

MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Role of Language Aptitude in First Language Attrition: A
Study on Greek-German Sequential Bilinguals.

verfasst von | submitted by

Eleni Zotou

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 867

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Allgemeine Linguistik:
Grammatiktheorie und kognitive
Sprachwissenschaft

Betreut von | Supervisor:

Univ.-Prof. Mag. Dr. Susanne Maria Reiterer

*To my sister,
whose passion for reading and learning lit the path long before I followed it.
You are my first role model.*

Acknowledgements

First, I would like to express my deepest gratitude to my supervisor and mentor, Univ.-Prof. Mag. Dr. Susanne Maria Reiterer, for her continuous guidance throughout the entire process. Her expertise, critical feedback, and inspiration were instrumental in shaping this thesis and in motivating my interest in language aptitude as a research topic.

I am also grateful to Mikel Santesteban for providing the methodological foundations that enabled me to design and analyze an empirical study. The practical training and insights gained in his course proved invaluable for this research.

Special thanks go to Lukas Nemesthothy for his extensive assistance, particularly with statistical analysis and data interpretation, as well as to Franz R. Schmid for his support during earlier stages of the project.

I am sincerely thankful to all participants and informants who generously contributed their time and effort, without whom this study would not have been possible.

Finally, I would like to thank my friends, family, and partner for their unwavering support and encouragement throughout my studies and during the completion of this thesis.

Abstract

First language attrition challenges the assumption that native language competence remains stable in adulthood, yet the contribution of individual cognitive differences remains unclear. In particular, language aptitude, a robust predictor of second language acquisition, has received limited and inconclusive attention in attrition research. This study examined whether language aptitude predicts L1 lexical maintenance in late bilingual adults. Using a quantitative, correlational design, data were collected from 34 late Greek–German bilinguals residing in German-speaking countries. Language aptitude was assessed with a vocabulary learning task from the LLAMA test battery, while L1 lexical maintenance was measured via a picture naming task that captured lexical retrieval accuracy and reaction times. The results showed no significant relationship between language aptitude and L1 lexical maintenance. Exploratory analyses indicated that experiential factors, particularly L2 frequency of use, were more strongly associated with L1 proficiency outcomes than aptitude. These findings support accounts of L1 attrition that emphasize processing demands and language use over cognitive traits and highlight the difficulty of isolating aptitude effects in attrition research. The study contributes to a nuanced account of individual variability in first language maintenance in bilinguals and motivates future work using larger samples, alternative operationalizations, and neurocognitive measures.

Kurze Zusammenfassung

Die Erstspracherosion (First Language Attrition) stellt die Annahme infrage, dass die muttersprachliche Kompetenz im Erwachsenenalter stabil bleibt, während der Beitrag individueller kognitiver Unterschiede bislang unzureichend geklärt ist. Insbesondere die Sprachbegabung (Language Aptitude), ein etablierter Prädiktor des Zweitspracherwerbs, wurde in der Forschung zur Erstspracherosion nur begrenzt untersucht und lieferte uneinheitliche Ergebnisse. Ziel dieser Studie war es, zu prüfen, ob Sprachbegabung den Grad des lexikalischen L1-Spracherhalts bei Spätbilingualen vorhersagt. In einem quantitativen, korrelativen Design wurden Daten von 34 spätbilingualen griechisch–deutsch Sprecher:innen im deutschsprachigen Raum erhoben. Die Sprachbegabung wurde mithilfe einer Vokabellernaufgabe aus der

LLAMA-Testbatterie erfasst, während der lexikalische L1-Spracherhalt durch eine Bildbenennungsaufgabe gemessen wurde, die sowohl die Korrektheit als auch die Reaktionszeiten erfasste. Die Ergebnisse zeigten keinen signifikanten Zusammenhang zwischen Sprachbegabung und Spracherhalt. Explorative Analysen deuteten hingegen darauf hin, dass erfahrungsbasierte Faktoren, insbesondere die Häufigkeit der L2-Nutzung, stärker mit dem L1-Spracherhalt assoziiert sind als die Sprachbegabung. Die Ergebnisse stützen Ansätze, die Erstspracherosion primär als verarbeitungs- und nutzungsbasiertes Phänomen begreifen, und unterstreichen die Schwierigkeit, Begabungseffekte in der Attritionsforschung eindeutig zu isolieren. Diese Studie trägt zu einem differenzierten Verständnis der individuellen Variabilität beim L1-Spracherhalt von bilingualen Sprecher:innen bei und weist auf den Bedarf weiterer Forschung mit größeren Stichproben, alternativen Operationalisierungen sowie neurokognitiven Messverfahren hin.

Table of Contents

Acknowledgements	2
Abstract	3
List of Figures	7
List of Tables	8
List of Abbreviations	9
1. Introduction.....	10
2. Theoretical Background.....	13
2.1. Bilingualism and Language Processing	13
2.1.1. Defining Bilingualism	13
2.1.2. Cross-Linguistic Influence	14
2.2. First Language Attrition	16
2.2.1. Defining Attrition	16
2.2.2. Linguistic Selectivity.....	17
2.2.3. Predictive Factors	18
2.2.3.1. Brain Mechanisms.....	19
2.2.3.2. Cognitive Processes.....	20
2.2.3.3. External Factors.....	21
2.3. Language Aptitude	21
2.3.1. Foundational Models.....	21
2.3.2. Rethinking Aptitude	23
2.3.3. Neural Correlates.....	24
2.4. Interfaces of Attrition and Aptitude.....	25
2.4.1. Language Aptitude and First Language Competence	26
2.4.2. The Role of Language Aptitude on Language Attrition	28
2.5. Present Study.....	30
3. Methodology	32
3.1. Experimental Design	32
3.2. Participants	34
3.2.1. Sample Selection Criteria.....	34
3.2.2. Participant Acquisition and Exclusion	36
3.3. Materials.....	37
3.3.1. Procedure.....	37
3.3.2. Task Overview.....	39

3.3.2.1.	Attrition Task.....	39
3.3.2.2.	Aptitude Task	49
3.3.2.3.	Proficiency Tasks	51
3.3.2.4.	Questionnaire	52
3.4.	Data Preparation	53
3.4.1.	Aptitude, Proficiency, and Questionnaire Data	53
3.4.2.	Attrition Data.....	54
4.	Results.....	56
4.1.	Descriptive Statistics	56
4.1.1.	Demographic Data.....	56
4.1.2.	Variable Overview	58
4.2.	Inferential Statistics.....	64
4.2.1.	Confirmatory Analysis	64
4.2.2.	Exploratory Analysis	67
5.	Discussion	72
5.1.	Interpretation of Findings.....	72
5.1.1.	The Role of Aptitude in L1 Lexical Maintenance.....	72
5.1.2.	The Influence of Other Factors on L1 Lexical Maintenance	73
5.1.3.	Theoretical and Methodological Implications	74
5.2.	Limitations	76
5.3.	Future Directions.....	78
5.4.	Conclusion.....	79
References		81
Appendices		90
Appendix A:	Recruitment and Information Materials	90
Appendix B:	Experimental Tasks and Materials	92
Appendix C:	Supplementary Data and Results	108

List of Figures

Figure 1. Example of a test trial from the picture naming task.	43
Figure 2. Microphone check for the picture naming task.....	44
Figure 3. Familiarization feedback for the picture naming task.....	45
Figure 4. Example of an attention question.....	46
Figure 5. Response to (a) wrong and (b) correct answers to attention questions.....	46
Figure 6. Familiarization phase of the aptitude task.	50
Figure 7. Example of a test trial from the aptitude task.	50
Figure 8. Example of a test trial from the L2 proficiency task.	51
Figure 9. Age of onset of L2 learning.	57
Figure 10. Languages spoken by the participants.	57
Figure 11. Distribution of participants by profession or field of study.	58
Figure 12. Boxplots of (a) accuracy (%) and (b) reaction times (ms).....	59
Figure 13. Boxplot of maintenance index.	59
Figure 14. Boxplot of aptitude scores.	60
Figure 15. Boxplots of (a) L1 and (b) L2 accuracy scores.....	61
Figure 16. Boxplots of (a) L1 and (b) L2 frequency of use.	61
Figure 17. Boxplots of (a) L1 and (b) L2 orientation.....	62
Figure 18. Boxplots of (a) L1 and (b) L2 judgement.	62
Figure 19. Boxplots of (a) retrieval and (b) affinity.	63
Figure 20. Boxplots of (a) migration and (b) repatriation.....	64
Figure 21. Scatterplot of (a) aptitude score and (b) maintenance index (%).	65
Figure 22. Scatterplot of (a) accuracy and (b) reaction time (z-scores).	66
Figure 23. Scatterplot of (a) aptitude score (%) and (b) reaction times (z-score).	66
Figure 24. Scatterplot of (a) aptitude score and (b) accuracy (%).	67
Figure 25. Heatmap of significant correlations.	68

List of Tables

Table 1. Questionnaire composite metrics.	54
Table 2. Descriptive statistics for accuracy (%) and reaction times (ms) (N = 34).....	59
Table 3. Descriptive statistics for maintenance index (z-score) (N = 34).	59
Table 4. Descriptive statistics for aptitude (N = 34).....	60
Table 5. Descriptive statistics for L1 and L2 accuracy (N = 34).....	60
Table 6. Descriptive statistics for L1 and L2 frequency of use (N = 34).	61
Table 7. Descriptive statistics for L1 and L2 orientation (N = 34).	62
Table 8. Descriptive statistics for L1 and L2 judgement (N = 34).	62
Table 9. Descriptive statistics for retrieval and affinity (N = 34).....	63
Table 10. Descriptive statistics for migration and repatriation (N = 34).....	63
Table 11. Linear regression predicting L1 maintenance index from aptitude score (%).	64
Table 12. Correlation between accuracy and reaction times (z-scores).	65
Table 13. Linear regression predicting reaction times (z-score) from aptitude score (%).	66
Table 14. Linear regression predicting accuracy from aptitude score (%).	67
Table 15. Spearman's correlations between age and maintenance index (%).	69
Table 16. Comparison of maintenance index across AoA and LoR groups.	69
Table 17. Kruskal–Wallis test for education and maintenance index (%).	70
Table 18. Mann–Whitney U test for aptitude scores (high vs. low maintenance).	70
Table 19. Mann–Whitney U test for maintenance scores (high vs. low aptitude).	71

List of Abbreviations

Acc	Accuracy
AoA	Age of Acquisition
ATI	Aptitude–Treatment Interaction
CEFR	Common European Framework of Reference for Languages
CLI	Cross-Linguistic Influence
CPH	Critical Period Hypothesis
IFG	Inferior Frontal Gyrus
L1	First Language
L2	Second Language
L3	Third Language
LA	Language Aptitude
LoR	Length of Residence
MLAT	Modern Language Aptitude Test
PNT	Picture Naming Task
PLAB	Pimsleur Language Aptitude Battery
RT	Reaction Time
SCCC	Social Cognitive Causal Chains
SD	Standard Deviation
SLA	Second Language Acquisition
WM	Working Memory

1. Introduction

“To lose your own language
was like forgetting your mother,
and as sad, in a way.”

