition than three-gendered ones. Similar tendency was found by Samuel and colleagues in their 2019 review, showing that higher support was present in studies with two-gendered languages, than with three-gendered ones (43% and 16% of supporting findings respectively). Importantly, the majority of studies on three-gendered languages have predominantly focused on German, raising the question of whether observed findings can be generalised across all three-gendered languages or if they are specific to German due to its typology (e.g., transparency of grammatical system in German, compared to other languages)? Studies on three-gendered Polish have recently provided less contradictory evidence in favour of grammatical gender effects (e.g., Haertlé, 2017; Maciuszek et al., 2019; Rączaszek-Leonardi, 2010, see Chapter 4). This discrepancy calls for further research exploring grammatical gender effects across diverse three-gendered languages with varying typological characteristics (e.g., presence or absence of neuter gender).

Additionally, another layer of complexity arises from the replication crisis in linguistic relativity research, particularly concerning the mixed or non-significant findings in more recent replications of seminal studies, such as those presented in Boroditsky and Schmidt (2000), which was also replicated for the current PhD project in Chapter 4. An example of such a failed replication is provided by Pavlidou and Alvanoudi (2013), and this study is also discussed in Chapter 4. Similarly, Elpers et al. (2022) conducted a direct replication of Phillips and Boroditsky's (2003) similarity judgement task (used in Chapter 3) which also failed to reproduce the original effect and is discussed further in Chapter 3. In another attempt to replicate the findings presented in Boroditsky et al. (2003), Mickan and colleagues (2014) replicated an adjective assignment task where Spanish and German monolingual participants were presented with a written list of nouns in their respective languages and were asked to assign three adjectives that come to mind. These adjectives were later rated by other two groups of Spanish and German speakers as more masculine or feminine. No effects of grammatical gender were found in either participant group.

Mickan et al. (2014) attributed this to the task itself, noting that most adjectives described the stimuli directly rather than reflecting gender associations. Additionally, the data showed considerable variability and resulted in about 200 Spanish and 150 German adjectives, indicating minimal repetition across participants. However, another possibility is that the study by Boroditsky et al. (2003) tested their participants in their genderless L2 (English), while Mickan et al. (2014) tested participants in their L1. Mickan et al. (2014) also designed a primed lexical decision task, where Spanish and German native speakers were presented with pictures of gendered primes (opposite grammatical genders in Spanish and German) and adjectives (words associated with male / female features and non-words). Participants had to decide whether the adjective was a word or a non-word. It was expected that when conceptual associations of adjectives and grammatical gender of primes were congruent, then the reaction times would be faster, compared to the incongruent pairs. However, no significant effects of gender were found in this experiment. This led Mickan et al. (2014, p. 47) to conclude that the original findings by

Boroditsky et al. (2003), were either “a statistical fluke” or a result of a methodological aspect that was unreported.

To sum up, one might argue that one possible explanation of the lack of significant gender effects, particularly in more recent studies, has more to do with the replication crisis, rather with the lack of effects of three-gendered languages. All of these reasons underscore the nuances of methodologies when designing a task that investigates grammatical gender effects. This is why I will turn to describing methodological issues in this line of research next.

2.4. Study design issues in grammatical gender research

In this section, I will summarise various methodologies used to investigate the effects of grammatical gender on cognitive processes and outline potential directions for future research. These directions aim to address the questions raised in this review and highlighted in previous calls for research (e.g., Athanasopoulos & Casaponsa, 2020; Bassetti & Filipović, 2022; Thierry, 2016). Specifically, the focus will be on methodologies that examine not merely whether grammatical gender influences perception, but rather the nature of these effects, as well as how and when they emerge. Additionally, I will connect these issues to those addressed in the current PhD project (Chapters 3-5). Some overlap between this subsection and subsequent chapters is unavoidable.

In their review, Samuel et al. (2019) identified six constrains in grammatical gender research that could influence the findings: (1) the salience of gender/sex in the task, (2) the salience of language in the task, (3) language of the testing, (4) differentiating between animate and inanimate stimuli, (5) different findings in two- and three-gendered languages, and (6) varying findings in adult and children. While the latter three parameters and their associated findings have been discussed to varying extents in earlier sections and Chapters 3-5, the first three elements are particularly relevant, as they directly relate to the methodological choices made for the experiments in the current PhD project.

2.4.1. Salience of gender/sex and language in the task and the transparency of grammatical gender manipulation

One of the key methodological considerations is the salience of gender/sex and language in experimental tasks. It is essential to consider the types of tasks used to investigate grammatical gender, as "the evidence for an influence of grammatical gender on conceptualisations is highly task- and context-dependent" (Samuel et al., 2019, pp. 1779-1780). Therefore, understanding these methodological nuances is crucial for evaluating past findings and designing robust experiments when investigating grammatical gender effects.

One factor contributing to the heightened salience of gender or sex in certain tasks is the absence of an objectively correct answer. Tasks lacking a definitive solution tend to provide a more gender- or sex-salient context compared to those where a definitive correct response is present. For instance, sex attribution and voice assignment tasks are most common in this area of research and account for the largest body of findings presenting evidence in favour of grammatical gender effects on categorisation (e.g., Belacchi & Cubelli, 2012; Flaherty, 2001; Kurinski & Sera, 2011; Lambelet, 2016; Vernich, 2017). However, these types of tasks are often subject to criticism, due to their high content of language and gender/sex. For example, Bender et al. (2011) argued that when participants are explicitly asked to assign gender or biological sex, particularly to inanimate objects, they may consciously and strategically use grammatical gender as a task-solving strategy. Therefore, this approach may reflect metalinguistic awareness rather than true conceptual change due to grammatical gender.

To address this criticism, a different set of studies used less explicit methodologies, such as the Extrinsic Affective Simon Task (EAST) (Bender et al., 2016), object-name memory task (Boroditsky & Schmidt, 2000; Kaushanskaya & Smith, 2016; Pavlidou & Alvanoudi, 2013, see Chapter 4 for the detailed description), and priming tasks with a clear correct answer (e.g., relying on stereotypes in Sato & Athanasopoulos, 2018, or semantic information in Boutonnet et al., 2012). For example, in the EAST task (Bender et al., 2016), participants respond to stimuli (e.g., words or images) presented in colour, where the colour is associated with a target category such as grammatical gender or biological sex. Participants respond to words using two keys (in one condition key are mapped to colours, in the other condition - to male/female sex), allowing to measure implicit activation of grammatical gender without making it task-relevant. While a small number of these studies showed an unconscious grammatical gender activation when it was not explicitly required (Boroditsky & Schmidt, 2000; Sato et al., 2013; Sato & Athanasopoulos, 2018; as well as neural evidence by Boutonnet et al., 2012 and Sato et al., 2020), a large body of work, particularly when looking at three-gendered languages, produced mixed results (e.g., Bender et al., 2011, 2016; Kaushanskaya & Smith, 2016) or found no effects at all (Bender et al., 2018; Pavlidou & Alvanoudi, 2013).

Additionally, we need to account for the degree of language salience in the task, particularly to distinguish between mere "thinking for speaking" effects and those that genuinely provide evidence in favour of

LR effects in non-linguistic tasks ("thinking *with* or *after* language"). Several studies, while showing significant effects of grammatical gender, used linguistic tasks and written stimuli (e.g., Boroditsky & Schmidt, 2000; Konishi, 1993; Rączaszek-Leonardi, 2010), leaving the interpretation of these findings a subject to criticism. Therefore, even though, the definition of a truly non-linguistic task in grammatical gender research remains a subject of debate (Lupyan, 2012; Saalbach et al., 2012), understanding the level of language salience in the task is critical for determining whether observed effects result from grammatical gender being recruited through language processing or reflect deeper conceptual change (Samuel et al., 2019).

Importantly, a review of the tasks employed to examine the effects of three-gendered languages, which is the focus of the current PhD project, revealed a predominant reliance on methodologies with both high language salience and high gender/sex salience in both studies that find support (Forbes et al., 2008; Konishi, 1993; Lambelet, 2016; Phillips & Boroditsky, 2003) and argue against the grammatical gender effects in speakers of three-gendered languages (Kousta et al., 2008; Sera et al., 2002; Vigliocco et al., 2005).

One way to avoid high gender salience in the task as well as research grammatical gender effects outside of the traditional focus on categorisation or object conceptualisation, is examining other perceptual mechanisms, such as odour perception (Speed & Majid, 2019). According to Speed and Majid (2019), the advantage of incorporating odour perception into this research lies in its departure from explicit lexical cues to gender. Instead of explicitly judging the referent of a noun, participants evaluate attributes or characteristics indirectly associated with the noun (e.g., fragrances linked to nouns with specific grammatical genders).

In their study, Speed and Majid (2019) presented German and French L1 speakers, whose additional L2s were mostly English or Dutch, with fragrances (4 male and 4 female) and their corresponding descriptions. The fragrance descriptions included ingredients with nouns of masculine and feminine gender, which were opposing in French and German (e.g., notes of "pumpkin" or "sage", masculine in German and feminine in French). Participants were asked to read each description and smell a corresponding fragrance. After that, they were instructed to rate it on various criteria using a Likert scale and complete a memory recognition task with distractor fragrances. It was hypothesised that when the "marketing gender" of the fragrance (i.e., whether it is designed for males or females) is congruent with grammatical gender of the nouns in the description, participants would recall them better, compared to the incongruent condition (Speed & Majid, 2019, p. 2066). Besides, grammatical gender effect on ratings was also predicted. The results confirmed the recall effect but primarily for male fragrances paired with masculine descriptors. This was attributed to the all-female participant group, where masculine attributes may have been more salient. In the analysis of the rating, significant grammatical gender effects were found among French speakers but not among German speakers. Specifically, French participants perceived the ingredients in male fragrances with masculine descriptors as more congruent, whereas no such effect was observed for German participants. Speed and Majid (2019) interpreted these findings as evidence of grammatical gender effects on odour perception. However, the variation in results between German and French participants was attributed to the differences between German and French grammatical systems. Echoing the earlier discussion, Speed and Majid (2019) highlight the two- vs three-gender distinction between French and

German, as well as different gender transparency in the two languages. Furthermore, they noted differences in the frequency distribution of masculine and feminine nouns: masculine nouns are more prevalent in French, whereas German has a more balanced distribution (Hopp, 2013). These linguistic differences may explain the observed discrepancies in effects between-group.

Another line of research investigating grammatical gender effects extends into the realm of gender personification in art (Segel & Boroditsky, 2011). This approach offers a unique perspective by examining how grammatical gender influences real-world, non-linguistic behaviour over long periods, such as the creation of allegorical artwork. Segel and Boroditsky (2011) analysed a large database of artwork containing depicted allegories (included 790 images) from Italy, Germany, France, and Spain to investigate whether artists' L1 and its grammatical gender system predict the gender personification depicted in their artwork. Despite its innovative approach, the study has several limitations. Firstly, the groups of artwork (based on the artists native language) were uneven, with Italian group having 422 paintings, and the second biggest group (French artists) having 213 items. Besides, the analysis relied on the modern grammatical gender of each language. However, as shown by Boutonnet et al. (2012), this category can change diachronically. Nevertheless, the analysis showed that the gender depicted in the paintings matched artists' grammatical gender in 78% cases, with significant effects for both masculine and feminine genders. Interestingly, there were more female personifications overall, but the authors did not explore whether this bias might relate to the artists' gender - a variable excluded from the analysis. Furthermore, adding to the discussion on three- vs two-gendered languages, Segel and Boroditsky (2011) found significant effects of grammatical gender across all languages but with the different percentage of accuracy: 82% of personifications made by Italian artists were congruent with Italian gender, 80% in French, 62% for German and 54% for Spanish. The two latter numbers are particularly interesting considering the earlier claims regarding gender transparency across German and Spanish. However, the small Spanish sample (24 paintings) and the moderate German sample (129 paintings) likely contributed to this variability. Despite these limitations, the study represents a critical step in examining grammatical gender effects outside the laboratory, aligning with the call for "Whorf in the Wild" approaches (see Athanasopoulos & Bylund, 2020).

2.4.2. Language of the testing

The choice of language for testing participants is another critical factor when examining grammatical gender effects. A substantial body of research has shown that even when participants are tested in their genderless L2, the effects of their gendered L1 often remain unchanged (Bassetti & Nicoladis, 2016; Forbes et al., 2008; Nicoladis & Foursha-Stevenson, 2012 among others). However, certain studies, such as Sato et al. (2013), indicate that increasing proficiency in a genderless L2 may diminish the influence of a gendered L1.

Early studies presented in this literature, while showing the effects of gender on participants' categorisation, tested their participants in their gendered L1 (e.g., Flaherty, 2001; Konishi, 1993). However, as Samuel et al. (2019) argue, this approach may limit the effects of grammatical gender on conceptual representations within that L1, rather than on concepts overall. For example, if a speaker's L1 modulates their perception and creates an understanding of the concept "table" as masculine, this understanding should remain even when tested in their genderless L2. This issue has been raised by multiple researchers (Boroditsky & Schmidt, 2000; Sedlmeier et al., 2016), further underscoring the need to include bilingual and multilingual speakers in this research. Despite this, some recent studies also test participants in their L1, for instance in three-gendered Polish (Haertlé, 2017; Maciuszek et al., 2019).

In contrast, studies testing their participants in genderless L2 provide more contradictory results. Some research shows no effects of the language of the testing with gender effects present in both conditions (e.g., Forbes et al., 2008). In other cases, such as Pavlidou and Alvanoudi (2013), the effects of grammatical gender in a sex attribution task were found in both L1 (Greek) and L2 (English), but they were more pronounced in participants' L2.

For bilingual speakers with more than one gendered L1 (as in case of the current PhD research), using a genderless L2 offers unique advantages. As Sedlmeier et al. (2016) suggest, using a third language such as English for German and Spanish speakers provides a more suitable framework for investigating the influence of language on language-independent thought. Additionally, translating instructions into a third language could help mitigate subtle interpretive differences. However, if participants lack fluency in the L2, other issues, such as varying comprehension levels, could arise. This also highlights the need to examine participants' linguistic profile in detail, including the proficiency levels in the language of testing.

When comparing grammatical gender studies based on the testing language (gendered vs genderless), Samuel et al. (2019) found a slightly stronger support for language effects in studies that tested participants in a genderless language than in gendered (32% vs 29%), raising question about the "thinking for speaking" account. However, there was no clear confirmation that grammatical gender effects are less likely to emerge in gendered language of the testing, when analysing different tasks individually. For instance, while in sex assignment tasks more support was found in gendered languages than in genderless (75% vs. 0%), for studies replicated in this current PhD the trend was opposite: for similarity tasks Samuel et al. (2019) found 100% support in genderless testing language and 0% in gendered, for object-name memory tasks - 54% support in genderless and 0% in genderless.

These findings raise further questions about bilinguals with two gendered L1s: how do the effects of grammatical gender manifest when participants are tested in one versus the other? For instance, in Bassetti's (2007) study on Italian-German simultaneous bilinguals, participants were only tested in Italian, leaving it unclear whether similar findings would emerge if they were tested in their second L1. Given the complexity of choosing the language of the testing and to attest effects of grammatical gender on concepts overall, rather than concepts within their L1s, across all studies of the current PhD project, participants were instructed and tested in English (their L2). Additionally, to further minimise confounding effects, participants who spoke other gendered languages were excluded to prevent interference from additional grammatical structures.

2.4.3. Alternative explanations for the found grammatical gender effects

In their review, Samuel and colleagues (2019) analysed 158 studies on grammatical gender and object conceptualisation and found that out of these studies 38% offered support for relativity, 28% - offered mixed support (e.g., marginally significant results or evidence for one language/condition, but not the other, etc.) and 34% - no support of the effects of grammatical gender on object conceptualisation. However, what possible factors could account for the results presented in these studies? Particularly, which factors could have enhanced or diminished the effect size?

There have been several alternative explanations put forward to explain the present and/or absent findings in grammatical gender research. One of the most prominent alternatives is the assumption that grammatical gender in Romance languages, particularly Spanish and Italian, is universal and covaries with other masculine or feminine attributes (Koch et al., 2007; Sera et al., 2002; Vigliocco et al., 2005; cf. Sedlmeier et al., 2016). What could potentially contribute to this universality status? One of the things that add up in this theory is “femininity scores” (Sedlmeier et al., 2016; Section 2.3.3 of this review). It includes four two-valued elements (first value - female qualities and second - male): artifact vs natural, angular vs curved, used by males vs by females, and dense vs not dense (see Sedlmeier et al., 2016).

Another possible explanation is the long-standing discussion whether there is a difference in perception of different kind of nouns, such as animate vs inanimate, suggesting that the former may have more pronounced features that are typically considered as masculine or feminine which covaries with the findings (Sedlmeier et al., 2016; Vigliocco et al., 2005). The findings regarding animacy are quite contradicting and seem to be task dependent. For example, Samuel et al. (2019) note that there was a higher overall support of grammatical gender effects for animate stimuli (50%) than for inanimate (27%), and in some tasks, such as sex assignment task there was a 100% accuracy for animate targets and 33% for inanimate. However, when looking at the voice attribution paradigm they did not find any evidence that responses were more consistent for animate than to inanimate nouns (38% and 68% of responses consistent with grammatical gender respectively).

Cultural factors and extralinguistic differences between participant groups also need to be considered as potential confounds. To account for these confounds researchers could either carefully select a homogenous sample of participants (Sedlmeier et al., 2016) or alternatively, employ a training paradigm (Eberhard et al., 2005; Phillips & Boroditsky, 2003). For example, in Experiments 4 and 5, Phillips and Boroditsky (2003) taught English monolingual participants new distinctions ("*soupative*" and "*oosative*" for a set of inanimate objects of different conceptual genders and girls and boy respectively) in a fictional Gumbuzi language and found that participants were relying on newly learnt categories, also with and without verbal interference. This suggests that even when controlling for conceptual and cultural distinctions, grammatical structures still influenced participants.

Finally, even after accounting for all these potential confounds, there is still a chance that grammatical gender is used strategically. Looking at the data provided by Samuel et al. (2019), the rate of support of grammatical gender effects is higher in those tasks where gender/sex content is high, compared to low (51% and 2% supporting evidence, respectively). Multiple scholars highlighted (Everett, 2013; Sato & Athanasopoulos, 2018 among others) the need to investigate these effects in such experimental paradigms where grammatical gender manipulation in not overt, however, designing tasks free from grammatical gender associations is still a challenge. One of the tools to use here is EEG, as was done by Sato et al. (2020) and Boutonnet et al. (2012), as well as in Chapter 5 of this thesis. While behavioural measures can reveal overt language effects, they often fail to capture subtle, early-stage influences of language on cognition. Moreover, ERP evidence allows researchers to differentiate between automatic and strategic processes, enabling a deeper understanding of how grammatical gender shapes perception and categorisation. As emphasised by Thierry and colleagues (2024), this distinction is essential for advancing linguistic relativity research, as it moves beyond surface-level observations to uncover the underlying neural mechanisms driving these effects. Overall, incorporating EEG into grammatical gender studies offers a more comprehensive approach to understanding how language unconsciously influences thought, particularly in areas where behavioural measures alone may yield ambiguous or inconsistent results. This also provided a rationale for EEG into the current PhD project.

To sum up, studies discussed in this literature review presented ample evidence in favour of both covert and potentially overt effects of grammatical gender on cognitive and perceptual processes. However, as I discussed above it is also important to point out the studies that argued to have found a more constrained evidence in this domain of linguistic relativity, for instance studies suggesting mere language for speaking effects (Kousta et al., 2008; Ramos & and Roberson, 2011) or that these effects are due to the gender-marked articles in the language (Imai et al., 2014), or that grammatical gender effects are limited to certain semantic categories (e.g., sex and gender hypothesis by Vigliocco et al., 2005). Additionally, it has also been discussed whether the phrasing of the instructions matter, i.e., if there is an explicit reference to gender (see Bender et al., 2011, 2016 for further evidence) and whether the number of grammatical genders also plays a role when finding grammatical gender effects or lack of thereof (Koch et al., 2007; Sera et al., 2002; Vigliocco et al., 2005). Finally, additional research is needed to assess whether transparency of the gender marking in a language (Sera et al., 2002) or linguistic profile of participants (e.g., whether they are monolingual or bilingual, see Bassetti, 2007) play a role. In the next section I will present how the current PhD project aimed to provide more testing ground for some of these constraints in the experiments from Chapters 3-5.

2.5. The current study

The current PhD project investigates the linguistic relativity hypothesis, focusing specifically on the effects of grammatical gender on cognitive processes (i.e., categorisation, memory recall and early perceptual discrimination) in simultaneous bilinguals.

2.5.1. The relevance of studying simultaneous bilinguals

The inclusion of simultaneous bilinguals addresses a significant gap in the literature, as reviewed in Chapters 3–5 and highlighted throughout the current literature review. Therefore, this PhD project aims to contribute to the field by investigating whether, and to what extent, grammatical gender effects emerge among adult simultaneous bilinguals with two partially contrasting three-gendered grammatical gender systems (Ukrainian and Russian). By employing a diverse set of experimental tasks, the study seeks to provide new insights into the interplay between two grammatical systems in a bilingual mind when these grammatical systems have been acquired implicitly and from an early age.

One of the reasons for investigating simultaneous bilinguals is the fundamental distinction in how linguistic properties, such as grammatical gender, are learned, compared to sequential bilinguals. Unlike the latter group, who typically acquires their L2 explicitly through formal instruction and with a conscious awareness of grammatical rules, simultaneous bilinguals acquire both of their L1s implicitly and naturally (Lambelet, 2016; Schmidt, 1990). This difference could have significant implications for how grammatical systems are integrated and how they influence cognition. By studying simultaneous bilinguals, I aim to explore the effects of grammatical gender in a context where both systems are deeply rooted and equally naturalised, potentially yielding unique patterns of language-cognition interaction compared to those observed in sequential bilinguals. Besides, turning to the multi-competence theory (Cook, 2002), simultaneous bilinguals offer a valuable test case for understanding how two grammatical gender systems coexist within a single cognitive framework. This approach allows to explore how these systems interact and influence conceptual representations in adult bilinguals.

Moreover, the study of simultaneous bilinguals enables me to examine the role of language proficiency and language dominance in shaping cognitive processes. This is particularly important, because previous research on sequential bilinguals primarily relied on self-reported proficiency rather than on standardised proficiency testing in their L1 (see discussion on limitations of previous studies in Chen & Faitaki, 2024) and generally lacked detailed analysis of the language profile of their participants, such as reporting whether participants spoke or were in the process of learning another grammatically gendered language. The current research investigates whether one L1 exerts greater influence on cognition than the other, using both standardised language proficiency tests (Cambridge University Press, 2024; Oxford University Press, 2001; Ukrainian Center for Educational Quality Assessment, 2020) and self-reported language dominance questionnaire (Bilingual Language Profile, BLP, Gertken et al., 2014).

It is important to note that in Chapters 3 and 4, I report findings that focus exclusively on language proficiency in participants' 2L1s and exclude language dominance. This decision was driven by the context of data collection, which occurred immediately after the war in Ukraine. Given that language dominance (BLP questionnaire from Gertken et al., 2014) is a self-assessment tool that requires participants to rate such factors as their language attitudes (e.g., How important it is for you to sound like a native speaker of [*language*]? I feel like myself, when I speak [*language*].), it was decided to rely solely on standardised proficiency tests to avoid the risk of inaccurate reporting. However, in Chapter 5, which presents data from an experiment conducted in Spring 2024, both language dominance and proficiency were included as variables.

Finally, unlike previous studies, I approached bilingualism of our participants as a continuum, rather than merely categorising them as Ukrainian- or Russian-dominant/proficient. To achieve this, I calculated language proficiency and dominance coefficients by subtracting participants' Russian scores from their Ukrainian scores. This produced a scale ranging from -100 (indicating exclusive proficiency/dominance in Russian) to +100 (exclusive proficiency/dominance in Ukrainian). By doing so, it was possible to capture the nuanced degrees of language proficiency and dominance among our bilingual participants.

2.5.2. Focus on Ukrainian-Russian simultaneous bilinguals

This research centres on speakers of Ukrainian and Russian due to the following reasons. First, both languages are three-gendered, providing a unique opportunity to address ongoing debates about whether three-gendered grammatical systems influence cognitive processes at all (e.g., Forbes et al., 2008; Lambelet, 2016; Sera et al., 2002, see Section 2.3.4.). Second, the inclusion of these languages allows us to examine the effects of grammatical gender in languages with typological structures that differ significantly from the languages most commonly studied in linguistic relativity research, such as German and Greek (see description of Ukrainian and Russian typology in Chapter 3).

Another key contribution of this research is the systematic investigation of the role of neuter gender, a factor often overlooked in previous studies. Many studies on three-gendered languages have either included or excluded neuter-gender stimuli without discussing how this manipulation might influence participants' responses, particularly in tasks with a high degree of gender or sex salience (e.g., voice assignment tasks). To address this issue, we conducted two identical experiments in Chapter 3, differing only in the inclusion or exclusion of neuter-gender stimuli. This design allows us to observe whether neuter gender stimuli modulate participants' responses.

Overall, by focusing on Ukrainian-Russian simultaneous bilinguals, this research not only contributes to the broader understanding of grammatical gender effects in three-gendered systems but also addresses key methodological and typological questions that have shaped debates in the field. It offers a more nuanced perspective on how typological and structural factors may account for the variability observed in previous studies, advancing the discussion on the cognitive impact of grammatical gender in bilingual minds.

2.5.3. Research questions and selected stimuli

Across all three studies (Chapters 3–5), we examined two types of stimuli, chosen to reflect the partially contrasting nature of the gender systems in our participants’ languages: Ukrainian and Russian as two L1s, with English as their L2. The stimuli comprised nouns with (1) matching grammatical gender in Ukrainian and Russian, and (2) mismatching grammatical gender in Ukrainian and Russian.

While hypotheses and research questions differ across studies (based on the methodologies employed in each study, described below), generally we aimed to address the following two key research questions:

(a) Do three-gendered languages overall influence bilinguals’ cognitive processes when they are presented with stimuli that have matching grammatical gender in Ukrainian and Russian?

(b) Does the more proficient or dominant L1 affect bilinguals’ cognitive processes when they are presented with stimuli that have mismatching grammatical genders across the L1s?

To adhere with the principles of open science, all materials and selected stimuli are available on Open Science Framework. These can be accessed via the links provided in each chapter.

2.5.4. Justification for the selected tasks

As noted by Sedlmeier et al. (2016), research on the effects of grammatical gender is often challenged by various methodological issues, including differences in the languages spoken by participants, variations in the tasks used, discrepancies in levels of analysis, and the selection of appropriate stimuli. While the detailed methodology of each study is described in its respective chapter, here I will provide an overall justification of the selected methods.

The main interest in this PhD was to investigate whether there is an influence of grammatical gender on various cognitive processes:

categorisation and conceptual representations in Chapter 3, memory recall in Chapter 4, and early perceptual processes (i.e., N1, P2/VPP, and N300 components) in Chapter 5. Additionally, the choice of paradigms in Chapters 3 (similarity ratings) and Chapter 4 (object-name memory task) was due to the ambiguity of previous findings in these tasks. Specifically, they provided more evidence in favour of gender effects in speakers of two-gendered languages, compared to three-gendered: "support from two-gendered languages over three-gendered languages came from the similarity tasks (52% vs. 27%) and object–name association tasks (42% vs. 30%)" (Samuel et al., 2019, p. 1779).

Additional methodological strengths of our experimental design are outlined as follows:

(a) Ukrainian-Russian bilingual participants were tested in a genderless L2 (English) to avoid any priming of their L1s or language-on-language effects.

(b) Including English control group in Chapters 3 and 4 allowed us to examine not only whether there is an effect of matching grammatical genders in 2L1 or mismatching grammatical gender in 2L1 on cognitive processes, but also to compare whether such effects are observed at all when compared to the responses from speakers of a genderless language.

(c) A pre-test (described in detail in Chapters 3 and 5) ensured that the stimuli used in the experiments were conceptually neutral, minimising the possibility that participants were influenced by conceptual associations rather than grammatical gender itself.

(d) After each experiment, we implemented a detailed linguistic profile with both standardised and self-rated measures[1], for both Ukrainian-Russian bilinguals and English monolinguals. This ensured that English speakers in our control group did not speak or learnt any other gendered languages that might influence their responses. The same criterion was applied to the main group of Ukrainian-Russian bilingual speakers, excluding individuals who were learning or speaking additional gendered languages.

(e) Conducting a debriefing was implemented to identify whether participants were aware of the grammatical gender manipulation, allowing us to exclude these individuals from the analysis. This also allowed us to understand what strategies participants used when completing the tasks.

Additionally, the choice of the tasks helped us to further address the replication crisis, particularly lack of replications of findings presented in

---

[1] Studies in the current PhD used the Bilingual Language Profile (BLP; Gertken et al., 2014) to assess participants' language dominance. The BLP was chosen for its modular design, which allowed flexible comparisons across multiple languages and enabled us to assess whether participants showed dominance in any language beyond Ukrainian, Russian, or English. While standardised instruments such as the Language History Questionnaire (LHQ; Li et al., 2020) and the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018) offer composite profiling based on proficiency and use, their formats are less suited to capturing relative dominance across more than two languages. Future work may benefit from combining such tools to provide broader composite indices or to complement pairwise analyses.

certain seminal studies (e.g., Boroditsky & Schmidt, 2000; Phillips & Boroditsky, 2003). While we only conceptually replicated these studies, adapting them to our population group, our findings also help to provide an overall understanding whether those findings can be replicated.

Finally, as mentioned above, we conducted two experiments in Chapter 3, that only differed in the inclusion of neuter gender in stimuli. This allowed us to contribute to understanding of the nature of grammatical gender effects, particularly when investigating grammatical properties, such as grammatical gender? Section 2.2.3 of the current literature review described the proposed mechanisms that drive language effects, including those of grammatical gender (e.g., structural-feedback hypothesis by Sato and Athanasopoulos, 2018; "language as a spotlight" by Wolff and Holmes, 2011). Yet, there is a lack of empirical evidence to fully support or challenge either of these ideas. To address this gap, we added a neuter gender manipulation to the similarity judgement paradigm in Chapter 3. While in Experiment 1 of that Chapter, we keep nouns that belong to all three grammatical genders in both L1s as stimuli, in Experiment 2, we only add nouns that are either masculine or feminine in Ukrainian and Russian. This allowed us to test if the presence of neuter gender dilutes the salience of differences between masculine and feminine conceptualisation of objects and weakens the effects (or causes them to disappear altogether). Therefore, if the effects are online, we expect stronger effects in Experiment 2, as removing the neuter gender highlighted the binary masculine-feminine distinction in time of the task completion. However, if the effects are offline, results should be consistent across both experiments, as a result of long-term and habitual use of linguistic patterns rather than immediate context of the presented task (see detailed rationale in Chapter 3).

To sum up, this literature review outlined the historical development of the linguistic relativity hypothesis and examined evidence both supporting and challenging it across various domains. Additionally, the review also highlighted the importance of inclusion of bilingual and multilingual participants, and the need to expand the methods of investigation, including both behavioural and neural measures. Besides, it traced the development of research in grammatical gender effects on cognition, focusing on various linguistic profiles of participants and different methodologies available till now. Finally, I presented the rationale and explanation behind the choices made for the current PhD thesis in our attempt to investigate the effects of two partially contrasting grammatical gender structures in simultaneous bilinguals of Ukrainian and Russian.

Chapter 3. The Influence of Three-Gendered Grammatical Systems on Simultaneous Bilingual Cognition: The case of Ukrainian-Russian Bilinguals

Linking statement: the empirical studies in the current thesis are interconnected through a systematic reduction in the potential for strategic use of grammatical gender, as well as through variations in the saliency of gender cues across tasks. Each subsequent study builds on the findings of the previous one by refining the experimental paradigm to limit conscious gender-based strategies and to better isolate implicit gender effects. The aim of this first study was to set the stage for those that follow by demonstrating the presence of grammatical gender effects in simultaneous bilinguals of two contrasting three-gendered languages (Ukrainian and Russian). Notably, these effects were shown to be context-dependent, as evidenced by the differing outcomes of Experiments 1 and 2.

3.1. Abstract

This paper examines the linguistic relativity principle (Whorf, 1956) by investigating the impact of grammatical gender on cognition in simultaneous bilinguals of three-gendered Ukrainian and Russian. It examines whether speakers of three-gendered languages show grammatical gender effects on categorisation, empirically addressing claims that such effects are insignificant due to the presence of the neuter gender (Sera et al., 2002). We conducted two experiments using a similarity-judgment paradigm while manipulating the presence of neuter gender stimuli (Phillips & Boroditsky, 2003). Experiment 1, including neuter gender, revealed no significant effects, compatible with earlier studies on three-

gendered languages. Conversely, Experiment 2, excluding neuter gender stimuli, showed significant language effects. Bilingual participants rated pairs as more similar when grammatical genders in both languages were congruent with the biological sex of a character. Significant effects were also found for pairs with mismatching grammatical genders in Ukrainian and Russian. Participants with higher proficiency in Ukrainian rated pairs as more similar when the grammatical gender of a noun in Ukrainian was congruent with the character's biological sex, and incongruent in Russian. Our findings thus provide the first empirical demonstration that the exclusion of neuter gender online induces grammatical gender effects in speakers of three-gendered languages.

3.2. Introduction

The majority of studies investigating linguistic relativity effects typically concentrate on the question "Does language influence our thoughts?" . While this question has been asked in a number of disciplines, such as philosophy, linguistics, anthropology and psychology, modern versions of the question can be traced to and more recent trans-disciplinary scholarly activity (Athanasopoulos et al., 2016; Gentner & Goldin-Meadow, 2003; Lucy, 1997), which has placed the question at the forefront of cognitive science. Various domains have been used as a testbed for the hypothesis, such as spaciotemporal metaphors, colour (Athanasopoulos, 2009; Winawer et al., 2007), and grammatical gender (Boroditsky & Schmidt, 2000; Boutonnet et al., 2012; Sato & Athanasopoulos, 2018).

The latest surge of attention led to more detailed explanations of the effects languages may have on cognitive processes, by including various experimental conditions, such as verbal interference, differentiating stimuli based on their perceptual characteristics, or manipulating the complexity of experimental design (Athanasopoulos & Casaponsa, 2020). Therefore, posing the aforementioned question as one that requires a binary answer seems outdated. Instead, the focus is moving away from providing evidence to a “yes-no” question towards investigating what circumstances lead to emerging language effects on cognitive processes (e.g., memory or categorisation), as well as how and why language-specific features form the groundwork for individual perceptual judgement, including multilingual speakers (Bassetti & Filipović, 2022; Casasanto, 2016). An illustrative example of the latter in our study pertains to the emergence of grammatical gender effects in speakers of three-gendered languages. Previous research on linguistic relativity (Sera et al., 2002; Vigliocco et al., 2005) has reported the absence of such effects, while more recent studies yield mixed results (Pavlidou & Alvanoudi, 2019). The primary factor contributing to these mixed or non-emergent outcomes has been hypothesised to be the presence of the neuter gender in these languages, which is thought to diminish the prominence of gender effects. Consequently, our research seeks to determine whether grammatical gender effects on cognitive processes, such as categorisation, are confined to two-gendered languages or can also be observed in speakers of three-gendered languages, and under what specific conditions these effects manifest.

We also focus on bilingual speakers who have two partially conflicting grammatical systems (where some nouns have matching and others mismatching grammatical gender in Ukrainian and Russian). Specifically, the impact two grammatical gender systems have on perception and categorisation, even when participants are not actively engaging with either language, as the testing was conducted entirely in

English, which unlike Russian and Ukrainian does not have a grammatical gender system.

Generally, research on language and cognition in bilinguals continues to be an important endeavour of the linguistic relativity theory complex, as Whorf (1956) himself pointed out that if language affects our thoughts, then learning other languages can free people from the shackles of their own language. Employing Ukrainian-Russian simultaneous bilinguals is of interest because the representation of two grammatical gender systems within an individual's mind and their effects on bilinguals' cognitive processes, such as memory or categorisation, have received little attention (e.g., the study by Bassetti, 2007). It remains unclear whether language effects would emerge only when grammatical gender matches in both languages or if they would also occur when grammatical gender mismatches, depending on the more proficient language. Additionally, there is uncertainty whether any effects would appear at all, given that both languages include a neuter gender in their grammatical system.

Here, we attempt to investigate the effects that two partially contrasting three-gendered grammatical systems (e.g., Ukrainian as L1 and Russian as 2L1) have on categorisation, as well as introduce simultaneous bilinguals with two distinct grammatical gender systems into linguistic relativity research. In addition, at a theoretical level, we aim to explore whether the presence of neuter grammatical gender mitigates language effects, as suggested previously (Sera et al., 2002; Vigliocco et al., 2005). To do so we employed a similarity judgement paradigm while manipulating stimuli with (Experiment 1) and without neuter gender (Experiment 2). Such manipulation would also allow us to investigate further into the nature of the gender effects, particularly whether (if found) the effects of grammatical gender arise online (in the moment of testing) or offline (entrenched in previous language experience) (Lupyan et al., 2020). If the effects arose online (Lupyan, 2012; Sato & Athanasopoulos, 2018), we anticipated observing more pronounced effects in Experiment 2, whereas if the effects were offline, comparable effects were expected across both Experiment 1 and Experiment 2.

3.2.1. Grammatical gender in language and mind

The empirical evidence of linguistic relativity effects can be found across various domains, such as colour categorisation/discrimination (Athanasopoulos, 2009; Roberson et al., 2005; Winawer et al., 2007), time and space (Athanasopoulos & Bylund, 2023; Boroditsky, 2001; Casasanto et al., 2004), motion (Athanasopoulos & Bylund, 2013), grammatical number and object classification (Athanasopoulos, 2006; Lucy, 1992), tactile perception (Miller et al., 2018) and even olfaction (Cao et al., 2024; Speed & Majid, 2019; Vanek et al., 2021). This evidence supports the idea that the structure of language can shape non-linguistic cognition, offering a compelling testbed for investigating how grammatical features, such as gender, influence thought.

Grammatical gender has been used as a subject of analysis by linguistic relativity researchers because of two primary reasons. Firstly, when grammatical gender is absent, no other lexicalisation pattern can replace it (Boutonnet et al., 2012). Secondly, the assignment of grammatical gender to inanimate nouns, and certain animals in the case of Ukrainian and Russian, is usually unpredictable and semantically illogical (Elpers et al., 2022). For instance, “parrot” in Ukrainian takes the feminine grammatical gender, while in Russian it is masculine. Besides, even though grammatical gender is superfluous for interaction in the case of many languages (e.g., English), for speakers of various languages, such as Russian and Ukrainian, it cannot be ignored. In such languages the gender of objects is mandatorily marked in a range of morphosyntactic constructions, such as demonstratives, pronouns, singular adjectives, and verbs in the past tense (Mitrofanova et al., 2018). Such morphosyntactic consequences of grammatical gender make it an ideal candidate for examining whether grammatical categories influence cognitive processes beyond lexical features (Sato & Athanasopoulos, 2018).

Despite extensive research, a notable gap exists in understanding the cognitive effects of grammatical gender across different grammatical systems, particularly three-gendered languages. Most studies have focused on German (Bassetti, 2007; Pavlidou & Alvanoudi, 2019; Sera et al., 2002; Vigliocco et al., 2005), which may yield less significant results due to inconsistencies in gender assignment (e.g., "*das Mädchen*" [a girl] being neuter) and the use of articles that do not always differentiate between genders (e.g., the dative case where both masculine and neuter use "dem"). In contrast, Ukrainian and Russian, both three-gendered, indicate gender primarily through noun endings, providing a more consistent gender-marking system. By extending research to these underrepresented languages, this study aims to offer new insights into how three-gendered grammatical systems influence cognitive processes.

A wide range of behavioural tasks have been developed to study the impact of grammatical gender on cognitive representation of concepts, with the most common one being the voice attribution task (i.e., asking participants to assign either a male or female voice to objects; see Samuel et al., 2019). Other methods include a sex assignment task (Belacchi & Cubelli, 2012), an object-name memory task (Boroditsky & Schmidt, 2000), and a similarity judgement task (Phillips & Boroditsky, 2003). The current study employs the similarity judgment task, where participants rate the similarity between pairs of depicted objects and characters with a clear biological sex using a Likert scale. The choice of this paradigm is rooted in its unique strengths, such as it requires using unlabelled stimuli that minimise active language processing that is a key element in testing whether language shapes non-linguistic representations (Casasanto, 2016). This methodology was first implemented in linguistic relativity research in in the seminal work of Phillips & Boroditsky (2003), who argued that Spanish-English and German-English sequential bilinguals perceived object-personified character pairs as more similar when the biological sex of the character and the grammatical gender of the object in their L1 were congruent, even when tested in English. This suggests that grammatical gender influences object categorisation even when grammatical gender is not explicitly used. Overall, the research has shown that when making gender-related judgments, individuals often take into account the object's grammatical gender (Flaherty, 2001; Konishi, 1993). Despite more recent studies that produced contrasting results and highlighted the issue of a replication crisis, including a failed replication by Elpers et al. (2022) and mixed findings by Sedlmeier et al. (2016), the study by Phillips and Boroditsky (2003) has nonetheless made a significant impact on the field.

One possible explanation for the mixed findings might be linked to the type of grammatical gender system present in a language, particularly the distinction between two-gendered and three-gendered systems. For instance, Sera et al. (2002) found that, unlike Spanish and French monolingual children, German children did not use grammatical gender to assign voices to objects during categorisation tasks, instead aligning their responded more closely to Spanish gender. The study suggests that two-gendered languages have a stronger association between grammatical and natural gender, leading to overgeneralisation of masculine and feminine traits to inanimate objects. In contrast, speakers of languages with a three-gender system, such as German, appear to rely less on gender and more on other conceptual distinctions when categorising objects. Similarly, Vigliocco et al. (2005) found significant gender effects in Italian but not in German during a similarity judgment task, arguing that the weaker link between grammatical gender and semantic properties in three-gender systems results in reduced gender effects on perception. Inconsistencies in gender assignment and a lack of clear correspondence with the sex of referents likely contribute to this difference. The authors suggest that the mapping between grammatical gender and semantic properties is weaker in three-gender systems like German compared to two-gender systems like Italian. They argue that three-gendered languages do not exhibit the same grammatical gender effects because the correspondence between gender and the sex of referents is less transparent. To address these criticisms and further examine the role of grammatical gender in three-gender systems, Pavlidou and Alvanoudi (2019) conducted a sex-attribution task (adapted from Sera et al., 2002) with speakers of German and Greek (both three-gendered languages). Participants were asked to assign names to depicted nouns for a preschool play, with nouns having masculine, feminine, or neuter gender. Their analysis revealed significant effects of grammatical gender on sex-attribution in both languages, challenging earlier claims by Sera, et al. (2002) and Vigliocco, et al. (2005).

Similarly, Bassetti (2007) – the only study to our knowledge that examined simultaneous bilinguals when looking at grammatical gender effects in linguistic relativity research - investigated how grammatical gender influences categorisation and representations of concepts in Italian-German simultaneous bilingual and Italian monolingual children using a voice attribution task. This is particularly relevant to the current study because objects were also chosen with opposite genders in Italian and German. Results showed that grammatical gender effects were only present in Italian monolinguals, echoing Sera et al. (2002), suggesting that Italian gender assignment may be more intuitive or 'natural' compared to German. The study also noted that bilinguals, who navigate two languages with mismatched grammatical gender systems, develop unique cognitive frameworks, integrating elements from both languages. Consequently, bilinguals may think differently from monolinguals, not because of bilingualism itself, but due to the specific characteristics of the grammatical systems embedded in the languages they speak, such as mismatching grammatical genders in Italian and German. This observation is particularly relevant to our study, as we also examine partially mismatching grammatical gender systems, albeit within two three-gendered languages.

3.2.2. Online vs offline nature of the grammatical gender effects

A central question in this line of research is whether grammatical gender effects operate online (as real-time, context-sensitive influences) or offline (as enduring impacts of long-term linguistic experience). According to Lupyan et al. (2020), online effects occur when language actively modulates perception and decision-making in the moment, often shaped by top-down feedback from linguistic labels and grammatical structures. Offline effects, in contrast, reflect long-term, habitual patterns ingrained by extensive language use that influence perception even outside linguistic contexts.

This study draws on two complementary theoretical frameworks to address this distinction. The label-feedback hypothesis (Lupyan, 2012) proposes that even when no explicit labels are presented, internal labelling processes may still influence perception and categorisation in real time. This reflects a top-down influence, where prior language knowledge actively shapes what features are noticed or emphasised during perception. Extending this idea, the structural-feedback hypothesis (Sato & Athanasopoulos, 2018) posits that the influence of grammatical gender extends beyond specific labels, stemming from the broader habitual patterns ingrained by the grammatical system itself. According to this hypothesis, grammatical gender activates unconsciously during the online categorical perception and by doing so, it modulates perception by emphasising the features associated with it.

The current study aims to directly engage with the online vs offline debate by designing two similar experiments with the main difference being that Experiment 1 includes objects of all three grammatical genders (masculine, feminine, and neuter), while Experiment 2 excludes neuter gender. This allowed us to test whether the presence of neuter stimuli dilutes the salience of masculine-feminine distinctions, potentially weakening online grammatical gender effects. If the effects are online, we expect stronger effects in Experiment 2, as removing neuter gender heightens the binary masculine-feminine distinction. Conversely, if the effects are offline, results should remain consistent across both experiments, reflecting the enduring impact of long-term linguistic patterns rather than immediate task context.

To sum up, given the mixed results demonstrated in studies involving speakers of three-gendered languages, it is important to note that no previous research has directly compared the strength of grammatical gender effects using the same task with and without the inclusion of neuter gender. The present study uniquely investigates the cognitive effects of bilingualism in two conflicting three-gendered languages, a topic that has not been previously explored. Besides, we extend research beyond typically used German to other three-gendered languages (Ukrainian and Russian). This approach provides a more comprehensive understanding of how grammatical gender influences cognition across diverse linguistic contexts.

3.2.3. Case of Ukrainian simultaneous bilingualism and typological differences in Ukrainian and Russian languages

Simultaneous bilingualism in Ukraine presents unique challenges and insights into the cognitive processing of language, particularly when the languages involved have distinct grammatical systems. This study focuses on Ukrainian-Russian bilingualism, specifically the typological differences between the languages, especially regarding grammatical gender.

Ukraine has a deep-rooted history of multilingualism (Poftak & Shykula, 2022), and the status of the Russian language has long been a subject of debate (Eberhard et al., 2019). According to the 2001 census, out of Ukraine's then-population of 48.5 million, 78% identified as Ukrainians and 17% identified as Russians when asked to choose one ethnic affiliation. However, linguistic preferences differed, with 68% selecting Ukrainian as their native language and 30% opting for Russian (Bilaniuk & Melnyk, 2008). Despite the historical stigmatisation of bilingualism even prior to the war (Pavlenko, 2012), it is clear that societal bilingualism is inherent in Ukraine (Csernicskó & Máté, 2017; Shumlianskyi, 2010).