— Alexander McCall Smith (The Full Cupboard of Life, 2003)

The quotation encapsulates the deep emotional dimension of language loss, comparing it to the severance of a maternal bond. The notion of “losing” a language encompasses a wide spectrum of linguistic phenomena. In a social context, language extinction occurs when the last speakers of a language pass away, a process with profound cultural and anthropological implications (Dorian, 1981). On an individual level, heritage language loss is experienced by heritage speakers, e.g., second-generation immigrants, who grow up with low proficiency in their ancestral language due to limited exposure and sociocultural pressures (Montrul & Polinsky, 2019), while language reversion can result from natural aging (Higby et al., 2019). In clinical contexts, language can be lost pathologically, as in cases of aphasia or dementia (Opitz, 2011).

The present study focuses on a subtler, often misunderstood form of language loss: first language (L1) attrition. Attrition refers to the gradual decline in native language proficiency, typically observed in bilingual speakers immersed in a second-language environment (Köpke & Schmid, 2004). This decline is often considered or experienced as unnatural, surprising, and disappointing, sentiments partly fueled by the misleading use of the term “loss” to describe a process that is neither total nor irreversible. Yet it challenges assumptions about the permanence of the native language, particularly in adulthood. First language attrition may manifest as disfluent and hesitant speech, characterized by long pauses, filler words, repetitions and self-corrections, difficulty in retrieving words, odd expressions and collocations, occasionally even subtle foreign accents. While it may seem rare or exceptional, it is in fact a frequent but much less transparent phenomenon (Schmid, 2011).

Research on bilingualism has traditionally focused on the influence of the first language (L1) on second language (L2) acquisition (Wen et al., 2017). However, growing evidence shows that this interaction is bidirectional. In bilingual speakers, the form, meaning, and distribution of lexical items in both languages influence one another (Schmid & Köpke, 2017). This cross-

linguistic interaction can lead to deviations from native-like usage in the L1, sometimes deliberately, but often without the speaker's awareness. Such interference is central to language attrition and highlights the need for a more nuanced understanding of bilingual language development and maintenance.

Numerous studies have identified sociolinguistic factors contributing to L1 attrition, such as the age of L2 acquisition, length of residence in an L2 environment, and the speaker's social identity. In contrast, cognitive factors, and particularly language aptitude, have received far less attention in attrition research (Köpke & Schmid, 2004; Schmid, 2011; Schmid & Yilmaz, 2018). Traditionally associated with foreign language learning, language aptitude (LA) is defined as an individual's natural ability or talent for acquiring new languages. It encompasses abilities such as phonetic coding, grammatical sensitivity, and working memory (Carroll, 1981; Skehan, 1998) and has been extensively studied within the field of second language acquisition (SLA) (Dörnyei & Skehan, 2003).

Empirical evidence on the relationship between language aptitude and first language attrition exists but is limited and yields conflicting results. Different theoretical approaches suggest that higher aptitude may lead to greater L1 maintenance, whereas others show the opposite effect or none at all. Therefore, the role of language aptitude in L1 attrition remains underexplored, yet it offers a promising line of inquiry for understanding individual variability in L1 retention and decline (Köpke, 2007).

The present study addresses this research gap by investigating the role of language aptitude in first language attrition, focusing on lexical retrieval performance among late bilinguals. Late bilinguals are individuals who acquired their second language after their first language was fully established (DeKeyser, 2000), thus minimizing the risk of conflating attrition with incomplete acquisition. Lexical retrieval is selected as the focal linguistic domain as it is both particularly susceptible to attrition (Schmid & Jarvis, 2014) and highly sensitive to individual differences, making it a robust indicator of changes in L1 proficiency. Since language aptitude remains largely underexamined in this context, this study aims to determine whether individual aptitude profiles can explain variability in L1 retention.

By investigating the role of aptitude in language attrition, this study aims to advance psycholinguistic theories addressing individual variability in bilingual language processing and to promote practical applications by raising awareness of the tangible implications of first language attrition. For example, such awareness could support the development of targeted language support programs that facilitate effective L1 reactivation and linguistic reintegration for individuals returning to their native linguistic environment after prolonged immersion in a

second-language context, where attrition may impact communicative competence and social integration. Additionally, increased understanding of attrition may inform research on multilingual aphasia (Goral, 2022), where attrition-related patterns are increasingly recognized, suggesting potential benefits for therapeutic strategies and for adapting speech-language therapy.

Moreover, this study seeks to alleviate the social stigma and the feelings of guilt and shame often attached to L1 attrition, which are frequently misunderstood as personal failure or a sign of inauthentic identity (Lazaridou-Chatzigoga & Karatsareas, 2022). It also aims to foster a more nuanced understanding of bilingual language dynamics. Bilingualism is not a static state but a dynamic spectrum, with individuals shifting across contexts and over time. Just as people differ in how they acquire additional languages, they also vary in how they retain or adapt their L1, shaped by situational, social, and cognitive factors. Recognizing this fluidity allows for a more accurate and compassionate understanding of language attrition and maintenance.

The following chapter outlines the theoretical background relevant to L1 attrition and language aptitude, reviewing key empirical studies and conceptual frameworks. Chapter 3 presents the methodological approach used to investigate the research question, including participant characteristics, materials, and testing procedures. Chapter 4 reports the results of the quantitative analyses, and Chapter 5 concludes with a discussion of the findings, their implications, and suggestions for future research.

2. Theoretical Background

This chapter provides the conceptual and empirical foundation for examining first language (L1) maintenance in late bilinguals, with a particular focus on the role of language aptitude. It situates the present study within research on bilingual language processing, language attrition, and individual differences, and identifies the theoretical gaps that motivate investigating aptitude as a potential factor of variability in L1 lexical outcomes.

Section 2.1 outlines core principles of bilingualism relevant to L1 attrition, including types of bilingualism, dominance patterns, and cross-linguistic influence. Section 2.2 reviews theoretical accounts and empirical findings on first language attrition, emphasizing its selective nature and established predictive factors. Section 2.3 introduces language aptitude, tracing its development from classical models to contemporary cognitive and neurobiological perspectives. Section 2.4 integrates these strands by examining existing evidence on the relationship between aptitude, L1 retention, and attrition. Finally, Section 2.5 formulates the rationale, research questions, and hypotheses of the present study.

2.1. Bilingualism and Language Processing

2.1.1. Defining Bilingualism

Since first language attrition presupposes bilingualism, a clear conceptualization of bilingualism and the fundamental properties of bilingual language systems is necessary. Bilingualism is broadly defined as the knowledge or regular use of two or more languages. It is not a uniform linguistic experience among speakers, nor does it involve equal competence across languages within a speaker. Bilingualism encompasses substantial variation in how languages are acquired, used, and accessed, making it a complex phenomenon that requires careful theoretical differentiation.

A core distinction along which bilinguals are commonly differentiated is the developmental stage of second-language acquisition, commonly defined in terms of age of onset. A core distinction along which bilinguals are commonly differentiated concerns the developmental stage of second-language acquisition, typically operationalized as age of onset. This distinction is commonly framed with reference to the Critical Period Hypothesis, which

proposes that the brain's sensitivity to naturalistic, implicit language acquisition peaks before puberty, around ages 12–13 (Lenneberg, 1967). Early bilinguals include simultaneous bilinguals, who acquire two languages from birth, as well as successive (or early sequential) bilinguals, who acquire a second language (L2) in early childhood after first language (L1) development has begun (de Bot & Bülow, 2021). In contrast, late (sequential) bilinguals acquire an L2 only after L1 development has reached full attainment, typically post-puberty (Kupisch, 2019). Late bilinguals may also be distinguished by the context of L2 acquisition in instructed bilinguals (e.g., classroom settings) and immersed bilinguals (e.g., naturalistic exposure).

Contemporary accounts reject the view of bilinguals as “*two monolinguals in one person*” (Grosjean, 1989). Instead, bilingualism is conceptualized as a dynamic and integrated cognitive system, often referred to as multicompetence, in which multiple languages coexist within a unified supersystem rather than as separate modules (Cook, 1991). A defining property of this system is non-selective activation, meaning that languages share a representational network and can co-activate during language processing, even in monolingual contexts (Jarvis, 2009), leading to interaction and competition between linguistic representations.

In the bilingual system, a high and balanced level of proficiency across a speaker's languages is not a prerequisite (de Leeuw, 2014). Over a given period, bilingual speakers typically use primarily one language, rendering it more readily accessible. This language is referred to as the dominant language and is independent of overall proficiency, being shaped mainly by patterns of use and processing efficiency (Köpke & Genevskaja-Hanke, 2018). Language dominance can shift repeatedly across the lifespan in response to changes in language exposure and use, such as relocation to a different linguistic environment (Schmid & Köpke, 2017). In late bilinguals immersed in an L2 environment, reduced L1 input often leads to increased L2 dominance, even when the L1 remains structurally intact (Schmid & Yılmaz, 2018). Such dominance shifts reflect changes in the speed and automaticity of language access, highlighting dominance as a dynamic, usage-driven property of the bilingual system (Schmid & Köpke, 2017; Doussias et al., 2021).

2.1.2. Cross-Linguistic Influence

In bilingual speakers, the integrated organization of multiple languages entails interaction between linguistic systems. Pavlenko (2000) devised the Cross-Linguistic Influence (CLI) framework accordingly, proposing that L2 influence on L1 is a distinct phenomenon.

Under specific circumstances, interference may constitute evidence of attrition, which is a necessary but not sufficient condition. CLI differentiates among five stages through which two linguistic systems interact.

Borrowing involves integrating L2 morphophonological and lexical elements into the L1 system. In particular, lexical borrowing occurs when bilinguals use L2-derived forms for concepts that lack L1 equivalents or to highlight subtle conceptual distinctions between existing items. For example, Russian L2-English speakers use *intruzivnost* for ‘intrusiveness’, a concept not lexicalized in Russian. Lexical borrowing thus enriches the L1 system by expanding its lexicon. It constitutes evidence of attrition only when an existing L1 equivalent is no longer accessible to the speaker in production or comprehension, even in monolingual L1 contexts.

Restructuring involves reanalyzing L1 items based on the scope or structure of the corresponding L2 items. This may involve the deletion or addition of L2 elements in the L1, resulting in morphosyntactic changes (e.g., case, gender, or number marking; preposition choice; word order) or semantic changes (e.g., extension or narrowing). A commonly cited example is the semantic extension of the Spanish verb *correr* (‘to run’), used by Cuban immigrants in the USA in expressions such as ‘running for office’. Restructuring is indicative of attrition when bilinguals make grammaticality or semantic judgments that diverge from those of monolingual L1 speakers and experience corresponding production or comprehension difficulties.

Convergence describes the merging of elements from both languages into a new, unified system distinct from either L1 or L2. It is typically observed in phonology, for example, when L1 voice onset time converges toward L2 norms, resulting in a foreign accent, or in conceptual representation, such as German speakers blending English *eventually* with German *eventuell* (‘conceivably’, ‘perhaps’). Convergence serves as evidence of attrition only when its effects are systematic and occur across multiple contexts.

Shift involves moving away from L1 structures or values in order to approximate L2 norms over time, most notably in phonology and the mental lexicon. It may manifest in pragmatic domains, such as typicality (e.g., politeness, apology) or in emotional expression. For instance, while Russian encodes emotions primarily through verbs, English predominantly uses adjectives.

Under sustained conditions of reduced use and exposure, such cross-linguistic interaction may give rise to more persistent changes, commonly referred to as attrition, which represents the most severe manifestation of this process.

2.2. First Language Attrition

2.2.1. *Defining Attrition*

While bilingualism expands a speaker's linguistic repertoire, it also initiates a dynamic process of native language restructuring known as first language (L1) attrition. A rigorous definition requires a clear delineation of attrition from other language change phenomena. First, attrition involves change at the individual level, which differentiates attrition from societal linguistic changes (e.g., language shift or language death) (Dorian, 1981; Schmid, 2011). Second, attrition is explicitly non-pathological, distinguishing it from language deterioration due to neurological impairment (e.g., aging, aphasia, or dementia) (Paradis, 2007; Schmid, 2011). Third, attrition presupposes a fully acquired language system before the onset of change, separating it from incomplete (e.g., international adoptees) or imperfect acquisition (e.g., heritage speakers) (Köpke & Schmid, 2004). First language attrition is therefore commonly defined as the “non-pathological decrease in a language that had previously been acquired by an individual” (Köpke & Schmid, 2004).

These changes in a neurologically unimpaired individual's established L1 system typically arise after a relocation to an L2 environment and are driven primarily by two catalysts. On the one hand, the prolonged reduction in L1 use weakens access to linguistic representations (Schmid, 2011; Paradis, 1993). On the other hand, the constant competition from the dominant L2 system further inhibits L1 processing (MacWhinney, 2019). Modern psycholinguistic accounts, therefore, conceptualize attrition not as the loss of the underlying competence, but as a cognitive reorganization triggered by disuse and cross-linguistic influence, leading to retrieval difficulties (Köpke, 2004).

This cognitive reorganization is marked by reduced accessibility and automaticity of L1 knowledge, making retrieval slower and more effortful. These changes manifest as performance difficulties (e.g., lexical retrieval problems, reduced fluency, and grammatical uncertainty) (Schmid & Jarvis, 2014), and their severity varies considerably across individuals, shaped by the interaction of internal and external factors (Köpke, 2007; Schmid, 2011). These effects are reversible with renewed L1 exposure, indicating that the underlying competence remains intact (Chamorro et al., 2016; Köpke & Genevska-Hanke, 2018). Crucially, a decline in linguistic performance itself defines the phenomenon of language attrition (Köpke, 2004; Gallo et al., 2021).

2.2.2. *Linguistic Selectivity*

First language attrition is a fundamentally selective process. Linguistic domains deteriorate at different rates and under distinct mechanistic pressures (Köpke & Schmid, 2004). Empirical research shows that this selectivity follows systematic patterns across domains rather than a uniform decline, reflecting differences in representational stability and the cognitive demands of online processing in a competitive bilingual architecture (Schmid, 2011; Schmid & Köpke, 2017).

The lexicon is consistently reported as the most vulnerable domain to show decline, largely due to its open-class nature and reliance on declarative memory (MacWhinney, 2019). Lexical attrition typically manifests as noticeable difficulties in word production, including slower retrieval, increased disfluency markers (e.g., hesitations, repetitions, self-corrections, pauses, tip-of-the-tongue phenomena) (Bergmann et al., 2015; Higby et al., 2019), and reduced lexical diversity and sophistication (Yilmaz & Schmid, 2012; Schmid & Jarvis, 2014). Experimental evidence from picture-naming studies reveals longer reaction times and higher error rates among attriters, especially for low-frequency items (Baus et al., 2013). These patterns are typically explained through the Activation Threshold Hypothesis, which posits that infrequent use during L2 immersion raises the activation thresholds of L1 lexical items, rendering them temporarily inaccessible (Paradis, 1993). High-frequency words and cognates are more resistant to attrition due to stronger or shared activation pathways (Schmid, 2011).

Morphological attrition is highly selective, consistently targeting markedness. Attrition affects inflectional morphology, both the nominal (i.e., case, gender) and verbal (i.e., agreement, mood) systems, with nominal morphology being more susceptible due to factors such as greater allomorphic variability and lower functional load (Montrul & Yoon, 2019). Irregular forms tend to undergo regularization, such as producing an irregular past-tense verb with the default suffix (Schmitt, 2019). Marked morphological categories also show instability, leading to the omission or variability of required markers, such as the use of an unmarked default form in contexts that require specific case marking (Schmitt, 2019). Complex inflections can trigger paradigmatic reduction, leading to a reduced productive inventory of inflectional patterns and contrastive alternations. Structurally demanding paradigms may shift toward analytic or periphrastic constructions, for instance, replacing a single inflected verb form with a multi-word tense construction (Montrul & Yoon, 2019). By contrast, highly entrenched, regular, and derivational morphology remains comparatively stable, reflecting its strong proceduralization and lower processing demands in adult speakers.

Phonology is a highly resilient domain, with segmental contrasts and phonological categories established in early childhood remaining stable even after decades of L2 immersion (de Leeuw et al., 2010). Observable changes are typically limited to phonetic drift: subtle shifts in the realization of sounds that do not cross phonemic boundaries (Chang, 2019). Common manifestations include convergence of Voice Onset Time toward L2 norms in stop consonants or minor adjustments to vowel space positions (Mennen, 2004). These shifts reflect gradient co-articulatory adjustments driven by bilingual co-activation and production adaptation, rather than restructuring the underlying phonological system (Chang, 2019; Schmid & Köpke, 2017). Drift magnitude correlates with extralinguistic factors such as L2 exposure and proficiency (Bergmann et al., 2015).

Syntax remains highly resilient in attrition, with its core properties, such as basic word order, phrase structure, and verb placement, remaining largely unaffected (Köpke & Schmid, 2004). Vulnerability instead emerges at the syntax–pragmatics interface, where structures rely on the integration of linguistic form with discourse and contextual cues (Sorace, 2011). These externally governed structures impose higher online processing demands and are therefore more susceptible to L2 influence, leading to non-native-like patterns such as the overuse of overt subject pronouns in null-subject languages like Italian and Greek (Tsimpili et al., 2004). By contrast, internal interfaces, namely syntax–semantics (e.g., direct object marker in Spanish) and purely structural grammatical mechanisms tend to remain stable (Chamorro, et al. 2016). This asymmetry indicates that attrition primarily reduces real-time processing efficiency.

2.2.3. Predictive Factors

Language attrition is a multicomponential process shaped by the interaction of internal and external influences. Drawing on Köpke’s comprehensive overview (2007), which synthesizes empirical data across multiple fields, this section discusses predictive factors for attrition. Although each factor can be described separately for analytical clarity, Köpke (2007) stresses that their effects are clustered and mutually reinforcing rather than independent. The possibility, extent, and type of attrition therefore depend not on isolated variables, but on how neurological, cognitive, and sociolinguistic pressures jointly shape bilingual processing.

2.2.3.1. Brain Mechanisms

Plasticity is a decisive biological factor in constraining the critical period for language acquisition. Premature synaptic connectivity in early life facilitates adaptation to new linguistic environments, meaning that a drastic change, such as immigration, interacts strongly with age of acquisition (AoA). L1 attrition in post-puberty migrants manifests primarily in online processing difficulties (i.e., performance), whereas pre-puberty migrants show profound restructuring or complete loss of the underlying linguistic system (i.e., competence). Reduced plasticity in older speakers limits adaptation to the L2 environment while simultaneously protecting the established L1 from attrition. Attrition research consistently confirms age as a central predictor of vulnerability.

Activation is a fundamental neural property of memory. According to the Activation Threshold Hypothesis (Paradis, 1993), the activation of items varies dynamically depending on the frequency of language use. Frequently and recently used items require fewer neural impulses for retrieval than infrequently used ones. In bilingual lexical access, highly used L2 items are activated more readily than less used L1 equivalents, promoting attrition through competitive disadvantage.

Inhibition regulates neural activity by suppressing the activation of competing systems and is crucial in contexts involving high cross-linguistic interference (e.g., L2 learning, L1 attrition, and polyglot aphasia). Its impact depends on the type of use (e.g., language mode, code-switching) and typological distance of the competing languages, leading to diverse attrition outcomes. For instance, immigrants with no L1 contact show significant processing difficulties as their L1 is dormant, yet they experience minimal cross-linguistic interference. Conversely, speakers actively operating in a bilingual mode experience fewer difficulties but greater interference. Furthermore, inhibition explains the L1 difficulties recent migrants face, as strong suppression of the L1 is required to establish the L2.

Subcortical structures contribute to language processing, as the limbic system is an integral part of human communication, especially in spontaneous, emotion-laden speech, and aspects of cognitive control. Language and emotion, therefore, share neuroanatomical substrates, with L1 tightly grounded in limbic, affective circuitry, whereas L2 processing relies more heavily on neocortical, controlled systems. This neurofunctional asymmetry underlies the phenomenon of language embodiment, in which bilinguals show distinct neurophysiological responses to emotionally charged expressions (e.g., prayers, taboo or vulgar expressions) across their languages. Consequently, L1 attrition can be accelerated in contexts of strong motivation

to acquire the L2 (e.g., assimilation, career goals) or when the L1 is actively rejected due to traumatic experiences (e.g., international adoption, political persecution, war-driven migration).

2.2.3.2. *Cognitive Processes*

Memory is related to attrition in multiple ways. Linguistic knowledge is stored in the structurally and functionally distinct components of long-term memory, namely declarative (i.e., explicit) and procedural (i.e., implicit) memory. Declarative memory supports the conscious recall of facts, events, and arbitrary information through associations. Procedural memory underlies automatized cognitive and motor skills and habits. From a linguistic standpoint, declarative memory primarily supports lexical storage and retrieval, whereas procedural memory supports grammar processing. Information that is not regularly activated can eventually be forgotten. Forgetting affects semantic information only minimally, slows over time, is more resistant in written comprehension, and is generally reversible. The age of acquisition, however, is decisive, since declarative memory takes over grammar processing in late L2 learning. Given that interference arises between similar structures, the lexicon remains susceptible to attrition regardless of age, whereas grammatical interference is more pronounced in early bilinguals and less likely in late bilinguals. The central executive of working memory is also implicated in cognitively demanding tasks, such as bilingual language processing. Under conditions of reduced L1 access and heightened L2 competition, cognitive load increases substantially, resulting in processing difficulties characteristic of attrition.

Literacy contributes to the cognitive organization and consolidation of language. It supports L1 maintenance by enabling continued contact with the language (e.g., through reading) and by motivating engagement in related activities. Critically, literacy strengthens memory representations by adding orthographic representations and creating additional synaptic connections, thereby reducing vulnerability to attrition.