The onset of the war in February 2022 dramatically altered these linguistic landscapes. There has been a sharp increase in the proportion of respondents who, according to self-reported questionnaires, speak predominantly Ukrainian in everyday life and a corresponding decrease of

Russian speakers. The most recent poll from December 2022 indicates that 41% of respondents claimed to communicate only in Ukrainian, another 17% reported using Ukrainian "in most situations", while only 6% speak only in Russian, and 9% predominantly in Russian, another 24% said they use both languages "equally" (Kulyk, 2023). Compared to 2017, the proportion of exclusive and predominant Ukrainian speakers increased by 8%, and the proportion of Russian speakers decreased by 11% (Kulyk, 2023). Given the fluid language attitudes and shifting language use among bilingual individuals in Ukraine, it is worth examining which languages have the most significant impact on cognitive processes of such speakers. It has been proposed in linguistic relativity research that language effects are found for the dominant native language, rather than for the second language (Bassetti, 2007; Phillips & Boroditsky, 2003). However, these assessments often relied on participants' self-evaluations of their language dominance and language proficiency. To address this issue, the current study includes proficiency tests for English (language of testing), Ukrainian, and Russian, as well as a self-rated Bilingual Linguistic Profile (BLP, Gertken et al., 2014) to comprehensively assess the proficiency differences.

Typologically, the two languages are linguistic cousins, both belonging to the East Slavic branch of the Indo-European language family (Kortmann & Auwera, 2011), which shares significant historical, lexical, and grammatical similarities. They have a considerable overlap in vocabulary, grammar, and pronunciation characteristics, setting them apart from other Slavonic languages. Various studies indicate that Ukrainian and Russian share about 55%-62% of their vocabulary, a lexical distance akin to that between Portuguese and French (Steinback, 2015). Like other Indo-European languages, Ukrainian and Russian incorporate grammatical gender, categorising nouns as feminine, masculine, or neuter. These languages are highly inflectional with overt gender systems, where gender influences noun declension and adjective endings (Budzhak-Jones, 1997). In Ukrainian, nouns are divided into three genders, with syntactic agreement indicating gender, except for invariably gender-neutral plural nouns (Rusanivskyj et al., 2004). Russian follows a similar division, but with an uneven distribution: 46% of nouns are masculine, 41% feminine, and 13% neuter. The masculine gender, being most prevalent, is often considered the default (Corbett, 1991; 2007). The lack of extensive research on Ukrainian gender distribution leaves the question of whether it follows a similar pattern open.

Although nouns in Ukrainian and Russian neither change according to genders nor have gendered articles, grammatical gender affects the declension of nouns and endings in both languages. In Ukrainian language, masculine gendered animate and inanimate nouns typically have consonant endings (e.g., дім [dim] – house), while feminine gender is predicted by -а / - я endings (e.g., кава [kava] – coffee, історія [istoriia] – history). Most abstract nouns are feminine (Pugh & Press, 1999), regardless of the ending (e.g., радість [radist'] – joy, тиша [tysha] – quiet). Neuter nouns have three possible endings: -о , -е, -ння / -ття (дерево [derevo] – tree, сонце [sontse]– sun, кохання [kokhannia] – love) (Bezpoiasko et al., 1993; Gorpynyč, 2004).

Similarly, in Russian, endings of nouns suggest their grammatical gender: masculine nouns end with a consonant or -й, feminine nouns end with -а or -я, while neuter nouns have -о / -е endings. There is also a large number of exceptions, such as nouns ending with a soft sign -ь, that can refer either to masculine or feminine nouns. In both languages, grammatical gender is semantically and morphologically assigned, affecting adjectives, pronouns, and determiners (Basova et al., 2003), and is a mandatory feature for nouns except in plural forms (Gorpynyč, 2004). The described grammatical gender distribution in Ukrainian and Russian provides a well-suited setting for investigating grammatical gender effects on cognitive processes. It presents an opportunity to go beyond investigating a three-gendered grammatical system but analysing language effects when 2L1s have contrasting three-gendered systems.

3.3. Aims and the scope of the current study

This study aims to shed light on what (if any) effects two partially contrasting three-gendered grammatical systems have on cognitive processes of simultaneous bilinguals. While research has examined the impact of single three-gendered systems (Konishi, 1993; Pavlidou & Alvanoudi, 2019; Sera et al., 2002), little is known about the cognitive implications of simultaneously acquiring two languages with differing grammatical features (Bassetti, 2007). We hypothesised that simultaneous bilinguals would demonstrate a language effect similar to that of sequential bilinguals – specifically, they would demonstrate the influence of grammatical gender on categorisation, despite prior research suggesting that gender effects are limited to speakers of two-gendered languages because the binary nature of the system makes grammatical gender more salient (Sera et al., 2002). We expect to observe a grammatical gender effect, by employing a more rigorous stimuli design encompassing grammatical genders both matching and mismatching across languages, coupled with the inclusion of languages where grammatical gender is manifested through diverse grammatical features rather than articles. Additionally, we aim to investigate whether the presence of neuter gender in the stimuli (Experiment 1) would affect the observed grammatical gender effects, compared to Experiment 2, where it was absent. If grammatical gender effects have an online nature, as shown in previous studies (Sato & Athanasopoulos, 2018), we would expect stronger effects in Experiment 2, compared to Experiment 1, as the absence of neuter gender would amplify the contrast between masculine and feminine gender, enhancing the observed effects in the real-time of task completion.

To investigate our hypothesis, we adapted a similarity judgment paradigm where participants rated the similarity of pairs of stimuli, comprising depicted conceptually neutral nouns (e.g., a notebook), presented alongside a picture of a male or female character (e.g., a ballerina) on a 9-point Likert scale (Phillips & Boroditsky, 2003). The tasks in both experiments were conducted in English (starting with the participant's information sheet in the first email until debriefing). This was done to prevent participant from actively using either of their L1s. The current paradigm was chosen for several reasons. Firstly, it has been used many times, yielding mixed results with speakers of three-gendered languages. However, it has never been used to our knowledge with a three-gendered language omitting the neuter gender as presented in Experiment 2. Using the same task ensures that any effects observed can be attributed to our experimental manipulation rather than any potential confounds of the task itself. Secondly, it was employed due to the high salience of gender/sex in the task (Samuel et al., 2019), laying the groundwork for subsequent exploration of more subtle, implicit effects of gender on cognitive processes.

Experiment 1 aims to provide initial understanding of the grammatical gender effects of Ukrainian and Russian on categorisation, in contrast to English monolingual controls. In the first part of this experiment, we look at the interaction between group (Ukrainian-Russian bilinguals vs English monolinguals) and condition (whether the noun's grammatical gender matches or mismatches the character's biological sex) and whether it had any influence on similarity ratings (Likert scores). Here we anticipate that Ukrainian-Russian bilinguals will show stronger effects of condition on the similarity ratings compared to English monolinguals. The stimuli include nouns with matching grammatical genders in Ukrainian and Russian (e.g., "pencil"- masculine in both, "candle" – feminine in both, "tree" - neutral in both). Confirming this prediction would reaffirm the original findings by Phillips & Boroditsky (2003) and demonstrate that presence of neuter gender does not negate the language effects. In the second part, when looking at the results of the bilingual group only, we analyse ratings based on participants' most proficient language (Ukrainian or Russian). Stimuli were chosen to include noun-character pairs with contrasting grammatical genders in Ukrainian and Russian languages (e.g., "a basket" – masculine in Ukrainian, feminine in Russian – paired with a ballerina (female character); "an iron" – masculine in Russian, feminine in Ukrainian – paired with a king (male character)). We predict that bilinguals will rate pairs as more similar when the grammatical gender of the object (masculine or feminine) in their more proficient language is congruent with the character's biological sex (male or female).

Experiment 2 contains only masculine and feminine nouns, investigating whether excluding neuter gender strengthens the grammatical gender effects. The manipulation here directly addresses a central question in the field regarding the possibility that the presence of neuter gender impairs language effects. The question is whether this happens at a general or a local level. In other words, does the presence of the neuter gender in the grammatical system of a language attenuates effects of gender on categorisation across the board, or are such attenuating effects only observable when the neuter gender is used as part of the similarity judgments that participants are asked to perform. Similar to Experiment 1, we anticipate to find grammatical gender effects on similarity ratings in the Ukrainian-Russian bilingual group but not in the English monolingual group. Within the Ukrainian group, the impact of language proficiency on ratings is also explored.

Overall, we expect to find a significant effect of grammatical gender on categorisation of simultaneous bilinguals, irrespective of the contrasting three-gendered systems of Ukrainian and Russian. The outcomes of this study are expected to highlight the influence grammatical gender has on cognitive processes, shedding more light on how complex and contrasting linguistic systems shape human cognition.

3.4. Method

Materials and analysis codes can be found on the Open Science Framework (OSF): https://osf.io/3xgaw/.

3.4.1. Experiment 1

3.4.1.1. Participants

63 Ukrainian-Russian simultaneous bilingual speakers (with English as a foreign language) and 37 English monolingual speakers completed the study online in exchange for time compensation in a form of a £10 Amazon voucher. After examining their linguistic profiles and responses, 51 Ukrainian-Russian bilinguals (48 females; $Mean_{age} = 32$, $SD = 10$) and 24 English monolinguals (9 females; $Mean_{age} = 30$, $SD = 13$) were included in the analysis. Exclusion criteria encompassed speaking other gendered

languages ($n = 22$) or consistently selecting a ‘1’ rating on the Likert scale, indicating inattention to instructions or lack of engagement ($n = 3$). Among the bilingual group, 66.7% ($n = 34$) had a postgraduate degree, 23.5% ($n = 12$) had an undergraduate degree, 2% ($n = 1$) had a college degree, and 7.8% ($n = 4$) had high school education or less. In contrast, among the monolingual group, 50% of participants ($n = 12$) had a postgraduate degree, 25% ($n = 6$) had an undergraduate degree, and 25% of participants ($n = 6$) had a college degree, with no participants having only finished high school.

The bilingual participants proficiency in Ukrainian, Russian, and English was assessed using standardised language tests. For Ukrainian and Russian, advanced ZNO Tests (External Independent Assessment) were used (Ukrainian Centre for Educational Quality Assessment, 2020). These standardised university entrance examinations evaluate participants' language skills up to the C2 proficiency level, thereby mitigating potential ceiling effects of L1 proficiency in our study. English proficiency was determined through the Oxford Quick Placement Test (Oxford University Press, 2001) or existing IELTS certification (Cambridge University Press, 2021). Acceptable scores were set at 67% for the OQPT and 5.5 for the IELTS, both equivalent to the B2 (Upper-Intermediate) level. ZNO tests classify Ukrainian and Russian proficiency levels between C1 (advanced) and C2 (proficient).

The bilingual participants reported an average age of 8.68 years ($SD = 3.21$) for acquiring English as a foreign language (L2), with a minimum proficiency level of Upper-Intermediate. The majority of participants demonstrated higher proficiency scores in Ukrainian (57.38%, $n = 29$), as opposed to Russian (22.95%, $n = 12$), or equal proficiency in both (19.67%, $n = 10$). The proficiency scores ranged widely, indicating no ceiling effects (see Table 3.1).

Table 3.1

Proficiency Scores and Distribution of Ukrainian-Russian Bilingual Participants in Experiment 1

Language	Mean Proficiency Score (100 maximum)	SD	Range	Percentage (Number) of Participants
Ukrainian	65.68	18.39	18.75 - 93.75	57.38% (29)
Russian	59.84	14.90	25.00 - 87.50	22.95% (12)
Equal proficiency in both	57.29	13.55	37.50 - 81.25	19.67% (10)

Participants completed the study online, after being recruited through social media or through posters at Lancaster University. The gender imbalance in bilingual participants, predominantly female, resulted from the data collection occurring after the onset of the war in Ukraine. However, as Flaherty (2001) notes, such a discrepancy in participants' gender is unlikely to significantly affect the responses. Besides, we used separate cumulative link mixed models for each experiment to investigate whether there was an effect of participants' gender (see Supplementary materials in Appendix A for full analysis and results). However, the absence of a significant three-way interaction between group (Ukrainian-Russian bilingual vs English monolingual), participant's gender (male vs female), and grammatical gender (masculine vs feminine vs neuter) suggested that the gender imbalance in the bilingual group did not appear to disproportionately affect the main findings of the study.

3.4.1.2. Materials

Pre-test. A pre-test was conducted to select conceptually gender-neutral items for the main experiment, following the approach of Sato & Athanasopoulos (2018). Ten Ukrainian-Russian-English speakers (5 females; *Mean age* = 26, *SD* = 4) and ten English monolinguals (4 females; *Mean age* = 31, *SD* = 10) were recruited. None of the participants took part in the main study. Participants were shown 137 black-and-white object images one by one and asked to rate each picture on a 7-point Likert scale ranging from "*very feminine*" (1) to "*very masculine*" (7). The objects were divided into five groups based on their grammatical genders in Ukrainian and Russian: (1) 20 nouns masculine in Russian and feminine in Ukrainian, (2) 24 nouns feminine in Russian and masculine in Ukrainian, (3) 31 nouns feminine in both languages, (4) 31 nouns masculine in both languages, and (5) 31 nouns neutral in both languages. All images, presented against a greyscale and white background to avoid colour biases, were sourced from the Bank of Standardised Stimuli (Brodeur et al., 2014).

The pre-test yielded 50 conceptually neutral items (*Mean* = 4.01*; SD* = 0.13), which were then divided into the five categories (see Table 3.2): (1) nouns with masculine grammatical gender in both Russian and Ukrainian languages, (2) feminine grammatical gender in both Russian and Ukrainian, (3) feminine in Russian, masculine in Ukrainian, (4) feminine in Ukrainian and masculine in Russian, and (5) neutral in both. A slight imbalance between stimuli (3) and (4) is not anticipated to impact our results, as they will be analysed collectively. This will yield a total of 20 nouns with matching grammatical gender in both languages, 20 nouns with mismatching grammatical gender, and 10 neuter fillers.

Table 3.2

Example of stimuli used for both Experiment 1 and 2

Type of stimuli	Example (Russian)	Example (Ukrainian)	English Translation	Number of Items
Masculine in both Russian and Ukrainian	миндаль (mindal)	мигдаль (myhdal)	almond	10
Feminine in both Russian and Ukrainian	свечка (svechka)	свічка (svichka)	candle	10
Feminine in Russian, Masculine in Ukrainian	лодка (lodka)	човен (choven)	boat	8
Feminine in Ukrainian, Masculine in Russian	муравей (muravei)	мураха (murakha)	ant	12
Neuter in both languages (Experiment 1 only)	яблоко (yabloko)	яблуко (yabluko)	apple	10

Main testing. In the main experiment, participants were presented with a hundred pairs, each consisting of one of the 50 selected conceptually neutral unlabelled black-and-white objects and one of the16 characters: 8 female images (a queen, a bride, a witch, a smurf, a ballerina, a girl, a pensioner, an ogre) and 8 male images (a king, a groom, a giant, a smurf, an architect, a boy, a man, an ogre). Each depicted noun was presented once with a male character and once with a female character, resulting in 100 pairs. Pairs were presented in a randomised order. Each participant had to provide a similarity rating on the Likert scale from 1 (*"not similar"*) to 9 (*"very similar"*) with each pair displaying the object on the left and the character on the right of the screen.

3.4.1.3. Procedure and design

To conduct the experiment, we utilised the Gorilla Experiment Builder software. Upon registration, participants received an introductory email containing the participant information sheet and a link to the experiment.

After signing a consent form, they were redirected to the main task, which they accessed on their personal laptops or computers.

Both groups undertook the same experimental task in English. The instructions were similar to those from Phillips and Boroditsky (2003, p. 929): “In this study, you will see pairs of pictures appear on the screen. In each pair, there will be a picture of a person on the left and a picture of an object or animal on the right. You will see a scale where 1 = not similar and 9 = very similar. For each pair of pictures, please choose a number between 1 and 9 to indicate how similar you think the two pictures are. Try to use the whole scale (give some 1’s and some 9’s and some of all the numbers in-between). Please respond with the first answer that comes to mind”.

Each object-person pair remained on the screen until participants selected “Next”. Once they moved on to the next pair, they could not change their answer. After completing the task, participants were asked what criteria were used to rate the pairs to determine whether they detected the experiment’s aim and used grammatical gender as a task-solving strategy. None of the participants reported reliance on grammatical gender or language in general. Instead, responses were reported to be influenced by associations with films or cartoons, shapes, or random guesses. Ukrainian-Russian bilingual participants then completed a Bilingual Language Profile (BLP, Gertken et al., 2014) questionnaire and two proficiency tests (Oxford University Press, 2001; Ukrainian Center for Educational Quality Assessment, 2020). The monolingual group only completed the BLP to identify any gendered language knowledge potentially affecting results. Additionally, we monitored the real-time completion of the experiment. In those instances where participants substantially exceeded the expected average response times or stopped during the task, their participation was manually excluded (6 bilingual and

9 monolingual participants), given the importance of capturing responses on the first-impression basis.

3.4.1.4. Analysis

For each experiment, data analysis involved cumulative link mixed models in RStudio (version 2022.07.22, R Core Team, 2022), using the ordinal package (Christensen, 2019), with similarity ratings as the dependent variable. Previous study that replicated the original experiment by Phillips and Boroditsky (2003) employed linear mixed-effects models (Elpers et al., 2022), highlighting their advantages, such as incorporating both fixed and random effects and analyzing non-averaged data (Baayen et al., 2008; Vasishth & Broe, 2011). However, as the analysis includes Likert scale and ordinal data, we used cumulative link mixed models (CLMMs) instead. Similarly to linear mixed effects models, CLMMs also accommodate multiple sources of error variance as random variables, such as participant variability and the gender of depicted characters (Bross, 2019). Yet, CLMMs are more suited for analysing ordinal data, as they account for possibility of varying distances between levels of the rating scale (Ackerman, 2018).

We divided the analysis into two parts. The first part involved a comparative analysis of responses from both Ukrainian-Russian bilingual and English monolingual participants. We focused on how the interaction between grammatical gender congruence of the pairs (grammatical gender of the object was congruent or incongruent in both Russian and Ukrainian with the biological sex of the character) and the participant group (Ukrainian-Russian or English) influenced the Likert scores. The maximal model that converged included random intercepts for participants and items. The detailed analysis is available on OSF (https://osf.io/3xgaw/).

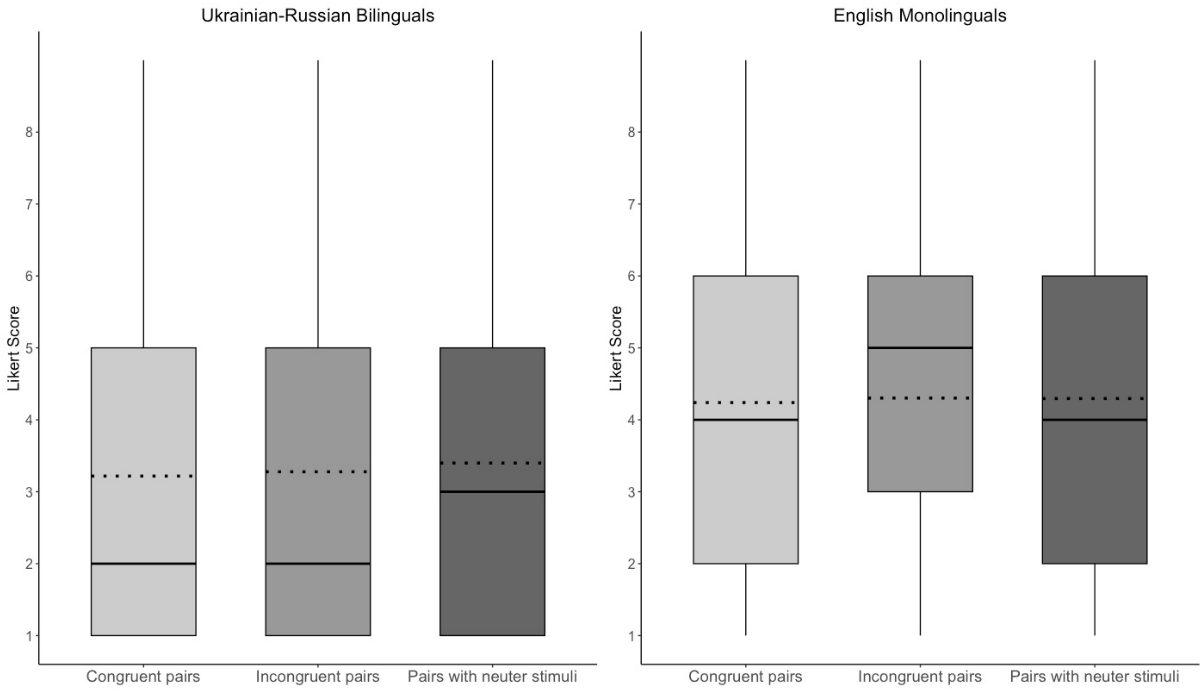
Secondly, to investigate deeper the effects of two contrasting three-gendered languages, we conducted an analysis comparing Ukrainian-Russian bilinguals only, based on their most proficient language. In the current study, we approached bilingualism as a continuum and measured it as a continuous variable by subtracting Russian proficiency from Ukrainian proficiency scores, resulting with the scale -100 being only proficient in Russian and +100 only proficient in Ukrainian. Participants with equal proficiency scores were included in the analysis with the coefficient score 0. Here, we examined how the congruence of an object's grammatical gender in L1 with the character's biological sex (and its incongruence in 2L1) interacted with language proficiency to affect similarity ratings. A maximal model in this part also included random intercepts for both participants and items.

3.4.1.5. Results

Comparing the Ukrainian-Russian bilingual and English monolingual participants. In this analysis, we included stimuli where the grammatical gender of nouns was either congruent or incongruent with the character's biological sex in both Ukrainian and Russian. An example of this would be "a ballerina" (female) and "a pen" (feminine in both Ukrainian and Russian) or "a ballerina" and "an almond" (masculine in both). Our expectation was that Ukrainian-Russian bilinguals would show stronger grammatical gender effects compared to English monolinguals. Specifically, we predicted that congruent pairs, where the character's biological sex is congruent with the object label's grammatical gender in 2L1s, would receive higher similarity ratings. For instance, Ukrainian-Russian bilinguals were anticipated to rate a congruent pair, such as "a ballerina" and "a pen", as more similar than incongruent pairs like "a king" and “a pen”. English monolinguals were not expected to show any significant trends.

Figure 3.1

Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent, incongruent, and neuter stimuli pairs in Experiment 1



Comparing the mean responses of Ukrainian-Russian bilingual participants in the congruent (*Mean* = 3.22, *SD* = 2.41) and incongruent (*Mean* = 3.28, *SD* = 2.41) pairs revealed nearly identical ratings, contrary to our predictions (see Figure 3.1). Notably, bilingual participants displayed slightly higher, but not statistically significant, average responses for stimuli with neuter grammatical gender (*Mean* = 4.00, *SD* = 2.55). In contrast, English monolingual participants consistently assigned similar ratings across all conditions (congruent: *Mean* = 4.24, *SD* = 2.38; incongruent: *Mean* = 4.30, *SD* = 2.17; neuter: *Mean* = 4.30, *SD* = 2.38), indicating that condition type did not notably influence their judgments of object-character similarity.

We built a cumulative link mixed model (clmm) to compare two groups of participants looking at the interaction between the group (Ukrainian-Russian bilinguals vs English monolinguals) and pair congruency in both L1s (congruent vs incongruent vs neutral), as a predictor for similarity ratings (Likert scores). Random intercepts were included for participants and items to account for variations specific to each.

The results revealed a statistically significant Group effect, with Ukrainian-Russian bilinguals exhibited lower similarity ratings compared to the English controls (SE = 0.3318, z = -2.771, p = 0.006). However, there were no statistically significant main effects for pair congruency (SE = 0.2194, z = 0.165, p = 0.869) or for the interaction between the two variables. Specifically, the lack of significant group - condition interaction (SE = 0.1376, z = 0.888, p = 0.3744) demonstrated that, in contrast to our hypothesis, Ukrainian-Russian bilinguals did not rate incongruent pairs as less similar compared to the congruent pairs[2].

---

[2] A Bayesian paired-samples t-test was conducted to assess whether Ukrainian-Russian bilinguals rated incongruent pairs as less similar than congruent pairs. The test yielded a Bayes Factor of BF_{01} = 5.65, indicating that the data were approximately 5.7 times more likely under the null hypothesis than under the alternative. This provides evidence in favour of the null hypothesis, suggesting that there is no meaningful difference in similarity ratings between the pairs within the Ukrainian-Russian bilingual group. This finding aligns with the

Comparing Ukrainian-Russian simultaneous bilinguals based on the Language proficiency in L1 and 2L1. To compare the results of Ukrainian-Russian bilinguals only and the investigate the effect of the more proficient first language (L1 or 2L1) on similarity ratings, we conducted a separate analysis with different stimuli. This included noun pairs where grammatical gender matched the character's biological sex in one language but not the other. For example, "a queen" and "an onion" (masculine in Russian, feminine in Ukrainian) were congruent in Ukrainian but incongruent in Russian. Conversely, "a king" and "a sock" (feminine in Ukrainian, masculine in Russian) were congruent in Russian and incongruent in Ukrainian.

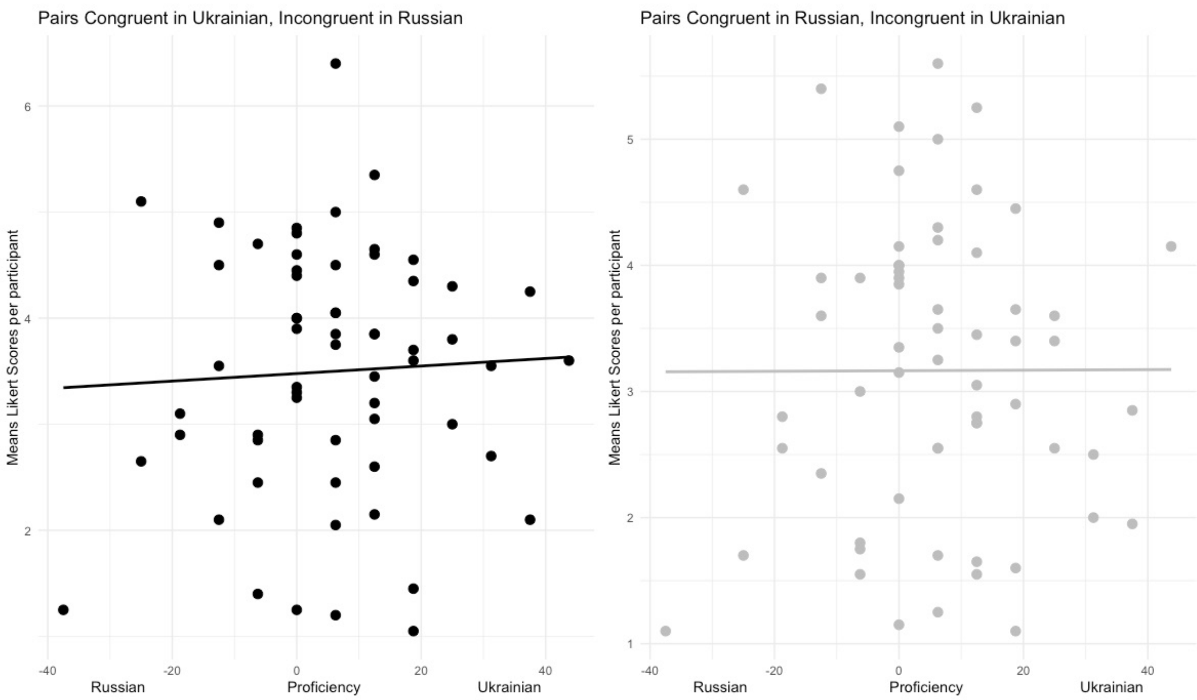
Ukrainian-Russian bilinguals assigned ratings to pairs congruent in Ukrainian (*Mean* = 3.50, *Range* = 2.92 - 4.08) and pairs congruent in Russian (*Mean* = 3.17, *Range* = 2.59 - 3.74) when their proficiency was higher in Ukrainian (Figure 3.2). However, the differences in ratings were minimal and statistically non-significant, against our expectations.

---

non-significant frequentist result ($t(60) = -0.686$, $p = 0.495$, 95% *CI* [-0.23, 0.11]), further supporting the conclusion that bilingual participants did not show differential sensitivity to pair congruency.

Figure 3.2

Mean Likert Scale Responses from Experiment 1 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent



In the second cumulative link mixed model, we explored whether Likert scores were influenced by the interaction between condition (biological sex and grammatical gender congruent in Ukrainian and incongruent in Russian vs congruent in Russian and incongruent in Ukrainian) and language proficiency (-100 to 100, with -100 being only Proficient in Russian, to 100 – only proficient in Ukrainian). The maximum convergence model included random intercepts for participants and items to account for participant-specific and item-specific variations. Contrary to our predictions, we found no significant effects for the condition-proficiency interaction (Condition "Congruent in Ukrainian, incongruent in Russian" as a reference level*: Estimate = -0.004, SE* = 0.005, z = -0.769, p = 0.442*;* "Congruent in Russian, incongruent in Ukrainian" as a reference level: *Estimate =0.004, SE* = 0.005, z = 0.769, p = 0.442), demonstrating that bilingual participants with higher proficiency in Russian did not assign higher ratings to the pairs that were congruent in Russian and incongruent in Ukrainian. Furthermore, no significant main effects for condition (*SE* = 0.3241, z = -0.741, p = 0.459) or language proficiency (*SE* = 0.0104, z = 0.725, p = 0.468) were found. Overall, our findings for the stimuli with mismatching grammatical gender in 2L1s suggest that neither the individual variables nor their interaction significantly contributed to participants' similarity ratings.

In summary, Experiment 1 revealed that gender congruence of noun-character pairs had no statistically significant impact on similarity ratings. Moreover, an unexpected pattern emerged, as Ukrainian-Russian bilinguals consistently rated objects as less similar than their English monolingual counterparts across all conditions. Our findings in this experiment align with the claims by Sera et al. (2002) that the presence of neuter grammatical gender may negate grammatical gender effects in speakers of three-gendered languages.

3.4.2. Experiment 2

The results from Experiment 1 suggest that including a neutral gender may have mitigated the significance of the language effect by diminishing the salience of grammatical gender. This raised the possibility that excluding neutral gender from the study design could affect the findings, particularly if the grammatical gender effects are online in nature and arise from real-time language effects. Therefore, in this study, we largely retained the methodology used in Experiment 1 but excluded the neuter gender from the stimuli.

3.4.2.1. Participants

40 English monolinguals and 70 Ukrainian-Russian bilinguals were recruited. After analysing their linguistic profile and responses, 64 bilinguals (44 females; $Mean_{\text{age}} = 30$, $SD = 12$) and 34 monolinguals (18 females; $Mean_{\text{age}} = 26$, $SD = 6$) were included in the analysis. Exclusions were due to participants either knowing other gendered languages ($n = 6$) or consistently using a single value on the Likert scale ($n = 6$), suggesting a potential lack of engagement or failure to follow instructions. The demographic distribution of the bilingual group in Experiment 2 was consistent with that of Experiment 1. As in the previous experiment, the largest proportion of bilingual participants held postgraduate degrees: 42.2% ($n = 27$). This was followed by 31.3% ($n = 20$) with undergraduate degrees, 18.8% (n = 12) with a high school diploma, and 7.8% ($n = 5$) with a college degree. For the monolingual group, the distribution shifted slightly from Experiment 1. While postgraduate degrees remained the most common (35.3%, $n = 12$), the proportions for college and undergraduate degrees changed. In Experiment 1, college and undergraduate diplomas were equally represented, but in Experiment 2, 32.4% ($n = 11$) had a college diploma, 23.5% ($n = 8$) held an undergraduate degree, and 8.8% ($n= 3$) had a high school education. Similarly to Experiment 1, no effects of participants' gender on their ratings were found (see Tables A.4 and A.5 in Supplementary materials in Appendix A).

Analogously to the first experiment, we assessed bilingual participants' linguistic profiles and proficiency of Ukrainian, Russian, and

English. Participants were recruited online and via posters at Lancaster University. The bilingual participants reported acquiring English (L2) at an average age of 9 years (*Range* = 4-20) and had at least an upper-intermediate proficiency level. Among them, 72% of participants demonstrated higher proficiency scores in Ukrainian and 28% in Russian. None of the participants reported using grammatical gender as a conscious strategy. The proficiency scores varied widely (see Table 3.3), demonstrating that ceiling effects were absent.

Table 3.3

Proficiency Scores of Ukrainian-Russian bilinguals in Experiment 2

Language	Mean Proficiency Score (100 maximum)	SD	Range	Percentage (Number) of Participants
Ukrainian	65.2	19.3	12.5 – 100	72% (46)
Russian	51.7	12.7	25 – 81.2	17% (11)
Equal proficiency in both	55.4	8.41	43.8 - 68.8	11% (7)

3.4.2.2. Materials

As with Experiment 1, participants were asked to rate object-character pairs using a 1 (*"not similar"*) to 9 (*"very similar"*) Likert scale. The stimuli consisted of 40 conceptually neutral black-and-white objects, categorised as follows: 10 masculine in both Russian and Ukrainian, 10 feminine in both languages, 8 feminine in Russian but masculine in Ukrainian, and 12 feminine in Ukrainian but masculine in Russian. In addition, 16 characters (8 male, 8 female; the same as in Experiment 1) were used. To compensate for the reduction in stimuli due to the exclusion of neutral grammatical gender, we adjusted the number of trials in this experiment. Specifically, we paired each object with every character (rather than just one male and one female pairing per item as in Experiment 1), resulting in 640 unique pairs. This adjustment was made for two main reasons. First, the exclusion of neuter gender reduced the overall number of stimuli, which could have impacted the statistical power of the study, while increasing the number of trials helped to counterbalance this reduction. Second, in Experiment 1, pairings were pseudorandomised to minimise the risk of semantic associations (e.g., avoiding obvious pairings like “a broom” with “a witch”). In Experiment 2, to eliminate this potential confound entirely, each object was paired with every character, thus increasing variability and reducing the chance of unintended semantic associations. The trial order was randomised for each participant, with objects presented on the left and characters on the right of the screen.

To ensure the validity of the data, we adopted enhanced measures, including comprehensive guidelines detailing the necessary procedures and environment for successful task completion. Additionally, participants were observed during the experiment. Any participant observed becoming distracted or communicating in their native languages was excluded from the analysis (13 bilingual and 11 monolingual speakers).

3.4.2.3. Procedure and design

The approach for Experiment 2 closely followed that of Experiment 1, but with the inclusion of participant observation conducted via Zoom. An experimenter monitored each session to ensure that participants were focused, free from distractions, and not using their native language during the task. All interactions were done in English and if participants needed clarifications, they did so in English as well. In Experiment 2, we also modified the verbal instructions to emphasise the use of the entire response scale (1 to 9). This adjustment was made based on observations from Experiment 1, where some participants tended to limit their responses to a narrower range of the scale. The experimenter used intonation to explicitly highlight this request during the verbal instructions, while maintaining the original instructions from Experiment 1. The modified instructions, given in English, were as follows: “In this study, you will see pairs of pictures appear on the screen. In each pair, there will be a picture of a person on the left and a picture of an object or animal on the right. You will see a scale where 1 = not similar and 9 = very similar. For each pair of pictures, please choose a number between 1 and 9 to indicate how similar you think the two pictures are. Try to use the WHOLE scale (give some 1’s and some 9’s and some of all the numbers in-between). Please respond with the first answer that comes to mind. Please try not to be distracted and avoid communicating with anyone (unless necessary) until the experiment is complete.” The final sentence, instructing participants to avoid distractions and communication, was added specifically for Experiment 2 to help maintain task focus.

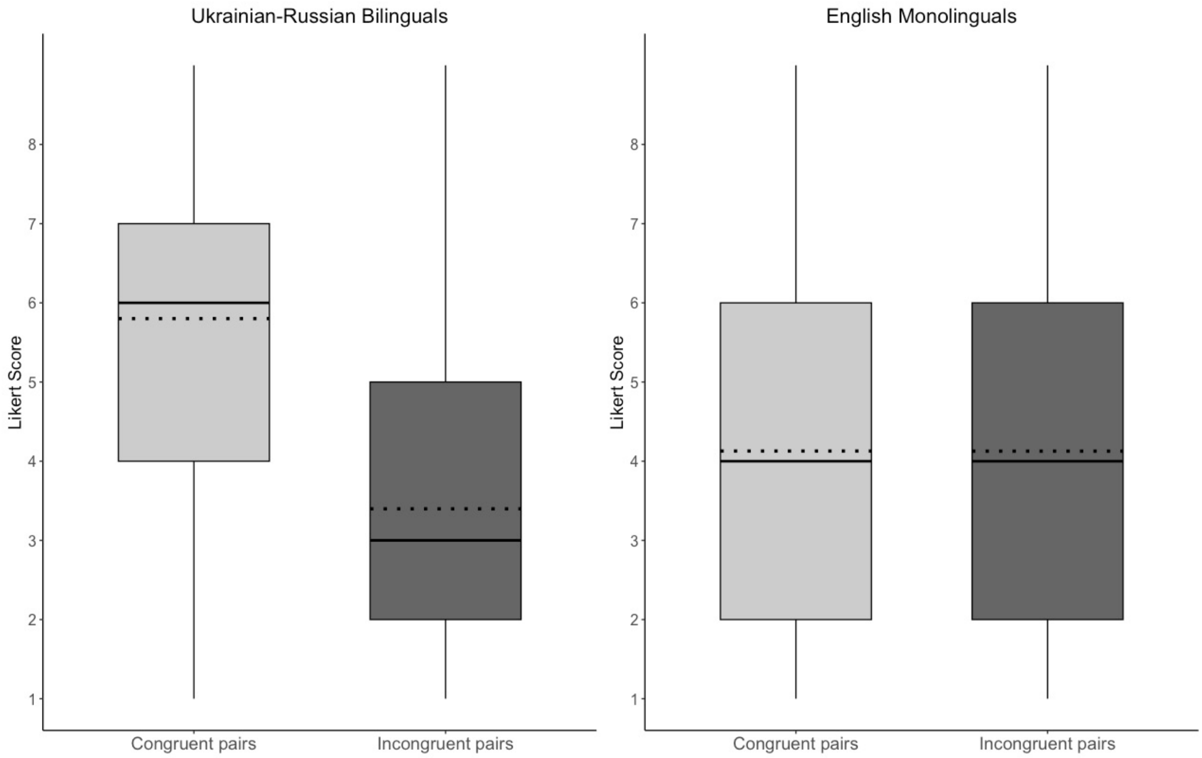
The analytical approach remained consistent with that of experiment 1, employing a similar structure for the cumulative link mixed models. The analysis comprised two parts. In the first part, we compared the responses of English monolinguals and Ukrainian-Russian bilinguals. This comparative analysis explored the effects of pair congruence (congruent vs incongruent in both Russian and Ukrainian) and group (Ukrainian-Russian bilinguals vs English monolinguals) interaction on Likert scores. The second part focused on the examining responses from Ukrainian-Russian bilinguals only, assessing the effect of pair congruence (congruent in Ukrainian / incongruent in Russian vs congruent in Russian / incongruent in Ukrainian) and language proficiency (-100 to 100, with -100 being only Proficient in Russian, to 100 –only proficient in Ukrainian) interaction on similarity ratings. In both parts of the analysis, the maximum convergence models included random intercepts for participants and items.

3.4.2.4. Results

Comparing the Ukrainian-Russian bilingual and English monolingual participants. Consistent with our predictions, bilinguals assigned significantly higher ratings to pairs with congruent biological sex and grammatical gender in both L1 and 2L1 (*Mean* = 5.8, *SD* = 2.0), as opposed to the incongruent pairs (*Mean* = 3.4, *SD* = 1.73). Besides, as confirmed by pairwise comparison, bilinguals rated congruent pairs significantly higher than monolingual participants (*Mean* = 4.12, *SD* = 2.4). As for the incongruent pairs (Figure 3.3), Ukrainian-Russian bilinguals tended to rate them significantly lower (*Mean* = 3.4, *SD* = 1.74) than English controls (*Mean* = 4.13, *SD* = 2.4). For the English monolingual group, there was no significant difference between the "congruent" and "incongruent" conditions (*SE* = 0.0245, z = 0.47, p = 0.639).

Figure 3.3

Comparison of Likert scores across conditions for Ukrainian-Russian Bilinguals and English Monolinguals: mean (dotted line) and median (solid line) differences in congruent and incongruent stimuli pairs in Experiment 2.



Analogously to the first experiment, a cumulative link mixed model examined the interaction between the proup (bilingual vs. monolingual) and condition (congruent vs. incongruent in both L1s), as a predictor for similarity ratings. The results revealed a statistically significant group effect for Ukrainian-Russian bilinguals (*SE* = 0.0888, *z* = 16.38, *p* < 0.001). We also found significant effects for the bilingual group - condition interaction, indicating that bilinguals assigned significantly lower rating to the incongruent pairs (*SE* = -1.9301, *z* = -55.15, *p* < 0.001) than English monolinguals[3]. These findings confirmed our hypothesis that matching grammatical gender in both languages of bilinguals significantly affects their categorisation once neutral gender is excluded from the testing conditions.

Comparing Ukrainian-Russian simultaneous bilinguals based on the language proficiency in L1 and 2L1. Figure 3.4 illustrates a clear difference in ratings, in line with our expectations. Ukrainian-Russian bilingual participants who were more proficient in the Ukrainian language gave significantly higher similarity ratings to object-character pairs where the object's grammatical gender in Ukrainian was congruent to the character's biological sex (*Mean* = 4.99, *SD* = 2.26), compared to pairs congruent in Russian (*Mean* = 4.56, *SD* = 2.38). Conversely, those with higher proficiency in Russian tended to give significantly higher ratings to pairs congruent in Russian (*Mean* = 5.12, *SD* = 2.24) than to incongruent ones (*Mean* = 4.54, *SD* = 2.26).

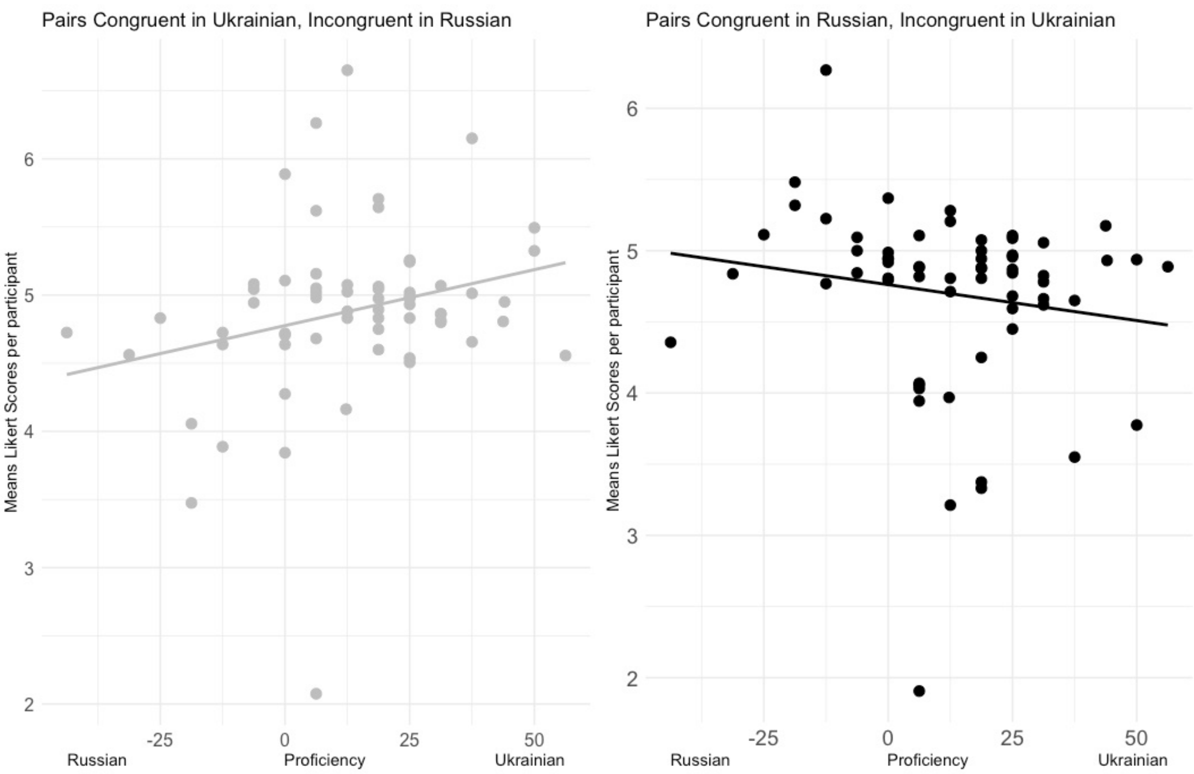
The designed cumulative link mixed model tested the impact of the interaction between condition (congruent with Ukrainian language and incongruent with Russian vs congruent with Russian language and incongruent with Ukrainian) and language proficiency (-100 to 100). While no significant main effect for Condition ($SE = 0.0292$, $z = 0.475$, $p = 0.635$), a significant main effect of Proficiency ($SE = 0.0022$, $z = -1.960$, p

[3] Follow-up contrasts showed that this effect was driven by bilinguals assigning substantially lower ratings to incongruent pairs compared to congruent ones (*Estimate* = –1.94, $SE = 0.025$, $z = –76.24$, $p < .001$), while monolinguals showed no such difference (*Estimate* = –0.01, $SE = 0.025$, $z = –0.45$, $p = .652$).

= 0.05) was observed. Besides, as predicted, a significant interaction was found between condition and proficiency (“Congruent in Russian, incongruent in Ukrainian” as a reference level: *Estimate* = 0.011, *SE* = 0.0013, $z = 8.622$, $p < 0.001$; “Congruent in Ukrainian, incongruent in Russian” as a reference level: *Estimate* = -0.011, *SE* = 0.0013, $z = -8.622$, $p < 0.001$). This suggests that the interaction between most proficient L1 of a simultaneous bilingual and condition had a significant impact on categorisation, and those bilingual participants that were more proficient in Ukrainian rated pairs that were congruent in Ukrainian and incongruent in Russian as more similar, and vice versa for those more proficient in Russian.

Figure 3.4

Mean Likert Scale Responses from Experiment 2 per participant (Ukrainian-Russian bilinguals only) by Language Proficiency for pairs of stimuli where characters' biological sex and objects' grammatical gender are (a) congruent in Ukrainian and incongruent



3.5. Discussion

The current study aimed to explore how language, grammatical gender in particular, affects cognitive processes of Ukrainian-Russian bilinguals in an all-English context. The group was chosen for several reasons. First, Ukrainian and Russian grammatical systems have nouns with both matching and contrasting grammatical gender across languages. Secondly, both languages have three grammatical genders (masculine, feminine, and neuter). Incorporating Ukrainian and Russian languages is beneficial for linguistic relativity research because, unlike previously studied languages such as Italian, Spanish, French, or German, they lack articles that could conflict with the biological sex of the referent. Instead, grammatical gender in Ukrainian and Russian is predominantly marked through noun, adjective, and sometimes verb endings. This distinct morphosyntactic feature — where gender is conveyed directly through morphological changes rather than through articles or fixed gender markers — has often been overlooked in existing research focused on languages with different gender-marking strategies.