Task dependency is a methodological factor with theoretical implications. That is, different tasks targeting the same linguistic structure place different cognitive demands and therefore do not yield comparable outcomes. For instance, lexical access is typically faster in picture–word matching tasks compared to picture-naming tasks, which require additional retrieval and production processes.

2.2.3.3. *External Factors*

Language use is a key variable in attrition (see also Activation and Inhibition). Both frequency (i.e., quantity) and type (i.e., quality) of use are equally important, although the latter has been comparatively understudied. Moreover, sustained engagement with both productive and receptive skills is essential for L1 maintenance.

Cultural context is shaped by Social Cognitive Causal Chains (SCCC), which establish the semantic, linguistic, and pragmatic norms of a community. Immigrants isolated from the L1 culture cannot continuously confirm these representations, leading to the erosion of pragmatic norms. Cultural contexts, therefore, yield different attrition outcomes. Isolated immigrants typically show weakened but structurally intact linguistic norms (i.e., performance attrition). By contrast, migrants embedded in bilingual communities receive regular yet variable input (e.g., attrited L1 speech, heritage speakers, code-switching), resulting in the development of a new SCCC aligned with the L2 culture and may trigger restructuring of L1 competence.

Attitude encompasses both community and individual values, shaping motivation for L2 acquisition or L1 retention. Motivation may be instrumental (e.g., financial reasons) or integrative (e.g., ideological) and interacts strongly with age. For adults, the L1 often forms a central component of identity, whereas school-aged children may prioritize integration into the L2 environment of their peers. Additionally, the speakers' perceptions of bilingualism, cultural affiliation, and degree of assimilation further modulate the likelihood and trajectory of attrition.

While attrition research has identified numerous factors affecting the attrition outcome, one cognitive factor that has received little attention in this context is language aptitude. Its potential relevance for L1 maintenance has been noted repeatedly, yet remains largely underexplored. Given its central role in L2 learning and its possible involvement in different cognitive mechanisms, aptitude warrants independent treatment.

2.3. **Language Aptitude**

2.3.1. *Foundational Models*

Language aptitude is commonly understood as an exceptional “talent” for learning languages. Dörnyei and Skehan (2003) captured this view in their definition of aptitude as “*a specific talent for learning foreign languages which exhibits considerable variation between*

individual learners”. The scientific study of aptitude began with John B. Carroll (1981), who operationalized it as a measurable, predictive construct. While this foundational view positioned aptitude as a stable trait, contemporary research has reconceptualized it as a dynamic and multifaceted attribute (Wen et al., 2017).

Carroll (1981) shifted the concept of language aptitude from a vague “talent” to a scientifically measurable construct, defined as “*an individual’s initial state of readiness and capacity for learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity.*” This framing established aptitude as a latent trait that can account for substantial individual differences in L2 achievement while remaining independent of other constructs, such as motivation and general intelligence. It also positioned aptitude as a fixed, possibly innate factor, resistant to training, and as a salient predictor of ultimate attainment.

Carroll’s central contribution was to conceptualize aptitude as a specialized cognitive faculty, a composite of four distinct components:

- Phonetic coding ability: The capacity to identify distinct sounds, associate them with written symbols, and retain these associations.
- Grammatical sensitivity: The ability to recognize the grammatical functions that words fulfill in a sentence structure.
- Rote learning ability: The skill of learning associations between sounds and meanings rapidly and retaining them efficiently.
- Inductive language learning ability: The capacity to infer or induce rules that govern a language from given samples.

This foundational theory was operationalized in the Modern Language Aptitude Test (MLAT), developed by Carroll and Sapon (1959). Designed to predict an individual's language learning potential, particularly in formal instructional settings, the MLAT was a groundbreaking psychometric test that deconstructed aptitude into specific, testable abilities. This influential test shaped aptitude research for over five decades. The other major instrument from this era, the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966), pursued the same predictive goal but broadened the construct by incorporating auditory ability and motivation. The enduring influence of this approach is evident in modern descendants, such as the computer-based LLAMA Language Aptitude Test (Meara, 2005), which, despite being language-neutral, assesses core cognitive skills that are conceptually rooted in Carroll's framework.

Building on Carroll's foundational work, later models sought to refine this framework. Peter Skehan (1998), for instance, proposed a processing-based model, re-categorizing aptitude

components according to their role in specific stages of language acquisition, moving the focus from purely predicting success to explaining the underlying cognitive mechanisms of learning. This view streamlined aptitude into three core components:

- **Phonetic Coding Ability:** The capacity to process auditory input efficiently and enter it into the processing system.
- **Language Analytic Ability:** The ability to infer linguistic rules, identify patterns, and generalize them.
- **Memory:** The capacity to acquire, store, and retrieve new lexical and grammatical information.

Both Carroll's and Skehan's frameworks underscore the multifaceted nature of aptitude. Together, these frameworks illustrate the foundational shift in aptitude research: from a psychometric tradition to a process-oriented cognitive model. Carroll's work, with its focus on stable traits and predictive tests, exemplifies the psychometric goal of predicting success. Skehan's model, by linking aptitude components to the cognitive mechanisms of acquisition, marks a crucial pivot toward the modern goal of explaining the learning process itself, setting the stage for the more dynamic reconceptualizations that followed.

2.3.2. *Rethinking Aptitude*

While Carroll's model provided the foundation for aptitude research, it has been increasingly challenged. Scholars have highlighted limitations in its treatment of memory (Wen, 2016), its neglect of implicit learning (De Graaff, 1997), its close association with general intelligence (Li, 2015), and, most notably, its assumption that aptitude is a stable, unchanging trait (Kormos, 2013), a view that has been revised across the broader field of individual differences in favor of more dynamic, context-sensitive models (Dörnyei & Ryan, 2015). Robinson's research on Aptitude-Treatment Interaction (ATI) and his "Aptitude Complexes" (2001) framework marked a turning point in the field. His work linked individual differences to specific processes of language acquisition, showing that different constellations of cognitive abilities, such as memory, attention, and processing speed, are optimal for learning different linguistic features under varied instructional conditions.

These theoretical advances are strongly supported by large-scale empirical synthesis. Li's (2015) meta-analysis of 33 studies spanning five decades provides a nuanced view of language aptitude in L2 grammar acquisition. The results showed a moderate correlation

between aptitude and grammar learning, lower than previously assumed. Language analytic ability emerged as the strongest predictor of grammatical success, surpassing rote memory and phonetic coding. Crucially, aptitude was most influential under explicit instruction and in the early stages of learning. Age effects were also context-dependent: younger learners benefited more from aptitude in long-term studies, whereas older learners relied on it in short-term interactional contexts, a pattern specified by research highlighting the role of executive functions, such as cognitive control, in how learners deploy their aptitude to meet different task demands (Li, 2015).

A major catalyst in this reframing has been the recognition of working memory (WM) as central to language aptitude. Working memory is the cognitive system responsible for the simultaneous storage and active processing of information needed for complex tasks (Baddeley, 1986). Its components are key for language learning: a phonological short-term store that retains new verbal and sound-based forms, and central executive functions that manage the manipulation and integration of complex information, such as novel grammatical structures (Wen, 2016). Crucially, WM has been reframed from a fixed trait to a capacity shaped by experience and training. This implies that if a central component of aptitude is malleable, then aptitude itself must be viewed as a dynamic, trainable construct, with working memory now widely considered its core cognitive substrate (Wen, 2016). This cognitive reconceptualization of aptitude provides the foundation for inquiry at the neural level, where researchers are now uncovering its biological correlates.

2.3.3. *Neural Correlates*

Neurocognitive research substantiates this theoretical shift, identifying tangible biological markers that distinguish high-aptitude learners. Turker et al. (2021) present their neurocognitive model based on a comprehensive review of existing neuroscientific research on language learning and aptitude by synthesizing findings on the structural and functional brain characteristics that distinguish high-aptitude learners. Structurally, high aptitude is associated with greater grey matter volume and more complex gyrification in the bilateral auditory cortices, supporting superior phonetic coding and novel speech sound learning. Increased cortical thickness in Broca's area (left IFG) correlates with enhanced language analytic abilities and grammar rule extraction. The integrity of white matter tracts, particularly a stronger left arcuate fascicle connecting temporal and frontal language regions, facilitates efficient speech

imitation and faster processing by enabling better communication between sound-processing and grammar-production centers.

Functionally, these structural advantages translate into more efficient and dynamic neural processing during linguistic tasks. High-aptitude learners often show more focal or decreased activation during tasks like speech imitation, reflecting optimized neural efficiency, while more complex tasks, such as novel grammar analysis, can elicit broader activation as additional resources are recruited. They also exhibit enhanced synchronization and connectivity, with stronger local coordination and more efficient global network patterns across language-related regions. Finally, heightened neuroplasticity allows faster and more effective formation of new neural pathways in response to language learning, consistent with their predisposed biological aptitude. Together, these findings provide a neurobiological basis for aptitude, anchoring its cognitive components in measurable brain structures and functions.

Taken together, research on language aptitude has moved from a static, trait-based conception to a dynamic, multidimensional understanding rooted in cognitive and neurobiological evidence. Language aptitude can then be described as a multi-faceted, working-memory-centered construct that is responsive to experience and instruction, has identifiable neural underpinnings, and remains predictive of learning outcomes. This synthesis integrates cognitive, empirical, and neuroscientific perspectives, providing a robust foundation for operationalizing aptitude within language attrition research.

2.4. Interfaces of Attrition and Aptitude

Having established the core concepts of language attrition and language aptitude, this section explores their intersection. Language aptitude is widely acknowledged as the most robust predictor of successful second language acquisition (Artieda & Muñoz, 2016). Yet, despite its prominence in L2 research, it remains one of the most underexplored variables in the study of language attrition (Bylund & Ramírez-Galán, 2016). Gardner (1982) was among the first to propose a possible link between aptitude and attrition, followed by Ginsberg (1986), who noted that the topic had received virtually no empirical attention. Although attrition research gained renewed momentum in the early 2000s, Köpke (2007) observed that aptitude remained on the margins of the discussion, despite its potential to account for individual differences in language retention. Since then, progress has been limited, with recent reviews

confirming that aptitude has not become a central focus and remains largely speculative in the field (Schmid & Köpke, 2019; Parafita Couto & Semplicini, 2024).

2.4.1. Language Aptitude and First Language Competence

Contrary to the long-standing assumption of uniform competence among native speakers, research increasingly reveals substantial variability in first language performance, shaped by both cognitive and experiential factors. These findings challenge the traditional view of language aptitude as a construct exclusive to second language acquisition and instead support a broader, domain-general capacity relevant to both native and non-native linguistic performance.

Empirical and correlational studies that directly examine the relationship between aptitude (or its components) and observable variation in native language performance, such as grammar, vocabulary, and phonological processing, provide foundational evidence that aptitude plays a role in L1 development. Complementing these findings, other studies drawing on theories and findings from implicit learning, memory systems, and neurobiology help illuminate the cognitive and biological mechanisms through which aptitude influences L1 outcomes.