Moreover, one of our research interests in the present study was to contribute to the discussion of whether gender effects arise online or offline, by examining whether having neuter gender embedded in the grammatical systems of both Ukrainian and Russian would lead to diminished grammatical gender effects. Therefore, we adapted one of the seminal studies on grammatical gender (Phillips & Boroditsky, 2003), while manipulating grammatical gender in Ukrainian and Russian, as well as presence (Experiment 1) and absence (Experiment 2) of neuter gender in testing conditions.

In Experiment 1, we observed a lack of significant effects of grammatical gender and group, as well as their interaction, when comparing the ratings of bilingual and monolingual participants. Additionally, we found no effects of the interaction between language proficiency and grammatical gender in Ukrainian-Russian bilinguals, indicating that their more proficient language had little to no effect on similarity judgements. Such findings align with previous research that reported lack of grammatical gender effects on speakers of three-gendered languages, such as German (Sera et al., 2002; Vigliocco et al., 2005).

However, after excluding neuter gender in Experiment 2, a significant interaction between group and condition was found when comparing bilingual and monolingual groups, indicating that Ukrainian-

Russian bilinguals rated higher those pairs where grammatical gender of an object in both Ukrainian and Russian was congruent with biological sex of a character, compared to the incongruent pairs. Additionally, a significant interaction between condition and language proficiency was observed, when only simultaneous bilinguals' results were analysed. The latter demonstrated that bilinguals with higher proficiency in Ukrainian rated those pairs as more similar where grammatical gender and biological sex were congruent in Ukrainian and incongruent in Russian. The analoguous effect was observed for speakers more proficient in Russian, as they perceived the pairs congruent in Russian to be more similar than those congruent in Ukrainian.

Before discussing differences between the experiments, we should first explore the possible reasons for the null results in Experiment 1. The absence of significant results in the first experiment might be attributed to several factors. Firstly, as suggested by Sera et al. (2002) and Vigliocco, et al. (2005), three-gendered grammatical systems may not show effects as strong as those in two-gendered languages with more direct and intuitive associations between grammatical gender and natural gender, which can lead to stronger perceptual biases. In contrast, three-gendered systems which include a neuter gender, introduce a level of grammatical complexity that may obscure the relationship between gender and categorisation. The neuter gender, in particular, could have reduced the salience of masculine and feminine distinctions, thereby weakening potential gender effects. Secondly, the broader lack of support for findings using this paradigm may reflect ongoing issues related to the replication crisis in linguistic relativity research. As mentioned earlier, most previous attempts to replicate Phillips and Boroditsky (2003) have not yielded significant results, except for Pavlidou and Alvanoudi (2019). For instance, Elpers et al. (2022), even with an increased sample size, failed to provide the significant results using the linear mixed effects models, though analysis using the t-tests showed significance. This issue is exacerbated by methodological variations and by the use of different statistical analyses across studies that employ the same paradigm, which makes it challenging to compare results consistently. Finally, the unique linguistic profiles of participants, which often differ across research contexts, add another layer of complexity. Previous studies that used a similarity judgement task also focused on bilingual participants, but there is limited consistency in how those participants were selected or their linguistic profiles were characterised. Key details, such as whether participants spoke other gendered languages and the criteria used for proficiency self-assessment are often not reported in sufficient detail. This variability makes it difficult to draw meaningful comparisons across studies, as differences in participant characteristics could significantly influence the observed effects – or the lack thereof - of grammatical gender. However, a key unifying factor between our study and those conducted by Sera et al. (2002) and Vigliocco et al. (2005) is the inclusion of neuter gender in the stimuli. This suggests that the presence of neuter gender may have influenced the absence of grammatical gender effects observed across these studies.

The discrepancy in language effects between Experiments 1 and 2 could be attributed to variations in experimental design, such as increased number of stimuli, variation in instructions or participant observations in Experiment 2, as well as lack of neuter gender in the task. While we initially hypothesised that the absence of neuter gender would be primarily driving the observed differences, it is important to consider that other methodological changes may also have contributed. First, the increased the number of stimuli in Experiment 2 likely enhanced statistical power, providing a clearer picture of language effects that might have been less detectable in Experiment 1. Besides, increased number of pairs allowed us to account for the possible semantic associations in Experiment 2 that could have emerged in Experiment 1 (e.g., pairing 'a broom and 'a witch' together). To examine the potential outcomes of using only the stimuli from Experiment 1 within the context of Experiment 2, an additional analysis was conducted with this subset. This analysis, which included 72 pairs of stimuli from Experiment 1, confirmed a robust and significant effect for both types of stimuli, consistent with the results obtained from the full stimuli set in Experiment 2. These findings strengthen the interpretation that the absence of neuter-gender stimuli in Experiment 2 may be a driving factor behind the observed grammatical gender effects, further validating our findings. Detailed analysis have been included in the supplementary materials (Appendix A). Second, the modified verbal instructions emphasised the use of the entire scale (1 to 9), which may have influenced participants to use a broader range of responses. Third, the addition of participant observation via Zoom allowed the experimenter to ensure that participants remained focused and did not revert to their native language.

However, it is also possible that the observed differences in the results were primarily due to the absence of the neuter grammatical gender, as hypothesised. This effect may be explained by considering the distinction between online and offline language processing discussed in the literature. According to both the label-feedback (Lupyan, 2020) and structural-feedback hypotheses (Sato & Athanasopoulos, 2018), online effects occur when language actively modulates perception and decision-making in real time, influenced by top-down feedback from specific linguistic labels and broader structural patterns respectivelly. In Experiment

1, the inclusion of neuter gender may have diluted the salience of masculine and feminine categories, reducing the immediate impact of gender cues on participants' judgments. Neuter nouns might have introduced a neutral, less distinctive category that disrupted the online processing of gender, as it did not align with the binary masculine-feminine distinction. This aligns with findings from previous research, which suggest that the presence of a third, neuter category can weaken the perceptual link between grammatical and natural gender - not in the offline manner as claimed by Sera et al. (2002), but during the process of task completion. In Experiment 2, by excluding neuter gender, the task environment emphasised over the course of the experiment the binary masculine-feminine distinction, creating a feedback loop where the structure of the gender system becomes more entrenched and influences real-time (online) processing more strongly. Without the neutral baseline provided by neuter nouns, participants were more inclined to use the salient gendered cues actively, resulting in more pronounced effects. This suggests that the grammatical gender effects observed in Experiment 2 were primarily driven by the immediate, context-sensitive use of gender information (i.e., online effects), but also by the reinforcing influence of the underlying linguistic structure on cognitive processing (i.e., a structural feedback effect).

In sum, our study shows that such an effect does not have its roots in the mere presence of the neuter gender in a languages's grammatical system, but rather arises online, as a function of the absence of the neuter gender in the task. Such an interpretation is compatible with modern accounts of the mechanisms underpinning linguistic relativity effects, such as the label-feedback hypothesis (Lupyan, 2012) and the structural-feedback hypothesis (Sato & Athanasopoulos, 2018). These findings also align with earlier research and demonstrate that three-gendered languages do indeed impact cognitive processes, such as categorisation. Furthermore, the language effects are present even when grammatical genders do not match in the two languages of simultaneous bilinguals, as they rely on the grammatical gender of their more proficient language.

The complexity of our findings underscores the necessity for more nuanced research methodologies. The similarity judgment task is merely the first step in analysing gender effects within our new group of participants. We suggest that future research employ more rigorous methodologies to further investigate these effects. For instance, incorporating neurophysiological measures, such as event-related potentials (ERPs), to better elucidate the effects of grammatical gender on bilingual cognition. This could be done by adapting previously used paradigms by Sato, et. al. (2020) or Boutonnet, et al. (2012) to investigate whether grammatical gender primes conceptual or semantic represention (looking at N300 or Left Anterior Negativity respectively) in speakers of three-gendered compared to speakers of two-gendered speakers that were used in these two studies. Additionally, we recommend expanding the range of stimuli used to test speakers of multiple three-gendered languages. For example, future research could include nouns that have masculine or feminine grammatical gender in one language (L1) and neuter gender in the second language (2L1). This expansion would provide further insights into the influence of grammatical gender on bilingual cognition, grammatical gender representation in simultaneous/early bilingual's mind, and contribute to the broader field of linguistic relativity.

3.6. References

Ackerman, L. (2018). *Dealing with ordinal data*. https://lmackerman.com/notebooks/ordinal_data.nb.html#what_is_%E2%80%9Cordinal%E2%80%9D_data

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96. https://doi.org/10.1017/S1366728905002397

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

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*(2), 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., & Bylund, E. (2020). Whorf in the wild: Naturalistic evidence from human interaction. *Applied Linguistics*, *41*(6), 947–970. https://doi.org/10.1093/APPLIN/AMZ050

Athanasopoulos, P., & Bylund, E. (2023). Cognitive restructuring: Psychophysical measurement of time perception in bilinguals. *Bilingualism: Language and Cognition*, *26*(4), 809–818. https://doi.org/10.1017/S1366728922000876

Athanasopoulos, P., Bylund, E., & Casasanto, D. (2016). Introduction to the Special Issue: New and Interdisciplinary Approaches to Linguistic Relativity: New Approaches to Linguistic Relativity. *Language Learning*, *66*, 482–486. https://doi.org/10.1111/lang.12196

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

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

Bassetti, B., & Filipović, L. (2022). Researching language and cognition in bilinguals. *International Journal of Bilingualism*, *26*(1), 3–12. https://doi.org/10.1177/13670069211022860

Belacchi, C., & Cubelli, R. (2012). Implicit knowledge of grammatical gender in preschool children. *Journal of Psycholinguistic Research*, *41*(4), 295–310. https://doi.org/10.1007/s10936-011-9194-y

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

Bezpoiasko, O. K., Gorodenska, K. G., & Rusanivskiy, V. M. (1993). *Grammar of Ukrainian language. Morphology* (M. S. Tymoshyk & L. L. Ščerbatenko, Eds.; 1st ed.). Lybid.

Bilaniuk, L., & Melnyk, S. (2008). A Tense and Shifting Balance: Bilingualism and Education in Ukraine. *International Journal of Bilingual Education and Bilingualism*, *11*, 340–372. https://doi.org/10.1080/13670050802148731

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Boroditsky, L., & Schmidt, L. A. (2000). Sex, Syntax, and Semantics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *22*(22). https://escholarship.org/uc/item/0jt9w8zf

Boroditsky, L., Scmidt, L. A., & Phillips, W. (2003). Sex, syntax and semantics. In *Language in mind: Advances in the study of language and thought* (pp. 61–79). Boston Review. https://doi.org/10.7551/mitpress/4117.001.0001

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

Bross, F. (2019). Using Mixed Effect Models to Analyze Acceptability Rating Data in Linguistics. Version 1.0. *Mimeo*. https://doi.org/www.fabianbross.de/mixedmodels.pdf.

Budzhak-Jones, S. (1997). Quantitative analysis of gender assignment in mono/bilingual discourse. *Journal of Quantitative Linguistics*, *4*(1–3), 67–91. https://doi.org/10.1080/09296179708590080

Cambridge University Press. (2021). *IELTS (International English Language Testing System) | Cambridge English*. https://www.cambridgeenglish.org/exams-and-tests/ielts/

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*(3), 714–730. https://doi.org/10.1111/lang.12192

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., &

Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26).

Christensen, R. H. B. (2019). *Cumulative Link Models for Ordinal Regression with the R Package ordinal* (Version R package version 2019.12-10) [Computer software]. http://www.cran.r-project.org/package=ordinal/

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*(2), 137–152. https://doi.org/10.1177/13670069060100020201

Cook, V. J. (2002). Background to the L2 user. In V. J. Cook (Ed.), *Portraits of the L2 user* (pp. 21–28). Multilingual Matters. https://www.degruyter.com/document/doi/10.21832/9781853595851/pdf?licenseType=restricted#page=9

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

Csernicskó, I., & Máté, R. (2017). *Bilingualism in Ukraine: Value or Challenge? 10*, 14–35. http://dx.doi.org/10.1515/sm-2017-0001

Eberhard, D., Simons, G., & Fennig, C. (2019). *Ethnologue: Languages of the World, 22nd Edition*.

Elpers, N., Jensen, G., & Holmes, K. J. (2022). Does grammatical gender affect object concepts? Registered replication of Phillips and Boroditsky (2003). *Journal of Memory and Language*, *127*, 104357. https://doi.org/10.1016/j.jml.2022.104357

Flaherty, M. (2001). How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. https://doi.org/10.1177/0022022101032001005

Gentner, D., & Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought*. The MIT Press. https://doi.org/10.7551/mitpress/4117.001.0001

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual language profile. In *Measuring L2 Proficiency: Perspectives from SLA* (pp. 208–225).

Gorpynyč, V. O. (2004). *Morfolohiya ukrayins'koyi movy: Pidruchnyk dlya studentiv vyshchykh navchal'nykh zakladiv [Morphology of the Ukrainian language: A textbook for students of higher educational institutions]* (1st ed.). Academy.

Grosjean, F. (Ed.). (2021). The Extent of Bilingualism. In *Life as a Bilingual: Knowing and Using Two or More Languages* (pp. 27–39). Cambridge University Press. https://doi.org/10.1017/9781108975490.003

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*(2), 417–426. https://doi.org/10.1016/j.cognition.2006.07.008

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1), 9–28. https://doi.org/10.1515/probus-2016-0002

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

Kortmann, B., & Auwera, J. V. D. (2011). *The Languages and Linguistics of Europe: A Comprehensive Guide*. De Gruyter, Inc.

http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=765858

Kulyk, V. (2023, January 7). *Мова та ідентичність в Україні на кінець 2022-го [Language and identity in Ukraine by the end of 2022]*. Zbruč. https://zbruc.eu/node/114247

Lucy, J. A. (1992). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis* (Vol. 13, Issue 13, p. xi 328). Cambridge University Press.

Lucy, J. A. (1997a). Linguistic Relativity. *Annual Review of Anthropology*, *2*, 291–312.

Lucy, J. A. (1997b). Linguistic Relativity. *Annual Review of Anthropology*, *26*(Volume 26, 1997), 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. https://doi.org/10.1016/j.cognition.2017.10.010

Mitrofanova, N., Rodina, Y., Urek, O., & Westergaard, M. (2018). Bilinguals' sensitivity to grammatical gender cues in Russian: The role of cumulative input, proficiency, and dominance. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01894

Oxford University Press. (2001). *Quick Placement Test*. Oxford University Press.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Pavlidou, T.-S., & Alvanoudi, A. (2013). Grammatical Gender and Cognition. *Versita/de Gruyter*, *2*(2), 109–124.

Pavlidou, T.-S., & Alvanoudi, A. (2019). Conceptualizing the world as 'female' or 'male': Further remarks on grammatical gender and speakers' cognition. *Selected Papers on Theoretical and Applied Linguistics*, *23*(0), Article 0. https://doi.org/10.26262/istal.v23i0.7351

Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). https://escholarship.org/uc/item/31t455gf

Pinker, S. (1994). *The language instinct: How the mind creates language* (22. Dr). HarperPerennial.

Poftak, N., & Shykula, D. (2022). The Ongoing Language Discussion in Ukraine. In *Being Ukraine: An Introduction to Europe's Eastern Vanguard*. Connecticut College. https://openpress.digital.conncoll.edu/beingukraine/chapter/chapter-7/

Pugh, S., & Press, I. (1999). *Ukrainian: A Comprehensive Grammar*. Routlege.

R Core Team. (2022). *R: The R Project for Statistical Computing*. The R Project for Statistical Computing. https://www.r-project.org/

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Rusanivskyj, V. M., Taranenko, O. O., & Zjabljuk, M. P. (2004). *Encyclopaedia of the Ukrainian language* (M. P. Bažan, Ed.; 2nd ed.). Kyiv: 'Ukrainian Encyclopaedia'.

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

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

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

Shumlianskyi, S. (2010). *Conflicting abstractions: Language groups in language politics in Ukraine*. *2010*(201), 135–161. https://doi.org/10.1515/ijsl.2010.007

Speed, L. J., & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention, Perception, & Psychophysics*, *81*(6), 2063–2077. https://doi.org/10.3758/s13414-019-01729-0

Steinback, S. (2015). *Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png 1,853×1,642 pixels*. Lexical Distance among the Languages of Europe. https://upload.wikimedia.org/wikipedia/commons/f/f3/Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png

Thierry, G., Rahman, R. A., & Athanasopoulos, P. (2024). An Introduction to the Cognitive Neuroscience of Language Embodiment and Relativity Special Issue of the Language Learning Cognitive Neuroscience Series. *Language Learning*, *74*(S1), 5–19. https://doi.org/10.1111/lang.12643

Ukrainian center for educational quality assessment. (2020). *Ukrainian center for educational quality assessment*. https://testportal.gov.ua/en/

Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, *206*, 104485. https://doi.org/10.1016/j.cognition.2020.104485

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology: General*, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Whorf, B. L. (1956). *Language, Thought, and Reality, second edition: Selected Writings of Benjamin Lee Whorf* (Caroll, J.B., Ed.). MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Chapter 4. Between Two Grammatical Gender Systems: Exploring the Impact of Grammatical Gender on Memory Recall in Ukrainian-Russian Simultaneous Bilinguals

Linking statement: Chapter 3 showed that grammatical gender effects on categorisation can emerge in simultaneous bilinguals of two three-gendered languages, only when neuter gender stimuli are excluded. These findings suggest that under specific conditions, the structural characteristics of three-gendered systems influence perceptual judgements even in the absence of overt language use. Building on these findings, the following chapter shifts focus from categorisation to memory recall, to examine whether grammatical gender similarly affects the encoding and retrieval of information. By exploring the extent to which matching and/or mismatching grammatical gender assignments in bilinguals' two first languages shape memory processes, this chapter extends the investigation of grammatical gender effects on bilingual cognition across cognitive domains.

4.1. Abstract

This study examines the impact of grammatical gender on memory recall among Ukrainian-Russian simultaneous bilinguals. Building on the foundational work of Boroditsky and Schmidt (2000), we adapted their methodology to explore whether grammatical gender in two three-gendered languages (Ukrainian and Russian) affects memory recall. Ukrainian-

Russian bilinguals and English monolingual controls were tested on their ability to remember names assigned to objects with either matching or mismatching grammatical genders across their two languages. Results showed that bilinguals recalled names more accurately when the biological sex of the names was congruent with the grammatical gender of objects in both languages (e.g., recalling a male name assigned to a noun with masculine grammatical gender in both L1s, rather than a female name). English monolinguals, in contrast, showed no difference in recall. However, when grammatical gender mismatched between Ukrainian and Russian, the expected influence of the more proficient language on recall accuracy was not observed (recalling a name when it is congruent with grammatical gender of the more proficient L1 and incongruent with the less proficient L1). These findings suggest that converging grammatical information from two L1s creates stronger memory associations, enhancing recall accuracy of simultaneous bilinguals. Conversely, mismatching grammatical genders appear to negate this effect. Taken together, these findings highlight the interconnected nature of bilingual conceptual representation.

4.2. Introduction

The principle of linguistic relativity posits that the languages we speak influence our thoughts in systematic ways (Casasanto, 2008; Lucy, 1997; Whorf, 1956). Various disciplines (i.e., linguistics, philosophy, and psychology, as well as interdisciplinary research) have put this hypothesis at the forefront of their investigations. Research in linguistic relativity has explored multiple areas, including grammatical number and object perception, spatial-temporal orientation, time, and grammatical gender. The current study focuses on grammatical gender in Ukrainian and Russian languages and its effects on cognitive processes, specifically how grammatical gender influences memory recall. Grammatical gender has received extensive attention, with some studies affirming its effects (e.g., Boutonnet et al., 2012; Sato et al., 2020; Sato & Athanasopoulos, 2018) and others showing evidence against its effect on cognitive processes or proposing alternative explanations (e.g., Bassetti, 2007; Pavlidou & Alvanoudi, 2013; Sera et al., 2002).

In the present study we focus on addressing two key gaps that we identified in linguistic relativity research on grammatical gender. Firstly, we aim to draw attention to the inclusion of the relatively underrepresented group of simultaneous bilinguals. The need to consider multilingualism was arguably put forward in some of Whorf's arguments (see Pavlenko, 2016). However, when it comes to focusing on bilingual individuals, researchers tend to focus on sequential bilinguals (e.g., Boutonnet et al., 2012; Sato et al., 2020; Phillips & Boroditsky, 2003). Meanwhile, simultaneous bilinguals, who acquire both languages (L1 and a second L1, henceforth 2L1) from birth, are scarcely represented in the research (Bassetti, 2007; Osypenko et al., 2025). Therefore, little is known about whether cognition of adult simultaneous bilinguals with two distinct grammatical genders embedded in their L1 and 2L1 is affected by language to the same degree as sequential bilinguals. Secondly, early research on linguistic relativity (Sera et al., 2002; Vigliocco et al., 2005) argued that language effects are present in speakers of two-gendered languages rather than three-gendered languages, as there is a stronger association with natural gender in the former group. Finally, when looking into the domain of grammatical gender, a large number of studies typically target categorisation mechanisms with a prominent grammatical gender present, e.g., voice-assignment task, where participants are asked to assign either a male or female voice to gendered objects (Kurinski et al., 2016; Sera et al., 2002). However, we intend to investigate the effects using a less gender-salient paradigm that involves recall memory (Boroditsky & Schmidt, 2000).

Our study aims to address these issues by looking at two three-gendered languages (Ukrainian and Russian) co-existing in the mind of simultaneous bilinguals. It will allow us to tackle another uncovered issue in linguistic relativity research: How do two grammatical systems that have been acquired since early childhood interact with each other? More importantly, are the effects of language on cognition enhanced when grammatical gender in L1 matches grammatical gender in 2L1? Alternatively, are the language effects negated or reduced when incorporating stimuli where grammatical gender in L1 mismatches with 2L1?

4.2.1. Effects of Grammatical Gender in Linguistic Relativity Research in Bilinguals

Grammatical gender is present in approximately 40% of the world's languages (Corbett, 2001), requiring speakers to mark gender through noun suffixes, as well as articles, adjectives, pronouns, and, in specific cases, within verb forms, in such languages as Ukrainian and Russian. This compels speakers of gendered languages to pay close attention to grammatical gender during language production and comprehension. This grammatical property has been extensively employed in linguistic relativity research for several reasons, including its cross-linguistic variability and inherent arbitrariness (Everett, 2013; Boutonnet et al., 2012). For instance, the noun "sun" is grammatically feminine in German *("die Sonne")*, masculine in Spanish (*"el sol"*), and neuter in both Ukrainian (*"сонце"*) and Russian (*"солнце"*), exemplifying the absence of any systematic relationship between grammatical gender and the semantic or biological attributes of the referent. Multiple studies with monolingual speakers provide evidence in favour of grammatical gender effects on cognitive processes (e.g., categorisation) and conceptual representations of nouns (Haertlé, 2017; Maciuszek et al., 2019; Vernich, 2017). A significant body of research has also investigated grammatical gender effects in bilingual speakers, specifically sequential bilinguals (Athanasopoulos & Boutonnet, 2016; Chen & Faitaki, 2024; Kurinski & Sera, 2011 among others). This allows for a deeper investigation of Whorfian effects, such as how two languages coexist in a bilingual mind, specifically whether bilinguals exhibit language effects comparable to monolingual speakers of their L1s or whether having two grammatical systems leads to differences in cognitive processes (Cook, 2006; Wang, 2020). Additionally, including bilinguals provides the opportunity to examine the stability of previously found gender effects. Specifically, testing bilinguals in their second language (L2) that does not have a specific grammatical/lexical property of their first language (L1), allows researchers to test whether the effects of L1's gender on bilinguals' responses can remain despite the presence of an L2 (Kousta et al., 2008; see Samuel et al., 2019 for a detailed review). Our study contributes to the growing body of research on grammatical gender effects in bilingual speakers by focusing on a relatively underexplored group within linguistic relativity studies - simultaneous bilinguals. These are individuals who acquire two languages from birth or very early in life and develop native-like proficiency in both first languages (henceforth, both L1s). In our case, the participants are simultaneous bilinguals of Ukrainian and Russian, two languages with partially contrasting grammatical gender systems.

Both Ukrainian and Russian have a three-gender grammatical system where all nouns are categorised as feminine, masculine, or neuter, which provides an interesting test-case for this line of research. In these languages, animate nouns generally align with biological sex of the referent, except for certain exceptions (e.g., *"мавпа"* (monkey) is feminine in Ukrainian regardless of gender, see Vakulenko, 2023). However, when the biological sex of the referent is unknown or irrelevant, speakers commonly use default grammatical gender assigned to the noun - for instance, *"слон"* (elephant) is typically used with masculine gender in Ukrainian, while *"лисиця"* (fox) takes feminine gender, regardless of the animal referent's actual sex (Vakulenko, 2023). Neuter forms also occur, most notably in diminutives in Ukrainian. In contrast, the gender assignment of inanimate nouns is arbitrary and unrelated to semantic meaning or biological sex (Corbett, 1991). Because of the arbitrariness of its application across languages and its detachment from conceptual-ontological meaning, grammatical gender is particularly relevant for discussions of linguistic relativity, as it exemplifies linguistic phenomena that are independent of real-world differences and are purely linguistic in nature (Bassetti, 2007).

The reason for including Ukrainian-Russian simultaneous bilinguals stems from the type of language pairing they provide and how this pairing can deepen our understanding of gender effects on cognition. Previously, studies on sequential bilinguals have explored various pairings of languages, such as speakers of a gendered L1 and a genderless L2 (Pavlidou & Alvanoudi, 2013; Sato & Athanasopoulos, 2008; see full discussion in Chen & Faitaki, 2024), a genderless L1 and a gendered L2 (Athanasopoulos & Boutonnet, 2016; Kurinski & Sera, 2011;), or a gendered L1 and gendered L2 (Lambelet, 2016). Each pairing allows to test for the variability of grammatical gender effects in bilinguals and in case of gendered L1-L2 pairings, to examine the interactions of two grammatical gender structures. Yet, simultaneous bilinguals with two gendered first languages (gendered L1 and 2L1) remain largely unexamined, with only two studies available to our knowledge (Bassetti, 2007; Osypenko et al., 2025). For instance, in the study by Bassetti (2007), Italian-German bilingual and Italian monolingual children were tested in the voice attribution task (i.e., participants assigning either a male or a female voice to gendered objects). All objects were selected so that their grammatical genders in Italian and German were mismatching (e.g., an object being masculine in Italian and feminine in German, and vice versa). The findings indicated that only the monolingual group exhibited grammatical gender effects on their responses, with Italian monolinguals assigning more male voice to nouns that are masculine in Italian, and female voice to feminine nouns. Italian-German bilinguals did not show any effects of either Italian or German grammatical gender on their responses, suggesting that gender mismatch across languages might have reduced grammatical gender effects. Importantly, the study did not examine the effects of grammatical gender for objects whose gender matched across both L1s. Therefore, it remains unclear whether the absence of predicted gender effects was specific to the set of mismatched stimuli, due to gender conflict between the two languages, or whether such effects are generally absent in the chosen participant group as a result of having two gendered L1s.

To address this, Osypenko et al. (2025) studied the effects of having two gendered L1s (Ukrainian and Russian, as in the current study), by investigating both matching and mismatching gendered objects across the two languages, using a similarity judgement task (adapted from Phillips & Boroditsky, 2003). Ukrainian-Russian adult simultaneous bilinguals with English as an L2 were presented with pairs consisting of a depicted object and a gendered character (a male or a female cartoon characters) and asked to rate how similar they are on a Likert scale from 1 (*"not similar at all"*) to 9 (*"very similar"*). The study had two experiments with the same task and experimental conditions; however, Experiment 1 included stimuli of all three genders represented in Ukrainian and Russian (i.e., masculine, feminine, and neuter), while Experiment 2 excluded those with neuter gender. First, stimuli that were matching in grammatical gender (e.g., "a fork" – feminine in both Ukrainian and Russian) paired with a male/female character resulted in two conditions: congruent/incongruent in both L1s. The prediction was that when pairs were congruent in both L1s, participants would rate them more similar, compared to the incongruent pairs. Second, stimuli with mismatching grammatical across the L1s ("a notebook" – masculine in Ukrainian, feminine in Russian) paired with male/female characters created conditions where pairs were either congruent in Ukrainian or congruent in Russian. Osypenko et al. (2025) predicted that participants would rate those pairs as more similar that are congruent in their more proficient L1. While Experiment 1 did not reveal the predicted effects of grammatical gender for either type of stimulus, Experiment 2 found these effects for both stimulus types. Specifically, simultaneous bilinguals rated pairs as more similar when the grammatical gender was congruent across both L1s, and when congruency aligned with their more proficient L1. Although alternative explanations for the discrepancy between the two experiments are considered, the findings overall suggest that grammatical gender can influence simultaneous bilinguals of two gendered languages. However, the manifestation of these effects appears to depend on the experimental context, i.e. when neuter gender is excluded.

4.2.2. Grammatical Gender Effects in Two- vs. Three-Gendered Languages

An additional factor motivating the current study is the ongoing debate over whether grammatical gender effects are present in speakers of three-gendered languages (i.e., languages with masculine, feminine, and neuter genders) or whether such effects are limited to speakers of two-gendered languages (i.e., those with only masculine and feminine genders).This discussion emerged in the early 2000s (Sera et al., 2002; Vigliocco et al., 2005) and has continued in more recent work (Osypenko et al., 2025). For instance, Sera et al.'s (2002) research on German-English bilingual, as well as French and Spanish monolingual children revealed that unlike their monolingual counterparts, German-English bilinguals did not use German grammatical gender as a basis for assigning voices to objects in a voice attribution task. The researchers speculated that the lack of effects could be attributed to the presence of neuter gender in German, suggesting that languages with a two-gender system have a strong association between grammatical and natural gender. The latter according to Sera et al. (2002) leads to overgeneralisation of masculine and feminine traits to inanimate objects. In contrast, speakers of languages with a three-gender system, such as German, appear to rely less on gender and more on other conceptual distinctions (artificial or natural entities) when categorising objects. Subsequently, Vigliocco et al. (2005) reached a similar conclusion, finding that grammatical gender effects are limited to two-gendered languages, as evidenced by comparative responses of Italian and German participants. The authors claimed that two-gendered systems have a high degree of transparent correspondence between the grammatical gender of nouns denoting humans and the biological sex of those humans.

This discussion has evolved with the emergence of evidence either fully (Beller et al., 2015; Bender et al., 2016; Haertlé, 2017; Maciuszek et al., 2019) or partially (Pavlidou & Alvanoudi, 2013, 2019; Osypenko et al., 2025) supporting the presence of grammatical gender effects in speakers of three-gendered languages, raising further questions about what factors (e.g., experimental design, language typology, etc.) contribute to the discrepancies observed in findings across studies involving speakers of these languages. This inconsistency with earlier findings may be attributed to typological differences between the languages examined. Specifically, the nature and transparency of grammatical gender systems - including how extensively grammatical gender is marked and the degree of interplay between cultural or conceptual associations and grammatical gender - varies across language families such as Germanic, Romance, and Slavic (see Kupisch et al., 2022 for further discussion on cross-linguistic gender transparency). These differences may influence the strength or presence of observed effects, rather than being reduced simply to whether a language has two or three grammatical genders. For instance, unlike Romance languages, German does not have a strong consistency in how grammatical gender of the nouns is referring to humans and their biological sex (e.g., "*das Mädchen*", translates as "the girl", yet has neuter gender assigned to it), while in Polish, Ukrainian, and Russian languages animate entities are referred to either with a masculine or a feminine gender, except for diminutive forms for animals in Ukrainian language (Gorpynyč, 2004). Furthermore, German articles in certain cases do not differentiate between genders. For instance, in the dative case both masculine and neuter genders would require the same article "*dem*" (e.g., "*der Mann"* (the man) – *"dem Mann"* (to the man), *"das Kind"* (the child) – *"dem Kind"* (to the child). These factors might have led to less pronounced effects of grammatical gender effects in German speakers, compared to speakers of two-gendered languages, where grammatical and semantic gender are more closely aligned (Kousta et al., 2008). Slavic languages (e.g., Polish, Ukrainian, Russian) do not have the aforementioned features, as they do not contain articles. Instead, grammatical gender is primarily indicated by the endings of nouns, adjectives, and in certain cases, verbs. Therefore, including three-gendered languages free from the constraints mentioned above, can lay out a good testing ground to determine whether language effects are indeed solely confined to two-gendered languages.

A few studies provided evidence in favour of gender effects with a Slavic three-gendered language, Polish, albeit with monolingual speakers (Haertlé, 2017; Maciuszek et al., 2019). Haertlé's (2017) study consisted of voice attribution and adjective assignment tasks, conducted in participants' L1, for 19 objects with mismatching grammatical gender in Polish and French (e.g., "a house" - masculine in Polish, feminine in French). Participants were French and Polish native speakers. Nouns that have neuter gender in Polish were not included in the stimuli. The findings showed significant interactions between language and grammatical gender in a voice attribution task for both Polish and French speakers, with stronger effects in French, suggesting that grammatical gender influences cognitive processes in three-gendered languages, though the effects were more pronounced in two-gendered French. However, it is unclear whether these effects would vary if neuter gender stimuli were included. A subsequent study investigating grammatical gender effects in Polish speakers used three different experimental designs: triadic similarity judgments, an implicit association test, and a voice attribution task (Maciuszek et al., 2019). While the triadic similarity judgment task did not show effects of grammatical gender, the other two tasks did. The study highlights that grammatical gender in Polish influences cognitive processes beyond simple categorization, affecting implicit cognition and the attribution of characteristics to objects. Similarly to Haertlé (2017), these findings suggest that while two-gendered languages might exhibit stronger grammatical gender effects, three-gendered languages like Polish still show significant influences on cognition.

Therefore, one can argue that the distinction between two- and three-gendered languages does not determine lack or presence of language effects; instead, the conditions under which grammatical gender effects emerge warrant closer examination. In line with this conclusion, Osypenko et al. (2025) reports supporting evidence. Although Experiment 1 did not reveal the predicted grammatical gender effects in speakers of two three-gendered languages, Experiment 2 provided evidence for such effects once the neuter gender was excluded from the experiment. This suggests that the presence or absence of neuter gender may influence whether effects are observed. Nonetheless, the study demonstrates that grammatical gender effects on categorisation can still be elicited in speakers of three-gendered languages, even when tested in a genderless second language (English, in the case of Osypenko et al., 2025).

Finally, a crucial aspect of all previously discussed studies is that all of them examined effects of grammatical gender on one cognitive process – categorisation. Therefore, the current study adds to this discussion by exploring whether such grammatical property as grammatical gender, plays a role in more complex cognitive functions, such as memory and objects' mental representations.

4.2.3. Memory recall effects in LR research

Studies examining how cross-linguistic structural and/or labelling differences impact recall also present contrasting evidence, both in favour (Boroditsky & Schmidt, 2000; Fausey & Boroditsky, 2011; Kirjavainen et al., 2020; Roberson et al., 2000; Tosun et al., 2013) and against Whorfian effects (Cibelli et al., 2016; Regier & Xu, 2017; Sakarias & Flecken, 2019; Ünal et al., 2016). For instance, Fausey and Boroditsky (2011) investigated how linguistic differences in describing videos of intentional and accidental events (a person pops balloon using tack vs a person reaches to put a tack in container and accidentally pops the balloon during reach) influence memory recall for agents in English and Spanish speakers. They found that while both groups described intentional events agentively and remembered agents equally well, differences emerged for accidental events. English speakers used more agentive language when describing accidents (e.g., “She popped the balloon”) and showed better memory for the agents involved in these events. In contrast, Spanish speakers, who often used non-agentive constructions (e.g., “The balloon popped”), were less likely to recall the agent responsible for accidental actions.

In a more recent study, examining effects of cross-linguistic differences on memory recall in a different domain of grammatical number, Kirjavainen et al. (2020) manipulated the presence/absence of compulsory number marking in monolingual speakers of English and Japanese. Across two experiments, participants viewed photos of either one or two objects/animals for two seconds, after which they answered questions about number information (e.g., “How many lions did you see? 1 or 2?”), along with control questions about other details. In Experiment 2, 20 “guessing” questions were added, referencing photos never shown, to assess whether participants were employing a guessing strategy. The results suggested that English speakers, whose language requires explicit singular/plural marking (e.g., “apple” vs. “apples”), better recalled plurality information. In contrast, Japanese speakers, whose language allows omission of number marking, showed significantly lower accuracy recalling plural items. Experiment 2 confirmed that this effect was not due to guessing or question wording.

Overall, a large body of research provides support for both lexical and grammatical properties influencing recall. However, existing evidence calls into question whether these findings show a true Whorfian effect (i.e., language affecting perception) or rather "language-on-language" effects (i.e., participants use language to complete a language-engaging task; Wolff & Holmes, 2011). To address these two possibilities, Sakarias and Flecken (2019) investigated how case markings in Estonian and Dutch influence attention allocation in verbal and non-verbal event encoding and memory recall of the event endings. In the current review, we limit our discussion to the recall-related findings, as they are most relevant here. Two types of events were chosen for the study: resultative events (where objects undergo a visually noticeable change in state during the event, e.g., peeling a potato) or non-resultative events (no or only partial change of object's state, e.g., stirring in a pan). Estonian language has obligatory case marking when objects sustain a partial / full change in state (e.g., a fully peeled potato marked by accusative case "*kartuli*", whereas a partially peeled potato - by partitive case "*kartulit*"). On the other hand, Dutch lacks such grammatical marking (e.g., "*een aardappel schillen*" can mean both partial and full peeling). Participants watched short video clips depicting everyday causative events and were then required either to verbally describe the videos or to complete a non-verbal distractor task involving detecting auditory cues. Afterward, participants performed a surprise forced-choice recognition memory task testing their memory for the event endings. Sakarias and Flecken (2019) found a language-specific boost on recall of event results among Estonian participants compared to Dutch, but only under the verbal task. Specifically, Estonian speakers exhibited superior memory recall for video endings only in the verbal condition. Therefore, the findings were interpreted as supporting not the true Whorfian effect, but rather thinking-for-speaking effects (i.e., case marking influenced event memory only within language-dependent contexts; see Slobin, 1996; Wolff & Holmes, 2011), as no significant language-specific differences emerged in the non-verbal encoding condition.

As for memory recall studies examining effects of grammatical gender, the evidence, to date, has been fairly scarce. In their review, Samuel et al. (2019) report that this paradigm comprises only 2% of all studies that were selected to analyse ways researchers can investigate cross-linguistic linguistic relativity effects of grammatical gender. In total, this task has been employed in three distinct research studies and provided a combination of mixed support and no support (Boroditsky & Schmidt, 2000; Kaushanskaya & Smith, 2016; Pavlidou & Alvanoudi, 2013). Given the centrality of memory in the human cognitive system (Baddeley & Hitch, 1974), and the robust manifestation of linguistic relativity effects on memory in other linguistic domains (Athanasopoulos & Bylund, 2013; Lucy, 1992; Roberson et al., 2005), the current study aims to redress the balance of evidence of possible Whorfian effects on memory recall in the domain of grammatical gender.

The chosen methodological approach was originally developed by Boroditsky and Schmidt (2000). In their study, sequential bilinguals (25 Spanish-English and 16 German-English) and 20 English monolinguals were tasked with memorising a male/female name placed next to an object possessing a distinct grammatical gender in the participant's native language. For example, "a chair" that is masculine in German (*"der Stuhl"*) and feminine in Spanish ("*la silla*") was paired with either a male name (e.g., Patrick) or with a female name (e.g., Patricia). All objects had opposite genders in Spanish and German. Half of the names had a biological sex that was congruent with the grammatical gender of the paired object in L1, and the other half was incongruent. Participants' ability to recall these word-name combinations was then assessed. Boroditsky and Schmidt (2000) found effects of grammatical gender in native speakers of Spanish and German while being tested in English. Specifically, both Spanish-English and German-English bilinguals recalled better those name-object pairs where the biological sex of a proper name was congruent with the grammatical gender of the object in their native language (82% and 74% correct responses respectively, $t = 2.55$, $p < .01$). Therefore, since the objects chosen for the study had opposite grammatical genders in German and Spanish (e.g., feminine in Spanish and masculine in German, and vice versa), participants show opposite memory biases. For objects that Spanish speakers were more likely to remember paired with female names, German participants remembered when they were paired with male names, and vice versa. While having certain limitations (i.e., effects of conceptual gender, sample size, etc.), the presented conclusion holds significant importance as it suggests that both two- and three-gendered languages exert comparable effects on cognition, despite the presence of neutral grammatical gender in German.

In a subsequent iteration, Pavlidou and Alvanoudi (2013) adapted this framework with Greek-English sequential bilingual speakers, Greek being a three-gendered language. The stimuli comprised 28 nouns, each accompanied by a unique proper Greek name (e.g., Vasilis/Vasiliki, Alekos/Aleka) and then automatically followed by another pair in a randomised order. It was hypothesised that participants' memory would be more effective when the grammatical gender of words/objects in their L1 matched with the gender of proper names compared to cases where such alignment was absent. However, the authors reported that the memory task did not show any effects of the congruence between the grammatical gender of nouns and biological sex of the names. Pavlidou and Alvanoudi (2013) attribute the non-replication of the memory task to methodological variations between their study and the original study by Boroditsky and Schmidt (2000). Although the tasks were similar, Pavlidou and Alvanoudi (2013) were unable to replicate the procedure exactly due to limited detail in the original methodological descriptions, emphasising the need for greater transparency in methodological design to facilitate replication. Additionally, going back to the discussion on three-gendered languages, Pavlidou and Alvanoudi (2013) do not speculate whether the lack of results can be explained by the Greek language being three-gendered.

Finally, Kaushanskaya and Smith (2016) used a similar experimental design as Boroditsky and Schmidt (2000) while looking at the reversed language pairing in their bilingual participant group (genderless English as an L1 and a two-gendered Spanish as an L2) to examine whether grammatical gender information from a second language could influence memory performance in a first language that lacks gender marking. Three groups of English L1 speakers were tested: monolinguals, emergent bilinguals with high exposure to Spanish, and those with low exposure. Analogously with the previous two studies (Boroditsky & Schmidt, 2000; Pavlidou & Alvanoudi, 2013), participants completed an associative learning task, pairing inanimate object names with gendered proper names. The Spanish translation of each object was either gender-congruent or gender-incongruent with the name (e.g., “corn – Patrick” vs. “beach – William”). Crucially, the task was conducted in English. The results showed that high-exposure bilinguals exhibited sensitivity to Spanish grammatical gender: they were significantly less accurate in recalling incongruent pairs compared to congruent ones. In contrast, monolinguals and low-exposure bilinguals showed no such effect. These findings suggest that grammatical gender information from a second language can be activated and influence memory performance, even during tasks conducted entirely in the native, genderless language. In doing so, Kaushanskaya and Smith (2016) provide evidence not only for grammatical gender effects on recall, but also for the potential of L2 grammatical properties to restructure bilinguals' cognitive processing.

4.2.4. The current study

The present study extends the examination of grammatical gender effects on cognition by employing a memory task adapted for Ukrainian-Russian simultaneous bilinguals. In addressing our hypotheses and research questions, we refined the methodology and the analysis from Boroditsky and Schmidt's (2000) original study in several ways. Firstly, instead of sequential bilinguals, we recruited simultaneous speakers of Ukrainian and Russian. Secondly, instead of comparing Spanish-English and German-English bilinguals, where one group has a two-gendered language and another group has a three-gendered language, we recruited speakers who have two three-gendered grammatical systems embedded in their L1s. Thirdly, we expanded the stimuli list from 24 to 46 nouns. Instead of solely relying on stimuli with opposite grammatical genders in German and Spanish for between-subject comparison, we opted for a dual approach with two types of stimuli. This allowed us to conduct between-subject analysis in the first part of the study, comparing Ukrainian-Russian bilinguals and English monolinguals, where chosen nouns had matching grammatical gender in both L1s (e.g., "key" – masculine in both Ukrainian and Russian, "strawberry" – feminine in both, "feather" – neuter in both). In the second part, analysing only the performance of Ukrainian-Russian bilinguals, we employed a within-participant design to explore mismatching grammatical genders across two L1s (e.g., "moon" – masculine in Ukrainian, feminine in Russian, "sock" – feminine in Ukrainian, masculine in Russian). Lastly, we added a more detailed linguistic profile analysis, as well as proficiency tests, to analyse the effect of language proficiency in bilinguals' both L1s on their performance.

Building on the previous research, we predicted that the effects of native language(s) on the memory recall of object-name pairs presented to participants will be observed. Specifically, the recall of the names by Ukrainian-Russian bilinguals is hypothesised to be enhanced when the grammatical gender of the noun is congruent with the biological sex of the name in participants' native language(s), compared to English monolinguals. The hypotheses were formulated based on the two types of selected stimuli. Firstly, for nouns with matching grammatical gender in both native languages, we expect to find a stronger language effect on the ability to remember the assigned names, compared to English-speaking controls. For instance, bilingual participants are expected to remember the pair "Patrick – key" better than "Patricia – key", as "key" is masculine in both Ukrainian and Russian and is congruent with male biological sex. Besides, we anticipate a more pronounced language effect for the stimuli with the matching gender across two languages, compared to the mismatching one, since the converging grammatical information from the two languages would lead to stronger memory associations.

Secondly, for nouns with mismatching grammatical gender among bilingual participants, we anticipate that participants will display an effect of their more proficient language when recalling the names. For example, if a participant is more proficient in Ukrainian rather than Russian, they will tend to remember those names where gender matches Ukrainian and mismatches Russian. Participants with greater proficiency in Ukrainian than Russian are expected to show higher accuracy when recalling the pair "Eric – moon" compared to "Erica – moon" (and vice versa for those more proficient in

Russian), as “moon” carries masculine grammatical gender in Ukrainian and feminine in Russian.

4.3. Methods

The materials, data, and analysis codes for this study can be retrieved from the OSF link: https://osf.io/xhs9v/.

4.3.1. Pre-test

To exclude conceptual gender from the analysis and focus solely on the effects of grammatical gender, we carried out a pre-test using the methodology outlined by Sato and Athanasopoulos (2018). This was done to select conceptually neutral items for the main experiment. We recruited ten Ukrainian-Russian-English speakers (5 females; $Mean_{\text{age}}$ = 26, *SD* = 4) and ten English monolinguals (4 females; $Mean_{age}$ = 31, *SD* = 10). None of the recruited participants were involved in the main study. Participants were asked to rate 137 black-and-white object images presented one by one against a greyscale and white background to minimise any bias related to colour. The objects were rated on a Likert scale ranging from (1) "*very feminine*” to (7) “*very masculine*”. The objects were divided into five groups based on their grammatical genders in Ukrainian and Russian: (1) 20 nouns masculine in Russian and feminine in Ukrainian, (2) 24 nouns feminine in Russian and masculine in Ukrainian, (3) 31 nouns feminine in both languages, (4) 31 nouns masculine in both languages, and (5) 31 nouns neutral in both languages. The images used in the study were obtained from the Bank of Standardised Stimuli (Brodeur et al., 2014) and Snodgrass and Vanderwart (1980) database.

4.3.2. Stimuli

From the pre-test, we obtained 46 conceptually neutral objects (*Mean* = 4.04; *SD* = 0.07; *Range* = 3.85 – 4.1). These objects were then divided into three groups (Table 4.1): (1) eighteen objects with matching grammatical gender in both Russian and Ukrainian (e.g., blender – masculine in both), (2) twenty objects with mismatching grammatical genders (e.g., tray – feminine in Ukrainian, masculine in Russian), and (3) eight filler objects with neuter grammatical gender in both languages. It is important to point out that given the typological proximity and lexical overlap between Ukrainian and Russian, the presence of cognates in the stimulus set was largely unavoidable. While the pre-test included a broader mix of cognates and non-cognates, the final selection was determined exclusively based on conceptual gender neutrality, which resulted in an uneven distribution of cognates across conditions. Specifically, 89% (23 out of 26, including fillers) of nouns in the matched-gender group were cognates (e.g., "cutting board" – "*дошка*" in Ukrainian and "*доска*" in Russian; "guitar" – "*гітара*" in Ukrainian and "*гитара*" in Russian), whereas only one noun (5%) in the mismatched-gender group was a partial cognate (e.g., "parrot" – "*папуга*" in Ukrainian and "*попугай*" in Russian). Although cognate status was not systematically manipulated in this study, we recognise that it may have influenced bilingual lexical processing.

We also acknowledge the slight imbalance between the two groups of stimuli (Table 4.1). Given the constraints in selecting conceptually neutral objects with mismatching grammatical gender in both Russian and Ukrainian, we prioritised internal balance within each analysed category. Specifically, the matching grammatical gender group was designed to be balanced (9 masculine, 9 feminine), while the mismatching group was constructed with the most conceptually neutral items available, which resulted in a slight difference in total count. Importantly, in the mismatching group for the variable “Condition” (i.e., Congruent in Russian / Incongruent in Ukrainian vs. Congruent in Ukrainian / Incongruent in Russian), we examined the effects of participants’ most proficient L1 (Ukrainian or Russian) across all mismatching nouns (9 masculine-feminine and 11 feminine-masculine) combined. Therefore, the slight numerical imbalance is not expected to affect the interpretation of our results. Finally, descriptive statistics for word length across all three languages are provided in Table 4.1. These show that word lengths are comparable across stimulus types and languages.

Table 4.1

Example of the selected stimuli types

Types of stimuli	Ukrainian grammatical gender	Russian grammatical gender	Number of nouns	Example (English / Ukrainian / Russian)	English Word Length	Ukrainian Word Length	Russian Word Length
Matching grammatical gender in both L1s	Masculine		9	Wineglass / Келих / Бокал Tomato / Помідор / Помидор	Mean=6.78, SD=2.60 Range = 3-13	Mean=6.72, SD=2.30 Range = 4-13	Mean=7.11 SD=3.20 Range = 4-17
	Feminine		9	Box / Коробка / Коробка Candle / Свічка / Свеча			
Mismatching grammatical gender in both L1s	Masculine	Feminine	9	Basket / Кошик / Корзина Notebook / Зошит / Тетрадь	Mean=7.05, SD=3.69 Range = 3-16	Mean=7.35, SD=3.31 Range = 4-15	Mean=7.10 SD=3.55 Range = 3-16
	Feminine	Masculine	11	Umbrella / Парасолька / Зонт Onion / Цибуля / Лук			
Fillers	Neuter		8	Apple /Яблоко/Яблуко Feather/Перо/Перо	Mean=5.12, SD=1.25 Range = 3-7	Mean=5.12, SD=1.13 Range = 4-7	Mean=4.62 SD=1.06 Range = 3-6

Each of the nouns was paired up with either a male or female name, with names counterbalanced per participant. The names in the study were retained from the original study (see Table 4.2) due to the limited availability of names with a comparable number of syllables in both Ukrainian and Russian languages.

Table 4.2

Names used in the study by Schmidt and Boroditsky (2000)

Male names	Female names
Christopher	Christina
Daniel	Danielle
Paul	Paula
Brandon	Brenda
Eric	Erica
Karl	Karla
Claude	Claudia
Phillip	Phyllis
Harry	Harriet
Donald	Donna
Alexander	Alexandra
Patrick	Patricia

Overall, for the current experiment, four experimental conditions were established based on the two types of stimuli. For the stimuli with matching grammatical gender, two conditions were delineated based on the congruence between the biological sex of the name and the noun's grammatical gender in bilinguals' both L1s: (1) Congruent in both L1s and (2) Incongruent in both L1s. Analogously, for the second type of stimuli (with mismatching grammatical gender), we defined two conditions: (1) Congruent in Ukrainian & Incongruent in Russian and (2) Congruent in Russian & Incongruent in Ukrainian.

4.3.3. Participants

We recruited 100 Ukrainian-Russian bilinguals and 40 English monolinguals in exchange for an inconvenience allowance of £10 in the form of an Amazon voucher. After analysing the responses and linguistic profiles of the participants, our final sample consisted of 94 Ukrainian participants (70 females, $Mean_{age}$ = 32, SD = 12.1) and 38 English monolinguals (21 females, $Mean_{age}$ = 23, SD = 2.5). Participants were removed for reasons such as speaking/learning another gendered language (n = 7) or showing unusually slow reaction times between stimuli (n = 1). The imbalance between male and female bilinguals was due to data collection occurring after the start of the war in Ukraine, resulting in skewness of available sample. Nevertheless, Flaherty (2001) reported, based on statistical analysis in a sex assignment task, that while the sex of participants influenced the responses in the younger group (5- to 7-year-olds and 8- to 10-year-olds for Spanish participants, and 5- to 7-year-olds for English participants), both for Spanish and English adults the sex of the participants did not affect the choices of male or female gender for the nouns (χ^2 = .8606, *ns,* and χ^2 = 2.88, *ns,* for Spanish and English adults respectively).

Proficiency levels in Ukrainian, Russian, and English for Ukrainian participants were gauged through standardised tests. The ZNO Tests (Ukrainian Center for Educational Quality Assessment, 2020) were used to evaluate advanced language skills in Ukrainian and Russian (on a scale from C1 to C2 levels). Participants could score a maximum of 100 points for each language. To calculate a continuous variable for language proficiency, scores from the Russian proficiency test were subtracted from those of Ukrainian. Consequently, this coefficient could range from a maximum of +100, indicating exclusive proficiency in Ukrainian, to a minimum of -100, signifying exclusive proficiency in Russian. English language proficiency was measured using the Oxford Quick Placement test, OQTP (Oxford University Press, 2001) or by evaluating existing valid IELTS scores (Cambridge University Press, 2021). The minimum acceptable scores were set at 67% for the OQPT and an IELTS score of 5.5, corresponding to a B2 (upper-intermediate) proficiency level. Both groups also completed a modified Bilingual Language Profile questionnaire (BLP, Gertken et al., 2014) to determine if they spoke any other languages.

Ukrainian participants included in the analysis reported acquiring English as a foreign language at an average age of 10 (SD = 4.21) with a minimum of upper-intermediate proficiency level. Most participants showed the highest proficiency scores in Ukrainian (55%), followed by Russian (27%), and equal proficiency in both languages (18%). The proficiency scores also varied greatly ($Range_{\text{Ukrainian}}$ = 6.25-93.75, $Range_{\text{Russian}}$ = 6.25-87.5, with 100 being a maximum score), indicating the absence of ceiling effects (see Table 4.3 for more details).

Table 4.3

Proficiency scores and distribution of Ukrainian-Russian bilingual participants

Language	Mean Proficiency Score (100 maximum)	SD	Range	Percentage (Number) of Participants
Ukrainian	66.69	14.96	6.25 - 93.75	55% (52)
Russian	59.11	14.45	6.25 – 87.5	27% (25)
Equal proficiency in both	62.13	14.03	31.25 - 87.5	18% (17)

All participants were recruited either online or via posters distributed at Lancaster University. The Research Ethics Committee at Lancaster University approved the study protocol and the data collection measures.

4.3.4. Procedure

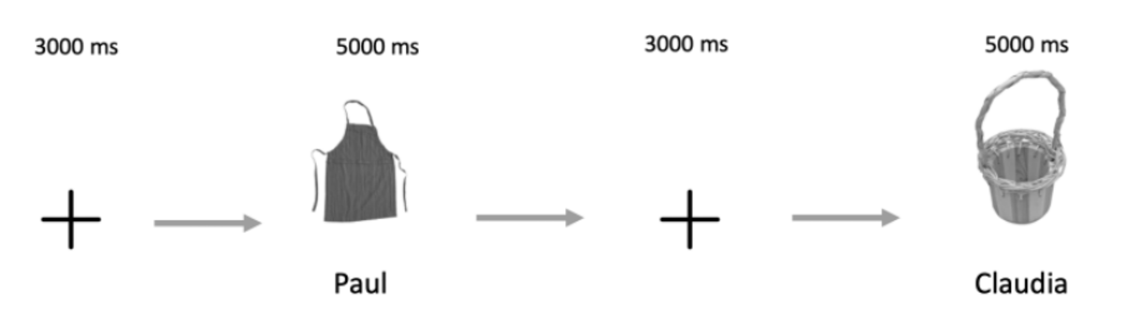
To conduct the current experiment, we used the Gorilla Experiment Builder software (Anwyl-Irvine et al., 2019). Participants were monitored online to ensure the integrity of their performance on the memory tasks.

The experiment was conducted in English and consisted of two phases: learning and testing, repeated twice. Following the study by Boroditsky and Schmidt (2000), participants were presented with the following instructions: *"For this experiment, we have given names to a bunch of objects. For example, we may have decided to call a chair 'Mary'. You will see objects and their names appear on the screen (e.g., chair Mary), and your task is to try to memorize the name we have given to each object as well as you can. Your memory for these names will be tested later in the experiment."*

Then, participants were presented with twenty-three object – name pairs. The pairs appeared on the screen for five seconds each, with the object in black-and-white presented in the centre of the screen and the name displayed below (Figure 4.1). Each object was shown only once per participant. Crucially, the gender of the name associated with each depicted noun was counterbalanced across participants. For instance, one participant might view "apple" paired with the name "Patrick," whereas another was presented the same object paired with "Patricia." This between-subjects counterbalancing ensured that each object was paired with both a masculine and a feminine name across the sample, but never more than once per participant. As such, there were no within-subject repetitions of objects, thereby minimising potential carryover effects from earlier exposures.

Figure 4.1

Example of the stimuli used in the learning phase



Afterward, participants completed an unrelated distractor task, which typically lasted 2 to 3 minutes. Since the original study did not specify the distractor task, we included a Thatcher task and a semantic priming task[4]. During the testing phase, object names were shown on the screen, and participants had to select the gender of the proper name from the learning phase (e.g., choosing between "Daniel" and "Danielle", see Figure 4.2). Since our study had twice the number of stimuli as in the original study, all participants repeated this process twice (23 pairs per trial). Each of the

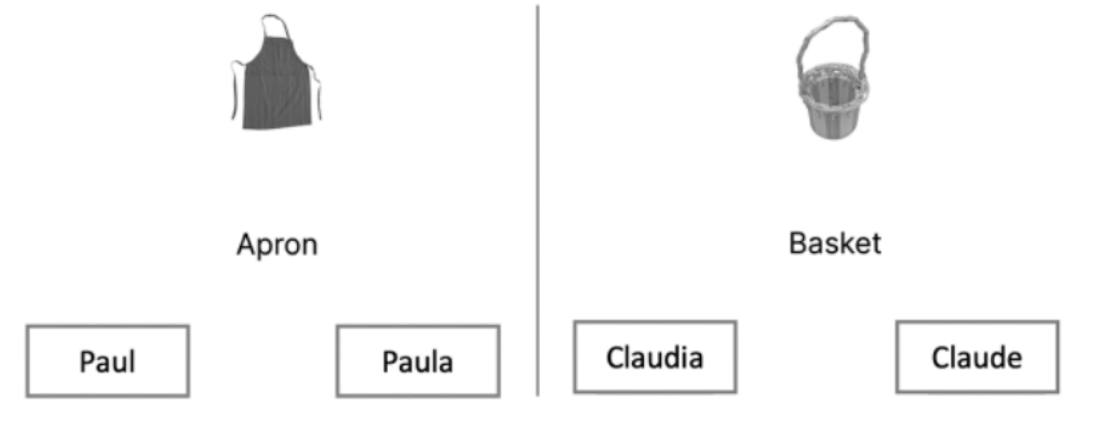
---

[4] The Thatcher task involved viewing upright and inverted faces with altered or unaltered features to assess participants' sensitivity to facial configuration. In the semantic priming task, participants were first shown a positive or negative adjective (e.g., "kind" or "cruel"), followed by a smiley face (happy or sad). They were instructed to press "F" if the face was happy and "J" if the face was sad.

names was repeated only once per trial. After the second session, participants completed language proficiency tests and a BLP questionnaire.

Figure 4.2

Example of the stimuli used in the testing phase



4.3.5. Analysis

Considering the intricacies of our study, we divided the analysis into two parts, based on the two types of stimuli. The first part focused on a comparative analysis of responses from Ukrainian-Russian bilingual and English monolingual participants, using the group of nouns that had matching grammatical gender in Ukrainian and Russian. In this part of the analysis, our aim was to replicate the findings of Boroditsky and Schmidt (2000), providing evidence that the ability to recall the human name by simultaneous bilinguals is enhanced if its biological sex is congruent with the grammatical gender of the object in question in both L1 and 2L1, compared to the English monolinguals. To achieve this, we designed a generalised linear mixed-effects (lmer) model (Linck & Cunnings, 2015) in R software (R Core Team, 2022) to determine whether the accuracy of responses (correct vs. wrong, coded as 1 and 0) was influenced by condition (Congruent in both L1s vs Incongruent in both L1s, contrast coded as 0.5 and -0.5) and the participant group (bilingual vs. monolingual, contrast coded as 0.5 and -0.5). The parsimonious model included by-participant random intercepts and slopes for condition, and by-item random intercepts. Additionally, we analysed whether there was a difference in the recall accuracy within Ukrainian-Russian bilingual group depending on the condition. For that we designed a lmer model with accuracy ("1" for accurate and "0" for inaccurate responses) as a dependent variable, and the condition as a predictor. The maximal model included random intercepts and slopes for condition both by participant and item. However, due to a singular fit and near-zero variance for participant-level random effects, the parsimonious model retained only random intercepts for items.

For the second part of the analysis looking at the group of nouns where grammatical gender was mismatching across languages, we explore further the effects of two contrasting three-gendered systems on memory and whether more proficient L1 (Ukrainian or Russian) will affect the accuracy of the responses, compared to the less proficient L1. To do this, we analysed how the accuracy of responses is affected by the interaction between Proficiency (measured from -100 to +100 for Russian and Ukrainian respectively) and Condition (Congruent in Russian & Incongruent in Ukrainian vs Congruent in Ukrainian & Incongruent in Russian, contrast coded as -0.5 and 0.5). The parsimonious model for this analysis also included a random intercept for Item to account for variability across stimuli. Participant-level variability was not included due to zero variance and convergence issues in the maximal model.

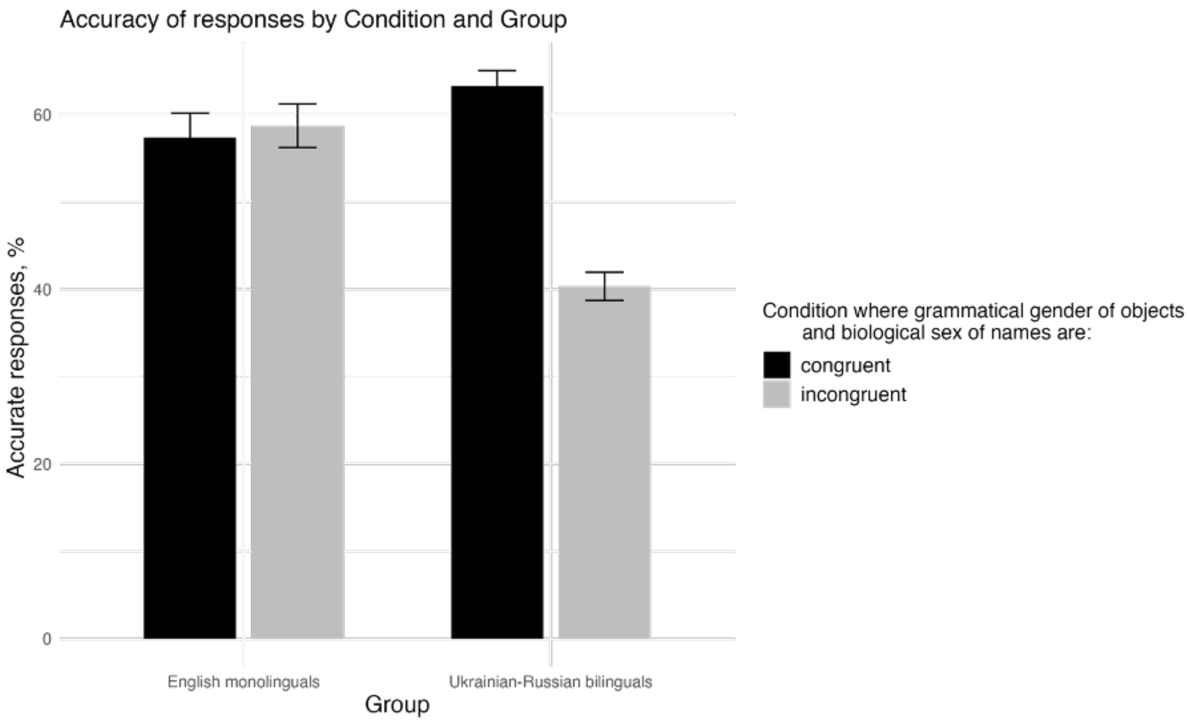
4.4. Results

4.4.1. Comparison of English Monolingual and Ukrainian-Russian Bilingual Participants

Aligning with our predictions, bilingual participants recalled names more accurately when the grammatical gender of objects in both Ukrainian and Russian languages aligned with the biological sex of the names assigned to them during the learning phase (*Mean* = 63%, *SE* = 1.75). This was compared to the accuracy of responses for objects whose grammatical gender in both languages misaligned with the biological sex of the names (*Mean* = 40%, *SE* = 1.61), as well as comparing with the responses of English monolinguals in both conditions (Figure 4.3). As expected, the English control group did not display any significant trends, regardless of whether the grammatical gender and biological sex of the object names were congruent (*Mean* = 57%, *SE* = 2.8) or incongruent (*Mean* = 58.7%, *SE* = 2.5) in Ukrainian and Russian. This suggests that congruency between the grammatical gender and biological sex of the names improves recall accuracy in Ukrainian-Russian bilinguals, compared to monolingual participants, supporting our hypothesis.

Figure 4.3

Accuracy of responses (%) based on condition (Congruent in both L1s and Incongruent in both L1s) by group (English monolinguals and Ukrainian-Russian bilinguals). Error bars indicate 95% confidence intervals.



In the linear mixed model, the dependent variable (Accuracy: correct vs wrong response) was significantly affected by both main effects of participant group (*Estimate* = -0.247, *SE* = 0.010, *z* = -2.48, *p* = .013) and the condition (*Estimate* = -0.687, *SE* = 0.277, *z* = -2.48, *p* = .013). As predicted, the interaction between group and condition was also significant (*Estimate* = -1.231, *SE* = 0.377, *z* = -3.27, *p* = .001). Additionally, the comparison of the models with and without the interaction between the predictors revealed a significant improvement in fit when the interaction term between the two variables was included ($\chi^2(1) = 10.40$, *p* = .001).

When comparing the accuracy of responses for the two conditions (Congruent in both L1s vs Incongruent in both L1s) only withing a Ukrainian-Russian bilingual group, we also found a significant effect of the condition (*Estimate* = -1.080, *SE* = 0.244, z = -4.44, $p < .001$), suggesting that bilinguals recalled significantly less accurate those pairs in the "Incongruent in both L1s" condition.

4.4.2. Comparison of Ukrainian-Russian Bilingual Participants based on Language Proficiency

For the second part of the analysis, we investigated whether bilingual participants would show improved recall of object names when those objects were paired with items whose grammatical gender in their more proficient language (Ukrainian or Russian) was congruent with the biological sex of the name. This was compared to the recall of names where the biological sex of the names was congruent with the grammatical gender of the paired object in the less proficient language.

However, no significant effects were found for the main effects of group proficiency (*Estimate* = -0.0003, *SE* = 0.003, z = -0.08, p = .935), condition (*Estimate* = -0.032, *SE* = 0.482, z = -0.07, p = .948), or group proficiency - condition interaction (*Estimate* = -0.006, *SE* = 0.006, z = -0.93, p = .352). Moreover, when comparing the two models using ANOVA, we found that the model including the two predictors did not provide a significantly better fit ($\chi^2(1) = 0.00$, p = 1.00), compared to the null model.

4.5. Discussion and conclusion

In the current study, our aim was to provide a deeper understanding of language effects on the mental representation of objects, investigate the effects of three-gendered languages, as well as introduce participants who have two three-gendered grammatical systems acquired simultaneously.

The latter would allow us to provide deeper insights into how such languages interact in a bilingual mind, as well as the effects they have on human cognition.

In the first part of our analysis, comparing the performance of Ukrainian-Russian bilingual and English monolingual groups, we confirmed our hypothesis. Particularly, we found that during the testing phase, the ability of bilingual participants to recall the names assigned to objects in the learning phase significantly improved when the grammatical gender of objects and the biological sex of the names were congruent in both of their native languages. In this part of the analysis, our results aligned with our initial predictions, notwithstanding the inclusion of neutral stimuli as fillers and our attempt to minimise conceptual relatedness by exclusively including conceptually neutral objects.

However, when analysing the results of the bilingual group using stimuli where objects had mismatching grammatical genders in Ukrainian and Russian, no effect of the more proficient language was found. This suggests that in the case of simultaneous bilinguals the effects of language may be negated when there is a misalignment in grammatical gender between the L1 and 2L1. Nevertheless, despite not finding effects for the second type of stimuli, we were able to replicate the original study (Boroditsky & Schmidt, 2000) and find similar results in simultaneous bilingual participants as those of sequential bilinguals, despite both L1 and 2L1 being three-gendered.

The discrepancy in our findings with the first and second types of stimuli, as well as with earlier studies indicating no effect of three-gendered languages (Sera et al., 2002; Vigliocco et al., 2005) can be explained in various ways. Firstly, as mentioned before, previous studies mainly compared speakers of two- and three-gendered languages with each other or with a monolingual control group. However, none of these studies investigated the potential language effects that arise when individuals acquire two three-gendered languages simultaneously. Our findings aligning with the first prediction can be attributed to the chosen first group of stimuli that had matching grammatical gender in Ukrainian and Russian. Therefore, there was no conflict or interference between the gender representations, as opposed to the second type with mismatching genders. Secondly, when interpreted through the lens of Baddeley's Working Memory Model (Baddeley & Hitch, 1974), which suggests that the central executive, phonological loop, and visuospatial sketchpad work together to process and integrate information, the enhanced recall accuracy observed in bilingual participants when grammatical gender and biological sex were congruent across both L1s suggests that converging grammatical information created stronger memory associations. This likely reduced the cognitive load on the central executive, allowing for more efficient retrieval. However, when grammatical gender and biological sex were congruent in one language and incongruent in another, the cognitive load may have increased, leading to weaker memory associations and lower recall accuracy. This is particularly supported by the fact that memory recall for these items was worse than that of the English monolingual controls (see Figure 4.3). However, an alternative explanation, is that in cases where bilingual participants were unsure of the correct answer, their guessing behaviour - whether conscious or unconscious - may have been influenced by the grammatical gender of the object presented on screen during the recall phase. In pairs where grammatical gender of the noun was matching across Ukrainian and Russian, this incidental influence could have increased the likelihood of a correct guess, inflating accuracy relative to English monolinguals, who lack such grammatical associations.

Importantly, this account is also supported by the lack of a significant reaction time difference (see Supplementary Materials in Appendix B) between bilingual Ukrainian-Russian and monolingual English participants. Specifically, if bilinguals were benefitting from more efficient memory retrieval due to gender congruency, we might expect faster reaction times in those trials. However, the absence of a reaction times advantage suggests that their improved accuracy may not stem from faster recall, but rather from a bias in guessing behaviour. Additionally, in pairs where grammatical gender mismatched across L1s, such guessing (conscious or unconscious) strategy was not accessible, as the presented objects activated both masculine and feminine gender information.

Another possibility to explain our findings, specifically the discrepancy between stimuli with matching and mismatching genders across two L1s, as suggested by an anonymous reviewer, is that the observed accuracy advantage for congruent pairs using nouns with matching gender in Ukrainian and Russian among bilingual participants may have arisen not from enhanced memory retrieval per se, but from an unconscious (or even strategic) influence of grammatical gender as a result of guessing. In other words, when bilinguals were uncertain of the correct name, they may have defaulted, consciously or unconsciously, to choosing the gendered name that was congruent with the grammatical genders across both L1s of the object presented on screen. Whereas for stimuli with mismatching grammatical genders in Ukrainian and Russian this strategy was unavailable as the object might have activated both masculine and feminine gender information, making guessing a more difficult task. Given the number of stimuli per trial (n = 23), such guessing strategies cannot be ruled out entirely. However, several aspects of our data suggest that while it could be an unconscious effect of grammatical gender, it is unlikely to be a conscious strategy or guessing.

First, none of the participants mentioned grammatical gender or language as a tool they used during the task in post-experimental debriefing. Instead, they referred to associations or visual mnemonic strategies, suggesting that any influence of grammatical gender likely occurred implicitly. Second, as mentioned above, our reaction time data revealed no significant differences between bilingual and English monolingual participants, even in trials involving congruent items. In a separate model, we also compared the response times for stimuli with matching grammatical genders across Ukrainian and Russian (conditions "Congruent in both L1s" and "Incongruent in both L1s") only for Ukrainian-Russian bilinguals and no significant differences were observed. If participants had relied on grammatical gender as a conscious cue during guessing, we would expect to see faster responses for the "congruent in both L1s" condition, particularly among bilingual group only – yet, no such pattern emerged. The absence of such effects further supports the idea that grammatical gender influenced recall at an implicit, conceptual level, rather than through deliberate response strategies. Nonetheless, we acknowledge that it is not possible to fully rule out that our findings reflect a conscious strategy or a guess in behavioural experiments (see Samuel et al., 2019). To fully separate conscious language manifestation from unconscious pre-linguistic Whorfian effects, future studies could incorporate neural measures (e.g., electroencephalography, EEG) and go beyond behavioural findings that are commonly facing such critique.

In addition, previous studies comparing the results of bilingual speakers fail to mention whether the presented stimuli were cognates or noncognates. As detailed in Section 2.2., the majority of nouns with matching grammatical gender across Ukrainian and Russian were cognates, whereas cognates were largely absent in the mismatched-gender condition. This asymmetry may have contributed to the observed differences in recall performance between conditions. While studies in linguistic relativity have not addressed the influence of cognates on grammatical gender effects in bilingual speakers, prior research on bilingual language processing shows a clear advantage of cognates when comparing the speed of translation of words by bilinguals (Degroot et al., 1994; Kroll & Mendoza, 2022; Sáchez-Casas et al., 1992; Salamoura & Williams, 2007). Therefore, we propose that the presence of cognates may have contributed to the enhancement of the language effects found in the recall of names in pairs where nouns had matching grammatical gender, compared to the mismatching ones. The latter also fits into the predictions based on the Working Memory Model (Baddeley & Hitch, 1974), specifically, because cognates facilitate the retrieval of the top-down information even more when participants categorise the stimuli.

Moreover, our findings are consistent with studies examining Polish grammatical gender (Haertlé, 2017; Maciuszek et al., 2019). This raises the question: what do these languages - Polish, Ukrainian, and Russian - have in common that sets them apart from other three-gendered languages such as Greek or German? All three are Slavic languages, with Ukrainian and Russian belonging to the East Slavic branch, and Polish to the West Slavic branch. Ukrainian and Polish have the shortest lexical distance at 30%, followed by Ukrainian and Russian at 38%, and Russian and Polish at 50% (Steinback, 2015). They also do not have articles and grammatical gender is inferred from the noun itself, contrary to Greek and German. In all three languages verbs in the past tense agree with the gender of the subjects, which is not present in many Indo-European languages, including Greek and

German. For instance, in Russian, “he went” is “*он пошёл*” (on poshol), “she went” is “*она пошла*” (ona poshla), and “it went” (for neuter) is “*оно пошло*” (ono poshlo). This broad grammatical distribution of gender marking may increase its salience and facilitate top-down processing in memory tasks, as gender is marked on prominent grammatical constituents like nouns and verbs rather than modifiers.

While Ukrainian and Russian are less gender transparent than Romance languages such as Spanish - which occupies the high end of the gender transparency continuum (Kupisch et al., 2022; Sá-Leite & Lago, 2024) and have been suggested to elicit stronger grammatical gender effects in earlier linguistic relativity studies (Sera et al., 2002) - they are more transparent than German. The latter has been argued to produce weaker gender effects in prior studies (e.g., Sera et al., 2002; Vigliocco et al., 2005), possibly due to its high opacity and the presence of neuter gender, which reduces alignment between grammatical and natural gender.

Crucially, these differences may affect how grammatical gender information becomes integrated into conceptual memory. According to the AUSTRAL model (Sá-Leite & Lago, 2024), gender can be accessed both via form-based (e.g., consistent morphological endings) and lemma-based (e.g., repeated and syntactically distributed agreement patterns) routes. However, lemma-level activation in Ukrainian and Russian may place higher cognitive demands on the language user, as gender information must be maintained and retrieved across multiple syntactic constituents - and in the absence of overt morphological cues like determiners (as in German or Spanish). This greater processing load may in turn enhance encoding into memory. In contrast, while German also stores gender at the lemma level, its more limited grammatical embedding of gender (e.g., use of invariable determiners) may lead to weaker conceptual integration. This interpretation aligns with broader findings that higher cognitive, or memory load tends to increase reliance on language as a resource (e.g., Bylund & Athanasopoulos, 2017; Winawer et al., 2007). Finally, future research could examine this proposal more directly through cross-linguistic studies comparing languages of varying transparency, and by isolating regular vs. ambiguous noun types (see Sá-Leite & Lago, 2024) within comprehension and memory paradigms.

Finally, it is also important to acknowledge the limitations of our study, such as the presence of the "*surzhyk*" dialect in the Ukrainian language. "Surzhyk" is an oral, non-standard mixed idiom that involves a blend of Ukrainian and Russian languages, and its usage could lead to mislabelling the grammatical gender of objects (Kostiučenko, 2023). To our knowledge, it is not possible to detect the usage of this dialect using proficiency tests in Ukrainian or Russian. It can only be observed in oral communication or if a participant reports it in their linguistic profile.

To conclude, potential limitations notwithstanding, bilingual participants with two distinct three-gendered grammatical systems, but not monolingual speakers of the genderless language, showed a grammatical gender effect (i.e., better recall for names congruent with the object's grammatical gender). This suggests that the effect of language on mental representations of objects might be observed even in speakers of two three-gendered languages when those objects have matching grammatical gender in both of those languages. However, proficiency did not modulate the grammatical gender effect when the objects had contrasting grammatical genders across Ukrainian and Russian. This indicates that the grammatical gender of the most proficient language of a simultaneous bilingual did not affect mental representations when it mismatched the less proficient language. Rather, the likelihood of a three-gendered grammatical system influencing memory recall may rest on the mechanisms by which top-down retrieval is facilitated, such as gender congruency and cognate status across languages.

4.6. References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Athanasopoulos, P., & Boutonnet, B. (2016). Learning grammatical gender in a second language changes categorization of inanimate objects: Replications and new evidence from English learners of L2 French. In R. Alonso Alonso (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 173–192). Multilingual Matters. https://doi.org/10.21832/9781783094837-011

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

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

Boroditsky, L., & Schmidt, L. A. (2000). Sex, Syntax, and Semantics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *22*(22). https://escholarship.org/uc/item/0jt9w8zf

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

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Cambridge University Press. (2021). *IELTS (International English Language Testing System) | Cambridge English*. https://www.cambridgeenglish.org/exams-and-tests/ielts/

Casasanto, D. (2008). *Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought* (SSRN Scholarly Paper 1402163). https://papers.ssrn.com/abstract=1402163

Chen, Z., & Faitaki, F. (2024). Effects of the French grammatical gender system on bilingual adults' perception of objects. *Bilingualism: Language and Cognition*, 1–14. https://doi.org/10.1017/S1366728924000464

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf Hypothesis and Probabilistic Inference: Evidence from the Domain of Color. *PLOS ONE*, *11*(7), e0158725. https://doi.org/10.1371/journal.pone.0158725

Degroot, A. M. B., Dannenburg, L., & Vanhell, J. G. (1994). Forward and Backward Word Translation by Bilinguals. *Journal of Memory and Language*, *33*(5), 600–629. https://doi.org/10.1006/jmla.1994.1029

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157. https://doi.org/10.3758/s13423-010-0021-5

Flaherty, M. (2001). How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. https://doi.org/10.1177/0022022101032001005

Haertlé, I. (2017). Does Grammatical Gender Influence Perception? A Study of Polish and French Speakers. *Psychology of Language and Communication*, *21*(1), 386–407. https://doi.org/10.1515/plc-2017-0019

Kaushanskaya, M., & Smith, S. (2016). Do grammatical-gender distinctions learned in the second language influence native-language lexical processing? *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *20*(1), 30–39. https://doi.org/10.1177/1367006915576830

Kirjavainen, M., Kite, Y., & Piasecki, A. E. (2020). The Effect of Language-Specific Characteristics on English and Japanese Speakers' Ability to Recall Number Information. *Cognitive Science*, *44*(12), e12923. https://doi.org/10.1111/cogs.12923

Kostiučenko, A. (2023). *Surzhyk in Ukraine: Between Language Ideology and Usage*. *1*, 15–17. https://doi.org/10.3929/ethz-b-000623475

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 843–858. https://doi.org/10.1037/0278-7393.34.4.843

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

Kupisch, T., Geiss, M., Mitrofanova, N., & Westergaard, M. (2022). Structural and phonological cues for gender assignment in monolingual and bilingual children acquiring German. Experiments with real and nonce words. *Glossa: A Journal of General Linguistics*, *7*(1), Article 1. https://doi.org/10.16995/glossa.5696

Kurinski, E., Jambor, E., & Sera, M. D. (2016). Spanish grammatical gender: Its effects on categorization in native Hungarian speakers. *International Journal of Bilingualism*, *20*(1), 76–93. https://doi.org/10.1177/1367006915576833

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

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Lucy, J. A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, *2*, 291–312.

Maciuszek, J., Polak, M., & Świątkowska, N. (2019). Grammatical Gender Influences Semantic Categorization and Implicit Cognition in Polish. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02208

Osypenko, O., Brandt, S., & Athanasopoulos, P. (2025). The influence of three-gendered grammatical systems on simultaneous bilingual cognition: The case of Ukrainian-Russian bilinguals. *Language and Cognition*, *17*, e25. https://doi.org/10.1017/langcog.2024.73

Oxford University Press. (2001). *Quick Placement Test*. Oxford University Press.

Pavlenko, A. (2016). Whorf's Lost Argument: Multilingual Awareness. *Language Learning*, *66*(3), 581–607. https://doi.org/10.1111/lang.12185

Pavlidou, T.-S., & Alvanoudi, A. (2013). Grammatical Gender and Cognition. *Versita/de Gruyter*, *2*(2), 109–124.

Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). https://escholarship.org/uc/item/31t455gf

R Core Team. (2022). *R: The R Project for Statistical Computing*. The R Project for Statistical Computing. https://www.r-project.org/

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *WIREs Cognitive Science*, *8*(6), e1440. https://doi.org/10.1002/wcs.1440

Roberson, D., Davies, I., & Davidoff, J. (2000). Color Categories Are Not Universal: Replications and New Evidence From a Stone-Age Culture. *Journal of Experimental Psychology. General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Sáchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, *4*(4), 293–310. https://doi.org/10.1080/09541449208406189

Sakarias, M., & Flecken, M. (2019). Keeping the Result in Sight and Mind: General Cognitive Principles and Language-Specific Influences in the Perception and Memory of Resultative Events. *Cognitive Science*, *43*(1), e12708. https://doi.org/10.1111/cogs.12708

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*(3), 257–275. https://doi.org/10.1017/S1366728907003069

Sá-Leite, A. R., & Lago, S. (2024). The role of word form in gender processing during lexical access: A theoretical review and novel proposal in language comprehension. *Psychonomic Bulletin & Review*, *31*(5), 1934–1953. https://doi.org/10.3758/s13423-023-02426-8

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

Sato, S., Casaponsa, A., & Athanasopoulos, P. (2020). Flexing Gender Perception: Brain Potentials Reveal the Cognitive Permeability of Grammatical Information. *Cognitive Science*, *44*(9), e12884. https://doi.org/10.1111/cogs.12884

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language*, *69*(2), 121–134. https://doi.org/10.1016/j.jml.2013.03.004

Ukrainian center for educational quality assessment. (2020). *Ukrainian center for educational quality assessment*. https://testportal.gov.ua/en/

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language*, *87*, 157–176. https://doi.org/10.1016/j.jml.2015.10.009

Vernich, L. (2017). *Does learning a foreign language affect object categorization in native speakers of a language with grammatical gender? The case of Lithuanian speakers learning three languages with different types of gender systems (Italian, Russian and German)*. https://journals.sagepub.com/doi/full/10.1177/1367006917728593

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology: General*, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Whorf, B. L. (1956). *Language, Thought, and Reality, second edition: Selected Writings of Benjamin Lee Whorf* (Caroll, J.B., Ed.). MIT Press.

Chapter 5. Early Perceptual Effects (or lack thereof) of Conflicting Grammatical Genders: ERP Evidence from Simultaneous Bilinguals

Linking statement: Empirical studies in the preceding chapters expanded our understanding of how grammatical gender in two simultaneously acquired languages can influence both categorisation and memory recall, albeit with context-dependent variations in effect magnitude. However, behavioural paradigms, including those used in Chapters 3 and 4, are inherently limited in their ability to distinguish between unconscious perceptual effects and conscious, metalinguistic processing. To address this, the following chapter turns to neural methods, using event-related potentials (ERPs) to explore whether grammatical gender congruency is accessed automatically during early stages of perception. This shift allows for a more fine-grained examination of whether the cognitive influence of grammatical gender in simultaneous bilinguals operates at a pre-attentive level.

5.1. Abstract

Previous studies revealed that grammatical gender may shape perception and categorisation (Sato et al., 2020), yet the extent of this influence remains unclear, specifically in simultaneous bilinguals with two gendered languages. This pre-registered ERP study investigates how two partially contrasting grammatical gender systems modulate perception in Ukrainian-Russian bilinguals. Participants completed a non-verbal categorisation task assessing associations between primes (depicted nouns with matching or mismatching grammatical genders across L1s) and target (male/female faces). Behavioural results (response types and reaction times) showed that bilinguals were not affected by the prime-target gender congruency for matching primes or by more dominant/proficient L1 for mismatched primes. ERP analyses showed no significant modulations of predicted components (N1, P2/VPP, N300) by grammatical gender for either type of primes. These findings suggest that grammatical gender alone may not independently modulate categorisation process in bilinguals with two gendered L1s, especially when explicit conceptual or semantic activation is not required.

5.2. Introduction

Over the past few decades, linguistic relativity studies have grown exponentially, leading to more nuanced investigations of how language affects perception and cognition. Although linguistic relativity research has extensively addressed *lexical* properties and their effect on thought and perception (e.g., colour discrimination; Athanasopoulos et al., 2010; Drivonikou et al., 2007; Gilbert et al., 2006; Roberson et al., 2005; Thierry et al., 2009; Winawer et al., 2007; spatio-temporal metaphors;

Athanasopoulos & Bylund, 2023; Bylund & Athanasopoulos, 2017), Whorf (1956) himself focused on the effects that *grammatical* properties have on one's thought. One grammatical property that has received a lot of attention is grammatical gender (see Fatemi, 2024; Samuel et al., 2019 for most recent reviews). The early studies in this line of research focused on finding a yes/no answer to whether grammatical gender of one's native language affects thought, and mostly recruited monolingual speakers (Flaherty, 2001; Sera et al., 1994; 2002; Vigliocco et al., 2005). However, recently these studies have expanded the scope of research and included bilingual speakers to investigate if these grammatical gender effects on perception remain when participants are tested in a genderless L2 (e.g., Chen & Faitaki, 2024; Pavlidou & Alvanoudi, 2013, 2019; Sato & Athanasopoulos, 2018), or whether learning a gendered L2 (while having a genderless L1) can result in a shift in perception (Athanasopoulos & Boutonnet, 2016; Kurinski & Sera, 2011), as summarised in more detail below.

5.2.1. Grammatical gender effects on perception: behavioural evidence across bilinguals

Research involving bilinguals with a gendered L1 and genderless L2 predominantly shows that the influence of grammatical gender from the native language persists even when participants perform tasks in a genderless L2 (cf. Sato et al., 2013). Conversely, studies involving bilingual participants with a genderless L1 and a gendered L2 reveal that exposure to a gendered L2 begins to affect object categorisation as early as 10 weeks after learning commences (Athanasopoulos & Boutonnet, 2016; Kurinski & Sera, 2011). Additionally, Athanasopoulos and colleagues showed that individual differences (e.g., length of residence, language use, etc.) can influence bilinguals' categorical perception linked to grammatical constructs (Athanasopoulos, 2006; Athanasopoulos & Bylund, 2023). For example, looking at grammatical number, Athanasopoulos (2006) showed that increased language proficiency, as measured by a standardised proficiency test, led to shifts in categorisation patterns among Japanese-English bilinguals. Specifically, participants with intermediate English proficiency displayed categorisation patterns resembling those of Japanese monolinguals, whereas participants with advanced English proficiency exhibited patterns similar to English monolinguals. Yet, in the domain of grammatical gender, the role of individual differences remains under explored (see however, Osypenko et al, 2025).

An important consideration when comparing findings across languages and experimental paradigms is how bilinguals' language profiles are assessed. Earlier studies examining the effects of grammatical gender in bilingual populations predominantly relied on self-reported proficiency (e.g., Bassetti, 2007; Boutonnet et al., 2012; Sato et al., 2020), while a number of more recent investigations (e.g., Athanasopoulos & Bylund, 2023; Chen & Faitaki, 2024; Kurinski & Sera, 2011; Osypenko et al., 2025) provide a more comprehensive assessment. While self-assessed measures yield valuable insights into participants' linguistic backgrounds - such as those captured by the Bilingual Language Profile (BLP; Gertken et al., 2014), which assesses language history, use, self-rated proficiency, and attitudes - they may not always provide a fully accurate representation, particularly in populations with complex language histories such as Ukrainian-Russian bilinguals. Standardised proficiency tests, therefore, offer crucial objective data regarding bilinguals' language profiles. By combining standardised and self-rated measures, researchers can obtain critical information on both the standardised levels of language proficiency

- found to influence grammatical gender effects (Athanasopoulos, 2006; Athanasopoulos & Bylund, 2023; cf. Chen & Faitaki, 2024) - and self-reported measures, such as age of acquisition, which have also been shown to impact bilinguals' perception of gendered stimuli (Athanasopoulos et al., 2010; Boroditsky et al., 2003).

Importantly, existing behavioural research has primarily focused on sequential bilinguals (Boutonnet et al., 2012; Forbes et al., 2008; Sato et al., 2020; Sato & Athanasopoulos, 2018 among others), with limited evidence of grammatical gender effects in simultaneous bilinguals (Bassetti, 2007; Osypenko et al., 2025). Inclusion of simultaneous bilinguals can provide further insights not only whether there are effects of grammatical gender on perception of speakers of two gendered languages (L1s) but also into how two grammatical systems interact with each other in one mind. Early behavioural research on simultaneous bilinguals (i.e., Bassetti, 2007) used stimuli exclusively with mismatching grammatical gender assignments across bilinguals' two languages, making it unclear if the lack of grammatical gender effects was due to co-activation of two opposite grammatical genders or other methodological reasons (e.g., small sample size).

To address this, Osypenko et al. (2025) conducted two behavioural experiments examining grammatical gender effects on categorisation in Ukrainian-Russian simultaneous bilinguals, while tested in a genderless L2 (English). One of the key reasons to employ simultaneous bilinguals of Ukrainian and Russian languages in particular is their typological distinctions. Both languages are three-gendered and assign masculine, feminine, and neuter grammatical gender to nouns (Budzhak-Jones, 1997). Importantly, while some nouns have the same grammatical gender across both languages (e.g., "a pencil" tis masculine in both Ukrainian ["olivets"] and Russian [“karandash”]), others differ in their grammatical gender assignment (e.g., “a notebook” is feminine [“tetrad”] in Russian and masculine [“zoshyt”] in Ukrainian). This feature allows us to test the effects of grammatical gender when they match or mismatch across two L1s.

Osypenko et al. (2025) employed a similarity judgement paradigm, where participants are presented with object-character pairs and asked to rate how similar they are on a 9-point Likert scale. The characters paired with the objects were of either male (e.g., a prince) or female biological sex (e.g., a queen). Stimuli were chosen to be conceptually neutral, and participants’ language proficiency in Ukrainian, Russian, and English was assessed using standardised tests. Experiment 1 included object nouns from all three grammatical gender categories (masculine, feminine, and neuter), while Experiment 2 included only masculine and feminine object nouns. This manipulation aimed to evaluate whether the presence of neuter grammatical gender weakened associations between grammatical and biological gender, potentially diminishing grammatical gender effects. Consistent with this hypothesis, Experiment 1, which included neuter nouns, revealed no significant grammatical gender effects. In contrast, Experiment 2, without neuter-gender nouns, produced significant grammatical gender effects. Specifically, bilingual participants rated object-character pairs as significantly more similar when the grammatical gender of the noun and the biological sex of the character were congruent across both Ukrainian and Russian. Moreover, for nouns with mismatching grammatical gender, participants’ similarity judgements aligned closely with grammatical gender in their more proficient language.

5.2.2. Limitations of behavioural findings

A crucial limitation of all behavioural studies, including those discussed above, is the inability to determine whether observed effects genuinely reflect perceptual changes driven by grammatical gender or whether they reflect participants' strategic use of explicit metalinguistic knowledge (Samuel et al., 2019; Sedlmeier et al., 2016). For instance, studies that use voice attribution (Bassetti, 2007) or similarity judgement (Osypenko et al., 2025) have been characterised by Samuel et al. (2019) as having high gender/sex salience, making it possible for participants to consciously engage grammatical gender as a task-solving strategy. Additionally, such studies are particularly vulnerable to verbal interference, where participants may implicitly or explicitly rely on verbal information (see Athanasopoulos & Casaponsa, 2020). Consequently, findings from behavioural studies often reflect language activation in the moment of task completion rather than providing evidence supporting long-lasting effects of language on perceptual encoding, which stem from neural adaptations over time.