In a longitudinal study, Skehan and Ducroquet (1988) found strong correlations between early L1 developmental milestones and later language aptitude scores in monolingual adolescents, suggesting that aptitude manifests early in native language acquisition. Dąbrowska (2018, 2022) examined grammatical, lexical, and collocational knowledge among adult native speakers and found that aptitude test scores significantly predicted variation in these domains. Sparks et al. (2019, 2023) investigated the relationship between general aptitude, L1 attainment, and cognitive processing in adolescent L2 learners. Their findings showed that individuals with higher L1 reading and writing skills were more successful in L2 learning, supporting the view that common cognitive abilities underlie both L1 and L2 performance. Reiterer et al. (2011, 2013) focused on phonological processing and demonstrated that aptitude predicts individual differences in L1 phonetic imitation and prosodic sensitivity. Taken together, these studies provide converging evidence that aptitude influences multiple dimensions of native language competence.

Variation in implicit learning ability helps explain individual differences in core native language skills, particularly grammar and phonological processing. Granena (2013) found that

learners with stronger implicit learning abilities performed better in grammatical tasks across both L1 and L2, highlighting the relevance of unconscious pattern detection. Building on this, Artieda and Muñoz (2016) examined distinct components of aptitude and identified phonemic coding as especially influential in early stages of L2 acquisition. Their findings align with Bylund et al. (2012), who argued that aptitude shapes outcomes in both formal and informal learning contexts. Together, this evidence suggests that implicit learning ability is a fundamental component of language aptitude and helps explain individual variation in native language skills.

Differences in L1 achievement can be largely explained by variation in procedural memory, a cognitive system responsible for implicit learning and the acquisition of skills such as grammar (Ullman, 2004). Supporting this view, Kaufman et al. (2010) found that native speakers' grammatical proficiency was predicted by performance on procedural learning tasks. Further evidence comes from Lum et al.'s (2014) meta-analysis, which linked procedural learning deficits in children with specific language impairment (SLI) to grammatical difficulties, underscoring the role of procedural memory in typical L1 development. These findings indicate that procedural memory may be a core cognitive mechanism underlying both language aptitude and native grammar skills.

Genetic, functional, and structural evidence increasingly supports a biologically grounded, domain-general view of language aptitude. Rimfeld et al. (2015) conducted a large-scale twin study showing that most individual differences in L2 achievement are explained by heritability rather than environmental influence. Crucially, they found a strong genetic correlation between L1 and L2 grammatical achievement, suggesting that a shared biological basis underlies broad linguistic aptitude across native and non-native domains. Neuroimaging research corroborates these genetic findings. Hu et al. (2012) investigated behavioral and neurobiological predictors of advanced L2 pronunciation. Their findings revealed that individuals with high aptitude exhibited strong activation patterns in language-related brain areas during both L1 and L2 tasks. Notably, their neural activation was not specific to a single language but rather reflected enhanced auditory and articulatory processing across linguistic contexts. Extending this work, Ramoser et al. (2024) used structural MRI with automated segmentation of Heschl's gyrus and found that individuals with higher aptitude exhibited fewer duplications but greater surface area in specific Heschl's gyri. These anatomical features also correlated with greater multilingual experience, linking variation in auditory cortex morphology directly to both native and foreign language performance.

Taken together, these findings indicate that language aptitude is not confined to L2 acquisition but plays a crucial role in shaping L1 competence as well. Individual variation in

grammar, vocabulary, and phonology among native speakers reflects underlying aptitude-related differences in cognitive and neural processing. Recognizing aptitude as a domain-general construct has important implications for attrition research, as it may help explain why some individuals are more resilient to L1 attrition than others. The next section explores this possibility in more depth.

2.4.2. The Role of Language Aptitude on Language Attrition

As previously mentioned, the implications of language aptitude for L1 attrition have barely been investigated to date. However, given its established role in both L1 competence and L2 acquisition, it is reasonable to assume that language aptitude may also be relevant for language attrition. Based on theoretical considerations, three hypotheses can be proposed regarding the role of aptitude in attrition: higher aptitude may (a) reduce attrition by supporting L1 retention, (b) increase attrition by facilitating L2 dominance, or (c) have no effect. The first two hypotheses align with distinct theoretical frameworks: (a) is supported by aptitude-based accounts, which assume a stabilizing role of aptitude in maintaining linguistic structures, while (b) resonates with proficiency-based accounts, where high L2 proficiency may suppress or override L1. The third hypothesis, that aptitude has no effect, is an open possibility in the absence of clear theoretical backing. Despite early theoretical references, only a handful of empirical studies have addressed the role of aptitude in the context of language attrition.

The pioneering study by Bylund et al. (2009) investigated the role of language aptitude in L1 attrition among pre-pubescent bilinguals, paving the way for subsequent research investigating this relationship. They emphasize that puberty represents a critical period for language acquisition and attrition, given that reduced exposure to L1 before puberty can lead to substantial attrition, while after puberty, the impact is minimal. The study involved Spanish L1 speakers who immigrated to Sweden before puberty and acquired Swedish as their L2. Participants completed an L1 grammaticality judgment test alongside a standardized language aptitude assessment. A monolingual Spanish control group was also tested for comparison. Results indicated that participants with above-average language aptitude scores maintained high levels of L1 proficiency regardless of the frequency of L1 use. Based on these findings, the authors argue that language aptitude has a compensatory function in early L1 attrition, mitigating the negative effects of reduced L1 exposure.

Opitz (2011) analyzed the interplay of L1 maintenance and L2 acquisition in late bilinguals after immersion in an L2 environment in adulthood, and attempted to identify the implications of language aptitude (among other factors) on these two processes. To that end, L1 German – L2 English bilinguals and one control monolingual group per language participated in a study consisting of multiple, extensive language tests in both languages, a language aptitude test, and questionnaires regarding their sociolinguistic background. The results revealed a significant correlation between lexical learning ability and the level of L1 retention. In particular, they indicate that apart from its role in L2 acquisition, aptitude, at least verbal skills, can potentially impact L1 maintenance as well, though to a lesser degree than other factors.

In an effort to expand this investigation, Bylund and Ramírez-Galán (2016) examined the role of language aptitude in L1 retention in late bilinguals. The L1 Spanish–L2 Swedish bilingual subjects immigrated to a Swedish-speaking environment in adulthood with no prior knowledge of the L2. The experiment involved an L1 grammaticality judgment test and a language aptitude test, and a detailed linguistic profile was obtained. The grammar and aptitude scores of the bilinguals were then compared to those of an L1 monolingual control group. The findings demonstrated that the control group scored significantly higher in grammatical judgment. Language aptitude was deemed non-significant for grammatical intuition, whereas L1 identification accurately predicted grammaticality judgment performance. Consequently, the authors argue that the lack of aptitude effects on L1 attrition is an age-related issue, concluding that aptitude does not affect L1 attrition in late bilinguals.

Lehner (2018) investigated the role of aptitude on L2 attrition by examining correlations between L2 proficiency level, maintenance, and attrition in high- vs. low-aptitude groups of L1 English – L2 German bilinguals. The participants completed a questionnaire regarding their personal, educational, and linguistic backgrounds, L1 and L2 proficiency tests, and a language aptitude test. In line with earlier findings, the results showed a significant correlation between aptitude and L2 proficiency. More specifically, above-average aptitude not only facilitates the achievement of high L2 proficiency but also prevents attrition in cases of reduced L2 contact. Additionally, age of acquisition, language use, and exposure were not significant for the high-aptitude group. The results, however, did not show any apparent association between aptitude and L1 proficiency, which was attributed to the limited sample size (i.e., 29 participants).

Empirical studies on the role of language aptitude in attrition therefore provide evidence supporting two of the three hypotheses: aptitude may act as a protective (Bylund et al., 2009; Lehner, 2018) or a neutral (Bylund & Ramírez-Galán, 2016; Opitz, 2011) factor in L1 attrition. However, firm conclusions remain elusive due to the limited number of studies and the

substantial variation in their scope and design. Differences in the type of bilingualism (early vs. late), the linguistic domain under investigation (grammar vs. lexicon), the language affected by attrition (L1 vs. L2), and the type or reporting of aptitude tests (e.g., LLAMA, MLAT, or unspecified) further complicate generalizations. Sample sizes also vary significantly, further limiting statistical power. As such, while the hypothesis that aptitude may protect against attrition finds some support, current evidence is inconclusive, and the question remains open for further empirical investigation.

2.5. Present Study

Despite growing interest in individual differences in first language (L1) attrition, the role of language aptitude remains empirically unresolved. As reviewed in the preceding section, existing studies have yielded mixed, often null findings, and no clear consensus has emerged on whether aptitude supports L1 maintenance, facilitates L2 dominance, or has no effect at all. Consequently, aptitude continues to occupy a marginal position in attrition research, despite its well-established relevance for foreign language learning and accumulating evidence that it contributes to variability in native language performance.

This lack of convergence can be attributed to both conceptual and methodological variation across studies, including differences in participant populations, linguistic domains examined, and the operationalization of variables. To expand the investigation of the role of language aptitude in attrition, the present study focuses on the lexical retrieval efficiency in late bilingual adults immersed in an L2 environment, thereby reducing confounds related to incomplete acquisition and targeting a domain particularly susceptible to attrition. Given the selective nature of attrition and the reduced neurocognitive plasticity in adulthood, attrition-related effects are more likely to manifest as subtle changes in processing efficiency. Importantly, a significant association between lexical learning ability and L1 retention has been reported in late bilinguals. Building on this finding, the present study further examines this relationship in greater detail by focusing on online lexical retrieval performance and efficiency measures, which are particularly sensitive to processing-based attrition outcomes.

The study addresses this topic through the following research questions:

RQ 1: Does language aptitude predict the degree of L1 lexical retrieval maintenance in late bilingual adults?

RQ 2: Does language aptitude interact with experiential or sociolinguistic factors to influence individual variability in L1 lexical maintenance outcomes?

The following hypotheses are tested:

H₀: Language aptitude has no effect on L1 lexical maintenance outcomes.

H₁: Higher language aptitude is associated with better L1 lexical maintenance outcomes.

H₂: Language aptitude interacts with sociolinguistic and experiential variables to modulate L1 lexical maintenance outcomes.

The primary aim of this study is to examine whether language aptitude predicts the degree of first-language lexical maintenance in late bilinguals, as well as the direction of this potential relationship. Beyond this central objective, the study considers a range of sociolinguistic and experiential factors to identify patterns that may account for variation in bilingual language maintenance. By focusing on lexical retrieval in late bilinguals, the present study targets a domain and population that provides favorable conditions for observing how language aptitude modulates attrition-related outcomes. The following chapter details the methodological approach adopted to address these research questions empirically.

3. Methodology

This chapter outlines the methodological framework for examining whether language aptitude is associated with variability in first language (L1) maintenance among late bilinguals. A controlled, quantitative, cross-sectional study was conducted online, focusing on lexical retrieval performance in participants' L1 following extended immersion in a second-language environment. Methodological control and data reliability were prioritized throughout the study's design and implementation.