To address this issue, researchers have increasingly turned to neural measures, particularly event-related potentials (ERPs). ERPs offer high temporal resolution, enabling precise tracking of automatic and unconscious cognitive processes that are time-locked to specific stimuli. As Thierry et al. (2024) point out, most language effects on cognition occur outside of conscious awareness, but to fully assess the strength of these effects, two key research design elements are needed. First, experimental paradigms should focus on non-verbal tasks that examine perception and conceptualisation mechanisms, as these processes are expected to be influenced by language. Second, it is crucial to incorporate neural methods that can track the temporal unfolding of effects. By differentiating between effects arising at early perceptual stages and those occurring later in the processing stream - when linguistic activation may influence processing - ERPs help distinguish unconscious language effects at the pre-attentive and pre-verbal stages of processing from the overt language use (Casaponsa et al., 2024; Maier & Abdel Rahman, 2024; Thierry et al., 2009; Xue & Williams, 2024).

5.2.3. Grammatical gender effects on perception: ERP evidence and ERP components related to grammatical gender processing

While ERPs have been widely used in linguistic relativity research on domains such as colour (see review by Thierry, 2016) and motion events (Flecken et al., 2015), EEG studies examining the influence of grammatical gender on perceptual categorisation remain comparatively scarce (but see Boutonnet et al., 2012; Sato et al., 2020). In the study by Boutonnet et al. (2012), Spanish-English bilinguals and English monolinguals completed a semantic categorisation task involving triplets of depicted objects in English. Participants had to determine whether the third object (target) belonged to the same semantic category as the first two objects, while ERPs were being recorded. Critically, the target's grammatical gender in Spanish was covertly manipulated, being either gender-congruent or gender-incongruent with the first two objects. Grammatical gender congruency was expected to modulate the left anterior negativity (LAN), a marker of morphosyntactic processing (Friederici et al., 1993; Friederici & Jacobsen, 1999) in the bilingual group only. The presence of LAN would indicate automatic and unconscious retrieval of grammatical gender, rather than a strategic use of such information. As predicted, in the Spanish-English bilingual group, but not the English monolingual group, LAN showed a statistically significant effect of gender congruency, being more negative in gender-incongruent trials compared to congruent ones, providing evidence in favour of spontaneous and unconscious retrieval of grammatical gender information in bilinguals regardless of semantic category.

Similarly to Boutonnet et al. (2012), Sato et al. (2020) investigated the effects of grammatical gender on perception and categorisation but shifted the focus from semantic relatedness to conceptual gender. French-English bilinguals and English monolinguals were presented with object-face pairs and tested in English. Each object had both a grammatical gender in French (masculine or feminine) and a conceptual gender (stereotypically male or female). Each object was paired with either a male or female face, resulting in four conditions: (1) conceptually related and grammatically congruent (e.g., “poupée” [doll], grammatically feminine and conceptually female, paired with a female face); (2) conceptually related and grammatically incongruent (e.g., “collier” [necklace], grammatically masculine and conceptually female, paired with a female face); (3) conceptually unrelated and grammatically congruent (e.g., “cravate” [tie], grammatically feminine and conceptually male, paired with a female face); and (4) conceptually unrelated and grammatically incongruent (e.g., “cigare” [cigar], grammatically masculine and conceptually male, paired with a female face). Participants were tasked with responding “yes” or “no” to whether the object made them think of the face, while their EEG data were recorded. N1, P2/VPP, and N300 components were examined to assess whether grammatical gender influenced perceptual processes in bilingual speakers during early face processing, even though it was irrelevant to the task. The N1 component, an early negative-going ERP waveform, has two distinct subcomponents: anterior (N1a) and posterior. Sato et al. (2020) focused on the anterior N1 (N1a) subcomponent, which typically peaks over fronto-central electrodes around 100 milliseconds after stimulus onset. This subcomponent is associated with the initial sensory processing of visual or auditory information, as well as perceptual expectations and evaluations (Marzecová et al., 2018; Vogel & Luck, 2000). The N1 component is known to be modulated by the expectations of concurrent stimuli (e.g., Federmeier & Kutas, 2001; Lee et al., 2012). In their study, Sato et al. (2020) found that grammatical gender congruency modulated N1 amplitude for bilingual participants but not for monolinguals. Specifically, the greater negativity was found for grammatically incongruent trials compared to congruent ones, suggesting an early, automatic attentional shift driven by grammatical gender. Interestingly, conceptual gender relatedness did not significantly influence the N1 component in either participant group, indicating that this component was solely sensitive to grammatical, rather than conceptual, gender information.

Another component examined in the study by Sato et al. (2020) is the P2 or vertex positive potential (VPP). This positive-going potential is observed around 150 milliseconds post-stimulus, particularly at anterior-central sites (Jeffreys, 1989; Joyce & Rossion, 2005). The P2/VPP is widely regarded as a key ERP marker for categorical face perception (i.e.,. N170 counterpart when bi-mastoids are use as reference electrodes); it is thought to reflect the structural encoding of faces and is modulated by categorical aspects such as facial expression, race, and gender (Ito & Urland, 2005; Kecskés-Kovács et al., 2013). Sato et al. (2020) found that bilinguals, but not monolinguals, showed greater P2/VPP positivity for grammatically congruent object-face pairs compared to incongruent pairs, suggesting facilitated early face processing when grammatical gender matched the biological sex of the face.

Lastly, Sato et al. (2020) predicted an N300 amplitude modulation in the later stages of visual processing. The N300, a late visual negative-going component, has an anterior and central-parietal region distribution, peaking approximately 300 milliseconds after the onset of a visual stimulus (Holcomb & McPherson, 1994; Kumar et al., 2021). The N300 has been linked with mechanisms of perceived object identification, particularly when semantic expectations are violated (Federmeier & Kutas, 2001) and is sensitive to the global features of visual stimuli (McPherson & Holcomb, 1999; Schendan & Kutas, 2002, 2003). This component typically shows greater negative amplitude for images that conflict with expected semantic categories compared to those that align with them. As expected, Sato et al. (2020) found N300 modulations for monolinguals with conceptually unrelated stimuli showing greater negativity than conceptually related stimuli pairs. However, bilingual speakers failed to show these modulations. Instead N300 component was modulated by grammatical gender congruency. Hence, grammatical gender seems to have exerted a stronger influence than conceptual gender stereotypes, suggesting that grammatical gender overrides conceptual associations.

Beyond the ERP components examined by Sato et al. (2020), the current study also sought to investigate an additional ERP effect associated with early gender information processing. Previous ERP evidence suggested that gender discrimination of faces can occur as early as 45 ms after stimulus onset, regardless of the task demands, i.e., whether the identification is intentional or incidental (Mouchetant-Rostaing et al., 2000; Mouchetant-Rostaing & Giard, 2003). Mouchetant-

Rostaing et al. (2000) examined early ERP responses to explicit and implicit gender processing using a forced-choice gender categorisation task (i.e., participants explicitly identified the gender of a face) and a passive viewing condition (i.e., participants viewed stimuli without performing a task). Participants were presented with images of male and female faces, as well as control stimuli (e.g., hands), while their neural responses were recorded. The study found significant ERP differences at mid-parietal electrodes (i.e., O1, PO3, T5, P3, CP1, POz, and Pz) between 45 ms and 70 ms post-stimulus. These early ERP effects were interpreted as either an automatic gender distinction between faces or, alternatively, an early, broad categorisation process distinguishing faces from non-face stimuli based on visual characteristics. Importantly, this effect was observed regardless of whether gender identification was intentional or incidental, indicating that gender discrimination occurs rapidly and automatically in early visual processing stages. These findings were later corroborated by Mouchetant-Rostaing and Giard (2003), who observed similar effects between approximately 45 and 90 ms after stimulus onset across all conditions, comparing processing of age and gender of faces. Mouchetant-Rostaing and Giard (2003) suggested that these findings, in combination with earlier studies on humans and monkeys, may reflect early, task-independent perceptual processes for the global extraction of visual differences (VanRullen & Thorpe, 2001). Furthermore, when considered alongside their previous results, these early ERP effects may indicate coarse, automatic, low-level categorisation processes that allow for rapid differentiation between broad stimulus categories. Such rapid visual processes challenge traditional models of visual system organisation and may involve feed-forward pathways that quickly reach higher-order visual cortical areas, acting in parallel (Mouchetant-Rostaing & Giard, 2003).

Overall, previous research using ERPs suggests that grammatical gender is unconsciously activated even in tasks when it is not explicitly required. However, to our knowledge, no ERP study has explored these early categorical effects in bilingual speakers with two gendered L1s. Furthermore, a methodological question arises from previous findings. Specifically, were the effects amplified by the presence of conceptual gender (Sato et al., 2020) or semantic associations (Boutonnet et al., 2012)? If so, would grammatical gender still be accessed if the stimuli were conceptually and semantically neutral? Building on the statement by Samuel et al. (2019, p. 1773), while studies like Sato et al. (2020) and Boutonnet et al. (2012) show evidence in favour of access to grammatical information when it is not required, “this is not the same as demonstrating that objects are conceptualised as more masculine or feminine as a function of their gender assignment. Such results might be explained in terms of an effect of membership in the same grammatical category, independently of biological sex information”. Therefore, designing an EEG study looking at the unconscious access to grammatical gender on its own without any additional manipulations, could provide insights to answer these questions.

5.2.4. Aims and scope of the current study

In the current pre-registered study (https://osf.io/2vr7k/), we recorded ERPs in Ukrainian-Russian simultaneous bilinguals as they were completing an association judgement task involving an object and a gendered face (male or female). The task was completed fully in English (participants' genderless L2) to avoid any associations with grammatical gender of L1. The experimental paradigm was adapted from Sato et al. (2020). However, unlike Sato et al. (2020), where participants judged a set of stereotypical gender-associated objects, our aim was to analyse whether grammatical gender effects persist in both behavioural and neurophysiological measures when using conceptually neutral items.

Two key research questions were formulated. First, we sought to determine whether the effects of matching grammatical gender across two L1s would be present in both behavioural (mean accuracy and reaction times) and ERP measures (N1, P2/VPP, N300, and early gender discrimination effect). Specifically, we predicted faster reaction times and/or gender-biased choices in pairs where the prime's grammatical gender in the two L1s is congruent with biological gender of the face (e.g., faster reaction times when a "pencil", masculine in both Ukrainian and Russian, is paired with a male face), compared to incongruent pairs. Additionally, we predicted to find decreased N1 and N300 amplitudes, as well as a greater P2/VPP modulation, for pairs where the grammatical gender of the object is congruent with the biological gender of the face in the two L1s, compared to incongruent pairs. For the early ERP effect (Mouchetant-Rostaing et al., 2000; Mouchetant-Rostaing & Giard, 2003), we expected to observe greater early negativity for incongruent pairs, compared to the congruent ones.

Second, we aimed to investigate potential L1 effects when grammatical gender of an item differs across languages (Bassetti, 2007). Specifically, we hypothesised that participants would exhibit priming effects of the grammatical gender of their more proficient/dominant L1 (Ukrainian or Russian). For behavioural measures, we expected to observe faster reaction times and/or biased choices towards the pairs where the prime's grammatical gender and the target's biological gender are congruent in participants' more proficient/dominant L1. If grammatical gender in gendered-languages impacted categorisation, we would expect modulation of N1, P2/VPP and N300 for the most proficient/dominant L1s.

Furthermore, for each research question we examined effects of English proficiency/dominance on participants' responses, based on earlier findings (Athanasopoulos, 2006; Athanasopoulos & Bylund, 2023). We hypothesised that higher English dominance/ proficiency would reduce grammatical gender effects, reflecting the influence of a genderless L2 (English). Specifically, we predicted that participants would exhibit weaker grammatical gender effects reflected in both reaction times and response types (i.e., smaller differences in reaction times/responses between congruent and incongruent pairs). For each ERP component, amplitudes for congruent pairs were expected to be less pronounced in participants with higher English proficiency/dominance, compared to those with lower English proficiency/ dominance.

In summary, the present study investigates the role of grammatical gender in shaping cognitive and perceptual processes in simultaneous Ukrainian-Russian bilinguals using both behavioural and neurophysiological measures. By examining congruent and incongruent grammatical gender conditions across two distinct grammatical systems, this study aims to disentangle the effects of grammatical gender from potential covert verbal strategies and contribute to the broader understanding of linguistic relativity in simultaneous bilingual cognition.

5.3. Methods

5.3.1. Participants

26 Ukrainian-Russian simultaneous bilinguals[5] with English as a foreign language (L2) participated in the current experiment at Lancaster University. Proficiency levels were measured using standardised tests, such as the advanced university-entry level proficiency tests for Ukrainian and Russian proficiency (Ukrainian Center for Educational Quality Assessment, 2020) and the Cambridge English Proficiency test (Cambridge University Press, 2024). Proficiency test scores were measured on a scale from 0 to 100. Additionally, participants completed the Bilingual Language Profile (Gertken et al., 2014) to determine their language dominance in all three languages. To adapt the BLP questionnaire to the current socio-political context in Ukraine, we excluded the "language attitudes" section, while retaining language use, language history, and self-assessed proficiency. Participant responses were recorded on a scale of 0 to 194. Proficiency tests and BLP questionnaire were completed prior to the main testing. Details of the participants' language profiles are provided in Table 5.1. Following data collection, we calculated scores for Ukrainian and Russian separately, and then derived two coefficients to reflect relative proficiency and

---

[5] Please note that the pre-registration protocol specified a target sample size of 30 participants, while also acknowledging that recruitment would rely on a convenience sample. At the time of data collection, the available population of eligible Ukrainian-Russian bilinguals in Lancashire, UK had significantly decreased compared to the time of pre-registration. Therefore, as a result the final sample included 26 participants.

dominance. Specifically, the L1 Proficiency coefficient was calculated by subtracting the Russian proficiency score from the Ukrainian proficiency score, resulting in values ranging from –100 (only proficient in Russian) to +100 (only proficient in Ukrainian). Similarly, the L1 Dominance coefficient ranged from –194 (fully dominant in Russian) to +194 (fully dominant in Ukrainian), based on the difference between language dominance scores in the BLP.

All participants reported little to no knowledge of additional gendered languages (two participants reported learning German and Polish at school, but not using it at the time of testing; another participant reported using learning applications to recreationally learn Spanish, German and Italian, but chose 0% of time in the "language use" section).

As reported in the pre-registration (https://osf.io/2vr7k/), we excluded data from participants who exhibited response biases, such as consistently responding "yes" or "no" or showing alternating patterns (e.g., "yes–no–yes–no") for 90% or more of their responses within a block. Data were also excluded due to poor electroencephalogram (EEG) signal quality, heavy artifact contamination (see "EEG analysis"), or fewer than 25 usable trials per condition after artifact rejection. This led to the exclusion of 7 participants (2 due to response biases and 5 due to poor EEG signal quality), resulting in a final sample of 19 Ukrainian-Russian bilinguals (18 females; $Mean_{age} = 39.0$, $SD = 8.70$, $Range = 22–50$). All participants had normal or corrected-to-normal vision. Informed consent was obtained prior to the experiment, and participants received an inconvenience allowance of £10 in the form of an Amazon voucher. The study protocol was approved by the Ethics Committee at Lancaster University (reference number FASSLUMS- 2023-3489-AmendPaper-1).

Table 5.1

Participants' language profiles

	Ukrainian			Russian			English		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**	**Mean**	**SD**	**Range**
Language Proficiency	40.40	13.96	25–67	41.70	10.87	17–67	58.20	19.87	24–92
Language Dominance	105.45	31.08	41–148	118.05	26.49	52–164	56.90	20.02	15–87

Note. The maximum score for Language Proficiency across all languages was 100, for Language Dominance for Ukrainian and Russian – maximum was 194 (without language attitudes) and for English - 218 (with language attitudes)

The study was advertised via fliers and emails sent out to Ukrainian-Russian speakers living in Lancashire, UK. A convenience sampling strategy was employed due to the specific linguistic and demographic characteristics of the target population. Given the unique profile of this group and the logistical constraints associated with accessing this community, the sample size was limited. The predominance of female participants (*n* = 18) can be attributed to the timing of data collection, which occurred after the onset of the war in Ukraine. However, based on the analysis by Flaherty (2001) and Osypenko et al. (2025), the gender imbalance is not expected to have a significant impact on the findings, as both studies found no significant differences between responses from male and female participants, suggesting that their gender did not affect the ratings.

5.3.2. Materials

The stimuli consisted of 320 image pairs, each featuring an object alongside either a male or female face (a complete list of materials is available on the Open Science Framework: https://osf.io/2vr7k/). For this study, we used the conceptually neutral stimuli from Osypenko et al. (2025), originally normed for a behavioural study. However, given the increased trial requirements for ERP research, we conducted an additional pre-test to obtain a larger set of stimuli.

5.3.2.1. Pre-test

A pre-test following the approach of Sato and Athanasopoulos (2018) and Osypenko et al. (2025) was carried out to select additional conceptually neutral stimuli. Fourteen Ukrainian-Russian bilinguals (10 females; $Mean_{age}$ = 24.0, *SD* = 4.33, *Range* =18-33) and eleven English monolinguals (7 females; $Mean_{age}$ = 28.0, *SD* = 3.32, *Range* = 23-33) were recruited online. None of these participants took part in the main experiment. Participants rated a set of 200 black-and-white images of inanimate objects and animals on a 7-point Likert scale for conceptual gender association (1 = "*Very feminine*", 7 = "*Very masculine*"). Of the 200 images, 80 depicted items with masculine grammatical gender in both Ukrainian and Russian, 80 depicted items with feminine grammatical gender in both languages, and 44 depicted items with differing grammatical gender between Ukrainian and Russian. Specifically, 28 items were feminine in Ukrainian but masculine in Russian, and 16 items were masculine in Ukrainian but feminine in Russian. Based on these ratings, 34 additional conceptually neutral images were selected (*Mean* = 4.48, *SD* = 0.98), which, combined with the 46 images from our previous study (Osypenko et al., 2025), resulted in a total of 80 conceptually neutral stimuli.

These nouns were divided into two groups (Table 5.2). The first group included 40 items with matching grammatical gender across both languages, such as “strawberry”, which is feminine in both Ukrainian and Russian, and “telephone”, which is masculine in both languages. The second group consisted of 40 items with mismatching grammatical gender across Ukrainian and Russian, such as “moon” (masculine in Ukrainian and feminine in Russian) or “sock” (feminine in Ukrainian and masculine in Russian). Full details about stimulus presentation and pairings are provided in the next section.

5.3.2.2. Main testing

Eighty depicted black-and-white objects selected during the pre-test were each paired with an image of a female face and with a male face resulting in 160 pairs. To prevent repetition effects, an additional set of images of the selected eighty objects was included and paired with a different male and a different female face. This resulted in 320 unique pairs in total. The images depicting eighty nouns were only presented once per block (four separate experimental blocks, 80 critical trials per block). Congruency between grammatical gender and biological sex of the face was fully counterbalanced across four experimental conditions (see Table 5.2): (1) 80 grammatically congruent trials in both Ukrainian and Russian (CR-CU), (2) 80 grammatically incongruent in both languages (IR-IU), (3) 80 grammatically congruent in Ukrainian and incongruent in Russian (IR-CU), and (4) 80 grammatically congruent in Russian and incongruent in Ukrainian (CR-IU). To simplify condition labels throughout the text, we use the following abbreviations: C = congruent, I = incongruent, R = Russian, and U = Ukrainian. For example, CR–IU refers to trials where the noun's grammatical gender is congruent with the face's sex in Russian but incongruent in Ukrainian.

Additionally, 8 conceptually gendered nouns were selected as fillers: 4 objects rated as conceptually masculine (*Mean* = 4.93, *SD* = 0.18) and 4 conceptually feminine (*Mean* = 3.07, *SD* = 0.67). These fillers resulted in a total of 32 conceptually gendered trials, which were equally distributed across the blocks (8 filler trials per block). Each object appeared only once in each block, with block order counterbalanced across participants. All trials were randomised individually for each participant. This procedure was repeated twice using different male and female faces, and a different picture of each object. In addition, 8 pairs with conceptually gendered nouns (e.g., "lipstick", "briefcase") were used for practice trials. The purpose of the practice trials was to encourage participants to focus on stereotypical associations rather than on grammatical gender.

Table 5.2

Examples of Stimuli based on Grammatical Gender Congruency and Incongruency in Ukrainian and Russian

Stimuli type		Grammatical gender in 2L1s	English translation	Ukrainian translation	Russian translation	Conceptual gender
1. Target stimuli	1.1. Grammatical gender matching across both languages	a) masculine in both	garlic	часник	чеснок	neutral
		b) feminine in both	cello	віолончель	віолончель	neutral
	1.2. Grammatical gender mismatching across both languages	a) masculine in Ukrainian, feminine in Russian	tent	намет	палатка	neutral
		b) feminine in Ukrainian, masculine in Russian	sock	шкарпетка	носок	neutral
2. Filler stimuli	2.1. Grammatical gender matching across both languages	a) masculine in both	convertible	кабріолет	кабриолет	male
		b) feminine in both	cooking pot	каструля	кастрюля	female
	2.2. Grammatical gender mismatching across both languages	a) masculine in Ukrainian, feminine in Russian	hat	капелюх	шляпа	female
		b) feminine in Ukrainian, masculine in Russian	axe	сокира	топор	male

5.3.3. Procedure

Prior to testing, we assessed participants' language proficiency and language dominance in Ukrainian, Russian, and English. On the testing day, all experimental sessions were conducted in English to minimise any cues suggesting the experiment's connection to participants' native languages. Participants were tested individually in a dimly lit booth, seated approximately 80 cm from a 17' CRT monitor. Stimuli were presented using E-Prime software (Version 2.0) foveally in a greyscale against a grey background. Responses to the target face were recorded via a button box while EEG activity was monitored.

Following the procedure of Sato et al. (2020), each trial began with a 1000 ms pre-stimulus fixation point presented at the centre of the screen, followed by the object image displayed for 500 ms (Figure 5.1). Then, a 500 ms interstimulus interval (ISI; blank screen) preceded the presentation of the target face image. Participants were instructed to respond whether the object made them think of the face, by pressing a “yes” or “no” button located on the outer edges of the button box. The target face remained on the screen until the participants’ responses were registered or 3000 ms had elapsed. While participants were not explicitly instructed to respond quickly, a feedback display prompted them to react faster if no response was registered within the allotted time. Each trial finished with a green prompt (+) for 1500 ms where participants were encouraged to blink if needed to minimise eye movements. A self-timed break was provided every 22 trials, and eight practice items were presented prior to the main experiment task.

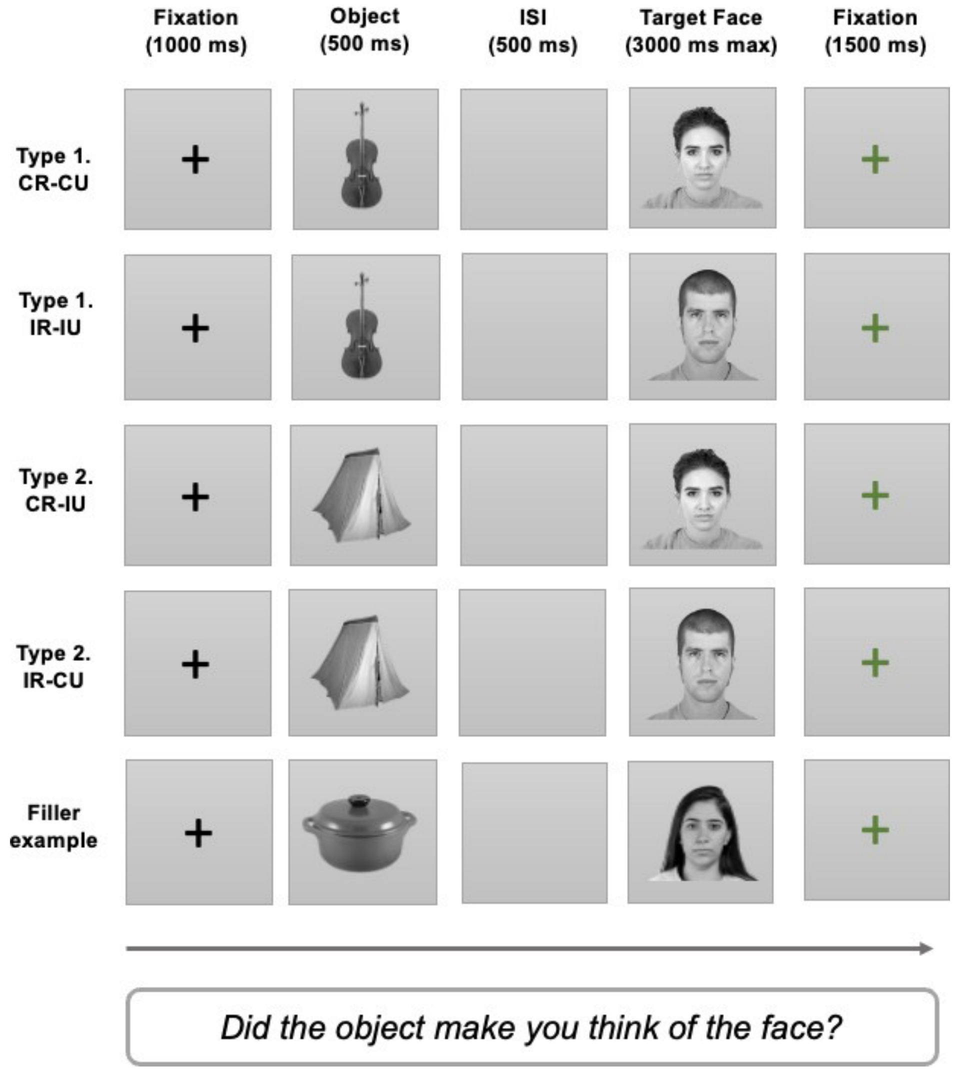
A debriefing session followed the main experimental task to assess participants’ comprehension of the experiment’s purpose and to determine whether they had relied on grammatical gender as a strategy for task completion. Participants who reported explicit awareness of the study’s objective would have been excluded from the analysis. However, no exclusions were necessary, as none of the participants identified the underlying aim.

5.3.3.1. Electroencephalography recording

EEG recordings were obtained using Neuroscan Curry7 software (Compumedics Neuroscan, NuAmps amplifier Compumedics, Charlotte, NC, USA) attached to a 38-channel elastic cap (Electro-Cap International, n.d.) positioned according to the standard 10-20 system. The electrode configuration included the following sites: FP1, FP2, Fz, F3, F4, F7, F8, FCz, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, Oz, O1, O2, POz, PO1, PO2, two additional electrodes to capture the horizontal eye movements, two off-line bi-mastoid reference electrodes, and the ground electrode. Two additional electrodes were used above and below the right eye to capture blinks and vertical eye movements. EEG recordings were amplified and digitised with NuAmps amplifier (Compumedics Neuroscan NuAmps amplifier Compumedics, Charlotte, NC, USA) at a sampling rate of 1 kHz. Electrodes were referenced to the average of left and right mastoids offline, and data down sampled to 250 Hz. The impedance of the mastoid and scalp electrodes was maintained below 5 kΩ and the eye electrodes below 10 kΩ throughout the recordings.

Figure 5.1

Sequential stimulus presentation in each trial



Note. Type 1 refers to stimuli with matching grammatical gender across Ukrainian and Russian, while Type 2 refers to stimuli with mismatching grammatical gender across both languages. The four experimental conditions included: (1) CR-CU – congruent in both Ukrainian and Russian, (2) IR-IU – incongruent in both languages, (3) CR-IU – congruent in Russian but incongruent in Ukrainian, and (4) IR-CU – incongruent in Russian but congruent in Ukrainian. A filler condition was also included. We selected filler items based on pre-test ratings, choosing only those rated as highly masculine or highly feminine.

5.3.4. Data analysis

The behavioural and EEG analyses were divided into two parts based on our research questions. In the first part, we analysed stimuli with matching grammatical gender across Ukrainian and Russian (CR-CU and IR-IU conditions). In the second part, we examined stimuli with mismatching grammatical gender across two languages (CR-IU and IR-CU conditions).

5.3.4.1. Behavioural analysis

Across all conditions, responses below 200 ms and above 3500 ms as well as time outs were excluded from the analysis (3.56%). In addition, responses below and above 2.5 SD from the mean for each intra-participant (2.65%) and intra-item (2.49%) condition (total = 3.80%) were also excluded. Importantly, in the pre-registration, we aimed to examine reaction times and accuracy of participants' responses. However, during the analysis stage, we found that participants had a response bias, and they were more likely to press "yes" for all conditions (see Figure 5.4). Therefore, to make sure that no biases influence our findings[6], we adapted the analysis based on the proportion of yes/no responses across

---

[6] As detailed in the Supplementary materials in Appendix C, accuracy in the pre-registration was dummy coded as "1" for responses aligning with Ukrainian grammatical gender and "0" for responses aligning with Russian gender. However, a general bias towards "yes" responses was later observed, which may have skewed these accuracy-based findings. To account for this, the final analyses focused on the proportions of "yes" / "no" responses instead.

conditions. To do so, variable “Response Type” was dummy coded with possible values of 1 (response “yes") or 0 (response “no") across all conditions. Reaction times were analysed as reported in pre-registration. Data analysis with accuracy can be found in Supplementary materials in Appendix C.

For both parts of the analysis, Reaction times and Response types were analysed using linear and logistic mixed effect models in R (R Core Team, 2022) with the lme4 package (Linck & Cunnings, 2015). Maximal models were fitted for all the analyses. When these failed to converge, the random structure was simplified by removing the randoms effects that contributed the least variance. All categorical predictors were dummy coded (-0.5 and 0.5) in all models.

In the second part of the analysis (conditions CR-IU and IR-CU), to quantify participants’ more proficient/dominant L1 (Ukrainian or Russian), we used L1 Proficiency and Dominance coefficients (see “Participants” section). The coefficients exhibited a wide range of variability (see Figure 5.2), with no ceiling effects observed (Table 5.3). Because of high correlation between the two variables, we analysed the coefficients in separate models.

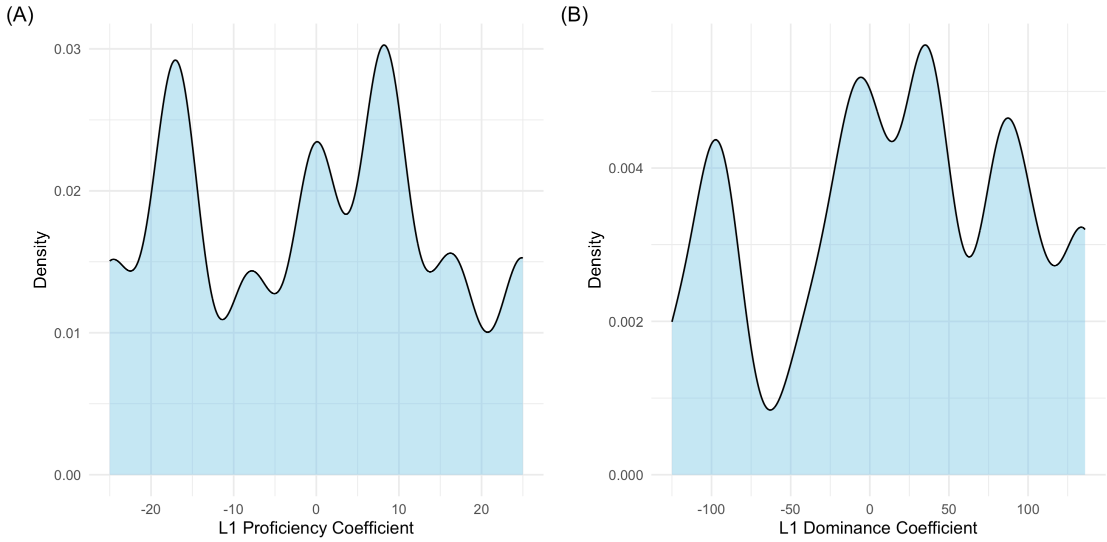
Table 5.3

L1 Proficiency and L1 Dominance Coefficients Across Participants

	Mean	SD	Range
L1 Proficiency Coefficient	-0.6	15.6	-25 to 25
L1 Dominance Coefficient	16.84	77.97	-125 to 136

Figure 5.2

Density plots for (a) L1 Proficiency and (b) L1 Dominance coefficients



Note. Each plot illustrates the distribution of participants' L1 coefficient scores, reflecting a wide range of variability without ceiling effects.

Additionally, in both parts of the analysis, we explored whether English proficiency and English dominance affected the grammatical gender influence of Ukrainian and Russian. We assessed the correlation between the two variables to determine whether they could be included in the same model. A Pearson correlation analysis revealed a low correlation between English proficiency and dominance scores ($r = 0.082$). Consequently, both variables were included in the same model. Additionally, to facilitate model convergence, both variables were scaled using the base "scale" function in R. The maximal model included random intercepts and slopes for grammatical gender congruency across both participants and items. For CR-CU and IR-IU conditions, a generalised linear mixed-effects model (glmer) was then built to analyse Response Type as a function of grammatical gender congruency, scaled

English proficiency, and scaled English dominance. For CR-IU and IR-CU conditions, to investigate English proficiency effects, we analysed a three-way interaction between English proficiency, L1 Proficiency coefficient, and Condition. Similarly, for English dominance, we analysed a three-way interaction between English dominance, L1 Dominance coefficient, and Condition. To understand the direction and nature of the effects, we designed separate glmer models using each condition as a baseline level.

5.3.4.2. EEG analysis

For the analysis of EEG data, we followed the approach of Sato et al. (2020), focusing on the N1, P2/VPP, and N300 components. We additionally analysed the early negative gender discrimination effect, emerging at 45-70 ms after stimulus onset in Mouchetant-Rostaing et al. (2000) and 45-90 ms in Mouchetant-Rostaing and Giard (2003). The data were analysed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). All EEG data pre-processing steps were scripted and run in MATLAB (v. R2022b). Obtained EEG data were down-sampled offline at 250 Hz and the average of two mastoids was used for re-referencing. Ocular artefacts were corrected using independent component analysis (ICA; Makeig et al., 1996), specifically with the "runica" function, excluding ocular electrodes from the analysis (see Casaponsa et al., 2024; Sato et al., 2020). Components associated with horizontal and vertical eye blinks were identified and manually removed using ICLabel (Pion-Tonachini et al., 2019). A high-pass filter of 0.5 Hz was applied before ICA, and a low-pass filter of 30 Hz was applied after ICA. While the pre-registration specified a high-pass filter of 0.01 Hz, the adjustment to 0.5 Hz was made based on recommendations from Delorme (2023) to optimise data quality. We computed mean ERPs time-locked to target onset off-line from trials free of ocular artefacts; epochs with activity exceeding ±75 lV at any cap electrode site were automatically discarded. Baseline correction was applied using the averaged EEG activity in the 100 ms preceding the onset of the stimuli.

Similar to the behavioural analysis, the EEG analysis was divided into two parts. However, parsimonious EEG models only included random structures for participants. In line with our pre-registration protocol, we examined EEG activity in regions where previous studies reported relevant effects. Specifically, for the N1, P2/VPP, and N300 components (based on findings reported by Sato et al., 2020), we focused on the anterior region (Fz, F3, F4, FC1, FCz, FC2 electrodes). For the early gender discrimination effect (based on Mouchetant-Rostaing & Giard, 2003), we analysed the central-parietal region (C3, Cz, C4, CP1, CP2, CPz electrodes). Temporal windows for each component were selected based on the latency peak. For the N1 and P2/VPP, a 60 ms time window (30 ms before and after the peak) was selected for amplitude analyses. For the N300, a 100 ms time window (50 ms before and after the peak) was selected for amplitude analyses. For early discrimination effect, we selected a 40 ms time window (20 ms before and after the peak).

For all selected components in anterior and central-parietal regions, we designed linear mixed-effects models (lmer) models. In the first part, we examined the effect of condition (CR-CU and IR-IU, contrast coded as 0.5 and -0.5) on the modulations of selected components and the early effect. In the second part of the analysis, we also investigated whether the interaction between condition (CR-IU and IR-CU, contrast coded as -0.5

and 0.5) and proficiency/dominance coefficient influenced the modulations of selected components and the early effect. For all models, random intercepts and random slopes for condition were included per participant. Additinoally, an exploratory analysis was conducted to examine the interaction between condition and participants' English proficiency/dominance on the modulation of the components of interest. These models also included random intercepts and random slopes for condition per participant.

5.4. Results

5.4.1 Analysis for Research Question 1 (stimuli with matching grammatical gender; conditions CR-CU and IR-IU)

For CR-CU and IR-IU conditions, we predicted faster reaction times (RTs) and/or biased choices that align with congruent grammatical gender across participants' two L1s. For ERP measures, we hypothesised that unconscious access to grammatical gender will manifest in decreased N1 and N300 amplitudes, as well as a greater P2/VPP modulation, for pairs in the anterior region from the CR-CU condition, compared to IR-IU. Additionally, for the early gender discrimination effect, we expected to observe greater early negativity in the central-parietal region for IR-IU pairs, compared to CR-CU.

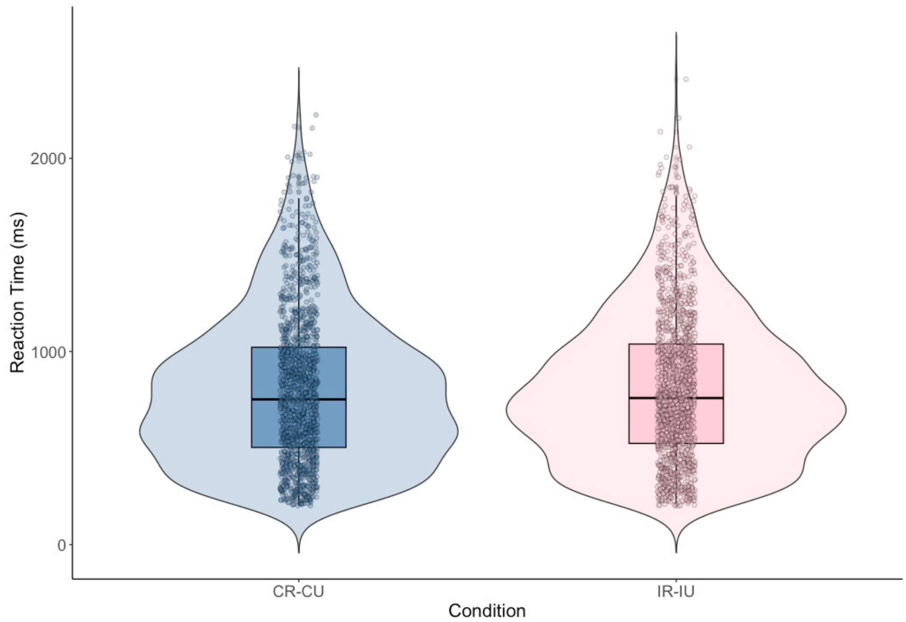
5.4.1.1. Behavioural analysis

Reaction times. The mean reaction time for the CR-CU condition was 797 ms (*SD* = 267.0), while for the IR-IU condition, it was 810 ms (*SD* =

246.7) (see Figure 5.3). The linear mixed effects model was built to determine whether the congruency between primes' grammatical gender and targets' biological sex (condition CR-CU) yielded faster reaction times compared to the incongruent condition (IR-IU). The parsimonious model included random intercepts for subjects and items, as well as by-subject random slopes for Condition. The model did not reveal a significant effect of grammatical gender congruency on reaction times (*Estimate* = -12.90, *SE* = 16.12, t = -0.80, p = .434, 95% CI[-46.90 21.10]), suggesting that participants' response times were comparable across the two conditions. These findings did not support our hypothesis that grammatical gender congruency facilitates faster responses.

Figure 5.3

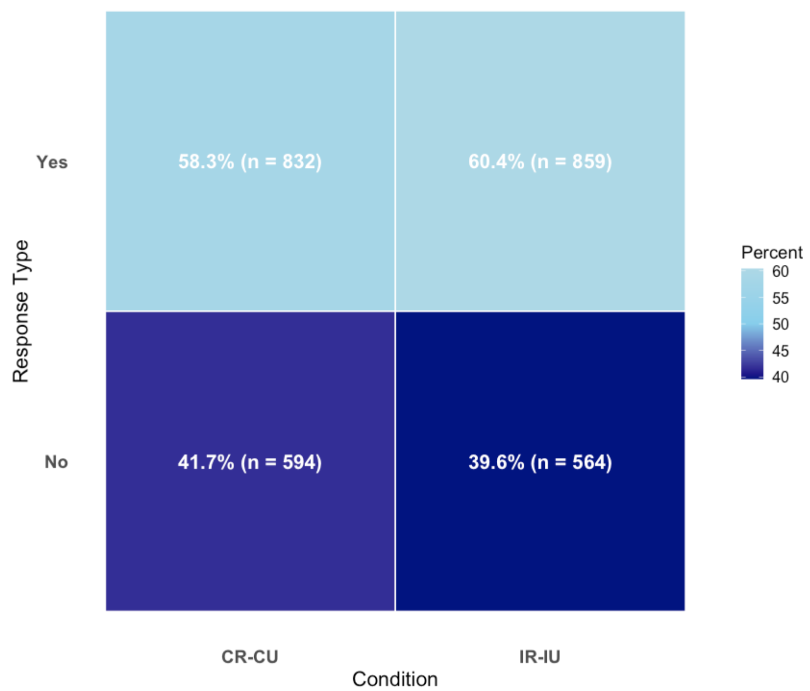
Distribution of reaction times by Condition (CR-CU vs IR-IU)



Response Types. Similarly to the reaction times, analysis of the response types suggested no sensitivity to grammatical gender congruency. The parsimonious generalised linear mixed effects (glmer) model included fixed effects of condition, random intercepts for subjects and face images, as well as by-subject and by-item random slopes for condition. The model revealed no significant effect of Condition (Estimate = -0.128, SE = 0.257, z = -0.496, p = .620, 95% CI[-0.631 0.376]), indicating no significant difference in the CR-CU condition compared to the IR-IU condition. Furthermore, as shown in Figure 5.4 participants displayed a bias in their responses and tended to press "yes" more frequently than "no" when judging associations between object nouns and faces regardless of the congruency of the noun's grammatical gender and the face's biological sex.

Figure 5.4

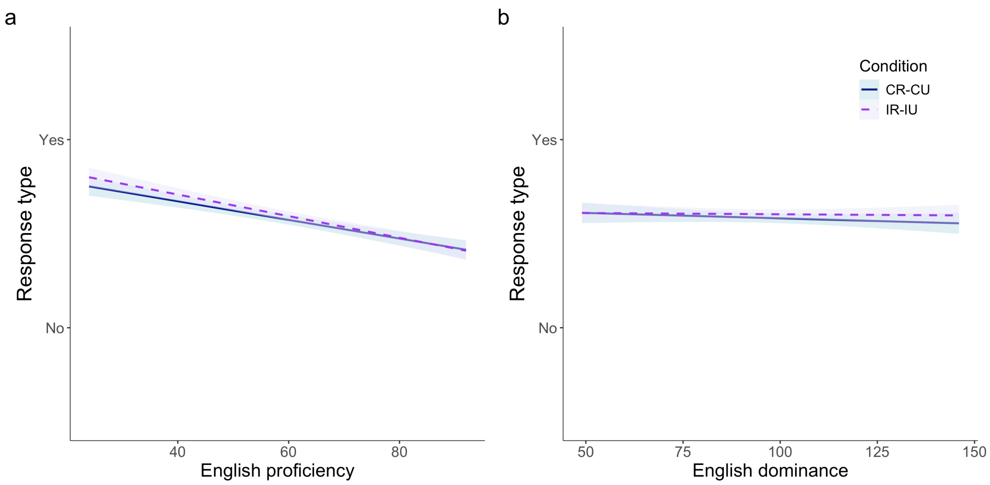
Proportion and Count of Yes and No responses by Condition (CR-CU vs IR-IU)



English proficiency and dominance. To examine the effects of English proficiency and dominance, we built a glmer model that included the interaction of Condition, English Proficiency, and English Dominance (all scaled), as well as random intercepts and by-condition slopes for both subjects and items. Contrary to our predictions, we found no significant interactions. Neither the interaction between Condition and English Proficiency (*Estimate* = -0.040, *SE* = 0.257, *z* = -0.157, *p* = .875), nor the interaction between Condition and English Dominance (*Estimate* = -0.025, *SE* = 0.230, *z* = -0.110, *p* = .912) was significant (see Figure 5.5). However, there was a significant main effect of English Proficiency (*Estimate* = -0.641, *SE* = 0.261, *z* = -2.459, *p* = .013), suggesting that, regardless of the Condition, participants with higher English proficiency scores, tended to press "no" significantly more often.

Figure 5.5

Effects of (a) English Proficiency and (b) English Dominance on Response Types (Yes vs No) across Conditions (CR-CU vs IR-IU)



5.4.1.2. ERP analysis

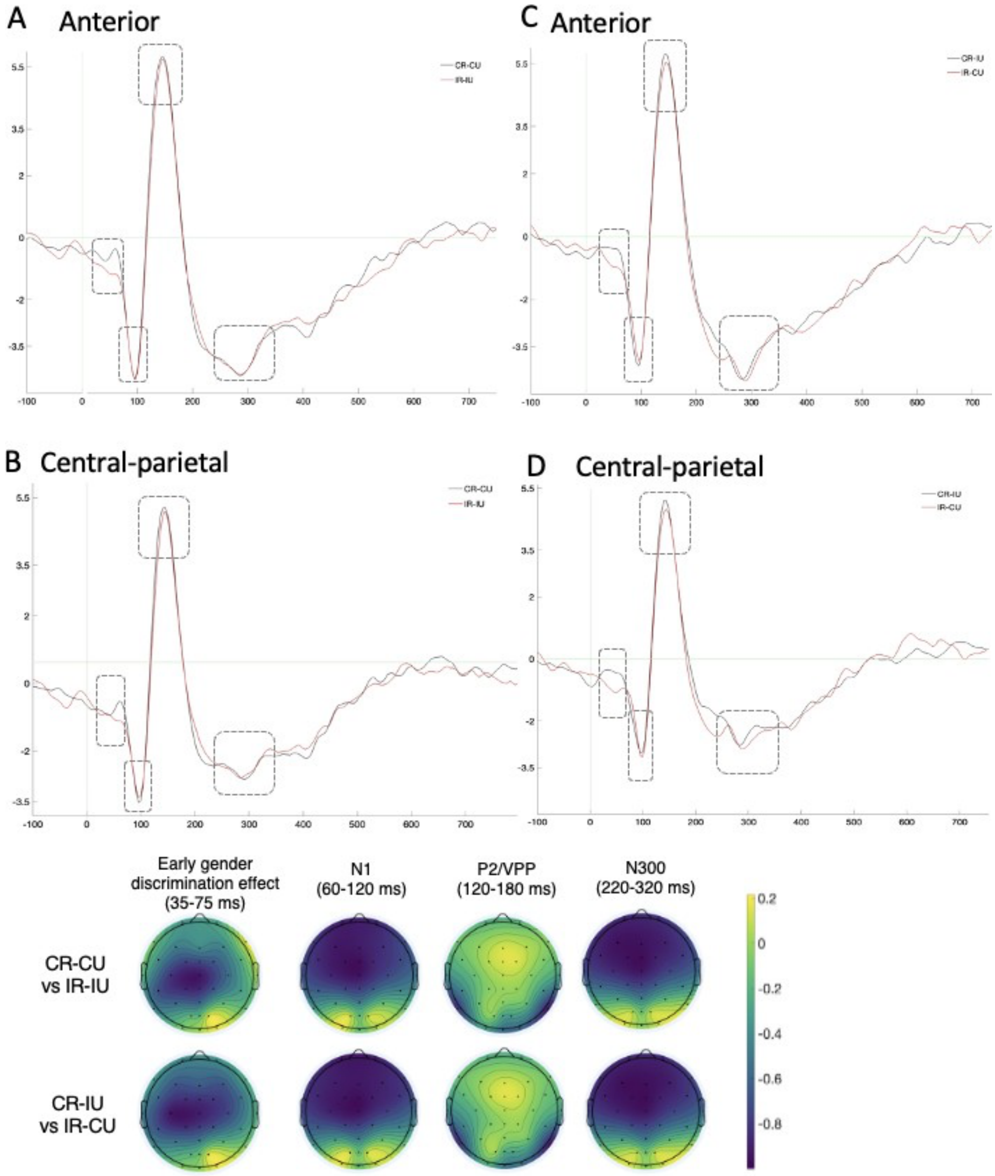
For each ERP component, we built a linear mixed-effects model with condition as a fixed effect and random intercepts and by-condition slopes for each participant. The model structure was consistent across all components.