Section 3.1 provides a brief overview of the study design and the variables under investigation. Section 3.2 then outlines the participant selection criteria and recruitment procedures that guided the sample composition. Section 3.3 presents the materials and experimental tasks in detail, together with the procedural framework governing participant preparation, task sequencing, and online implementation. Section 3.4 describes the data preparation and scoring procedures through which task performance was quantified and prepared for statistical analysis.

3.1. Experimental Design

This study employs a behavioral, correlational design to examine whether language aptitude predicts the extent and severity of first language attrition in late bilinguals. The independent variable is language aptitude, and L1 attrition is the dependent variable. To ensure the validity of the findings, other factors known to influence language attrition (Schmid, 2011; Köpke & Schmid, 2004) were measured and accounted for. These predictors include participants' attitudes and motivation towards each language and culture, type and frequency of language use, age of onset of L2 acquisition (AoA), L2 proficiency, educational background, and length of residence (LoR) in an L2 environment. This approach isolates the specific role of aptitude, while also considering the influence of other factors, thereby strengthening the reliability of the findings and offering nuanced insights into the dynamics of first language attrition.

A control group was not included in this study due to the unique nature of L1 attrition, which specifically affects expatriates with prolonged exposure to L2 environments. Including a control group of non-expatriate monolinguals or bilinguals would not provide a relevant

comparison, as these groups do not experience the same attritional influences. Therefore, the study focuses on within-group comparisons among expatriates to investigate how individual differences, such as aptitude levels, influence the degree of L1 attrition. This approach allows for a more nuanced understanding of the factors that affect L1 maintenance among those living in non-native language environments.

Prior to launching the study, quality assurance measures were implemented to test and refine the experimental procedures. In a preliminary evaluation, two L1 native speakers, both educated in linguistics, were engaged as informants and provided feedback on the test formats (e.g., aptitude test, proficiency test, questionnaire), clarity and formulation of instructions (e.g., poster, leaflet, within-experiment), the selection of the items and the scoring scale for the picture naming task. Additionally, two trial runs were conducted with participants who matched the target population to ensure that the study was properly designed and functioning as intended. This involved examining the outputs to determine the reliability of the equipment, setup, and instructions. As a result, a few minor adjustments were made.

Language production was not a central focus in psycholinguistics until the last decade, largely due to the lack of online methods for studying production in controlled conditions. Early research, therefore, relied on offline techniques, which limited its scope. While online data collection has long been established for language perception studies, it was not until the post-pandemic era that significant attention was given to developing online solutions for speech production research. Recent studies have shown that reliable and accurate data on overt articulation can be gathered through online platforms, and that internet-based experiments can reliably replicate effects typically observed in lab-based speech production studies (Fairs & Strijkers, 2021; Vogt et al., 2021).

Conducting studies online offers several notable advantages. It provides broader outreach by accessing a larger, more diverse participant pool, which enhances the generalizability of findings. Data collection is also more time- and cost-efficient, being quicker, more flexible, and involving lower logistical expenses. Additionally, online studies have been shown to yield reliable data quality (Fairs & Strijkers, 2021). However, there are challenges to consider. Technical variability, such as differences in programming language, operating system, internet speed, and audio quality, can impact data consistency. Furthermore, the informal online setting may lead to less controlled participant behavior, potentially affecting the quality of the data collected (Vogt et al., 2021).

3.2. Participants

3.2.1. *Sample Selection Criteria*

To ensure internal validity and limit external variability, key sample selection criteria were established, aligned with standard practices in language attrition research (Schmid, 2011). This rigor in sample selection enhances the study's relevance to the research question, minimizes confounding factors, and strengthens the replicability of findings, allowing for clearer insights into the role of language aptitude in L1 attrition.

a) Age

Age-related considerations are crucial for the sample selection. A minimum age of acquisition (AoA) of 15 years for the L2 was initially set to ensure that participants did not benefit from the enhanced abilities associated with early language acquisition. The study therefore specifically targets late sequential bilinguals who acquired their second language outside this window of heightened neuroplasticity. This ensures that the focus remains on explicit learning processes, which involve conscious attention to language rules and do not rely on the natural advantages of the critical period (DeKeyser, 2000). For the same reason, a lower AoA was accepted for participants who engaged in purely explicit and non-intensive language learning. Moreover, an age limit of 55 years was initially set due to observed age-related declines in cognitive functioning as early as the 50s (Connor et al., 2004)

b) Typical Language Development

Participants with diagnosed language disorders or learning disabilities were excluded, as these cognitive impairments can significantly alter language production and retention patterns, thereby impacting key mechanisms such as memory and linguistic structure (Ullman & Pullman, 2015; Leonard, 2014). This exclusion criterion is common in studies of L1 attrition and language production to ensure that observed outcomes reflect typical language abilities without interference from cognitive deficits (Schmid, 2011).

c) Linguistic and Cultural Homogeneity

Sample homogeneity in linguistic and cultural backgrounds is an essential factor, as it can significantly impact language processing and attrition. A sample of Greek participants in a

German-speaking environment was selected for its methodological and practical advantages, ensuring that any observed attrition effects could be attributed primarily to cognitive factors rather than to typological proximity or cultural motivations (Schmid & Cherciov, 2019).

Firstly, the Greek (L1) – German (L2) linguistic pair ensures consistency in cross-linguistic effects by limiting variability that could arise from different L1-L2 pairs. Additionally, Greek and German, belonging to the Hellenic and Germanic language families, respectively, albeit both Indo-European languages, have considerable typological differences. This linguistic distance reduces the confounding effects of language similarity and enhances the role of language aptitude, as managing typologically different languages requires greater cognitive effort and are more resistant to attrition (Schmitt & Sorokina, 2024; De Bot, 1997). Finally, the shared language combination leverages the experimenter's sociolinguistic background, which ensures more accurate and nuanced assessments of the outcomes.

A shared cultural background controls for sociocultural influences on language attrition, aligning participants' motivations and attitudes toward L1 maintenance and L2 proficiency (Schmid & Köpke, 2017). Research on Greek migrants consistently shows a pattern of integration that prioritizes adaptation without full assimilation, allowing them to engage actively in the host society while maintaining a strong connection to their linguistic and cultural heritage as a core aspect of identity (Chalari, 2024). In parallel, Germany and Austria emphasize structured language proficiency requirements in their integration policies, underscoring language as essential for social and economic inclusion (Lang, 2022; Konrad et al., 2024).

d) Length of Residence

A five-year minimum length of residence (LoR) in an L2 environment was chosen for both theoretical and practical reasons. Research indicates that this threshold is sufficient to observe significant patterns of L1 attrition, as cognitive processing shifts in favor of the L2, reallocating brain resources to the more frequently used language (Schmid, 2011). Continuous exposure to the L2 over extended periods accelerates L1 decline as individuals reduce L1 use and experience increased cognitive interference from the L2. Furthermore, five years of immersion supports adequate L2 proficiency, which enhances cognitive interference and, in turn, more pronounced L1 attrition effects.

e) L2 Proficiency

Given that this study examines late bilinguals immersed in an L2 environment, participants' L2 proficiency was a crucial criterion in the sample selection process. Higher L2 proficiency, especially in immersive contexts, is closely linked to greater first language (L1) attrition, as increased L2 use often leads to reduced L1 fluency due to less frequent use and cognitive interference (Köpke & Schmid, 2004). In this study, participants were required to have a high level of L2 proficiency to ensure sufficient L2 immersion, which could potentially reveal L1 attrition. Proficiency was assessed using a combination of self-reports, standardized language assessments and language usage questionnaires, ensuring consistent measures across participants. This criterion controls for variability in L2 proficiency, allowing for a clearer analysis of how language aptitude interacts with attrition.

f) Sample Size

Sample size considerations for psycholinguistic research depend on the study's objectives and design. Sample sizes of around 30 participants per experimental and control group have been used in L1 attrition research, showing that meaningful effects can be detected without overextending resources (de Leeuw et al., 2010). An optimal sample size of 40 participants is recommended for online speech production studies to ensure reliable results comparable to those obtained in traditional laboratory settings (Fairs & Strijkers, 2021). For studies focusing on language aptitude, sample sizes generally range from 30 to 50 participants to adequately capture variability in individual language learning abilities. For this study, a minimum sample size of 30 participants was chosen to balance statistical power with practical constraints, while, as previously discussed, a control group for the present study was deemed irrelevant.

3.2.2. Participant Acquisition and Exclusion

A poster was created to attract potential participants, providing brief information about the study topic, participation requirements, procedures, and contact details (see Appendix A1). The design included cultural references to captivate the target audience, and a 20€ voucher incentive was offered to increase motivation. The poster was primarily shared on social media, particularly in Facebook groups for Greek expats, which proved to be the most effective method. Additionally, Greek embassies and associations were contacted via email (see

Appendix), emphasizing the study's significance for the Greek community. The email highlighted that the study would enhance the representation of the Greek language and diaspora in the international scientific community, where it is currently underrepresented. Finally, participants were also contacted personally or referred through acquaintances.

Despite efforts to control participation and maintain engagement, participant withdrawal was inevitable. A substantial total of 26 candidates dropped out of the study, representing approximately 43.3% of the initial participants. Some of these candidates withdrew immediately before beginning any tasks, while others finished some tasks but never completed the study. This high dropout rate highlights the challenges of retaining participants throughout the study.

Participant exclusion after the study commenced was necessary in some cases. One participant was excluded due to technical issues that resulted in significant data loss, namely, missing audio files from the picture naming task, rendering their participation invalid. Another participant was excluded due to evident bad-faith participation, such as repeatedly pressing the same button or engaging in conversations during audio recordings, despite explicit instructions to refrain from doing so.

Following data collection, all responses were reviewed and prepared for analysis. Data were collected anonymously, with no personal identifiers, in accordance with ethical guidelines. Participants were assigned anonymized IDs consisting of 'p' plus a sequential number (p1–p36), reflecting the order of participation and ensuring confidentiality throughout data processing.

3.3. Materials

3.3.1. Procedure

a) Participant Preparation

After initial contact, participants received a pre-study briefing in the form of a short guide containing information about the study. Therein, sufficient information was provided on the study's purpose, participation requirements, procedure, task formats and duration, reward, and data protection, without disclosing the goal to avoid subject bias. All instructions were provided in Greek to ensure full comprehension and to establish a consistent monolingual context (Opitz, 2011), aligning with the language environment of the upcoming test.

Participants were informed only that language use was being investigated, as attrition is an emotionally loaded topic that could lead to overcompensation or negative feelings towards the study (Schmid, 2011). They were encouraged to ask for clarification after reading the instructions, and were given the opportunity to do so. They were also asked questions to verify that they fulfilled the criteria and understood the procedure. Once participants agreed to the terms, they were sent the study link and reminded that they could reach out with any questions at any point during the process. The link was not made public to prevent potential exclusions later on due to participant withdrawal, failure to meet criteria, or unreliable participation.

b) Study Instructions

Throughout the study, instructions are designed to be clear and concise, providing all essential information while maintaining participant focus. All instructions are delivered in Greek to preserve a consistent monolingual environment. The experiment begins with a welcome message in which the experimenter introduces herself and briefly outlines the study's purpose. This is followed by a consent form detailing the terms and conditions of participation. Upon granting informed consent, participants receive general information about the study. They are informed that each task differs in format and that instructions must be read carefully. It is emphasized that there is no option to return to a previous task, and participants are advised to complete the study in a quiet, distraction-free environment. They are reminded that, despite the informal setting, this remains a scientific study and that appropriate conduct is expected, and that their data will be used only for research purposes. Before each task, participants are presented with specific instructions outlining its purpose, format, response method, number of trials, and any time constraints. The study concludes with a short thank-you message, an option to enter the voucher raffle, and a request for information about participants' technical setup (e.g., mouse, browser, operating system), enabling later evaluation of data consistency, if needed.

c) Task Sequence

The tasks are administered in a predetermined sequence, beginning with those that demand significant cognitive effort and are most relevant to the research question, gradually progressing in descending order. This arrangement minimizes the risk of overstimulation and ensures that participants' cognitive resources are at their peak for the most critical components of the study (Chen, Paas & Sweller 2023). The first task requires prompt and authentic responses, while the subsequent task demands more sustained concentration. Following these,

proficiency tests merely establish a baseline of participants' language competencies. Finally, the questionnaire, being non-performance-based and self-referential, is less cognitively taxing and more engaging for participants at the end of the session. Despite the variety of tasks, the study was designed to be completed within roughly 30 minutes to prevent fatigue and reduce dropout rates.

d) Experimental Platform

The study was created entirely on and administered via Gorilla Experiment Builder (gorilla.sc), a cloud-based, “*fully tooled experiment authoring and deployment platform*” (Anwyl-Irvine et al., 2019) specifically designed for conducting online behavioral research. Gorilla provides a user-friendly interface, wide range of experimental tools and customizable features at a low cost, enabling researchers of varying expertise to design and deploy transparent and reliable experiments online. The platform provided comprehensive control over both task design and sequencing and supported various task formats, including picture naming, multiple-choice questions, and text input, enabling diverse data outputs. Gorilla’s integrated media tools ensured accurate presentation of visual stimuli and recording of participants' vocal responses. Timing functions controlled the duration of each trial, ensuring consistency in task performance across participants. Automated data logging captured responses, minimizing human error and safeguarding data integrity. Additionally, Gorilla’s compatibility with various operating systems ensured seamless accessibility for participants.

3.3.2. *Task Overview*

3.3.2.1. *Attrition Task*

To assess the presence and extent of first language attrition, a Picture Naming Task (PNT) was administered. Picture naming is a widely used and reliable experimental paradigm in speech production research (Fairs & Strijkers, 2021; Schmid, 2011) and serves as a behavioral measure of lexical access. In this task, participants are presented with visual stimuli of familiar objects or concepts to be named promptly in the L1. The speed and accuracy of their responses are recorded, providing comprehensive insights into their lexical retrieval processes. Reaction times (RTs) and accuracy serve as key indicators of automaticity and lexical retention, respectively (Schmid, 2013). Delayed naming latencies often indicate slower or less automatic

lexical retrieval, a hallmark of attrition. The picture naming task provides a fine-grained analysis of naming speed and word access, making it particularly effective for investigating L1 attrition. The PNT is especially sensitive to latency-related deficits, meaning that although attriters often maintain high accuracy rates, reaction times can reveal subtle deficits in lexical access (Gallo et al., 2021). Additionally, the PNT provides consistent results in measuring lexical retrieval when stimuli are rigorously controlled, ensuring uniformity across participants and minimizing extraneous factors.

a) Stimuli

The stimulus material for the picture naming task comprises 60 items. Research suggests that a stimulus set of around 60 items is optimal for picture-naming tasks, as it provides a balance between reliability and participant engagement, while minimizing task fatigue (Schmid, 2011). All items were selected from the MultiPic databank (Duñabeitia et al., 2018), which provides 750 illustrations of concrete concepts (i.e., single objects) standardized for name agreement and visual complexity across six European languages. To ensure consistency and reliability in lexical retrieval, the selection process involved multiple criteria applied systematically, including visual and linguistic properties, word frequency, and length. The final stimuli were selected for their appropriateness to the task, with particular emphasis on minimizing cognitive load and ensuring valid measures of lexical access and retrieval in the context of L1 attrition in late bilinguals.

i) Visual Properties:

Since items in MultiPic are not normed for Modern Greek, the initial selection of items was based on the British English data provided in the database, specifically focusing on visual criteria. These criteria included a percentage of correct responses of at least 90%, a percentage of unknown items no greater than 10%, and a mean visual complexity rating of 3 or below on a 1 (very simple) to 5 (very complex) scale. The selected images were thus chosen for their visual simplicity, which minimizes cognitive load and ensures consistency across participants by allowing them to focus on linguistic processes rather than visual perception (Alario et al., 2004).

ii) Linguistic Properties:

The English target names of the selected items were independently translated into Modern Greek. The linguistic properties of the target names in Greek were carefully controlled

to ensure consistency and reliability in measuring lexical retrieval. The following criteria guided the selection process:

- Concrete nouns: All stimuli are concrete nouns, as concrete words are more readily processed and retrieved due to their strong mental representations. This reduces variability in participant responses, making them ideal for lexical retrieval studies (Ding et al., 2017).
- Inflection: To minimize grammatical complexity, all nouns were expected to be named in the nominative singular case. This baseline representation avoids variability introduced by morphological inflection, thereby enhancing the reliability of lexical retrieval measures (Regel et al., 2019).
- Single Words: Single words and closed compound words (e.g., ‘spaceship’) were chosen, as they function as cohesive lexical units in cognitive processing. The cognitive load resulting from semantic and syntactic complexities introduced by multi-word units like phrasal nouns (e.g., ‘bus stop’) is thus limited (Brooks & Garcia, 2015).
- Non-Cognates: Cognates with German and English, including internationalisms, were excluded to minimize the cognate facilitation effect resulting from shared phonological and semantic properties, which facilitate lexical retrieval (van Hell & Dijkstra, 2002). Their exclusion ensures that retrieval is solely based on first-language processing.
- Semantic Categories: The stimuli cover several semantic categories to minimize domain-specific biases (Banks & Connell, 2022) and include universally recognizable concepts. Each category contains a balanced number of items reflecting the natural variability of language use and engages distinct cognitive processes (e.g., tools evoke functional, food sensory, and professions sociocultural associations).
- Semantic Unambiguity: Ambiguity was considered during the stimulus selection, as it has been shown to affect processing speed and accuracy in lexical retrieval (Rodd et al., 2002). To ensure semantic clarity and invariance, ambiguity stemming from polysemy (‘baby’ - infant or a term of endearment), homonymy (‘bank’ - financial institution or side of a river), and synonymy (‘cup’ or ‘mug’) was significantly limited. The inclusion of such items was addressed by using clear, contextually defined images of concrete objects, ensuring participants would identify the intended meanings unambiguously. In the case of synonymy in the stimuli, only the target name was controlled, while synonymous answers were accepted as correct.

iii) *Word Frequency and Length:*

The next step in the stimulus selection process involved evaluating word frequency, a critical psycholinguistic factor influencing lexical retrieval. Word frequency significantly affects naming speed and overall performance in lexical tasks, with more frequent words being retrieved faster and more accurately (Alario et al., 2004). For this study, words with lower frequency values were chosen to better capture signs of attrition. The frequency values of the Greek target names were obtained from SUBTLEX-GR, a subtitle-based corpus containing over 23 million words in Modern Greek (Dimitropoulou et al., 2010). Following the recommendations of Dimitropoulou et al. (2010), frequency data were extracted from the corpus's restricted file, a refined version designed for accurate psycholinguistic research.

Two frequency metrics were considered during item selection to ensure validity and precision. The log-transformed contextual diversity (i.e., Lg10CD), which reflects the variety of contexts in which a word is used with four-digit precision, was chosen as it is considered the most reliable frequency measure for material selection (Dimitropoulou et al., 2010). On the other hand, SUBTLEX_WF measures raw word frequency per million words and facilitates comparison across different databases, as it is independent of corpus size. In this study, words with Lg10CD values ≤ 3.000 and SUBTLEX_WF values ≤ 35.000 were selected. Although the study focused on lower-frequency words, the selected stimuli were further grouped into three categories based on Lg10CD frequency values to allow for a more nuanced analysis of frequency effects on lexical retrieval. Syllables were chosen as the unit of measurement for word length to align with the phonological processing. Words were divided into syllables based on native-speaker intuition and in accordance with the syllabification rules of Modern Greek (Malikouti-Drachman, 2001). Based on this analysis, words were grouped into three categories, depending on the syllable count.

The word frequency and length categories were systematically combined to create nine distinct groups, each containing 6 or 7 items (see Appendix B.3.3). This procedure ensured balanced and controlled item selection across frequency and length dimensions. This approach not only strengthens the validity of the lexical retrieval measures but also provides a framework for future studies to explore these variables in greater detail.

b) Task Design and Procedure

This section details the technical setup, participant instructions, and procedures for administering the task, with a focus on standardization to ensure high-quality, reliable data. To address the challenges of unsupervised online testing, a combination of standardized trials,

practice runs, and strategically placed attention checks was implemented. These measures were crucial for maintaining participant engagement and ensuring consistency across sessions, in line with best practices in experimental design.

i) Trial Design and Setup:

Audio recordings were facilitated by Gorilla's Audio Recording Zone, which captures participants' oral responses directly from their devices. The record mode enabled audio, embedded data and metrics to be recorded and saved for analysis. Additionally, the echo cancellation feature was activated to prevent interference from external sounds.

The trial configuration (see Appendix Table C5.1) is designed to balance consistent temporal spacing, minimize distractions, and provide flexibility for participants, ensuring optimal conditions for data collection. Each trial followed the same standardized sequence, starting with a fixation cross displayed on a white screen for 250 ms, with a 10 ms inter-stimulus interval before and after its appearance. This duration effectively orients participants' attention without unnecessary delays or overexposure, while the brief pause ensures no perceptual overlap between stimuli (Liu et al., 2021; Shao et al., 2013).

Figure 1. Example of a test trial from the picture naming task.



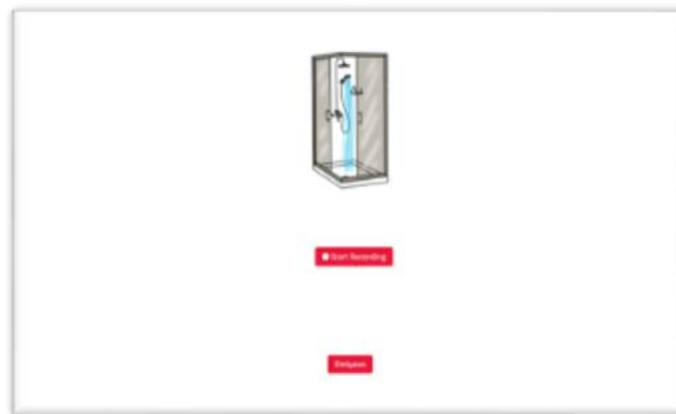
The presentation order of the test items was pseudorandomized to avoid sequence effects, ensuring that participants were not exposed to items in a predictable or biased manner. After the fixation cross, the test item is presented on a white screen for 10.000 ms, during which automatic recording begins immediately as the image appears. Automatic recording ensures no delay between stimulus onset and data collection, which is critical for speech latency measurements. The white screen minimizes visual distractions (e.g., by omitting an on-screen recording indicator or countdown timer) by isolating task-specific stimuli. The 10,000 ms response time provides participants with ample time to articulate their responses while encouraging prompt replies.