N1 time window (60-120 ms). No significant effect of Condition was found for N1 amplitudes (*Estimate* = 0.303, *SE* = 0.209, $df = 18$, $t = 1.452$, $p = .164$, 95% CI[-0.136 0.742]). This indicates that, contrary to our predictions, N1 amplitude was not significantly greater for congruent pairs compared to incongruent ones (see Figure 5.6).

Similarly, including English Proficiency or Dominance as factors revealed no significant effects in the anterior region. The interaction between English Proficiency and Condition was non-significant (*Estimate* = -0.003, *SE* = 0.010, $df = 17$, $t = -0.249$, $p = .807$), as was the interaction between English Dominance and Condition (*Estimate* = -0.003, *SE* = 0.008, $df = 17$, $t = -0.380$, $p = .709$). These results suggest that increasing English proficiency and/or dominance had no measurable effect on N1 amplitude in this task.

Figure 5.6

Grand average ERP waveforms of the early discrimination effects (35-75 ms), N1, P2/VPP, and N300 components (circled with a grey line) showing the effects of grammatical gender congruency for Ukrainian-Russian bilinguals across conditions



Note. A and B represent conditions CR-CU and IR-IU, indicated by black and red lines, respectively. C and D represent conditions CR-IU and IR-CU, also indicated by black and red lines, respectively. The analysis includes two regions of interest: the anterior region (A and C), encompassing electrodes Fz, F3, F4, FC1, FCz, and FC2; and the central-parietal region (B and D), encompassing electrodes C3, Cz, C4, CP1, CP2, and CPz. Average ERPs were calculated from pooled electrodes, with time zero marking the onset of the target face image. Scalp topographies represent the effects of the Grammatical gender congruency as differences obtained between the CR-CU and IR-IU condition (E) and CR-IU and IR-CU conditions (F)

P2/VPP time window (120-180 ms). No significant effect of Condition was observed for P2/VPP amplitudes (*Estimate* = 0.295, *SE* = 0.257, *df* = 18, t = 1.147, p = .266, 95% CI[- 0.245 0.835]). We had predicted that when the biological sex of the face and grammatical gender of the noun were congruent, P2/VPP amplitude would be significantly greater compared to incongruent pairs. However, our findings did not confirm this prediction.

Including interactions with English Proficiency/Dominance and Condition did not reveal any significant effects. Neither the interaction between Condition and English Proficiency (*Estimate* = -0.013, *SE* = 0.013, *df* = 17, t = -1.066, p = .301) nor the Condition - English Dominance interaction (*Estimate* = -0.004, *SE* = 0.010, *df* = 17, t = -0.417, p = .682) reached significance. These findings suggest that higher English proficiency or dominance did not reduce the P2/VPP amplitude for congruent pairs, contrary to our predictions.

N300 time window (220-320 ms). No significant main effect of Condition was found for the N300 amplitude in the anterior region (*Estimate* = -0.026, *SE* = 0.130, *df* = 208, t = -0.201, p = .841, 95% CI[- 0.283 0.23]). These results indicate that N300 amplitudes did not differ between congruent and incongruent pairs, contrary to our predictions. When English Proficiency was added as a factor, the interaction between Condition and

English Proficiency was only marginally significant (*Estimate* = 0.011, *SE* = 0.006, *df* = 17, *t* = -1.738, *p* = .084). Adding English Dominance revealed no significant effects for Condition (*Estimate* = 0.219, *SE* = 0.484, *df* = 17, *t* = 0.452, *p* = .652) or a Condition-Dominance interaction (*Estimate* = -0.002, *SE* = 0.005, *df* = 17, *t* = -0.413, *p* = .680).

Early gender discrimination effect time window (35-75 ms). As part of our secondary analysis, we examined the early gender discrimination effect reported by Mouchetant-Rostaing and Giard (2003) in the central-parietal region. No significant effect of Condition was found (*Estimate* = 0.27, *SE* = 0.27, *df* = 18*, t* = 1.01, *p* = .326, 95% CI[-0.29 0.83]). Furthermore, adding English Proficiency (*Estimate* = -0.006, *SE* = 0.013, *df* = 17, *t* = -0.46, *p* = .648) or English Dominance (*Estimate* = -0.01, *SE* = 0.01, *df* = 17*, t* = -0.74, *p* = .471) to the model did not yield any significant main effects or interactions with Condition.

5.4.2. Results for Research Question 2 (stimuli with matching grammatical gender; conditions CR-IU and IR-CU)

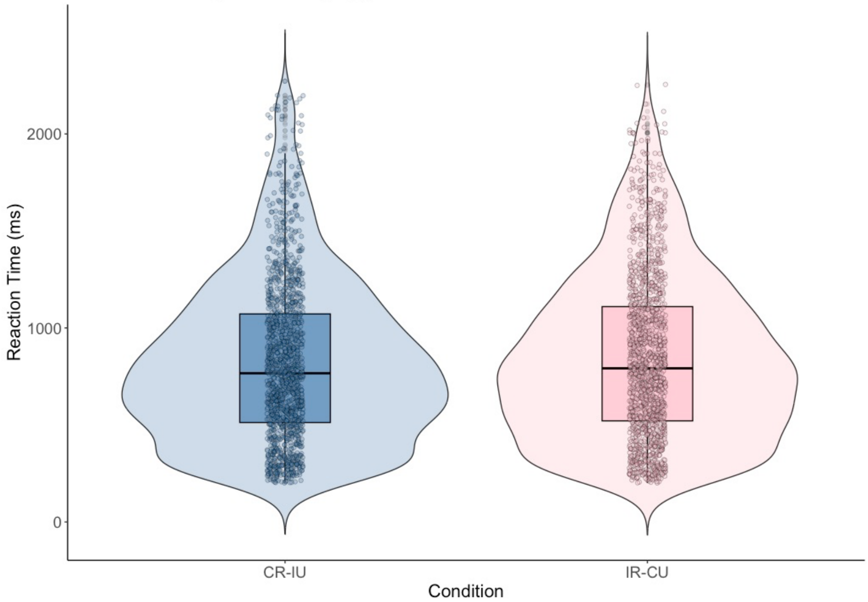
For the CR-IU and IR-CU conditions, we predicted that participants would show priming effects of the grammatical gender of their more proficient/dominant L1 (as determined by proficiency tests and the language dominance questionnaire). For behavioural measures, we predicted faster reaction times and/or biased choices towards the congruent grammatical/conceptual gender associations in participants' more proficient/dominant L1. For ERP measures, we predicted to find a decreased N1 and N300 amplitudes and greater P2/VPP modulation for pairs where the biological gender of the face aligns with the grammatical gender of the object noun in the participant's more proficient/dominant L1. Additionally, we expected a greater early negativity effect (45–90 ms) in the central-parietal region when the grammatical gender of the object noun is incongruent with the biological gender of the face in participant's more proficient/dominant L1 compared to when they are congruent.

5.4.1.1. Behavioural analysis

Reaction times. The mean reaction time for the CR-IU condition was 824.8 ms (*SD* = 266.4), while for the IR-CU condition, it was 846.4 ms (*SD* = 262.9) (Figure 5.7). Parsimonious fitted linear mixed-effects models including the interaction between Condition and L1 Proficiency/Dominance, as well as random intercepts and by-condition slopes for both subjects and items. Consistent with previous analyses, we found no significant interaction between the L1 Proficiency coefficient and Condition (*Estimate* = -0.598, *SE* = 0.845, t = -0.707, p = .489) or between L1 Dominance coefficient and Condition (*Estimate* = -5.91, *SE* = 13.29, t = -0.45, p = .662) on reaction times. This suggests that being neither more proficient nor more dominant in a particular L1 influenced participants' speed of responses when presented with pairs where the object nouns' grammatical gender was congruent with the biological sex of the face in that L1 (see full models' specifications on OSF: https://osf.io/2vr7k/).

Figure 5.7

Distribution of reaction times by Condition (CR-IU vs IR-CU)



Response Types. For the selected conditions, participants also displayed bias towards "yes" responses (see Figure 5.8), especially in the IR-CU condition. To examine the role of L1 dominance and proficiency on Response types, we built glmer models with the interaction between Condition and L1 Dominance or L1 Proficiency, and random intercepts and by-condition slopes for both subjects and items. As in the Reaction times analyses, there was no significant interaction between L1 Dominance coefficient and Condition (*Estimate* = -0.037, *SE* = 0.101, z = -0.363, p = .716) or between L1 Proficiency and Condition (*Estimate* = -0.004, *SE* = 0.006, z = -0.571, p = .568). However, we observed significant main effects of Condition with participants' bias to respond "yes" more often overall (*Estimate* = 0.441, *SE* = 0.141, z = 3.137, p = .002). In addition, there was a significant main effect of L1 Dominance (*Estimate* = 0.435, *SE* = 0.201, z = 2.170, p = .030, see Figure 9). It suggests that participants with higher L1

Dominance scores were more likely to give "yes" responses, regardless of condition.

Figure 5.8

Proportion and Count of Yes and No responses by Condition (CR-IU vs IR-CU)

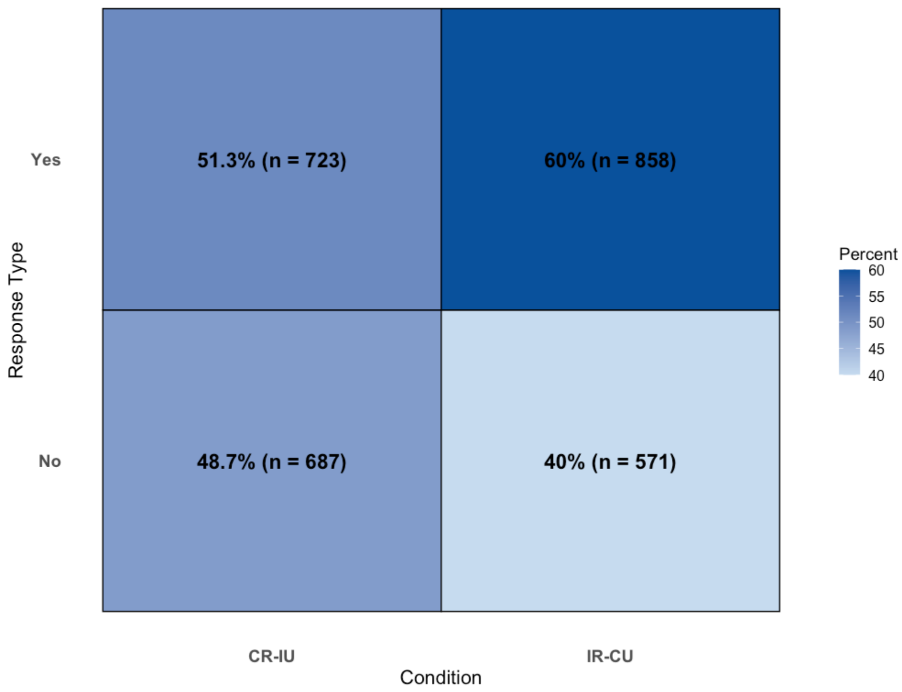
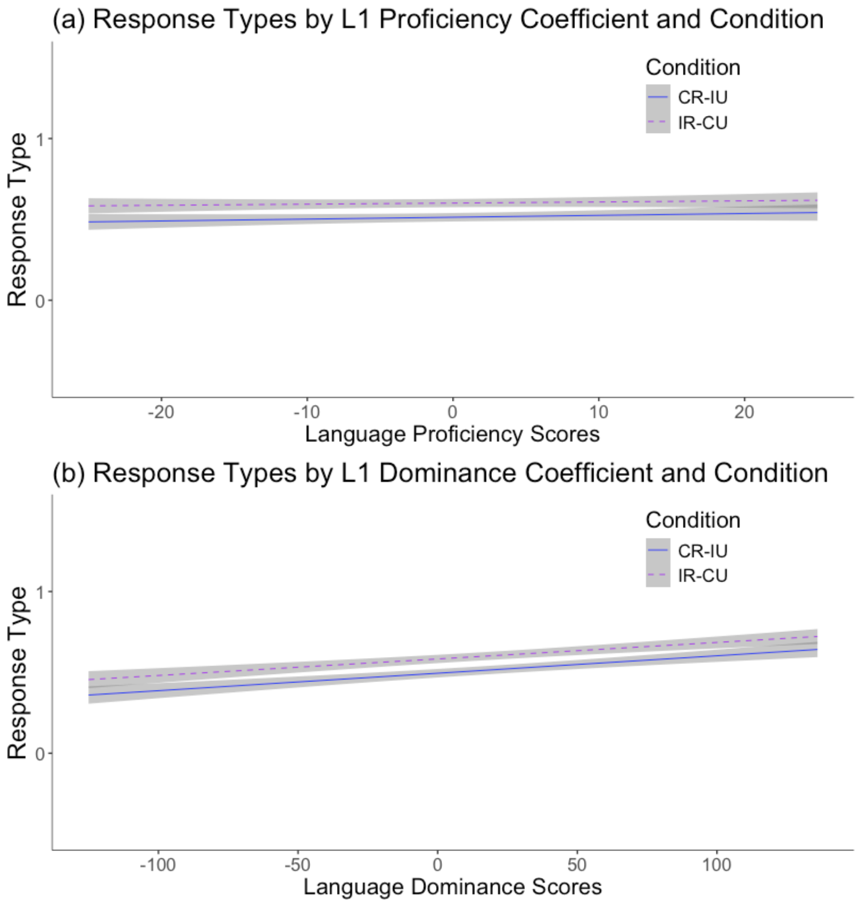


Figure 5.9

Response types (Yes vs No) as a function of (a) L1 proficiency coefficient and (b) L1 dominance coefficient across CR-IU and IR-CU conditions.



English proficiency and dominance. To examine how English proficiency and dominance influenced responses across conditions, we built a series glmer models. These models included the interactions between Condition, English Proficiency or Dominance, and their respective L1 measures (Proficiency or Dominance), along with random intercepts and by-condition slopes for both subjects and items.

When analysing the effects of English language proficiency on these types of stimuli, we did not find a significant interaction between English Proficiency, Condition, and L1 Proficiency coefficient (IR-CU used as a reference level: *Estimate* = -0.049, *SE* = 0.078, *z* = -0.623, *p* = .533). However, there was a significant interaction between Condition and English Proficiency (*Estimate* = -0.936, *SE* = 0.084, *z* = -11.082, *p* < .001). This indicates that participants with higher English proficiency were more likely to respond "yes" in the IR–CU condition (*Estimate* = 0.488, *SE* = 0.060, *z* = 8.191, *p* < .001) and "no" in the CR–IU condition (*Estimate* = -0.448, *SE* = 0.260, *z* = -7.483, *p* < .001), regardless of their L1 Proficiency.

Similarly, for English dominance, the analogous interaction between English Dominance, Condition and L1 Dominance was also not significant (IR-CU used as a reference level: *Estimate* = -0.061, *SE* = 0.108, *z* = -0.566, *p* = .571). Contrary to English Proficiency analyses, there were no significant interactions between Condition and English Dominance or between Condition and L1 Dominance (see full models' specifications and results on OSF: https://osf.io/2vr7k/).

5.4.1.2. ERP analysis

For each ERP component, we fitted linear mixed-effects models including the interaction between Condition and either the L1 Proficiency or L1 Dominance coefficient, with random intercepts and by-condition slopes for each ERP set.

N1 time window (60-120 ms). No significant interactions were observed between Condition and L1 Proficiency coefficient (*Estimate* < 0.001, *SE* = 0.02, *df* = 17, *t* = -0.004, *p* = .997), contradicting our predictions, showing no reduction in N1 amplitude for pairs where the biological sex of the face and the grammatical gender of the noun were congruent in participants' more proficient L1.

The results for L1 Dominance coefficient mirrored those for L1 Proficiency. No significant interaction was found in the anterior region (*Estimate* = 0.003, *SE* = 0.01, *df* = 17, *t* = 0.56, *p* = .581). This further indicates that participants' more dominant L1 had no measurable influence on N1 amplitude, regardless of congruency. For example, participants dominant in Ukrainian did not exhibit reduced N1 amplitudes for pairs congruent in Ukrainian but incongruent in Russian.

P2/VPP time window (120-180 ms). No significant interactions were observed between Condition and L1 Proficiency coefficient (*Estimate* = 0.02, *SE* = 0.02, *df* = 17, *t* = 0.71, *p* = .487) or Condition and L1 Dominance coefficient (*Estimate* = 0.01, *SE* = 0.01, *df* = 17, *t* = 1.16, *p* = .261). A marginally significant main effect of L1 Proficiency coefficient was detected (*Estimate* = 0.09, *SE* = 0.05, *df* = 17, *t* = 1.92, *p* = .072). Yet, contrary to our predictions, these findings indicate no significant increase in P2/VPP amplitude for pairs congruent in participants' more proficient or dominant L1 compared to those congruent in their less proficient or dominant L1.

N300 time window (220-320 ms). For the anterior N300, no significant interactions were observed between Condition and L1 Proficiency coefficient (*Estimate* = 0.031, *SE* = 0.022, *df* = 17, *t* = 1.389, *p* = .183) or Condition and L1 Dominance coefficient (*Estimate* = 0.008, *SE* = 0.008, *df* = 17, *t* = 0.953, *p* = .354). These findings failed to support our prediction that N300 amplitude would decrease for pairs congruent in participants' more proficient or dominant L1 compared to their less proficient or dominant L1.

Early gender discrimination effect time window (35-75 ms). As with the previous analysis for CR-CU and IR-IU conditions, the only near-significant results were observed for this component in the central-parietal region. No significant interactions were found between Condition and L1 Proficiency coefficient (*Estimate* = 0.001, *SE* = 0.01, *df* = 17, *t* = 0.06, *p* = .953) or between Condition and L1 Dominance coefficient (*Estimate* = 0.01, *SE* = 0.01, *df* = 17, *t* = 1.08, *p* = .297). Nevertheless, significant main effects of Condition were observed in both models: for L1 Proficiency (*Estimate* = -0.429, *SE* = 0.202, *df* = 17, *t* = -2.127, *p* = .048) and for L1 Dominance (*Estimate* = -0.422, *SE* = 0.194, *df* = 17, *t* = -2.178, *p* = .044). These results also suggest that while there was no effect of the more dominant L1 on the amplitude of this effect, there was a decreased amplitude for the pairs congruent in Ukrainian (similar to findings in the proficiency model), irrespective of participant's dominance in their two L1s.

5.5. Discussion and conclusion

In the current study, we aimed to investigate the role of two grammatical gender systems in modulating categorisation and perception in simultaneous Ukrainian-Russian bilinguals. To achieve this, we analysed behavioural (response types and reaction times) and neural (ERPs) responses elicited when presented with gendered (female and male) faces paired with two types of primes: those with matching grammatical gender across participants' two L1s (conditions CR-CU and IR-IU) and those with mismatching grammatical gender (conditions IR-CU and CR-IU).

We first examined the overall effects of overlapping grammatical gender systems on categorical perception by looking at behavioural and

ERP outcomes in trials where the grammatical gender of the prime and the biological sex of the target were either congruent or incongruent in both their L1s. Behaviourally, we found no significant effects of congruency on reaction times and response types aligning with previous studies with sequential bilinguals with one grammatical gender system (e.g., French-English: Sato et al., 2020; Spanish-English: Boutonnet et al., 2012) . In terms of ERP results, contrary to our predictions, no significant effects of grammatical gender congruency were observed in the N1, P2/VPP, or N300 components. The N1 component, typically associated with early sensory discrimination, has been shown to reflect participants' ability to differentiate stimuli at a perceptual level (Marzecová et al., 2018). The absence of significant N1 modulations suggests that congruency between prime's grammatical gender and target's biological sex did not influence early perceptual discrimination in this task. The P2/VPP component, linked to attention allocation and stimulus evaluation (Sato et al., 2020; Yu et al., 2017), similarly showed no significant effects, indicating that prime-target congruency did not lead to enhanced attention or evaluation for congruent pairs. Finally, the N300, which is often associated with semantic integration and categorisation (Federmeier & Kutas, 2001), showed no significant modulations. This lack of differentiation implies that grammatical gender did not affect higher-order cognitive processing during categorisation process. The significant overlap in ERP waveforms across congruent and incongruent conditions supports the conclusion that grammatical gender congruency alone (i.e., conceptually neutral stimuli) was not sufficient to modulate neural processing in our task. Furthermore, no significant effects were observed for the early gender discrimination effect that has been interpreted as reflecting low-level categorisation processes and automatic, rapid gender differentiation of faces. The lack of significant findings suggests that both CR-CU and IR-IU conditions might require comparable cognitive effort when categorising pairs as either associated or not associated based on the prime-target congruency.

We then examined whether participants would exhibit a grammatical gender bias consistent with their more proficient or dominant L1, by looking at primes with mismatching grammatical gender across Ukrainian and Russian. Behaviourally, reaction times and response types revealed similar pattern of results independently of the relative participant's L1 proficiency or dominance. Similarly, ERP results for the N1, P2/VPP, and N300 components showed no significant modulations based on L1(s) dominance or proficiency coefficients. The absence of significant effects in these components suggests that presented pairs congruent in grammatical gender with the more proficient/dominant L1 and incongruent in the less proficient/dominant L1 did not influence early sensory discrimination (N1), attention allocation or stimulus evaluation (P2/VPP), or higher-order semantic integration and categorisation processes (N300). Notably, the only significant results emerged for the early gender discrimination effect (35-75 ms), where a persistent main effect of Condition was observed, consistent with findings reported by Mouchetant-Rostaing and Giard (2003). Regardless of participants' more proficient or dominant L1, there was a consistent decrease in amplitude for pairs congruent in Ukrainian but incongruent in Russian. One explanation for this result relates to participants' language use on the day of testing. During debriefing, all participants included in the final analysis reported speaking Ukrainian exclusively prior to arriving at the lab, which may have biased their perception and led to pairs congruent in Ukrainian being perceived as more similar overall.

The consistent null effects observed across the N1, P2/VPP, and N300 components and behaviourally contrast sharply with findings from our previous behavioural studies (Osypenko et al., 2025; see also Sato et al., 2020). One possible explanation for these discrepancies lies in differences between experimental paradigms. In our earlier behavioural experiments (Osypenko et al., 2025), which were conducted online, there may have been unmonitored linguistic input (e.g., participants could have had oral or written communication with someone in either L1 when completing a task) that influenced results and that we could not account for. Furthermore, those paradigms likely engaged grammatical gender processing more explicitly than the current study, due to the high gender saliency in the task (see Samuel et al., 2019 for detailed discussion). Additionally, while participants in Osypenko et al. (2025) were instructed to respond based on their first impressions, there was no strict time limit, which may have allowed for more deliberate processing. In contrast, the current study's design emphasised rapid categorisation, potentially reducing the opportunity for conscious grammatical gender effects to emerge.

A key consideration is how our results differ from those by Boutonnet et al. (2012) and Sato et al. (2020), which reported significant effects of grammatical gender on perception and categorisation in bilingual participants, even when grammatical gender activation was not explicitly required by the task. Unlike the present research, both studies introduced overt manipulations: Sato et al. (2020) used conceptual gender, while Boutonnet et al. (2012) employed semantic associations. Rather than directly testing whether grammatical gender alone modulates categorisation, these studies provided evidence in favour of unconsciously activation of grammatical gender when participants engage with semantic or conceptual associations. By contrast, the present study examined whether grammatical gender effects emerge independently when participants are exposed to conceptually neutral stimuli. Our findings, which revealed no significant effects, raise the question of whether grammatical gender effects are task-dependent, requiring co-activation with other processes, such as conceptual representations or semantic associations, to emerge.

Besides, language typology may also explain the discrepancy between findings. Both French and Spanish, studied by Sato et al. (2020) and Boutonnet et al. (2020) respectively, are two-gender systems (with masculine and feminine genders) that maintain a more transparent relationship between grammatical and natural gender. By contrast, Ukrainian and Russian are three-gender systems (masculine, feminine, and neuter). Although our study excluded neuter stimuli, the mere presence of neuter gender in the language may have diluted the effects of grammatical gender. This aligns with the longstanding debate by Sera et al. (2002) and Vigliocco et al. (2005), who argued that grammatical gender effects are more evident in speakers of two-gendered languages than in speakers of three-gendered languages due to the more transparent masculine–feminine distinction in two-gendered languages. While comparing our results with those of Boutonnet et al. (2012) and Sato et al. (2020) provides some confirmation of this point, further research is needed to strengthen this claim. Specifically, it would be beneficial to recruit speakers of two-gendered languages to complete the task employed in the present study - without additional manipulations - and compare their findings with those of speakers of three-gendered languages in this paper. This approach would help assess the robustness of grammatical gender effects across typologically different language systems.

In addition, French and Spanish use articles as prominent gender markers, allowing faster access to grammatical gender information. Ukrainian and Russian, on the other hand, lack articles, and gender information is encoded at the semantic and morphological levels, such as in adjectives, pronouns, and determiners. Finally, a critical typological difference is the grammatical gender of the word "face." In French, "*le visage*" is masculine, which may have amplified incongruency effects in Sato et al.'s study (2020). In Ukrainian and Russian, however, the word "face" ("*обличчя*" [oblychchya] and "*лицо*" [litso], respectively) is neuter, which may have further weakened the salience of gender congruency effects.

Another factor that requires consideration is the role of cognates in Ukrainian and Russian, which were also present in our stimuli list. A post-hoc analysis (see Supplementary materials, Appendix C) revealed that cognates were more prevalent in matching-gender primes, with mismatching-gender primes showing a much lower proportion of cognates. However, ERP analyses showed no significant differences in processing of the two types of primes, irrespective of the cognate ratio. This suggests that the presence of cognates was unlikely to be the reason for the null effects observed in our experiment.

We also investigated whether participants' L2 proficiency or dominance in English, a genderless language, influenced the findings. Both behavioural and ERP analyses revealed limited effects. While a marginally significant interaction emerged for N300 amplitudes, with higher English proficiency reducing the difference between conditions, no other significant effects were observed, contrary to previous findings (Athanasopoulos, 2006; Athanasopoulos & Bylund, 2023). It is possible that English proficiency does influence participants' response types, as predicted;

however, it is challenging to determine whether it diminishes the effects of Ukrainian and Russian grammatical genders, given that these effects were not themselves significant. To further explore the role of L2 proficiency in studies using ERP, additional study designs are required. For example, longitudinal studies could examine whether increasing L2 proficiency and length of exposure to English modulate these effects over time, as observed in earlier research (Athanasopoulos et al., 2010). Additionally, alternative paradigms that more robustly elicit the effects of grammatical gender on modulation of ERP components in participants' L1s could provide greater insight into how L2 proficiency contributes to cognitive restructuring.

Additionally, the characteristics of our participant sample may have influenced the findings. Our participants were older and less familiar with experimental testing compared to typical student samples, which could have impacted their performance. However, the complete overlap of ERP waveforms across conditions suggests that increasing the sample size would likely yield similar results. This consistency suggests that the null effects are more likely due to the absence of grammatical gender effects on categorical perception in this task or may be attributable to other factors discussed earlier in this section.

Our null findings, especially in contrast to previous studies that reported significant grammatical gender effects, prompt us to reflect on a point raised by Athanasopoulos and Casaponsa (2020). They argue that rather than framing the question of language effects on cognitive processes as a binary "yes/no", researchers should focus on investigating when and how these effects emerge - or, as in the case of our findings, disappear. Following this approach, we are prompted to consider: Could having two three-gendered grammatical systems, which are also partially contrasting, reduce the salience of grammatical gender effects? By contrast, do two-gendered languages such as Spanish or French provide clearer connections to biological gender and a more transparent grammatical system (i.e., presence of articles), thereby amplifying these effects? Alternatively, could the absence of effects in our findings be due to the experimental paradigm used here, which did not involve any conceptual or semantic processing that might unconsciously activate grammatical gender effects?

To address the questions arising from this study, future research could include a monolingual Russian control group to determine whether the presence of two three-gender systems negates grammatical gender effects or whether the three-gender typology itself contributes to the observed pattern. Additionally, examining other simultaneous bilingual groups with greater typological differences between their languages could provide additional insights into the interplay between grammatical systems and cognitive processing. Finally, employing tasks that require Ukrainian-Russian bilinguals to engage in conceptual or semantic processing, rather than isolated categorisation, may provide a deeper understanding of the influence of grammatical gender on perceptual processes and our findings presented in the current study.

To sum up, our findings suggest that grammatical gender does not exert a significant influence on categorisation and perception when assessed in isolation. The null effects observed in this study may be attributable to a combination of factors, including the experimental paradigm, the structural properties of Ukrainian and Russian, and the linguistic profiles of the participants. These findings underscore the need for further research to elucidate the complex interplay between grammatical gender, task demands, and cognitive processing in bilinguals.

5.6. References

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96. https://doi.org/10.1017/S1366728905002397

Athanasopoulos, P., & Boutonnet, B. (2016). Learning grammatical gender in a second language changes categorization of inanimate objects: Replications and new evidence from English learners of L2 French. In R. Alonso Alonso (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 173–192). Multilingual Matters. https://doi.org/10.21832/9781783094837-011

Athanasopoulos, P., & Bylund, E. (2023). Cognitive restructuring: Psychophysical measurement of time perception in bilinguals. *Bilingualism: Language and Cognition*, *26*(4), 809–818. https://doi.org/10.1017/S1366728922000876

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, *116*(3), 437–443. https://doi.org/10.1016/j.cognition.2010.05.016

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

Boroditsky, L., Scmidt, L. A., & Phillips, W. (2003). Sex, syntax and semantics. In *Language in mind: Advances in the study of language and thought* (pp. 61–79). Boston Review. https://doi.org/10.7551/mitpress/4117.001.0001

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

Budzhak-Jones, S. (1997). Quantitative analysis of gender assignment in mono/bilingual discourse. *Journal of Quantitative Linguistics*, *4*(1–3), 67–91. https://doi.org/10.1080/09296179708590080

Cambridge University Press. (2024). *Test your English | Cambridge English*. https://www.cambridgeenglish.org/test-your-english/

Casaponsa, A., García-Guerrero, M. A., Martínez, A., Ojeda, N., Thierry, G., & Athanasopoulos, P. (2024). Electrophysiological Evidence for a Whorfian Double Dissociation of Categorical Perception Across Two Languages. *Language Learning*, *74*(S1), 136–156. https://doi.org/10.1111/lang.12648

Chen, Z., & Faitaki, F. (2024). Effects of the French grammatical gender system on bilingual adults' perception of objects. *Bilingualism: Language and Cognition*, 1–14. https://doi.org/10.1017/S1366728924000464

Delorme, A. (2023). EEG is better left alone. *Scientific Reports*, *13*(1), 2372. https://doi.org/10.1038/s41598-023-27528-0

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *27*(1), 202–224. https://doi.org/10.1037/0278-7393.27.1.202

Flaherty, M. (2001). How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. https://doi.org/10.1177/0022022101032001005

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484. https://doi.org/10.1023/A:1023264209610

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual language profile. In *Measuring L2 Proficiency: Perspectives from SLA* (pp. 208–225).

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*(2), 259–276. https://doi.org/10.1006/brcg.1994.1014

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21–36. https://doi.org/10.3758/CABN.5.1.21

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, *78*(1), 193–202. https://doi.org/10.1007/BF00230699

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology*, *116*(11), 2613–2631. https://doi.org/10.1016/j.clinph.2005.07.005

Kecskés-Kovács, K., Sulykos, I., & Czigler, I. (2013). Is it a face of a woman or a man? Visual mismatch negativity is sensitive to gender category. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00532

Kumar, M., Federmeier, K. D., & Beck, D. M. (2021). The N300: An Index for Predictive Coding of Complex Visual Objects and Scenes. *Cerebral Cortex Communications*, *2*(2), tgab030. https://doi.org/10.1093/texcom/tgab030

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lee, C.-Y., Liu, Y.-N., & Tsai, J.-L. (2012). The Time Course of Contextual Effects on Visual Word Recognition. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00285

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Maier, M., & Abdel Rahman, R. (2024). Transient and long-term linguistic influences on visual perception: Shifting brain dynamics with memory consolidation. *Language Learning*, *74*(Suppl 1), 157–184. https://doi.org/10.1111/lang.12631

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996). *Independent Component Analysis of Electroencephalographic Data*. *8*.

Marzecová, A., Schettino, A., Widmann, A., SanMiguel, I., Kotz, S. A., & Schröger, E. (2018). Attentional gain is modulated by probabilistic feature expectations in a spatial cueing task: ERP evidence. *Scientific Reports*, *8*(1), 54. https://doi.org/10.1038/s41598-017-18347-1

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65. https://doi.org/10.1017/S0048577299971196

Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience*, *15*(6), 900–910. https://doi.org/10.1162/089892903322370816

Mouchetant-Rostaing, Y., Giard, M.-H., Bentin, S., Aguera, P.-E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, *12*(1), 303–310. https://doi.org/10.1046/j.1460-9568.2000.00888.x

Pavlidou, T.-S., & Alvanoudi, A. (2013). Grammatical Gender and Cognition. *Versita/de Gruyter*, *2*(2), 109–124.

Pavlidou, T.-S., & Alvanoudi, A. (2019). Conceptualizing the world as ‘female’ or ‘male’: Further remarks on grammatical gender and speakers’ cognition. *Selected Papers on Theoretical and Applied*

Linguistics, *23*(0), Article 0. https://doi.org/10.26262/istal.v23i0.7351

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

R Core Team. (2022). *R: The R Project for Statistical Computing*. The R Project for Statistical Computing. https://www.r-project.org/

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

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

Sato, S., Casaponsa, A., & Athanasopoulos, P. (2020). Flexing Gender Perception: Brain Potentials Reveal the Cognitive Permeability of Grammatical Information. *Cognitive Science*, *44*(9), e12884. https://doi.org/10.1111/cogs.12884

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945. https://doi.org/10.1016/S0028-3932(01)00176-2

Schendan, H. E., & Kutas, M. (2003). Time Course of Processes and Representations Supporting Visual Object Identification and Memory. *Journal of Cognitive Neuroscience*, *15*(1), 111–135. https://doi.org/10.1162/089892903321107864

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

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

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567–4570. https://doi.org/10.1073/pnas.0811155106

Thierry, G., Rahman, R. A., & Athanasopoulos, P. (2024). An Introduction to the Cognitive Neuroscience of Language Embodiment and Relativity Special Issue of the Language Learning Cognitive Neuroscience Series. *Language Learning*, *74*(S1), 5–19. https://doi.org/10.1111/lang.12643

Ukrainian center for educational quality assessment. (2020). *Ukrainian center for educational quality assessment*. https://testportal.gov.ua/en/

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461. https://doi.org/10.1162/08989290152001880

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for

Language Learning and Language Use. *Journal of Experimental Psychology: General*, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Xue, Y., & Williams, J. (2024). Inducing Shifts in Attentional and Preattentive Visual Processing Through Brief Training on Novel Grammatical Morphemes: An Event-Related Potential Study. *Language Learning*, *74*(S1), 185–223. https://doi.org/10.1111/lang.12642

Yu, M., Li, Y., Mo, C., & Mo, L. (2017). Newly learned categories induce pre-attentive categorical perception of faces. *Scientific Reports*, *7*(1), 14006. https://doi.org/10.1038/s41598-017-14104-6

Chapter 6. General Discussion

The current thesis set out to extend existing knowledge of the linguistic relativity hypothesis, specifically in the domain of grammatical gender, by examining Ukrainian-Russian simultaneous bilinguals. This aimed to answer a key overarching question that unites all papers/experiments presented in the current body of work: *Do simultaneous bilingual speakers of two partially contrasting three-gendered languages exhibit grammatical gender effects from their first language(s), even when tested in a genderless L2?* In particular, the thesis first sought to determine if grammatical gender effects on cognitive processes (i.e., categorisation and memory recall) emerge similarly to those reported in earlier studies across monolingual and sequential bilingual speakers. Second, it explored whether grammatical gender in the more proficient/dominant L1 would influence cognitive processes when stimuli grammatical gender mismatched between the two L1s.

These research questions guided the three empirical studies presented in this thesis (Chapter 3-5), as well as the review that preceded them (Chapter 2). The purpose of this final chapter is to summarise the major contributions of this work, and to outline directions for future research.

6.1. Summary of major contributions, by chapter

Chapter three explored grammatical gender effects on categorisation in simultaneous bilinguals of Ukrainian and Russian adapting a similarity judgement task from Phillips and Boroditsky (2003). This chapter provided the first empirical evidence that grammatical gender effects can emerge in bilingual speakers of two three-gendered languages, but notably, only when neuter gender nouns were excluded. Specifically, Experiment 1 investigated whether participants rated object-character pairs as more similar based on congruency between the objects' grammatical gender and the characters' biological sex, either across both L1s or in participants' more proficient L1. This experiment included nouns from all three grammatical genders (masculine, feminine, neuter) and found no significant grammatical gender effects, aligning with earlier claims suggesting the presence of a neuter gender in three-gendered languages diminishes grammatical gender effects (Sera et al., 2002; Vigliocco et al., 2005). In contrast, Experiment 2, which excluded neuter gender nouns, revealed significant grammatical gender effects. Specifically, participants perceived object-character pairs as more similar when grammatical gender was congruent with biological sex in both L1s or with the participant's more proficient language. These results also provided empirical support for the online effects of grammatical gender, in line with the structural-feedback hypothesis (Sato & Athanasopoulos, 2018), showing that grammatical gender effects become observable primarily when the neuter gender is excluded from stimuli.

A major contribution of this study is providing the first empirical evidence that grammatical gender effects can emerge in simultaneous bilingual speakers of two three-gendered languages, specifically under conditions where the neuter gender stimuli are removed. This suggests that the presence of neuter gender might have diluted associations between grammatical and natural gender at-the-time of task completion.

Chapter four extended the findings from Chapter three by examining grammatical gender effects on memory recall in Ukrainian-Russian bilinguals. In this study I adapted the object-name memory task from Boroditsky and Schmidt (2000), with neuter-gendered nouns kept as fillers.

Participants were shown pairs of depicted gendered nouns alongside a male or female name, and they were asked to remember as much as possible. In the chosen pairs, grammatical gender of the prime and the implied sex of the name were either congruent/incongruent in both L1s or congruent/incongruent in participants' more proficient L1. After a brief distractor task, participants were asked to recall the assigned names. Results demonstrated that memory recall improved significantly for bilingual participants when the biological sex of proper names was congruent with grammatical gender in both languages. However, when grammatical genders mismatched across Ukrainian and Russian, the expected gender effect of the more proficient L1 on recall was not observed.

The main contribution of this study was the use of a less gender/sex-salient task with objectively correct answers, designed to avoid language-on-language effects previously discussed as a potential limitation of the similarity judgment task in Chapter three. Chapter four further contributed theoretically by integrating findings into discussions of bilingual conceptual representations, referencing Baddeley and Hitch's (1974) Working Memory Model to explain cognitive interactions between two grammatical systems.

Chapter five adapted a non-verbal categorisation task from Sato et al.'s (2020) study to investigate whether there are early ERP effects of grammatical gender alone in simultaneous bilinguals, as opposed to earlier studies that also explicitly required activation of conceptual or semantic information (Boutonnet et al., 2012; Sato et al., 2020). In this task, participants viewed object primes (depicted gendered nouns, excluding neuter gender entirely) followed by male or female faces (targets) and were asked to judge if there was an association between stimuli in each pair.

Across all pairs, grammatical gender of the prime and sex of the target were either congruent/incongruent in both L1s, or congruent/incongruent in participants' more proficient/dominant L1. This study was the first ERP investigation of grammatical gender effects on perception of simultaneous bilinguals, providing insights beyond behavioural data alone. Unlike previous ERP studies with sequential bilinguals with a gendered L1 and a genderless L2 (e.g., Sato et al., 2020), our findings revealed no significant modulation of the predicted ERP components (N1, P2/VPP, N300) by grammatical gender of conceptually neuter nouns. Consistent with the ERP findings, behavioural results revealed only near-significant effects of grammatical gender on perceived similarity between primes and targets. This suggests that grammatical gender alone may not independently influence early perceptual processing but rather requires conceptual or semantic activation.

Chapter five provides three major contributions. First, this is the first study to go beyond behavioural measures and look at the chosen bilingual group using neural measures (i.e., ERPs). Second, this is also the first examination of the grammatical gender effects on early perceptual effects when stimuli are free from conceptual or semantic associations (i.e., when grammatical gender is the sole manipulation). Third, we also examined whether there are effects of L2 (English) dominance/proficiency on perception, based on the earlier evidence in linguistic relativity research (e.g., Athanasopoulos & Bylund, 2023).

Overall, the studies presented in the current thesis allow me to provide the following theoretical and methodological contributions to our understanding of Whorfian effects:

1. Theoretical contribution:

1.1. *Inclusion of simultaneous bilinguals of two languages that have grammatical gender*. The inclusion of this understudied bilingual population allowed a deeper exploration into how two grammatical gender systems coexist and interact within a bilingual mind, directly addressing gaps identified in previous literature (e.g., Bassetti, 2007). Specifically, whether grammatical gender effects emerge for both stimuli with matching and mismatching grammatical gender across the two L1s.

1.2. *Effects of three-gendered grammatical systems.* This thesis provided the first empirical demonstration that grammatical gender effects can emerge in speakers of two three-gendered languages, specifically Ukrainian and Russian. These findings challenge previous claims regarding the negligible effects of grammatical gender in three-gendered languages due to the presence of a neuter gender (Sera et al., 2002; Vigliocco et al., 2005; cf. Haertle, 2017). By showing that gendered conceptual effects do emerge despite the presence of neuter gender, the results extend the "sex and gender" hypothesis (Vigliocco et al., 2005) beyond two-gendered systems, suggesting that the conceptual impact of grammatical gender is not necessarily diminished by a tripartite gender structure.

1.3. *Contributions to the structural-feedback hypothesis and discussion of online vs offline language effects.* Through manipulating the presence or absence of neuter gender in stimuli (Chapter 3), this thesis contributed directly to discussions on the structural-feedback hypothesis (Sato & Athanasopoulos, 2018), showing that grammatical gender effects are context-dependent and sensitive to the online processing demands of specific experimental tasks rather than being stable offline cognitive structures.

1.4. *Comprehensive investigation across tasks with varying gender/sex salience contributing to conscious vs unconscious emergence of gender effects.* A further theoretical contribution of this thesis lies in the systematic investigation of grammatical gender effects within the same bilingual population using tasks with varying degrees of gender and sex salience - ranging from high (similarity judgement, Chapter 3), to medium (memory recall, Chapter 4), to low (early neural activation in categorisation using EEG, Chapter 5). This design directly addresses a central theoretical concern in the literature examining grammatical gender effects: that these effects may arise primarily from task-specific strategies or explicit metalinguistic awareness, rather than reflecting genuine conceptual change (Samuel et al., 2019). By providing evidence that gender effects persist in tasks with the “medium” degree of gender salience (Chapter 4) and minimal linguistic prompts, the findings suggest that grammatical gender can influence cognition at a level that is not necessarily conscious or strategic. However, the absence of effects in Chapter 5, where gender information was implicit and responses were measured at both early perceptual (N1, P2/VPP) and later conceptual (N300) ERP components, points to limitations in the automaticity of these effects.

1.5. *Replication and transparency addressing the replication crisis.* This thesis contributed to the replication of seminal grammatical gender studies (Boroditsky & Schmidt, 2000; Phillips & Boroditsky, 2003), directly addressing the replication crisis in linguistic relativity research and offered transparent data analysis and providing stimuli in the open access.

2. Methodological contributions:

2.1. *Comprehensive linguistic profiling*. Unlike most prior studies relying either on self-reported language proficiency or various standardised measures, this thesis combined both kinds of assessments in every language spoken by participants to obtain a detailed and accurate assessment of participants' language backgrounds. While the first two studies only analysed proficiency tests and did not present the results of self-reported BLP questionnaire, as the data collection started shortly after the beginning of the war in Ukraine, the study in Chapter five presented both assessments and discussed the discrepancies between them.

2.2. *Treating language proficiency as a continuous variable.* Rather than categorising participants solely based on their native language, this thesis employed continuous measures of proficiency and dominance, allowing a nuanced exploration of how varying degrees of bilingual proficiency affect the responses.

As an integrated whole this thesis provides the comprehensive account to date of the grammatical gender effects in simultaneous bilinguals that have two contrasting grammatical systems. Returning to the questions initially posed:

> *1. Do simultaneous bilingual speakers of two contrasting three-gendered languages exhibit grammatical gender effects of their first languages when tested in a genderless L2?*
>
> *Yes*, though their presence or absence depends on various experimental factors (i.e., exclusion or inclusion of neuter gender, salience of gender/sex embedded in the task among others). For example, gender/sex salience was high in the similarity judgement task (Chapter 3), where participants were explicitly asked to rate similarity between objects and gendered characters; medium in the memory recall task (Chapter 4), where participants memorised name-object pairs without explicit attention to gender; and low in the EEG categorisation task (Chapter 5), where no linguistic cues were provided, and gender congruency had to be inferred. Notably,

grammatical gender effects were observed only in tasks with high or medium salience.

2. Do bilingual speakers show effect of grammatical gender on their conceptual representations when it matches across their two first languages?

Yes, partially. This was demonstrated in the similarity judgement task (Chapter 3, Experiment 2), where participants rated object-character pairs as more similar when grammatical gender and biological sex were congruent across both L1s. Similarly, in the memory recall task (Chapter 4), recall improved when grammatical gender and the sex of the name aligned in both languages. These findings suggest that when grammatical gender is consistent across a bilingual's two languages, it can shape conceptual associations, at least under conditions of moderate-to-high gender/sex salience.

3. Do bilingual speakers demonstrate grammatical gender effects on their conceptual representations when grammatical gender mismatches across languages, influenced by their more proficient or dominant L1?

Yes, partially. Participants showed stronger gender-congruency effects when grammatical gender aligned with their more proficient or dominant L1, suggesting that conceptual associations can be shaped by language dominance when gender categories conflict across languages. However, this effect did not replicate in the memory recall task (Chapter 4), possibly due to the co-activation of conflicting grammatical gender categories (e.g., masculine in one language and feminine in the other), which may have led to conceptual ambiguity, thereby weakening the effect.

Given these nuanced results, and the variability of grammatical gender effects on simultaneous bilingual cognition, there is a need for further systematic research. Directions for such investigations are presented next in section 6.2.

6.2. Directions for future research in linguistic relativity

The research presented in the current PhD research, while comprehensive, highlights numerous questions requiring further investigation. In this subsection I will present two possible directions that can be taken following the findings presented in the current thesis: (1) the directions for research focusing on grammatical gender, and (2) expanding the research with a simultaneous bilingual group to the whole research on linguistic relativity.

6.2.1. Future directions of grammatical gender research and taking it outside of the laboratory setting

One of the key discussions in Chapter 3 was the possibility of finding language-on-language (Samuel et al., 2019; Sato & Athanasopoulos, 2018) effects rather than genuine language-on-thought effects. While this study was necessary to establish the baseline effects in simultaneous bilingual speakers of two gendered languages, future research should further investigate this issue by employing tasks with lower explicit gender salience (e.g., adapted lexical decision and semantic categorisation tasks from Sato & Athanasopoulos, 2018). Furthermore, incorporating neural measures, such as ERPs, into these paradigms would allow researchers to directly assess whether grammatical gender effects emerge at an early perceptual level or only at a later processing stage. Such methodological improvements would also strengthen our understanding of the structural-feedback hypothesis by clarifying how linguistic structures affect cognitive processing.

Another crucial future direction arising from our findings throughout Chapters 3-5 involves the expansion of participant groups. Specifically, additional recruitment of Russian L1 speakers with English as an L2 to participate in the same tasks as presented in Chapters 3-5 would enable a more fine-grained analysis of the source of the observed effects or absence of such effects. This would help determine whether the variability in the found effects is driven primarily by (a) the co-activation and potential conflict inherent in bilingual processing of two contrasting grammatical gender systems, or (b) whether it stems from properties of three-gendered languages themselves and/or Ukrainian and Russian typology. Clarifying this distinction would allow for a more precise understanding of the conditions under which the examined Whorfian effects emerge, and whether acquiring two gendered languages simultaneously reshapes the conceptual impact of grammatical gender. Given the difficulty of finding Ukrainian monolinguals, Russian-speaking groups represent the next most suitable alternative. However, this was not possible to be done due to the lack of access to such participants during the current PhD.

Furthermore, providing more naturalistic, ecologically valid evidence (i.e., “Whorf in the Wild”, see Athanasopoulos & Bylund, 2020) is crucial to shed light on the grammatical gender effects outside of those found in the more artificial laboratory setting. For instance, future research should consider incorporating corpus data in grammatical gender studies. Specifically, corpus analysis could examine how grammatical gender influences adjective use in real life. For example, Boroditsky et al. (2003) previously found that speakers of gendered languages systematically used stereotypically gendered adjectives (e.g., describing grammatically masculine objects as “strong” or “powerful”, and feminine nouns as “beautiful” or “delicate”). Corpus-based studies could explore whether this phenomenon generalises to real-life examples from literature or news, comparing texts from speakers of gendered languages versus those of genderless languages. Similarly, corpus analyses could build upon the work by Mecit et al. (2020), who found shifts in risk perception depending on how COVID-19 was gendered (e.g., masculine as “virus” or feminine as “COVID-19”) in languages such as Spanish and French. Having access to news articles and using corpus, could also demonstrate whether this tendency was consistent in real life when comparing adjectives and judgements made about Covid-19 depending on the language of the text (gendered vs genderless) and term used to describe it (“Covid-19” vs “virus”).

Another way to extend the ecological validity of grammatical gender studies is using virtual reality (VR) technology. The latter could help bridging the gap between lab and naturalistic setting. For instance, putting a participant in an environment where they need to interact with gendered objects (e.g., adjective assignment in Boroditsky et al., 2003) or adapting memory recall task used in Chapter 4 to VR (i.e., showing people who say their names and hold a specific object. After that participants need to recall the name of the person holding an object). Additionally, using findings from Mecit et al. (2020), VR technology can help immerse participants into the situations where they need to make judgements regarding danger perception in a more realistic setting. In addition to provide more ecologically valid evidence in favour of absence or presence of grammatical gender effects, it would also connect this line of research with emerging research on foreign language effect (FLE).

Collectively, these suggested directions would further deepen the theoretical understanding of grammatical gender effects in bilingual cognition and better delineate the scope and conditions under which linguistic properties, such as grammatical gender, exert an influence on cognition within the frameworks of structural-feedback and multicompetence theories (Cook, 2002). Additionally, future research on LR would benefit from a more in-depth examination of simultaneous bilingual populations discussed next.

6.2.2. Future directions of bilingual research in linguistic relativity

Generally, linguistic relativity research does not commonly include simultaneous bilinguals, or at least does not explicitly differentiate between types of bilingualism. Therefore, systematically comparing simultaneous bilingual speakers of languages with contrasting lexical and grammatical properties across various domains frequently explored within linguistic relativity research could further illuminate how two L1s interact within a bilingual mind. For instance, research on motion, particularly aspect and non-aspect languages, which differ in how they encode the temporal structure of events (Athanasopoulos & Bylund, 2020; Flecken et al., 2015), has shown that such structural differences can influence how speakers perceive and remember events. Aspect languages (e.g., English, Russian) grammatically mark distinctions such as ongoing versus completed (imperfective vs perfective) actions, while non-aspect languages (e.g., Swedish, German) lack obligatory grammatical encoding for such distinctions. These differences have been shown to shape event perception and memory (Athanasopoulos & Bylund, 2013) with Swedish speakers more attuned to endpoints, compared to English speakers, due to the absence of grammatical aspect in Swedish. However, to my knowledge, no research has yet explored these phenomena in simultaneous bilinguals of both aspect and non-aspect languages, based on their language proficiency / dominance coefficients (e.g., Russian-German, Russian-Finnish, or Polish-German simultaneous bilinguals that are quite common pairing, where Russian and Polish languages are aspect language, and German and Finnish are non-aspect languages).

Similarly, studies on spatio-temporal cognition showed that language might shape our mental representations of time, with conceptual metaphors like “long time” influencing how duration is perceived (Lakoff & Johnson, 1980). Multiple evidence supports LRH and suggests that speakers of different languages, such as English and Swedish (languages with distance-based metaphors: “long”/ “short” time) versus Spanish and Greek (amount-based metaphors: “big”/ “small” time), conceptualise time differently, as abstract domains appear more susceptible to linguistic influence than concrete ones (Casasanto et al., 2004; Lakoff & Johnson, 1980; Núñez & Cooperrider, 2013). This direction also has not looked into the effects on simultaneous bilinguals. For instance, English-Spanish or English-Greek simultaneous bilinguals could provide more insights into the language effects when two L1s of speakers have contrasting lexical properties. The future directions described so far in this subsection can also provide more context into how two languages coexist in a bilingual mind and could help further develop the multicompetence theory (Cook, 2006), which views bilinguals not as two monolinguals in one but as possessing a unique and integrated language system (differing from either monolingual system) that affects cognition. Importantly, our findings in Chapter five highlight the need to combine standardised and self-reported assessments when examining bilinguals' language profiles. This is particularly crucial given the variety of measures used in earlier studies (as discussed in Chapter two) and the discrepancies observed in Chapter five. Combining assessments can provide researchers with a more nuanced understanding of the presence or absence of Whorfian effects across various domains.

Furthermore, longitudinal research on Ukrainian-Russian bilinguals represents another valuable future direction, given the ongoing process of language attrition and emerging evidence of cognitive restructuring in sequential bilinguals (e.g., Athanasopoulos & Boutonnet, 2016; Athanasopoulos & Bylund, 2023 among others). Over time, there will available Ukrainian-Russian bilingual speakers who maintain proficiency in both L1s, alongside those who have ceased using Russian since 2022, which provides an opportunity to investigate cognitive restructuring in simultaneous bilinguals with shifting language use and language attitudes. Ideally, future research could investigate the following groups: (1) Ukrainian-Russian bilinguals who have continued using both languages since 2022; (2) those who speak only Ukrainian; and (3) those who have resided abroad since 2022 and have subsequently acquired other gendered or genderless languages as foreign languages (L2/L3).

Finally, simultaneous bilinguals, particularly from post-Soviet regions, have remained underrepresented in linguistic relativity research despite their prevalence and relative accessibility for experimental testing. Specifically, Russian-Belarusian, Russian-Latvian, Russian-Kazakh, or Russian-Georgian bilingual populations remain widely available, yet understudied. Additionally, investigating pairs of languages across post-Soviet countries could offer insights into whether linguistic distance, between these languages, as well as the number of cognates between the two L1s modulate the strength of language effects on cognition.

To sum up, this thesis provides a comprehensive investigation of grammatical gender effects in a bilingual population novel to linguistic relativity research, contributing new empirical and theoretical insights to the field. By systematically manipulating grammatical gender congruency, experimental design, and L1 assessments in Ukrainian-Russian simultaneous bilinguals, this research has demonstrated that grammatical gender can shape cognitive processes - but not uniformly. Rather, these effects are task- or context-dependent. These findings call into question binary assumptions about grammatical gender effects - namely, that they are either universally present or entirely absent - and instead provide support for a more nuanced, context-sensitive account of linguistic influence, as proposed by the structural-feedback hypothesis. Ultimately, this body of work contributes to a growing shift in the field - from asking whether language shapes thought to examining when, how, and under what conditions such effects arise (Athanasopoulos & Casaponsa, 2020). Answering the latter questions is a key to refining our understanding of bilingual cognition and understanding the true scope of the effects both lexical and grammatical language properties, such as grammatical gender, have on one's cognitive processes.

Chapter 7. Consolidated Bibliography

Ackerman, L. (2018). *Dealing with ordinal data*. https://lmackerman.com/notebooks/ordinal_data.nb.html#what_is_%E2%80%9Cordinal%E2%80%9D_data

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96. https://doi.org/10.1017/S1366728905002397

Athanasopoulos, P. (2007). Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes*, *22*(5), 689–699. https://doi.org/10.1080/01690960601049347

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

Athanasopoulos, P., & Aveledo, F. (2012). Linguistic relativity and bilingualism. In *Memory, Language, and Bilingualism* (pp. 236–255). Cambridge University Press. https://doi.org/10.1017/CBO9781139035279.010

Athanasopoulos, P., & Boutonnet, B. (2016). Learning grammatical gender in a second language changes categorization of inanimate objects: Replications and new evidence from English learners of L2 French.

In R. Alonso Alonso (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 173–192). Multilingual Matters. https://doi.org/10.21832/9781783094837-011

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*(2), 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., & Bylund, E. (2020). Whorf in the wild: Naturalistic evidence from human interaction. *Applied Linguistics*, *41*(6), 947–970. https://doi.org/10.1093/APPLIN/AMZ050

Athanasopoulos, P., & Bylund, E. (2023). Cognitive restructuring: Psychophysical measurement of time perception in bilinguals. *Bilingualism: Language and Cognition*, *26*(4), 809–818. https://doi.org/10.1017/S1366728922000876

Athanasopoulos, P., Bylund, E., & Casasanto, D. (2016). Introduction to the Special Issue: New and Interdisciplinary Approaches to Linguistic Relativity: New Approaches to Linguistic Relativity. *Language Learning*, *66*, 482–486. https://doi.org/10.1111/lang.12196

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation. *Psychological Science*, *26*(4), 518–526. https://doi.org/10.1177/0956797614567509

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

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, *116*(3), 437–443. https://doi.org/10.1016/j.cognition.2010.05.016

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

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Basova, G., Kachura, A., Kihno, A., Musienko, V., Ozerova, N., Oleinik, G., Romanova, N., Snytko, E., & Tvoronovych, O. (2003). *Сопоставительная грамматика русского и украинского языков [Comparative grammar of Russian and Ukrainian languages]* (E. Maznyčenko, Ed.; 1st ed.). Naukova Dumka. https://www.academia.edu/10003894/СОПОСТАВИТЕЛЬНАЯ_ГРАММАТИКА_РУССКОГО_И_УКРАИНСКОГО_ЯЗЫКОВ

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

Bassetti, B., & Filipović, L. (2022). Researching language and cognition in bilinguals. *International Journal of Bilingualism*, *26*(1), 3–12. https://doi.org/10.1177/13670069211022860

Bassetti, B., & Nicoladis, E. (2016). Research on grammatical gender and thought in early and emergent bilinguals. *International Journal of*

Bilingualism, *20*(1), 3–16. https://doi.org/10.1177/1367006915576824

Beit-Hallahmi, B., Catford, J. C., Cooley, R. E., Dull, C. Y., Guiora, A. Z., & Paluszny, M. (1974). Grammatical gender and gender identity development: Cross cultural and cross lingual implications. *American Journal of Orthopsychiatry*, *44*(3), 424–431. https://doi.org/10.1111/j.1939-0025.1974.tb00895.x

Belacchi, C., & Cubelli, R. (2012). Implicit knowledge of grammatical gender in preschool children. *Journal of Psycholinguistic Research*, *41*(4), 295–310. https://doi.org/10.1007/s10936-011-9194-y

Beller, S., Brattebø, K. F., Lavik, K. O., Reigstad, R. D., & Bender, A. (2015). Culture or language: What drives effects of grammatical gender? *Cognitive Linguistics*, *26*(2), 331–359. https://doi.org/10.1515/cog-2014-0021

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

Bender, A., Beller, S., & Klauer, K. C. (2016). Lady Liberty and Godfather Death as candidates for linguistic relativity? Scrutinizing the gender congruency effect on personified allegories with explicit and implicit measures. *Quarterly Journal of Experimental Psychology*, *69*(1), 48–64. https://doi.org/10.1080/17470218.2015.1021701

Bender, A., Beller, S., & Klauer, K. C. (2018). Gender congruency from a neutral point of view: The roles of gender classes and conceptual connotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(10), 1580–1608. https://doi.org/10.1037/xlm0000534

Bezpoiasko, O. K., Gorodenska, K. G., & Rusanivskiy, V. M. (1993). *Grammar of Ukrainian language. Morphology* (M. S. Tymoshyk & L. L. Ščerbatenko, Eds.; 1st ed.). Lybid.

Bilaniuk, L., & Melnyk, S. (2008). A Tense and Shifting Balance: Bilingualism and Education in Ukraine. *International Journal of Bilingual Education and Bilingualism*, *11*, 340–372. https://doi.org/10.1080/13670050802148731

Bloom, A. H. (1981). *The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West*. Psychology Press. https://doi.org/10.4324/9781315802893

Boas, F. (1911). *The Mind of a Primitive Man*. The Macmillan Company.

Boas, F. (1920). The Methods of Ethnology. *American Anthropologist*, *22*(4), 311–321.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123–129. https://doi.org/10.1016/j.cognition.2010.09.010

Boroditsky, L., & Schmidt, L. A. (2000). Sex, Syntax, and Semantics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *22*(22). https://escholarship.org/uc/item/0jt9w8zf

Boroditsky, L., Scmidt, L. A., & Phillips, W. (2003). Sex, syntax and semantics. In *Language in mind: Advances in the study of language and thought* (pp. 61–79). Boston Review. https://doi.org/10.7551/mitpress/4117.001.0001

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

Boutonnet, B., Dering, B., Viñas-Guasch, N., & Thierry, G. (2013). Seeing Objects through the Language Glass. *Journal of Cognitive Neuroscience*, *25*(10), 1702–1710. https://doi.org/10.1162/jocn_a_00415

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Bross, F. (2019). Using Mixed Effect Models to Analyze Acceptability Rating Data in Linguistics. Version 1.0. *Mimeo*. https://doi.org/www.fabianbross.de/mixedmodels.pdf.

Budzhak-Jones, S. (1997). Quantitative analysis of gender assignment in mono/bilingual discourse. *Journal of Quantitative Linguistics*, *4*(1–3), 67–91. https://doi.org/10.1080/09296179708590080

Bylund, E., & Athanasopoulos, P. (2014). Linguistic Relativity in SLA: Toward a New Research Program. *Language Learning*, *64*(4), 952–985. https://doi.org/10.1111/lang.12080

Bylund, E., & Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General*, *146*(7), 911–916. https://doi.org/10.1037/xge0000314

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, *51*(5), 929–955. https://doi.org/10.1515/ling-2013-0033

Bylund, E., Samuel, S., & Athanasopoulos, P. (2024). Crosslinguistic Differences in Food Labels Do Not Yield Differences in Taste

Perception. *Language Learning*, *74*(S1), 20–39. https://doi.org/10.1111/lang.12641

Cambridge University Press. (2021). *IELTS (International English Language Testing System) | Cambridge English*. https://www.cambridgeenglish.org/exams-and-tests/ielts/

Cambridge University Press. (2024). *Test your English | Cambridge English*. https://www.cambridgeenglish.org/test-your-english/

Casaponsa, A., García-Guerrero, M. A., Martínez, A., Ojeda, N., Thierry, G., & Athanasopoulos, P. (2024). Electrophysiological Evidence for a Whorfian Double Dissociation of Categorical Perception Across Two Languages. *Language Learning*, *74*(S1), 136–156. https://doi.org/10.1111/lang.12648

Casasanto, D. (2008). *Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought* (SSRN Scholarly Paper 1402163). https://papers.ssrn.com/abstract=1402163

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*(3), 714–730. https://doi.org/10.1111/lang.12192

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., & Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*(26).

Chatterjee, R. (1985). Reading Whorf through Wittgenstein: A solution to the linguistic relativity problem. *Lingua*, *67*(1), 37–63. https://doi.org/10.1016/0024-3841(85)90012-9

Chen, Z., & Faitaki, F. (2024). Effects of the French grammatical gender system on bilingual adults' perception of objects. *Bilingualism: Language and Cognition*, 1–14. https://doi.org/10.1017/S1366728924000464

Christensen, R. H. B. (2019). *Cumulative Link Models for Ordinal Regression with the R Package ordinal* (Version R package version 2019.12-10) [Computer software]. http://www.cran.r-project.org/package=ordinal/

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf Hypothesis and Probabilistic Inference: Evidence from the Domain of Color. *PLOS ONE*, *11*(7), e0158725. https://doi.org/10.1371/journal.pone.0158725

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

Cook, V., & Bassetti, B. (with Cook, V., & Bassetti, B.). (2011). *Language and Bilingual Cognition* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203836859

Cook, V. J. (2002). Background to the L2 user. In V. J. Cook (Ed.), *Portraits of the L2 user* (pp. 21–28). Multilingual Matters. https://www.degruyter.com/document/doi/10.21832/9781853595851/pdf?licenseType=restricted#page=9

Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, G. G. (2013). Number of Genders (v2020.4). In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Zenodo. https://doi.org/10.5281/zenodo.13950591

Csernicskó, I., & Máté, R. (2017). *Bilingualism in Ukraine: Value or Challenge? 10*, 14–35. http://dx.doi.org/10.1515/sm-2017-0001

Degroot, A. M. B., Dannenburg, L., & Vanhell, J. G. (1994). Forward and Backward Word Translation by Bilinguals. *Journal of Memory and Language*, *33*(5), 600–629. https://doi.org/10.1006/jmla.1994.1029

Delorme, A. (2023). EEG is better left alone. *Scientific Reports*, *13*(1), 2372. https://doi.org/10.1038/s41598-023-27528-0

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Devitt, M., & Sterelny, K. (1999). *Language and Reality: An Introduction to the Philosophy of Language*. MIT Press.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, *104*(3), 1097–1102. https://doi.org/10.1073/pnas.0610132104

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—Third Edition*. https://doi.org/10.1037/t15145-000

Eberhard, D., Simons, G., & Fennig, C. (2019). *Ethnologue: Languages of the World, 22nd Edition*.

Eberhard, K. M., Heilman, M., & Scheutz, M. (2005). An empirical and computational test of linguistic relativity. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27*(27). https://escholarship.org/content/qt86j5k3tv/qt86j5k3tv.pdf

Elffers, E. (2012). *The Rise of General Linguistics as an Academic Discipline: Georg von der Gabelentz as a Co-Founder* (p. 18). https://doi.org/10.1017/9789048517336.004

Elpers, N., Jensen, G., & Holmes, K. J. (2022). Does grammatical gender affect object concepts? Registered replication of Phillips and Boroditsky (2003). *Journal of Memory and Language*, *127*, 104357. https://doi.org/10.1016/j.jml.2022.104357

Ervin, S. M. (1961). Learning and Recall in Bilinguals. *The American Journal of Psychology*, *74*(3), 446–451. https://doi.org/10.2307/1419752

Everett, C. (2013). *Linguistic Relativity: Evidence Across Languages and Cognitive Domains*. De Gruyter, Inc. http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=1130355

Fatemi, S. M. (2024). The influence of grammatical gender on the conceptualization of the world: A systematic literature review. *GENDER*, *16*(3), 168–186.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157. https://doi.org/10.3758/s13423-010-0021-5

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 202–224. https://doi.org/10.1037/0278-7393.27.1.202

Filipović, L. (2011). *Speaking and remembering in one or two languages: Bilingual vs. Monolingual lexicalization and memory for motion events*. *15*(4), 466–485. https://doi.org/10.1177/13670069114030

Flaherty, M. (2001). How a Language Gender System Creeps into Perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31. https://doi.org/10.1177/0022022101032001005

Flecken, M., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving Along the Road or Heading for the Village? Conceptual Differences Underlying Motion Event Encoding in French, German, and French–German L2 Users. *The Modern Language Journal*, *99*(S1), 100–122. https://doi.org/10.1111/j.1540-4781.2015.12181.x

Fodor, J. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A. (1983). *The Modularity of Mind*. The MIT Press. https://doi.org/10.7551/mitpress/4737.001.0001

Forbes, J., Poulin-Dubois, D., Rivero, M. R., & Sera, M. D. (2008). Grammatical Gender Affects Bilinguals' Conceptual Gender: Implications for Linguistic Relativity and Decision Making. *The Open Applied Linguistics Journal*, *1*, 68–76. https://doi.org/10.2174/1874913500801010068

Franklin, A., Pilling, M., & Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, *91*(3), 227–248. https://doi.org/10.1016/j.jecp.2005.03.003

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484. https://doi.org/10.1023/A:1023264209610

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Gentner, D., & Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought* (1st ed., pp. viii–viii). MIT Press. https://doi.org/10.7551/mitpress/4117.001.0001

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual language profile. In *Measuring L2 Proficiency: Perspectives from SLA* (pp. 208–225).

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gorpynyč, V. O. (2004). *Morfolohiya ukrayins'koyi movy: Pidruchnyk dlya studentiv vyshchykh navchal'nykh zakladiv [Morphology of the Ukrainian language: A textbook for students of higher educational institutions]* (1st ed.). Academy.

Grosjean, F. (Ed.). (2021). The Extent of Bilingualism. In *Life as a Bilingual: Knowing and Using Two or More Languages* (pp. 27–39). Cambridge University Press. https://doi.org/10.1017/9781108975490.003

Grotjahn. (1992). *Der C-test: Theoretische Grundlagen und praktische Anwendungen*. AKS-Verlag. https://scholar.google.com/scholar_lookup?title=Der+C-Test.+Theoretische+Grundlagen+und+Praktische+Anwendungen&author=Coleman+J.+A.&author=Grotjahn+R.&publication+year=1994&pages=217-237#d=gs_cit&t=1745934490992&u=%2Fscholar%3Fq%3Dinfo%3AsAWqsC71LK8J%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den

Guiora, A. Z. (1984). The Dialectic of Language Acquisition. *Language Learning*, *33*(s5), 3–12. https://doi.org/10.1111/j.1467-1770.1984.tb01321.x

Guiora, A. Z., & Acton, W. R. (1979). PERSONALITY AND LANGUAGE BEHAVIOR: A RESTATEMENT[1]. *Language Learning*, *29*(1), 193–204. https://doi.org/10.1111/j.1467-1770.1979.tb01059.x

Guiora, A. Z., & Sagi, A. (1978). A Cross-Cultural Study of Symbolic Meaning–Developmental Aspects. *Language Learning*, *28*(2), 381–386. https://doi.org/10.1111/j.1467-1770.1978.tb00141.x

Gumperz, J. J., & Levinson, S. C. (1996). *Rethinking linguistic relativity*. University Press.

Haertlé, I. (2017). Does Grammatical Gender Influence Perception? A Study of Polish and French Speakers. *Psychology of Language and Communication*, *21*(1), 386–407. https://doi.org/10.1515/plc-2017-0019

Haun, D. B. M., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, *119*(1), 70–80. https://doi.org/10.1016/j.cognition.2010.12.009

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*(2), 259–276. https://doi.org/10.1006/brcg.1994.1014

Holmes, K. J., & Wolff, P. (2010). Simulation From Schematics: Dorsal Stream Processing and the Perception of Implied Motion. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2704–2709). Cognitive Science Society.

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

Humboldt, W. (1836). *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind* (P. Heath, Ed.). Cambridge University Press.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, *98*(3), 377–389. https://doi.org/10.1037/0033-295X.98.3.377

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All Giraffes Have Female-Specific Properties: Influence of Grammatical Gender on Deductive Reasoning About Sex-Specific Properties in German Speakers. *Cognitive Science*, *38*(3), 514–536. https://doi.org/10.1111/cogs.12074

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 21–36. https://doi.org/10.3758/CABN.5.1.21

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2010). Does the grammatical count/mass distinction affect semantic representations? Evidence from experiments in English and Japanese. *Language and Cognitive Processes*, *25*(2), 189–223. https://doi.org/10.1080/01690960902978517

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*(2), 417–426. https://doi.org/10.1016/j.cognition.2006.07.008

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, *78*(1), 193–202. https://doi.org/10.1007/BF00230699

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology*, *116*(11), 2613–2631. https://doi.org/10.1016/j.clinph.2005.07.005

Kaushanskaya, M., & Smith, S. (2016). Do grammatical-gender distinctions learned in the second language influence native-language lexical processing? *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *20*(1), 30–39. https://doi.org/10.1177/1367006915576830

Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, *86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Kecskés-Kovács, K., Sulykos, I., & Czigler, I. (2013). Is it a face of a woman or a man? Visual mismatch negativity is sensitive to gender category. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00532

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*(4), 638–653. https://doi.org/10.1037/a0020507

Kirjavainen, M., Kite, Y., & Piasecki, A. E. (2020). The Effect of Language-Specific Characteristics on English and Japanese Speakers' Ability to Recall Number Information. *Cognitive Science*, *44*(12), e12923. https://doi.org/10.1111/cogs.12923

Koch, S. C., Zimmermann, F., & Garcia-Retamero, R. (2007). El sol—Die Sonne. *Psychologische Rundschau*, *58*(3), 171–182. https://doi.org/10.1026/0033-3042.58.3.171

Kodish, B. I. (2003). What We Do with Language—What It Does with Us. *ETC: A Review of General Semantics*, *60*(4), 383–395.

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

Kortmann, B., & Auwera, J. V. D. (2011). *The Languages and Linguistics of Europe: A Comprehensive Guide*. De Gruyter, Inc. http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=765858

Kostiučenko, A. (2023). *Surzhyk in Ukraine: Between Language Ideology and Usage*. *1*, 15–17. https://doi.org/10.3929/ethz-b-000623475

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 843–858. https://doi.org/10.1037/0278-7393.34.4.843

Kroll, J. F., & Mendoza, G. A. (2022). Bilingualism: A Cognitive and Neural View of Dual Language Experience. In *Oxford Research Encyclopedia of Psychology*. https://doi.org/10.1093/acrefore/9780190236557.013.900

Kulyk, V. (2023, January 7). *Мова та ідентичність в Україні на кінець 2022-го [Language and identity in Ukraine by the end of 2022]*. Zbruč. https://zbruc.eu/node/114247

Kumar, M., Federmeier, K. D., & Beck, D. M. (2021). The N300: An Index for Predictive Coding of Complex Visual Objects and Scenes.

Cerebral Cortex Communications, *2*(2), tgab030. https://doi.org/10.1093/texcom/tgab030

Kupisch, T., Geiss, M., Mitrofanova, N., & Westergaard, M. (2022). Structural and phonological cues for gender assignment in monolingual and bilingual children acquiring German. Experiments with real and nonce words. *Glossa: A Journal of General Linguistics*, *7*(1), Article 1. https://doi.org/10.16995/glossa.5696

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208.

Lambelet, A. (2016). Second grammatical gender system and grammatical gender-linked connotations in adult emergent bilinguals with French as a second language. *International Journal of Bilingualism*, *20*(1), 62–75. https://doi.org/10.1177/1367006915576832

Leavitt, J. (2011). *Linguistic Relativities: Language Diversity and Modern Thought*. Cambridge University Press.

Lee, C.-Y., Liu, Y.-N., & Tsai, J.-L. (2012). The Time Course of Contextual Effects on Visual Word Recognition. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00285

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior*

Research Methods, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. The MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Lucy, J. (2011). Language and cognition: The view from anthropology. In *Language and Bilingual Cognition*. Psychology Press.

Lucy, J. A. (1992). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis* (Vol. 13, Issue 13, p. xi 328). Cambridge University Press.

Lucy, J. A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, *2*, 291–312.

Lucy, J. A. (2016). Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning*, *66*(3), 487–515. https://doi.org/10.1111/lang.12195

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of Language on Visual Perception. *Trends in Cognitive Sciences*, *24*(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Maciuszek, J., Polak, M., & Świątkowska, N. (2019). Grammatical Gender Influences Semantic Categorization and Implicit Cognition in Polish. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02208

Macwhinney, B. (2005). The emergence of linguistic form in time. *Connection Science*, *17*(3–4), 191–211. https://doi.org/10.1080/09540090500177687

Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 859–876. https://doi.org/10.3758/s13415-018-00679-8

Maier, M., & Abdel Rahman, R. (2024). Transient and long-term linguistic influences on visual perception: Shifting brain dynamics with memory consolidation. *Language Learning*, *74*(Suppl 1), 157–184. https://doi.org/10.1111/lang.12631

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996). *Independent Component Analysis of Electroencephalographic Data*. *8*.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marzecová, A., Schettino, A., Widmann, A., SanMiguel, I., Kotz, S. A., & Schröger, E. (2018). Attentional gain is modulated by probabilistic feature expectations in a spatial cueing task: ERP evidence. *Scientific Reports*, *8*(1), 54. https://doi.org/10.1038/s41598-017-18347-1

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65. https://doi.org/10.1017/S0048577299971196

Mecit, A., Shrum, L. J., & Lowrey, T. M. (2020). *COVID-19 is Feminine: Grammatical Gender Influences Future Danger Perceptions and Precautionary Behavior*. OSF. https://doi.org/10.31234/osf.io/x8kp2

Mickan, A., Schiefke, M., & Stefanowitsch, A. (2014). Key is a llave is a Schlüssel: A failure to replicate an experiment from Boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association*, *2*(1), 39–50. https://doi.org/10.1515/gcla-2014-0004

Mikić Ljubi, J., Matkovič, A., Bon, J., & Kanjuo Mrčela, A. (2022). The effects of grammatical gender on the processing of occupational role names in Slovene: An event-related potential study. *Frontiers in Psychology*, *13*, 1010708. https://doi.org/10.3389/fpsyg.2022.1010708

Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space–time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review*, *18*(3), 598–604. https://doi.org/10.3758/s13423-011-0068-y

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. https://doi.org/10.1016/j.cognition.2017.10.010

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied*

Linguistics in Language Teaching, *56*(1), 1–24. https://doi.org/10.1515/iral-2017-0016

Mitrofanova, N., Rodina, Y., Urek, O., & Westergaard, M. (2018). Bilinguals' sensitivity to grammatical gender cues in Russian: The role of cumulative input, proficiency, and dominance. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01894

Monaghan, L. (2011). The Expanding Boundaries of Linguistic Anthropology: 2010 in Perspective. *American Anthropologist*, *113*(2), 222–234. https://doi.org/10.1111/j.1548-1433.2011.01326.x

Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience*, *15*(6), 900–910. https://doi.org/10.1162/089892903322370816

Mouchetant-Rostaing, Y., Giard, M.-H., Bentin, S., Aguera, P.-E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, *12*(1), 303–310. https://doi.org/10.1046/j.1460-9568.2000.00888.x

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

Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, *17*(5), 220–229. https://doi.org/10.1016/j.tics.2013.03.008

O'Grady, P. (2004). Wittgenstein and Relativism. *International Journal of Philosophical Studies*, *12*(3), 315–337. https://doi.org/10.1080/0967255042000243975

Osypenko, O., Brandt, S., & Athanasopoulos, P. (2025). The influence of three-gendered grammatical systems on simultaneous bilingual cognition: The case of Ukrainian-Russian bilinguals. *Language and Cognition*, *17*, e25. https://doi.org/10.1017/langcog.2024.73

Oxford University Press. (2001). *Quick Placement Test*. Oxford University Press.

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*(4), 477–493. https://doi.org/10.1037/0096-3445.131.4.477

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Park, H. I., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean–English bilinguals. *Bilingualism: Language and Cognition*, *17*(2), 410–430. https://doi.org/10.1017/S1366728913000400

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition*, *2*(3), 209–230. https://doi.org/10.1017/S1366728999000322

Pavlenko, A. (2011). *Thinking and speaking in two languages* (1st ed.). Multilingual Matters. https://doi.org/10.21832/9781847693389

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Pavlenko, A. (2014). *The Bilingual mind: And what it tells us about language and thought*. Cambridge University Press.

Pavlidou, T.-S., & Alvanoudi, A. (2013). Grammatical Gender and Cognition. *Versita/de Gruyter*, *2*(2), 109–124.

Pavlidou, T.-S., & Alvanoudi, A. (2019). Conceptualizing the world as ‘female’ or ‘male’: Further remarks on grammatical gender and speakers’ cognition. *Selected Papers on Theoretical and Applied Linguistics*, *23*(0), Article 0. https://doi.org/10.26262/istal.v23i0.7351

Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). https://escholarship.org/uc/item/31t455gf

Piaget, J. (2005). *Language and Thought of the Child: Selected Works vol 5* (4th ed.). Routledge. https://doi.org/10.4324/9780203992739

Pinker, S. (1994). *The language instinct: How the mind creates language* (22. Dr). HarperPerennial.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Poftak, N., & Shykula, D. (2022). The Ongoing Language Discussion in Ukraine. In *Being Ukraine: An Introduction to Europe’s Eastern Vanguard*. Connecticut College. https://openpress.digital.conncoll.edu/beingukraine/chapter/chapter-7/

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pugh, S., & Press, I. (1999). *Ukrainian: A Comprehensive Grammar*. Routlege.

Pullum, G. K. (1991). *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*. University of Chicago Press.

R Core Team. (2022). *R: The R Project for Statistical Computing*. The R Project for Statistical Computing. https://www.r-project.org/

Rączaszek-Leonardi, J. (2010). Multiple Time-Scales of Language Dynamics: An Example From Psycholinguistics. *Ecological Psychology*, *22*(4), 269–285. https://doi.org/10.1080/10407413.2010.517111

Rahimi, M., & Saadat, M. (2005). A Verbal Protocol Analysis of a C-Test. *International Journal of Applied Linguistics*, *8*(2), 55–85.

Ramos, S., & and Roberson, D. (2011). What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese. *Journal of Cognitive Psychology*, *23*(1), 102–111. https://doi.org/10.1080/20445911.2011.466795

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *WIREs Cognitive Science*, *8*(6), e1440. https://doi.org/10.1002/wcs.1440

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Roberson, D., Davies, I., & Davidoff, J. (2000). Color Categories Are Not Universal: Replications and New Evidence From a Stone-Age Culture. *Journal of Experimental Psychology. General*, *129*(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Rusanivskyj, V. M., Taranenko, O. O., & Zjabljuk, M. P. (2004). *Encyclopaedia of the Ukrainian language* (M. P. Bažan, Ed.; 2nd ed.). Kyiv: 'Ukrainian Encyclopaedia'.

Saalbach, H., Imai, M., & Schalk, L. (2012). Grammatical Gender and Inferences About Biological Properties in German-Speaking Children. *Cognitive Science*, *36*(7), 1251–1267. https://doi.org/10.1111/j.1551-6709.2012.01251.x

Sáchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, *4*(4), 293–310. https://doi.org/10.1080/09541449208406189

Sakarias, M., & Flecken, M. (2019). Keeping the Result in Sight and Mind: General Cognitive Principles and Language-Specific Influences in the Perception and Memory of Resultative Events. *Cognitive Science*, *43*(1), e12708. https://doi.org/10.1111/cogs.12708

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*(3), 257–275. https://doi.org/10.1017/S1366728907003069

Sá-Leite, A. R., & Lago, S. (2024). The role of word form in gender processing during lexical access: A theoretical review and novel proposal in language comprehension. *Psychonomic Bulletin & Review*, *31*(5), 1934–1953. https://doi.org/10.3758/s13423-023-02426-8

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

Sapir, E. (1929). The Status of Linguistics as a Science. *Language*, *5*(4), 207–214. https://doi.org/10.2307/409588

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

Sato, S., Casaponsa, A., & Athanasopoulos, P. (2020). Flexing Gender Perception: Brain Potentials Reveal the Cognitive Permeability of Grammatical Information. *Cognitive Science*, *44*(9), e12884. https://doi.org/10.1111/cogs.12884

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

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945. https://doi.org/10.1016/S0028-3932(01)00176-2

Schendan, H. E., & Kutas, M. (2003). Time Course of Processes and Representations Supporting Visual Object Identification and Memory. *Journal of Cognitive Neuroscience*, *15*(1), 111–135. https://doi.org/10.1162/089892903321107864

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning1. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

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

Segel, E., & Boroditsky, L. (2011). Grammar in art. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00244

Sera, M. D., Berge, C. A. H., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*(3), 261–292. https://doi.org/10.1016/0885-2014(94)90007-8

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

Shumlianskyi, S. (2010). *Conflicting abstractions: Language groups in language politics in Ukraine*. *2010*(201), 135–161. https://doi.org/10.1515/ijsl.2010.007

Slobin, D. I. (1996). *From "thought and language" to "thinking for speaking"*. https://philarchive.org/archive/SLOFTA

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, *06*(04), 537–558. https://doi.org/10.1142/S0219525903001055

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Speed, L. J., & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention,* *Perception, & Psychophysics*, *81*(6), 2063–2077. https://doi.org/10.3758/s13414-019-01729-0

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the Sexes in Language. In *Social communication* (pp. 163–187). Psychology Press.

Steinback, S. (2015). *Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png 1,853×1,642 pixels*. Lexical Distance among the Languages of Europe. https://upload.wikimedia.org/wikipedia/commons/f/f3/Lexical-distance-among-the-languages-of-europe-2-1-mid-size.png

Talmy, L. (1985). Lexicalization Patterns: Semantic Structure in Lexical Forms. In T. Shopen (Ed.), *Language typology and syntactic description III: Grammatical categories and the lexicon*. Cambridge University Press.

Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567–4570. https://doi.org/10.1073/pnas.0811155106

Thierry, G., Rahman, R. A., & Athanasopoulos, P. (2024). An Introduction to the Cognitive Neuroscience of Language Embodiment and Relativity Special Issue of the Language Learning Cognitive Neuroscience Series. *Language Learning*, *74*(S1), 5–19. https://doi.org/10.1111/lang.12643

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language*, *69*(2), 121–134. https://doi.org/10.1016/j.jml.2013.03.004

Ukrainian center for educational quality assessment. (2020). *Ukrainian center for educational quality assessment*. https://testportal.gov.ua/en/

Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language*, *87*, 157–176. https://doi.org/10.1016/j.jml.2015.10.009

Underhill, J. W. (2009). *Humboldt, Worldview and Language*. Edinburgh University Press. https://www.jstor.org/stable/10.3366/j.ctt1r29b9

Vakulenko, M. (2023). Normalization of Ukrainian letters, numerals, and measures for natural language processing. *Digital Scholarship in the Humanities*, *38*(3), 1307–1321. https://doi.org/10.1093/llc/fqac090

Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, *206*, 104485. https://doi.org/10.1016/j.cognition.2020.104485

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461. https://doi.org/10.1162/08989290152001880

Vasishth, S., & Broe, M. (2011). The Foundations of Statistics: A Simulation-Based Approach. *CHANCE*, *24*(4), 59–61. https://doi.org/10.1080/09332480.2011.10739891

Vernich, L. (2017). *Does learning a foreign language affect object categorization in native speakers of a language with grammatical* *gender? The case of Lithuanian speakers learning three languages with different types of gender systems (Italian, Russian and German)*. https://journals.sagepub.com/doi/full/10.1177/1367006917728593

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology: General*, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Wang, Y. (2020). Cognitive restructuring in the multilingual mind: Motion event construal in Cantonese-English-Japanese multilingual speakers [Doctoral, UCL(University College London)]. In *Doctoral thesis, UCL(University College London).* UCL(University College London). https://discovery.ucl.ac.uk/id/eprint/10100838/

Wang, Y., & Wei, L. (2019). Cognitive restructuring in the bilingual mind: Motion event construal in early Cantonese–English bilinguals. *Language and Cognition: An Interdisciplinary Journal of Language and Cognitive Science*, *11*(4), 527–554. https://doi.org/10.1017/langcog.2019.31

Whorf, B. L. (1956). *Language, Thought, and Reality, second edition: Selected Writings of Benjamin Lee Whorf* (Caroll, J.B., Ed.). MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of*

Sciences, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Wittgenstein, L. (1922). *Tractatus logico-philosophicus.* Harcourt, Brace.

Wittgenstein, L. (1953). *Philosophical Investigations* (G. Anscombe, Ed.). Wiley-Blackwel.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Xia, T., Xu, G., & Mo, L. (2019). Bi-lateralized Whorfian effect in color perception: Evidence from Chinese Sign Language. *Journal of Neurolinguistics*, *49*, 189–201. https://doi.org/10.1016/j.jneuroling.2018.07.004

Xue, Y., & Williams, J. (2024). Inducing Shifts in Attentional and Preattentive Visual Processing Through Brief Training on Novel Grammatical Morphemes: An Event-Related Potential Study. *Language Learning*, *74*(S1), 185–223. https://doi.org/10.1111/lang.12642

Yu, M., Li, Y., Mo, C., & Mo, L. (2017). Newly learned categories induce pre-attentive categorical perception of faces. *Scientific Reports*, *7*(1), 14006. https://doi.org/10.1038/s41598-017-14104-6

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, *107*(22), 9974–9978. https://doi.org/10.1073/pnas.1005669107

Appendix A. Supplementary Materials for Chapter 3

A.1. The role of participants' gender on the similarity ratings

To address the potential effects of gender imbalance in the bilingual participant groups in both Experiment 1 and 2 (see Table A.1), we have conducted a separate analysis examining gender as a potential predictor on similarity ratings. Except for "female" and "male", our participants also had choices "prefer not to say" and "other", however, none of our participants chose either of these options.

Table A.1

Number of participants in Experiments 1 and 2, based on their gender

	Experiment 1		Experiment 2	
	Ukrainian-Russian bilinguals	**English monolinguals**	**Ukrainian-Russian bilinguals**	**English monolinguals**
Male	3	15	20	16
Female	48	9	44	18
Total number of participants	51	24	64	34

In Experiment 1, we designed two cumulative link mixed models for each part of the analysis. In the first part where we compare the similarity ratings between Ukrainian-Russian bilinguals and English monolinguals, the maximum convergence model included a three-way interaction between group (Ukrainian-Russian vs English), participant's gender (male vs female) and grammatical gender of an item in both L1s (feminine vs masculine vs neuter) and whether there were any effects on Likert scores.

By-item and by-participant random intercepts were also added to the model. The results of the clmm model are presented in Table A.2.

Table A.2

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 1, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, including Participant's Gender as a predictor

	Estimate	Std. Error	z-value	Pr(>|z|)
Group (Ukrainian)	**-1.349**	**0.392**	**-3.445**	**0.001*******
Participant's gender (male)	-0.659	0.560	-1.178	0.239
Grammatical gender of an item in both languages (masculine)	-0.141	0.236	-0.598	0.550
Grammatical gender of an item in both languages (neuter)	-0.087	0.236	-0.369	0.712
Group (Ukrainian) * Participant's gender (male)	**1.605**	**0.704**	**2.279**	**0.023***
Group (Ukrainian) * Grammatical gender of an item in both languages (masculine)	0.187	0.167	1.120	0.263
Group (Ukrainian) * Grammatical gender of an item in both languages (neuter)	0.205	0.168	1.226	0.220
Participant's gender (male)*Grammatical gender of an item in both languages (masculine)	**0.486**	**0.236**	**2.064**	**0.039***
Participant's gender (male)*Grammatical gender of an item in both languages (neuter)	0.303	0.237	1.282	0.200
Group (Ukrainian) * Participant's gender (male)*Grammatical gender of an item in both languages (masculine)	-0.450	0.299	-1.505	0.132
Group (Ukrainian) * Participant's gender (male)*Grammatical gender of an item in both languages (neuter)	-0.175	0.302	-0.579	0.562

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

If there were an effect of participant gender specifically within the bilingual group, we would expect a significant three-way interaction between group (Ukrainian-Russian vs. English), participant's gender (male vs. female), and grammatical gender of the item in both L1s (masculine vs. feminine). This interaction would indicate that male Ukrainian-Russian bilingual participants rated items with masculine grammatical gender significantly higher than those with feminine grammatical gender. However, this three-way interaction was not significant ($SE = 0.302$, $z = -0.579$, $p = 0.562$), suggesting that participant gender did not have a unique effect within the bilingual group based on grammatical gender.

However, the only significant fixed effect was group (Ukrainian-Russian bilinguals), which was unrelated to the prediction about participant gender effects. In addition, two interactions reached significance. Firstly, group and participant's gender significant interaction ($SE = 0.704$, $z = 2.279$, $p = 0.023$) indicated that, on average, male Ukrainian-Russian bilingual participants rated all items higher than female bilingual participants. Secondly, significant interaction between participant's gender and grammatical gender in both L1s ($SE = 0.236$, $z = 2.064$, $p = 0.039$) suggested that male participants from both groups (Ukrainian-Russian bilinguals and English monolinguals) rated items with masculine grammatical gender significantly higher than those with feminine grammatical gender. However, while male participants tended to rate masculine-gendered items higher overall, this pattern was consistent across both groups (including English monolinguals that were not aware of grammatical gender systems in Ukrainian and Russian), with no evidence of a unique gender effect specific to the bilingual group.

In the second part of the analysis, we examined a model with a four-way interaction involving condition (congruent in Ukrainian vs. congruent in Russian), grammatical gender in Ukrainian (masculine vs. feminine), participant's gender (male vs. female), and group proficiency (ranging from -100 for full proficiency in Russian to +100 for full proficiency in Ukrainian). A significant four-way interaction would indicate that male Ukrainian-Russian bilingual participants with higher proficiency in Ukrainian rated stimuli with masculine grammatical gender in Ukrainian higher than those with feminine grammatical gender. And vice versa for female participants. However, no significant interactions confirming this prediction was found (Table A.3).

As can be seen in Table A.3, no significant effects were found for either interaction/ main effect, suggesting that participant's gender did not significantly impact their responses when rating nouns that have mismatching grammatical gender in Ukrainian and Russian.

Table A.3

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 1, Comparing Ukrainian-Russian bilingual group based on Language Proficiency, including Participant's Gender as a predictor

	Estimate	**Std. Error**	**z-value**	**Pr(>|z|)**
Condition (congruent in Russian)	-0.369	0.438	-0.842	0.4
Participant's gender (male)	0.731	0.549	1.331	0.183
Grammatical gender of an item in Ukrainian (masculine)	-0.280	0.462	-0.607	0.544
Language Proficiency	0.007	0.011	0.59	0.555
Condition (congruent in Russian) * Participant's gender (male)	0.036	0.319	0.113	0.91
Condition (congruent in Russian) * Grammatical gender of an item in Ukrainian (masculine)	0.316	0.653	0.484	0.628
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	0.169	0.332	0.507	0.612
Condition (congruent in Russian) * Language Proficiency	-0.004	0.007	-0.568	0.57
Participant's gender (male) * Language Proficiency	0.006	0.038	0.165	0.869
Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.003	0.008	-0.427	0.669
Condition (congruent in Russian) * Participant's gender (male) * Language Proficiency	-0.071	0.468	-0.152	0.879
Condition (congruent in Russian) * Participant's gender (male) * Language Proficiency	-0.022	0.022	-1.008	0.313
Condition (congruent in Russian) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.004	0.011	0.356	0.722
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.019	0.023	-0.823	0.411
Condition (congruent in Russian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.024	0.032	0.726	0.468

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To check if participants' gender affected their ratings in Experiment 2, we build clmm models analogously to Experiment 1. In the first part of the analysis where we analysed ratings for nouns with matching grammatical gender across Ukrainian and Russian (Table A.4), the maximum convergence model included a four-way interaction between Condition (congruent in both L1s vs incongruent in both L1s), Grammatical gender of an item in both languages (masculine vs feminine), Participant's gender (male vs female), Group (Ukrainian-Russian bilinguals vs English monolinguals). It also included random intercepts for participants and items.

We did not find any significant interactions either for participant's gender – grammatical gender in 2L1s ($SE = 0.070$, $z = 0.309$, $p = 0.757$), or for the four-way condition-participant's gender – group – grammatical gender interaction ($SE = 0.145$, $z = 0.683$, $p = 0.495$), indicating that participants' gender did not have a significant effect on ratings of items with masculine grammatical gender.

Table A.4

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, including Participant's Gender as a predictor

	Estimate	Std. Error	z value	Pr(>|z|)
Condition (incongruent in both L1s)	-0.051	0.048	-1.057	0.291
Grammatical gender of an item in both languages (masculine)	-0.098	0.059	-1.669	0.095
Participant's gender (male)	**0.736**	**0.236**	**3.121**	**0.002****
Group (Ukrainian-Russian bilinguals)	**1.760**	**0.193**	**9.142**	**< 0.001*****
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine)	0.080	0.068	1.181	0.237
Condition (incongruent in both L1s) * Participant's gender (male)	0.020	0.070	0.287	0.774
Grammatical gender of an item in both languages (masculine)* Participant's gender (male)	0.022	0.070	0.309	0.757
Condition (incongruent in both L1s) * Group (Ukrainian-Russian bilinguals)	**-1.897**	**0.063**	**-29.997**	**< 0.001*****
Grammatical gender of an item in both languages (masculine)* Group (Ukrainian-Russian bilinguals)	0.122	0.063	1.949	0.051
Participant's gender (male)* Group (Ukrainian-Russian bilinguals)	**-0.789**	**0.304**	**-2.597**	**0.009****
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine) * Participant's gender (male)	-0.044	0.099	-0.442	0.658
Condition (incongruent in both L1s) * Grammatical gender of an item in both languages (masculine) * Group (Ukrainian-Russian bilinguals)	-0.029	0.089	-0.333	0.739
Condition (incongruent in both L1s) * Participant's gender (male) * Group (Ukrainian-Russian bilinguals)	-0.114	0.102	-1.113	0.266
Grammatical gender of an item in both languages (masculine) * Participant's gender (male)* Group (Ukrainian-Russian bilinguals)	-0.021	0.102	-0.202	0.840
Condition (incongruent in both L1s) * Participant's gender (male) * Group (Ukrainian-Russian bilinguals) * Grammatical gender of an item in both languages (masculine)	0.099	0.145	0.683	0.495

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

In the part 2, the maximum convergence model condition (congruent in Ukrainian vs. congruent in Russian), grammatical gender in Ukrainian (masculine vs. feminine), participant's gender (male vs. female), and language proficiency (ranging from -100 for full proficiency in Russian to +100 for full proficiency in Ukrainian). Similarly, to the second clmm model in Experiment 1, no significant effects were found for either interaction/ main effect (Table A.5), suggesting that participant's gender did not significantly impact their responses when rating nouns that have mismatching grammatical gender in Ukrainian and Russian.

Overall, results from both experiments indicate that participants' gender did not significantly impact ratings or impact the main findings. The gender imbalance in the bilingual groups did not seem to introduce systematic bias into the results, and rating patterns were consistent across male and female participants.

Table A.5

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian bilingual group based on Language Proficiency, including Participant's Gender as a predictor

	Estimate	Std. Error	z-value	Pr(>|z|)
Condition (congruent in Ukrainian)	**0.108**	**0.052**	**2.085**	**0.037***
Grammatical gender of an item in Ukrainian (masculine)	-0.086	0.087	-0.982	0.326
Participant's gender (male)	-0.096	0.108	-0.895	0.371
Language Proficiency	-0.002	0.011	0.59	0.555
Condition (congruent in Ukrainian) * Grammatical gender of an item in Ukrainian (masculine)	0.036	0.003	-0.644	0.520
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	0.035	0.076	0.464	0.643
Condition (congruent in Russian) * Participant's gender (male)	**-0.250**	**0.081**	**-3.073**	**0.002****
Condition (congruent in Ukrainian) * Participant's gender (male)	-0.004	0.007	-0.568	0.57
Grammatical gender of an item (masculine)*Participant's gender (male)	-0.024	0.085	-0.289	0.773
Condition (congruent in Ukrainian) * Language Proficiency	**0.005**	**0.002**	**2.753**	**0.006****
Participant's gender (male) * Language Proficiency	-0.007	0.005	-1.412	0.158
Condition (congruent in Ukrainian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine)	0.004	0.120	0.033	0.974
Condition (congruent in Ukrainian) * Grammatical gender of an item in Ukrainian (masculine)* Language Proficiency	0.002	0.003	0.715	0.475
Condition (congruent in Ukrainian) * Participant's gender (male) * Language Proficiency	**0.015**	**0.004**	**3.702**	**<0.001*****
Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	0.004	0.004	0.839	0.402
Condition (congruent in Ukrainian) * Participant's gender (male) * Grammatical gender of an item in Ukrainian (masculine) * Language Proficiency	-0.005	0.006	-0.925	0.355

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A.2. Reanalysis of Experiment 2 results using stimuli from Experiment 1 (excluding neuter-gender stimuli)

In this subsection, we examine whether significant results would still emerge for the second experiment when using only a subset of stimuli from Experiment 1. Experiment 1 included 50 conceptually neutral items, 10 of which had neuter grammatical gender in both Ukrainian and Russian. After excluding the neuter stimuli, the subset consisted of 40 items, each paired once with a male and once with a female character (object-character pairs the same as in the Experiment 1). This yielded a total of 80 pairs chosen from 640 pairs.

The current analysis was divided into two parts, following the structure of the analyses in both Experiments 1 and 2: (1) the analysis of stimuli with matching grammatical gender in both L1s and (2) the analysis of stimuli with mismatching grammatical gender in the two L1s.

For the first part, we focused on the similarity ratings of Ukrainian-Russian bilinguals and English monolinguals, considering stimuli with matching grammatical gender in both Ukrainian and Russian. A cumulative link mixed model (CLMM) was employed, identical to the model used in prior analyses, to test whether the interaction between Group (Ukrainian-Russian bilinguals vs. English monolinguals) and Condition (congruent in both Ukrainian and Russian vs. incongruent in both Ukrainian and Russian) had a significant effect on similarity ratings (Likert scores). The maximum convergence model also included random intercepts for participants and items.

The results (Table A.6) revealed a robust statistically significant effect of the interaction (*SE* = 0.105, $z = -16.620$, $p < 0.001$ for Ukrainian-

Russian group and incongruent condition), consistent with the findings from the full analysis in Experiment 2. These results indicate that, even when using only the stimuli from Experiment 1 without adding additional pairs to mitigate potential semantic associations, Ukrainian-Russian bilinguals rated incongruent pairs in both of their L1s as significantly less similar compared to congruent pairs.

Table A.6

Results of the Cumulative Link Mixed Model (CLMM) for Similarity Ratings from Experiment 2, Comparing Ukrainian-Russian Bilinguals and English Monolinguals, using pairs of stimuli from Experiment 1 (without neuter gender)

	Estimate	Std. Error	z-value	Pr(>|z|)
Group (Ukrainian-Russian bilinguals)	**1.291**	**0.165**	**7.806**	**<0.001*****
Condition (incongruent in 2L1s)	0.075	0.074	1.008	0.313
Group (Ukrainian-Russian bilinguals) * Condition (incongruent in 2L1s)	**-1.750**	**0.105**	**-16.620**	**<0.001*****

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Similarly, in the second part of the analysis, we focused on the similarity ratings of pairs containing items with mismatching grammatical gender in the two L1s, as rated by Ukrainian-Russian bilinguals. To do so, we developed a clmm model to examine the interaction between language proficiency (ranging from -100 for exclusive proficiency in Russian to +100 for exclusive proficiency in Ukrainian) and condition (congruent in Ukrainian vs. congruent in Russian). By-item and by-participants random intercepts were also included in the model.

The results (Table A.7) revealed a statistically significant interaction between proficiency and condition ($SE = 0.004$, $z = 2.684$, $p = 0.007$ for pairs congruent in Ukrainian), also consistent with the findings from the full analysis in Experiment 2. This suggests that participants with higher proficiency in Ukrainian rated pairs congruent in Ukrainian (but incongruent in Russian) as more similar, whereas participants with higher proficiency in Russian showed the opposite pattern, rating pairs congruent in Russian as more similar.

Table A.7

Results of the cumulative link mixed model (CLMM) for similarity ratings from Experiment 2, comparing Ukrainian-Russian bilinguals based on the Language Proficiency, using pairs of stimuli from Experiment 1

	Estimate	**Std. Error**	**z-value**	**Pr(>|z|)**
Language Proficiency	-0.004	0.003	-1.260	0.208
Condition (congruent in Ukrainian)	0.101	0.087	1.155	0.248
Language Proficiency * Condition (congruent in Ukrainian)	**0.010**	**0.004**	**2.684**	**0.007****

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Overall, this analysis confirms that even when using pairs from Experiment 1 without expanding the stimuli to include pairings with all eight characters (as opposed to just two), significant effects of language on bilinguals' categorisation are still observed for the subset of data from the second experiment.

A.3. Distribution of Likert scores in Experiments 1 and 2

To explore the potential effects of the modifications in instructions between experiment 1 and 2, we analysed the differences in distributions of Likert scores. While the main body of instructions remained unchanged, in Experiment 2 participants received the instruction verbally in addition to seeing it on the screen, with the experimenter emphasising the need to use the whole range of scores from 1 to 9.

A.3.1. Comparing Ukrainian-Russian bilinguals and English monolinguals (using stimuli with matching grammatical gender in both L1s)

First, we examined the histograms of Likert scores for the two groups of participants using pairs that contain stimuli that had matching grammatical gender. In Experiment 1, the histograms for the two groups (English and Ukrainian-Russian) show distinct patterns (Figure A.1). For instance, the English monolingual group displays a fairly uniform distribution across the Likert scale with no clear peaks. The responses are spread quite evenly, although there is a slight increase in frequency around the middle scores (4-6). On the other hand, the Ukrainian-Russian bilingual group shows a different pattern, with a noticeable concentration of responses at the lower end of the Likert scale (1-3), that was also reflected in Figure A.1 in the manuscript. The distribution is positively skewed (Table A.8), indicating that participants from this group tended to select lower scores more frequently. In Experiment 2, the histograms illustrate a shift in the response patterns for both groups. The English group exhibits a more left-skewed distribution compared to Experiment 1, with a higher frequency of responses at the lower end (1-4). This change is supported by an increase in skewness (from 0.056 in Experiment 1 to 0.391 in Experiment 2) and a slight increase in variance (5.737) and standard deviation (2.395), as shown in Table A.8.

Conversely, the Ukrainian-Russian group shows a more balanced distribution with a peak around the middle of the Likert scale (scores 4-6). The responses appear less skewed than in Experiment 1 (from 0.847 in Experiment 1 to 0.177 in Experiment 2), indicating a broader spread of scores and a more symmetric pattern, potentially due to the emphasis in the modified instructions. Additionally, the group's variance (4.941) and standard deviation (2.223) were lower than in Experiment 1, suggesting a more consistent use of the scale. The interquartile range (IQR) also narrowed from 4 to 3, reflecting a more concentrated central tendency.

Figure A.1

Distribution of Likert scores in Experiments 1 and 2, by participant group and using stimuli with matching grammatical gender in 2L1s

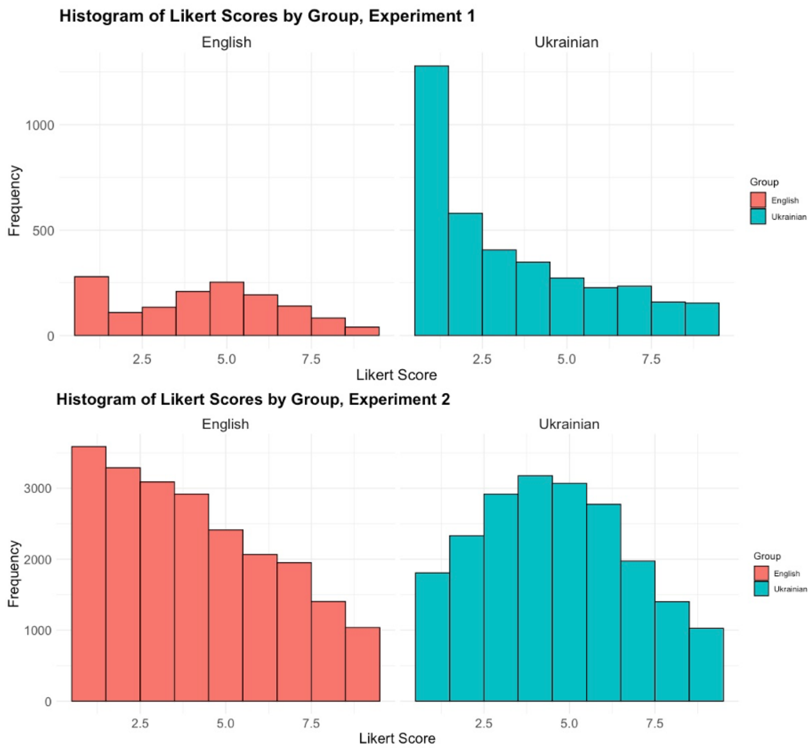


Table A.8

Descriptive statistics for Likert scores across Experiments 1 and 2 by participant group

Variable	Scores for Experiment 1		Scores for Experiment 2	
	English	Ukrainian-Russian	English	Ukrainian-Russian
Range	1-9	1-9	1-9	1-9
Interquartile Range (IQR)	4	4	4	3
Variance	5.269	6.044	5.737	4.941
Standard Deviation	2.295	2.459	2.395	2.223
Skewness	0.056	0.847	0.391	0.177
Kurtosis	2.023	2.516	2.060	2.157

A.3.2. Results of Ukrainian-Russian bilinguals (using stimuli with mismatching grammatical gender in both L1s)

The distribution of Likert scores for Ukrainian-Russian bilinguals responding to stimuli with mismatching grammatical gender also displayed notable shifts between Experiment 1 and Experiment 2 (Figure A.2).

In Experiment 1, the histogram shows a pronounced skew toward the lower end of the Likert scale, with the majority of responses concentrated between scores 1 and 3. The descriptive statistics (Table A.9) further support this observation, with a positive skewness of 0.846, reflecting the asymmetry of the distribution. The variance (6.347) and standard deviation (2.519) highlight substantial variability in the scores, though the distribution is less spread out than in Experiment 2. The kurtosis value of 2.468 suggests a heavier tail compared to a normal distribution, indicating some extremity in responses.

Conversely, in Experiment 2, the histogram illustrates a more balanced distribution, with a peak around the middle of the scale (scores 4–6). This indicates a broader use of the Likert scale, likely influenced by the emphasis in the modified instructions to use the full range of scores. The descriptive statistics (Table A.9) show a reduction in skewness to 0.072, reflecting a more symmetric response pattern compared to Experiment 1. With the decrease of variance and standard deviation (to 5.338 and to 2.311 respectively), less variability was observed in participant responses. The kurtosis also decreased to 2.018, indicating a less peaked and more evenly distributed set of responses.

Overall, these results suggest that the modification in instructions had a significant effect on response behaviour, encouraging participants to utilise the entire Likert scale more evenly.

Figure A.2

Distribution of Likert scores of Ukrainian-Russian bilinguals in Experiments 1 and 2, using stimuli with mismatching grammatical gender in 2L1s

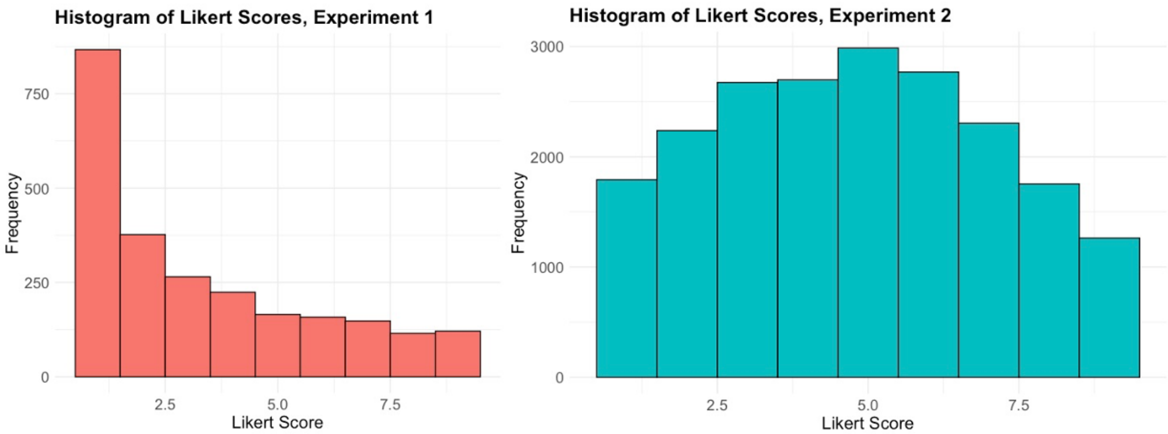


Table A.9.

Descriptive statistics for Likert scores across Experiments 1 and 2 for Ukrainian-Russian bilingual group

Variable	**Experiment 1**	**Experiment 2**
Range	1-9	1-9
Interquartile Range (IQR)	4	4
Variance	6.347	5.338
Standard Deviation	2.519	2.311
Skewness	0.846	0.072
Kurtosis	2.468	2.018

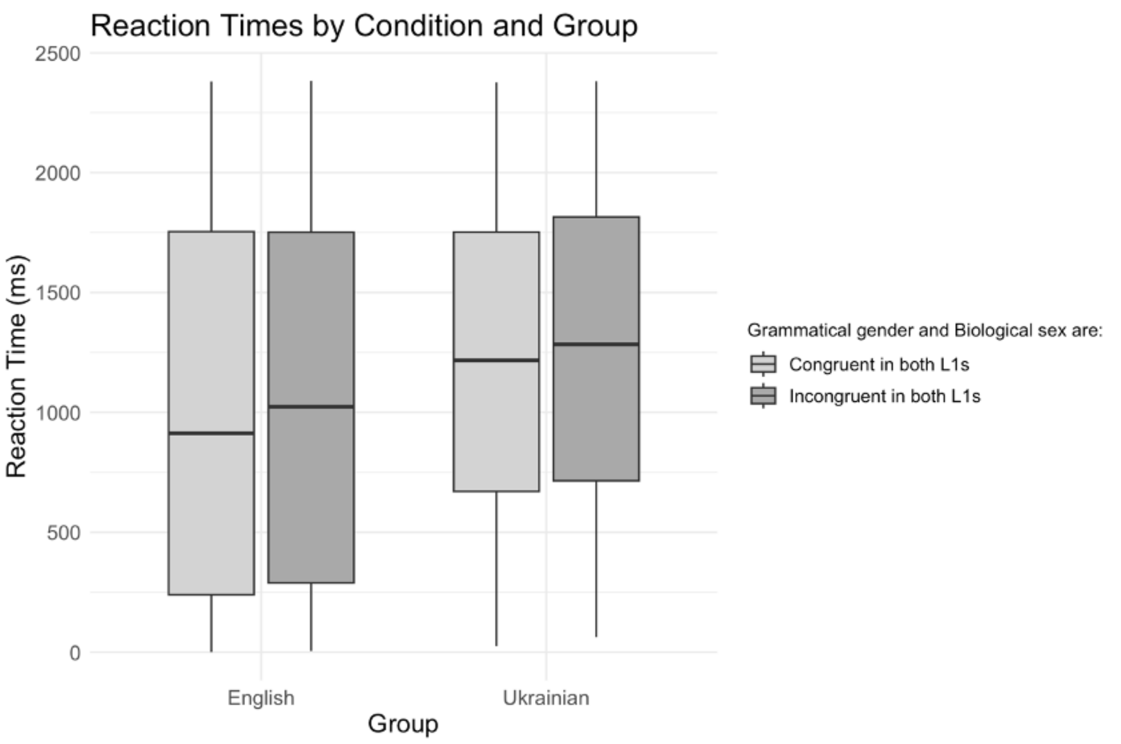
Appendix B. Supplementary Materials for Chapter 4

In addition to analysing accuracy, we also explored reaction times to investigate whether there was any significant difference between Ukrainian-Russian bilinguals and English monolinguals, as well as within the Ukrainian-Russian bilingual group based on language proficiency. For this, we constructed linear mixed-effects models with reaction time as the dependent variable. Specifically, for the first part of the analysis, we modelled reaction times as influenced by the interaction between the condition (Congruent in both L1s vs Incongruent in both L1s) and participant group (bilingual vs. monolingual). Random intercepts and random slopes were included for the interaction between condition and group for both participant and item. In the second part of the reaction time analysis, we also examined whether reaction times within the Ukrainian-Russian bilingual group were influenced by the L1 proficiency. We modelled reaction times as a function of the interaction between language proficiency and condition (Congruent in Russian & Incongruent in Ukrainian vs. Congruent in Ukrainian & Incongruent in Russian), including random intercepts and random slopes for the interaction for both participant and item.

However, when analysing reaction times for the responses, the difference between Ukrainian-Russian bilinguals and English monolinguals was nonsignificant (Figure B.1). Specifically, the reaction times (RT) for bilingual participants were somewhat slower but not significantly different from those of monolinguals (*Estimate* = -26.92, *SE* = 176.49, t =-0.15, p = .879). This indicates that while accuracy was influenced by the congruency between grammatical gender and biological sex, this congruency did not lead to faster response times. Main effects of Condition and Group were also non-significant (Condition: *Estimate* = -41.62, *SE* = 193.12, t =-0.22, p = .832; Group: *Estimate* = 136.04, *SE* = 275.76, t =0.49, p = .623). Additionally, when looking at bilingual group alone, there was no significant difference in reaction times based on the condition (*Estimate* = -49.53, *SE* = 183.11, t = -0.27, p = .790).

Figure B.1

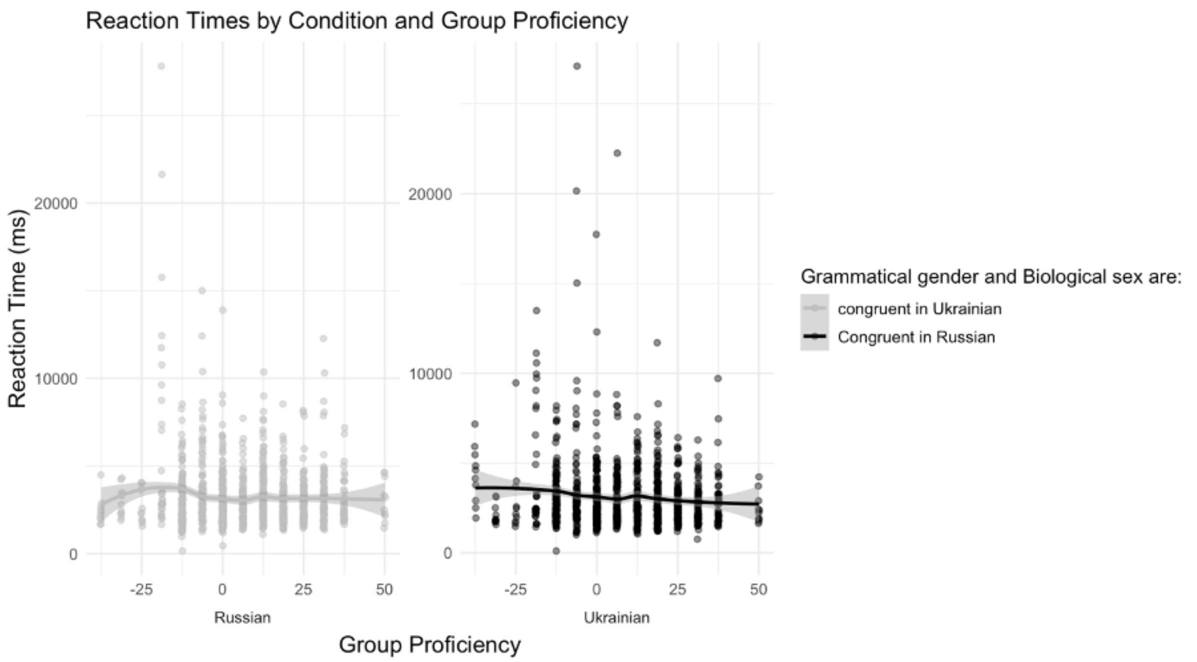
Reaction Times by Condition (Congruent in both L1s vs Incongruent in both L1s) and Participant Group (English Monolinguals vs. Ukrainian-Russian Bilinguals)



Similarly, as can be seen in Figure B.2, the results showed no significant difference in RTs based on the interaction between group proficiency and the condition (*Estimate* = -3.73. *SE* = 4.64, t = -0.80, p = .422). Main effects of condition and group proficiency were also non-significant (Condition: *Estimate* = -119.00, *SE* = 171.04*, t* =-0.70, *p* = .495; Group Proficiency: *Estimate* = -11.15, *SE* = 7.15, *t* =-1.56*, p* = .122). These findings suggest that language proficiency did not significantly impact the speed of responses, consistent with the lack of effect on recall accuracy.

Figure B.2

Reaction Times by Condition (Congruent in Ukrainian & Incongruent in Russian vs Congruent in Russian & Incongruent in Ukrainian) and Language Proficiency (Ukrainian-Russian Bilinguals)



Appendix C. Supplementary Materials for Chapter 5

C.1. Examining the role of cognates on the findings

To explore potential influences on our findings, we examined the role of cognates in our stimuli. Given that some of our nouns in Ukrainian and Russian are cognates, particularly nouns that are matching in grammatical gender in both languages, due to the language proximity of the two languages, we also analysed whether there was any significant difference in the processing of primes. As shown in Table C.1, the highest number of cognates appeared in primes with matching grammatical gender (37.5% had the same spelling, and 45% had similar spelling, differing by only one letter). In contrast, primes with mismatching grammatical gender contained far fewer cognates: 2.5% with the same spelling and 5% with similar spelling.

For this supplementary analysis, we focused on the N1, P2/VPP, and N400 components, rather than the N300 component analysed in the main study. This allowed us to examine ERP modulations immediately following the onset of primes, in the absence of explicit targets. If the N400 amplitude were modulated by the type of prime (Matching vs. Mismatching), this would suggest that processing effects are driven by the prime itself, rather than by (in)congruency of the prime-target pairs.

Table C.1

Distribution of primes based on spelling and grammatical gender matching

Types of primes (based on spelling)	Matching grammatical gender		Mismatching grammatical gender	
	N	%	N	%
Cognates (same spelling)	15	37.50	1	2.50
Cognates (similar spelling)	18	45.00	2	5.00
Non-cognates	7	17.50	37	92.50
Total number of primes	40	100.00	40	100.00

For this analysis, we examined ERPs focusing on the prime-onset time window instead of after the onset of subsequent triggers. The full analysis code and materials are available on OSF (https://osf.io/2vr7k/).

We restricted our analysis to ERP data, as our aim was to determine whether participants processed primes with matching and mismatching grammatical gender differently in their two first languages (2L1s). To do so, we applied linear mixed-effects models similar to those used in the main analysis. The maximal convergence model included type of prime (Matching vs. Mismatching grammatical gender) as a fixed effect, with the dependent variable reflecting the amplitudes of the N1, P2/VPP, and N400 components. Random intercepts and slopes for the type of prime were included, grouped by participant.

N1 time window (85-145 ms). In both the anterior and central-parietal regions, no significant interaction effects were observed for the type of prime. For the anterior N1, there was no significant effects of prime (SE = 0.154, df = 19.000, t = 0.786, p = .442), and similarly, no significant effect was found in the central-parietal region (SE = 0.163, df = 19.000, t = 1.083, p = .292).

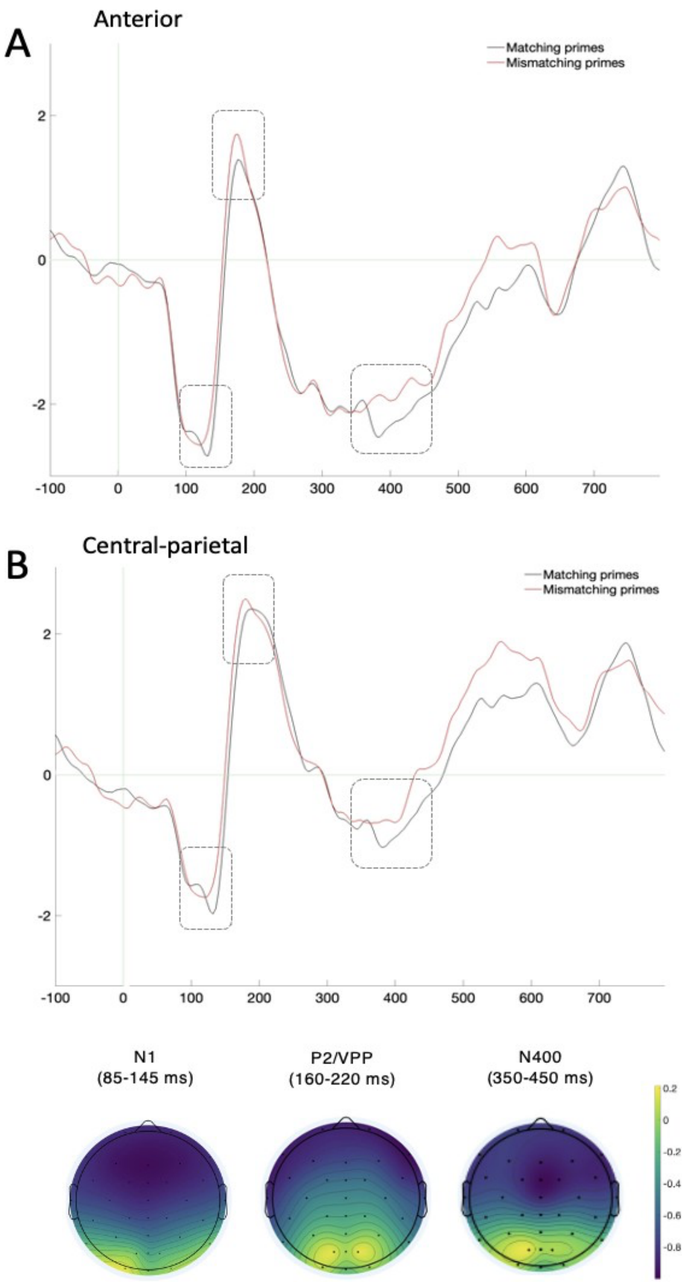
P2/VPP time window (160-220 ms). Similar to the N1 results, no significant interaction effects were found for the type of prime in either region. In the anterior region, results were non-significant (SE = 0.187, df = 19.000, t = 0.929, p = .365), as was the case for the central-parietal region (SE = 0.181, df = 19.000, t = 0.153, p = .880).

N400 time window (350-450 ms). Once again, no significant effects were observed. In the anterior region, the type of prime did not modulate the N400 amplitude (SE = 0.253, df = 19.002, t = 1.295, p = .211), and the same pattern was observed for the central-parietal region (SE = 0.196, df = 19.000, t = 1.580, p = .131).

Overall, our supplementary analysis revealed no significant differences in the processing of primes whether they matched or mismatched in grammatical gender across Ukrainian and Russian (see Figure C.1). The lack of significant effects across the N1, P2/VPP, and N400 components suggests that the type of prime alone did not influence early perceptual or semantic processing in this task.

Figure C.1

Grand average ERP waveforms of the N1, P2/VPP, and N400 components (circled with a grey line) showing the processing of primes by Ukrainian-Russian bilinguals



Note. A represents processing of primes with matching and mismatching grammatical genders in Ukrainian and Russian, indicated by black and red lines, respectively, in the anterior electrodes. B represents processing of primes with matching and mismatching grammatical genders in Ukrainian and Russian, indicated by black and red lines, respectively, in the central-parietal electrodes. The analysis includes two regions of interest: the anterior region (A), encompassing electrodes Fz, F3, F4, FC1, FCz, and FC2; and the central-parietal region (B), encompassing electrodes C3, Cz, C4, CP1, CP2, and CPz. Average ERPs were calculated from pooled electrodes, with time zero marking the onset of the target face image. Scalp topographies represent the effects of the gender congruency in primes as differences obtained between primes with matching and mismatching grammatical genders.

C.2. Analysis of the variable "Accuracy" as reported in the pre-registration

In our pre-registration (link: https://osf.io/2vr7k/), we specified plans to analyse both reaction times and accuracy of participants' responses. In the Research Question 1 (Conditions CR-CU and IR-IU), accuracy was operationalised as a binary variable, coded as 1 for "accurate" responses (e.g., pressing "yes" for congruent pairs or "no" for incongruent pairs), and 0 for "inaccurate" responses (e.g., pressing "no" for congruent pairs or "yes" for incongruent pairs). However, during the analysis phase, we identified a response bias whereby participants were more likely to press "yes" across all conditions. To mitigate the potential impact of this bias on the interpretation of our findings, we adjusted the analytical approach and instead examined the proportions of "yes" and "no" responses across conditions. The updated variable ("Response Types") was dummy coded, with 1 indicating a "yes" response and 0 indicating a "no" response (see "Data analysis" in the current paper).

The same adjustments were applied to the Research Question 2 (CR-IU and IR-CU conditions). In the pre-registration, we aimed to examine whether accuracy and reaction times were influenced by the interaction of participants' more proficient/dominant L1 (Ukrainian or Russian) and condition (CR-IU and IR-CU), using stimuli with mismatching grammatical gender in 2L1s. Accuracy was also dummy coded as a binary variable with possible values of 1 or 0. A value of 1 was assigned in two cases: (1) when participants pressed "yes" for pairs congruent with the Ukrainian grammatical gender, or (2) when they pressed "no" for pairs congruent with the Russian grammatical gender. A value of 0 was assigned when participants pressed "no" for pairs congruent with Ukrainian gender or "yes" for pairs congruent with Russian gender. However, the observed "yes" response bias prompted a shift towards analysing response proportions in the final analysis. Here, in the Supplementary Materials, we present the originally pre-registered accuracy analyses for completeness and transparency (all codes can be accessed on OSF: https://osf.io/2vr7k).

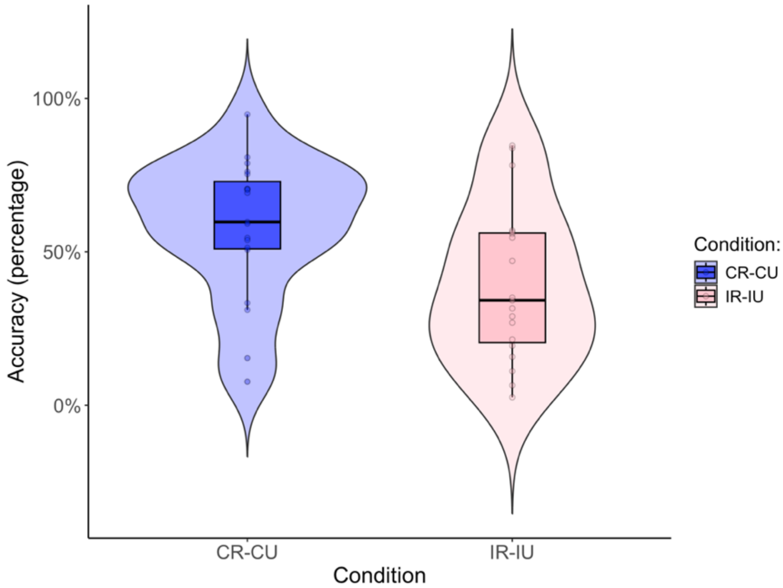
C.2.1. Response Accuracy for Research Question 1 (CR-CU and IR-IU)

Prior to finding the response bias, accuracy results suggested a potential sensitivity to grammatical gender congruency. The linear mixed effects model revealed a marginally significant effect of condition on accuracy (*Estimate* = 1.01, *SE* = 0.562, z = 1.803, p =.071, 95% CI[-0.088 2.11]), indicating a trend toward higher accuracy in the CR-CU condition compared to the IR-IU condition.

As shown in Figure C.2, participants exhibited higher accuracy in the CR-CU condition (compared to IR-IU) when judging associations between nouns and faces based on the congruency of the noun's grammatical gender and the face's biological sex. This indicated that participants were more likely to perceive an association when grammatical gender and biological sex were congruent in both L1s, compared to the incongruent pairs.

Figure C.2

Accuracy of the responses per condition (CR-CU vs IR-IU)

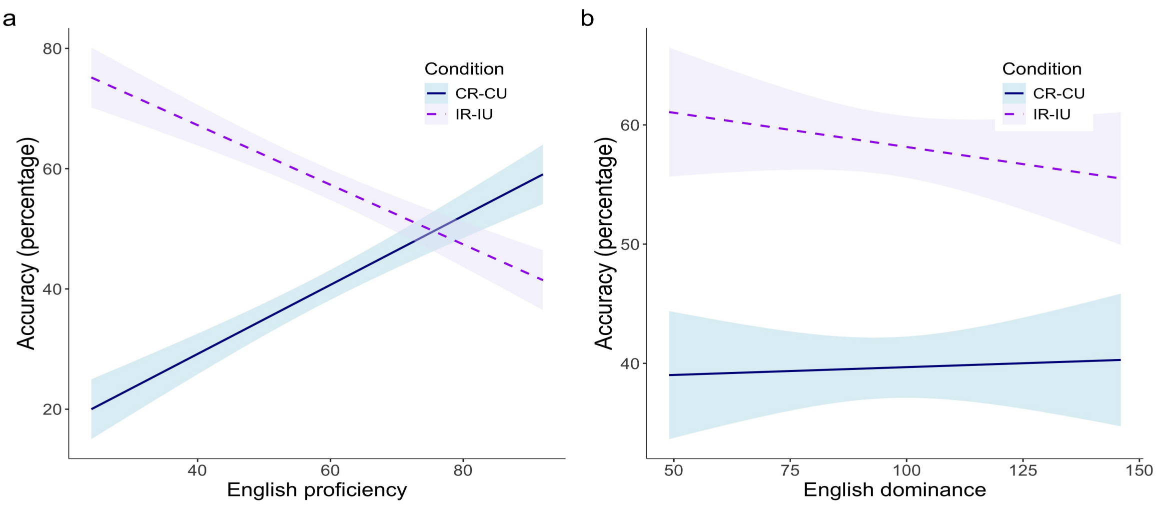


English proficiency and dominance. When examining effects of English (L2) proficiency and dominance as a function of Accuracy, we found a significant interaction between Condition and English proficiency (*Estimate* = 1.282, *SE* = 0.521, *z* = 2.460, *p* = .014). For the CR-CU condition further comparisons showed that as English proficiency increased, participants reported having less associations between primes and targets (*Estimate* = -0.563, *SE* = 0.238, *z* = -2.365, *p* = .018). Analogously, for condition IR-IU the rend was reversed, with participants found significantly more associations between prime and target in this condition as their English proficiency increased (*Estimate* = 0.647, *SE* = 0.277, *z* = 2.339, *p* = .019). No significant interaction was found for

English dominance and Condition (*Estimate* = -0.031, *SE* = 0.473, *z* = -0.065, *p* = .949) (see Figure C.3).

Figure C.3

Effects of (a) English Proficiency and (b) English Dominance on Accuracy Across Conditions (CR-CU vs IR-IU)



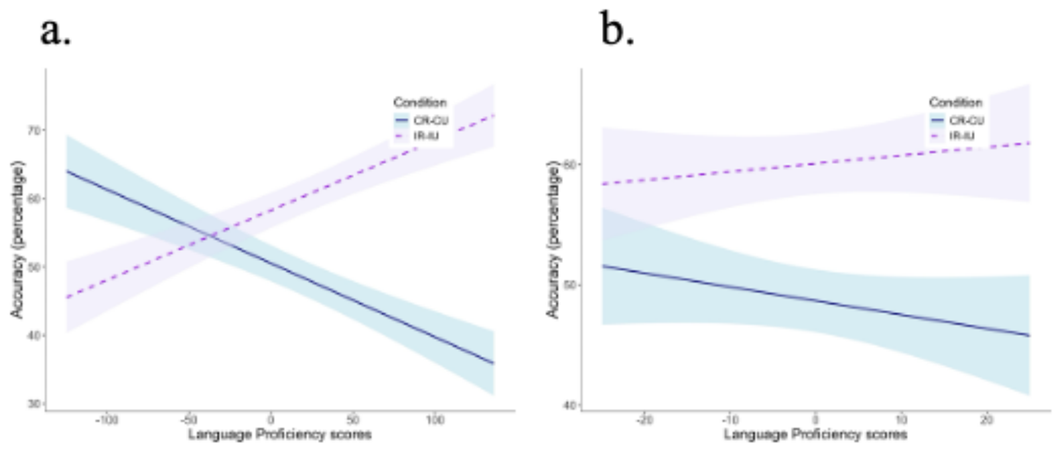
C.2.2. Response Accuracy for Research Question 2 (CR-IU and IR-CU Conditions)

In contrast to reaction times, linear mixed effects model identified a significant interaction between L1 Dominance coefficient and Condition on response accuracy (*Estimate* = 0.872, *SE* = 0.402, *z* = 2.170, *p* = .030). Further analysis showed that higher dominance in Ukrainian was associated with reporting more associations for pairs in the IR-CU condition (*Estimate* = 0.416, *SE* = 0.202, *z* = 2.063, *p* = .039), and fewer associations for pairs in the CR-IU condition (*Estimate* = -0.477, *SE* = 0.222, *z* = -2.149, *p* = .032). However, no significant interaction was found between L1 Proficiency coefficient and Condition (*Estimate* = 0.213, *SE* = 0.448, *z* =

0.476*, p* = .634). Figure C.4 further illustrates these effects. Participants with higher Ukrainian dominance were more accurate when the biological sex of the face matched the noun's grammatical gender in Ukrainian but not in Russian (IR- CU condition). Conversely, participants with higher Russian dominance showed greater accuracy when the grammatical gender corresponded to the biological sex in Russian but not in Ukrainian (CR-IU condition). However, Ukrainian-Russian bilinguals with higher Ukrainian proficiency scores did not exhibit increased accuracy in the IR-CU condition.

Figure C.4

Interaction of (a) L1 Dominance and (b) L1 Proficiency Coefficients with Accuracy by Condition (CR-IU vs IR-CU)



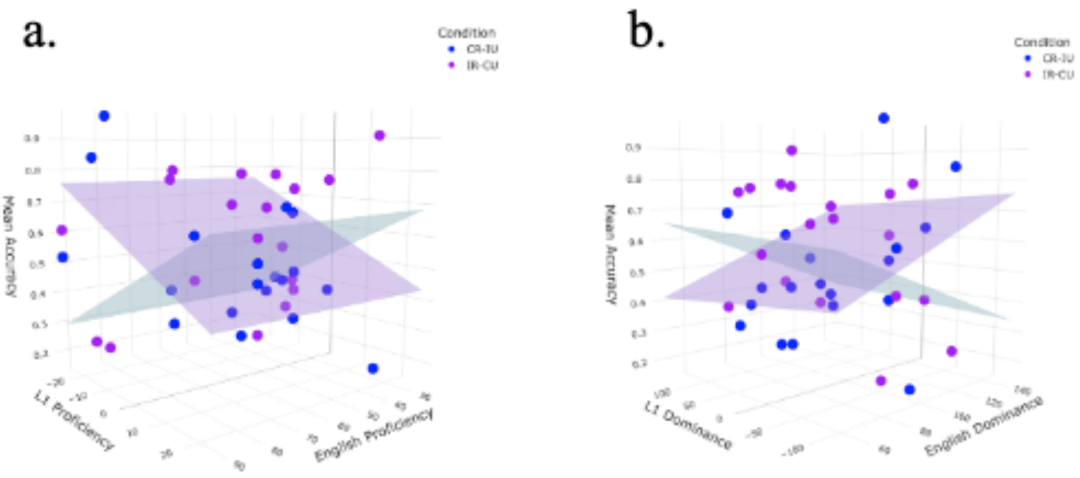
English proficiency and dominance. When analysing the effects of English language proficiency on these types of stimuli, we did not find a significant interaction between English proficiency, Condition, and L1 Proficiency coefficient (CR-IU as a baseline: *Estimate* = -0.049, *SE* = 0.078*, z* = -0.623, *p* = .533; IR-CU as a baseline: *Estimate* = 0.049, *SE* =

0.078, $z = 0.623$, $p = .533$). This indicates that English proficiency did not significantly modulate grammatical gender effects.

In contrast, for English dominance, the analogous interaction with English dominance, Condition and L1 Dominance was significant (CR-IU as a baseline: *Estimate* = 0.808, *SE* = 0.108, $z = 7.481$, $p < .001$; IR-CU as a baseline: *Estimate* = -0.808, *SE* = 0.108, $z = -7.481$, $p < .001$). This suggests that higher English dominance was associated with weaker grammatical gender effects in the CR-IU condition, while the pattern was reversed in the IR-CU condition (see Figure C.5).

Figure C.5

Interaction of (a) English Proficiency - L1 Proficiency- Accuracy and (b) English Dominance -L1 Dominance – Accuracy by Condition (CR-IU vs IR-CU)



While the accuracy analyses yielded some statistically significant effects - including a marginal effect of condition in RQ1 and interactions between condition and both English proficiency (RQ1) and L1 dominance (RQ2) - these findings should be interpreted with caution. When considered alongside the analyses of response types (“yes” vs “no”) and ERP data, both of which showed null results, the observed effects in accuracy are likely driven by a general bias towards “yes” responses across conditions. As such, the combined evidence suggests that the accuracy-based effects do not reflect genuine sensitivity to grammatical gender congruency, but rather an overarching response tendency.