To accommodate individual pacing preferences, trial progression occurs either automatically when time elapses or manually when the participant presses the *Next* button. This dual method grants participants agency, ultimately improving engagement and data quality. Additionally, it is not possible to repeat a trial, ensuring that responses remain spontaneous.

ii) Preparatory Phases:

Prior to the main task, a microphone check was conducted to establish proper functionality of the recording equipment, which was critical to avoid data loss or interference with the quality of the speech data. This check was facilitated using the Audio Recording Zone's test mode, specifically designed for such purposes. Initially, a browser notification appears requesting permission to access the microphone, which participants must accept. An image was displayed on a white screen for naming, and participants initiated the recording by pressing *Start Recording*.

Figure 2. *Microphone check for the picture naming task.*



The test mode does not store any data or audio but allows participants to hear a playback of their recording instantly. This step was essential in confirming the quality of recordings before proceeding with the experiment. Additionally, participants could re-record their responses if necessary. By using this setup, potential technical issues, such as distorted audio or silent recordings, were addressed, ensuring the integrity of the data for accurate response analysis.

The task began with a familiarization phase consisting of five example items representative of the stimuli, using the same configurations to ensure consistency in task presentation and response conditions. At the end, participants received feedback by being shown the correct answers, and key instructions were reiterated to reinforce task expectations. This phase helped participants become accustomed to the experimental setup, reducing

response variability and enhancing their comfort. Familiarization is commonly integrated into the picture-naming paradigm (Alario et al., 2004), as task expectations and response formats are clarified, ultimately improving data reliability.

Figure 3. *Familiarization feedback for the picture naming task.*



In online studies, where direct supervision is absent, ensuring participants' attentiveness is essential, as they often overestimate their understanding of instructions and may fail to fully engage with them (Oppenheimer et al., 2009). To mitigate this, two strategically placed attention check questions are used to assess comprehension and encourage careful reading. Before the practice phase, participants were asked, “When does recording start?” and before the main task, “What do you do if you do not know an answer?”, each with two possible answers. Correct answers prompt positive reinforcement, while incorrect responses lead to a review of the instructions, presented in a lighthearted manner with engaging phrasing and imagery.

Figure 4. *Example of an attention question.*



The unexpected nature of these checks helps prevent superficial engagement, ensuring participants remain focused. Additionally, they allow for retrospective assessment of compliance, serving as a valuable measure of data quality. The 93.75% success rate on the second question, following an initial error, demonstrated their effectiveness. Research corroborates that attention checks reduce inattentive responses and improve the validity of experimental data (Berinsky, Margolis, & Sances, 2014). After correctly answering the questions, they were immediately redirected to the testing phase.

Figure 5. *Response to (a) wrong and (b) correct answers to attention questions.*



iii) Task Instructions:

At the beginning of this part of the study, participants were briefly informed that the task aims to assess their speed and accuracy in recognizing and producing words in Greek, with all responses being audio-recorded. Participants were first guided through a technical check to ensure their microphone was functioning properly. They were instructed to allow microphone access to their device, record a test response by naming a displayed image, and listen to playback to confirm audio quality, with the option to re-record if necessary. In the event of technical issues, participants were advised to discontinue the study and inform the researcher.

Following the technical check, participants were informed they would be presented with 60 images for naming in Greek, each preceded by a fixation cross. The image would appear on the screen for 10 seconds, during which the audio recording would begin automatically. Participants were instructed to articulate the corresponding word loudly, clearly, and spontaneously. Progression to the next image occurred automatically once the 10-second limit expired, or manually by pressing the *Next* button. However, participants were cautioned not to skip images by pressing *Next* once the time limit had expired.

The images depicted a range of nouns, including animals, professions, foods, body parts, and tools. The target words could be simple nouns (e.g., paper) or closed compounds (e.g., newspaper), but open compounds (e.g., paper cutter) were excluded, meaning only one-word answers were required. If uncertain, participants were instructed to say ‘pass’. Additional instructions emphasized avoiding fillers (e.g., ‘uhm’) and minimizing background noises (e.g., rustling headphone cables). Participants were also informed that they could change their response within the allotted time by simply saying a new word.

To familiarize themselves with the task, participants were asked to complete five example items and then received the correct answers to confirm their understanding of the task requirements. Key instructions were reiterated at this stage. Before beginning the task, they were warned that clicking *Next* on the final instruction screen would immediately start the experiment. Upon completing all 60 trials, participants received encouraging feedback and had the option to take a short break before proceeding to the next part of the study.

The instructions were designed to optimize data collection and ensure reliable evaluation. Emphasizing clear and spontaneous responses aimed to capture genuine lexical retrieval, providing an authentic assessment of the words participants instinctively produce, which is essential in detecting first language attrition. Allowing response changes within the allotted time offered insight into lexical access and retrieval processes, helping assess accuracy more thoroughly. The inclusion of a ‘pass’ option ensured that missing responses reflected uncertainty rather than accidental skips or technical issues, and that every trial is accounted for. Additionally, warning participants against skipping images prevented unnecessary data loss. Discouraging fillers and minimizing background noise helped maintain audio recording clarity, preventing distortions that could interfere with reaction time measurements and transcription accuracy. To reinforce these guidelines, key instructions were repeated throughout the process, ensuring participants remained consistently aware of the task requirements. These considerations were integral to maintaining high-quality speech data for analysis.

c) Data Collection and Scoring

To evaluate responses in the picture naming task, both reaction times (RT) and accuracy were measured. Specialized software captured precise RTs, while a custom 4-point scoring system assessed varying levels of lexical retrieval success. This combined approach enabled a nuanced analysis of language attrition, considering both response speed and accuracy. Herein, data collection, scoring methodology, and handling of invalid or missing data are outlined

i. Reaction Times:

Each participant's audio data was stored as .weba files on Gorilla, then downloaded and converted to .wav files for further processing. The .wav files were uploaded to Chronset (Roux, Armstrong, & Carreiras, 2016), a computational tool designed to precisely detect speech onset latencies through automated analysis of acoustic signals, ensuring millisecond-level accuracy and minimizing human error. After processing, the results were emailed as a document, with a column for the items and another for the corresponding reaction times. When Chronset failed to produce results or non-verbal sounds (e.g., coughing) were detected, Praat (Boersma & Weenink, 1992) was used as a fallback method. Praat, a widely used software for speech and acoustic analysis, allows for manual measurement of speech onset latency by visually inspecting waveforms and spectrograms, ensuring accurate RT data collection. Therefore, the RTs for each participant and item were processed manually in those cases. Although more time-consuming, Praat offers high precision, complementing Chronset's automated process.

ii. Accuracy:

To assess PNT accuracy, a custom 4-point scoring system (0–3) was developed to capture subtle differences in lexical retrieval success (Bates et al., 2003). Unlike a binary correct/incorrect classification, this graded scale allows for a more fine-grained assessment of lexical access in language attrition, including approximations and self-repairs that binary measures would overlook. The scoring system was developed after an initial analysis of participant responses, which revealed recurring patterns such as internationalizations, lexical approximations, self-corrections, and possible misrecognitions. By incorporating these factors into the scoring criteria, the system ensures that the classification of responses reflects the range of lexical retrieval difficulties observed in the dataset. This systematic approach ensured consistency in scoring while maintaining transparency and replicability. Furthermore, since reaction times inherently capture retrieval delays and hesitations, these were not factored into the accuracy scoring. Below is a breakdown of the criteria used to assign each score, providing a detailed framework for evaluating participants' response accuracy.

- **0 points - Incorrect response:** This category captures complete lexical retrieval failures, which may include non-responses (e.g., ‘pass’ or no answer), unrelated answers (e.g., ‘letter box’ instead of ‘ballot’), or internationalizations (e.g., ‘computer’ instead of the Greek equivalent).
- **1 point - Partially inaccurate response:** This includes lexical approximations, such as oversimplifications (e.g., ‘athlete’ instead of ‘football player’) or semantic inaccuracies (e.g., ‘bird’ instead of ‘seagull’).
- **2 points - Partially accurate response:** This score includes self-corrections, near-synonyms, and minor perceptual errors as a result of naming an item with a slightly inaccurate label (e.g., ‘jacket’ instead of ‘coat’).
- **3 points - Fully accurate response:** This represents target responses or exact synonyms given without repair and overspecifications (e.g., ‘mountaintop’ instead of ‘mountain’).

The audio recordings for every trial of each participant were reviewed and transcribed, and then individually evaluated based on the 4-point scale. After scoring all trials, each participant’s total accuracy score was calculated by summing their individual trial scores.

3.3.2.2. *Aptitude Task*

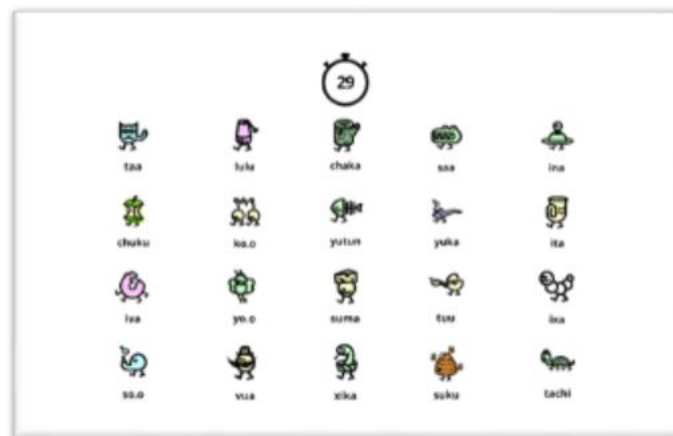
To measure the independent variable, namely the participants' language aptitude, the increasingly cited LLAMA test battery was employed (Meara & Rogers, 2020; Meara, 2005). Although the LLAMA tests have not yet been standardized, they have demonstrated resilience against various factors, such as being language-neutral (i.e., performance is not influenced by one's first language) and independent of gender, education, or logic training (e.g., chess) (Rogers et al., 2017). However, it is worth noting that bilinguals generally outperform monolinguals, and adults tend to score better than children. Moreover, second language instructed learners tend to perform better on this subset, as they have developed vocabulary learning strategies (Rogers et al., 2017). These elements do not impact the present study, as it focuses exclusively on bilingual adults.

For this research, participants were administered the LLAMA_B test, a subcomponent specifically designed to assess vocabulary learning abilities. While the test was adapted for this study by transferring it onto the Gorilla platform, all other parameters remained unchanged. The LLAMA_B test consists of 20 items in which participants are required to pair unfamiliar

names with unfamiliar objects. The test includes a familiarization phase followed by a testing phase.

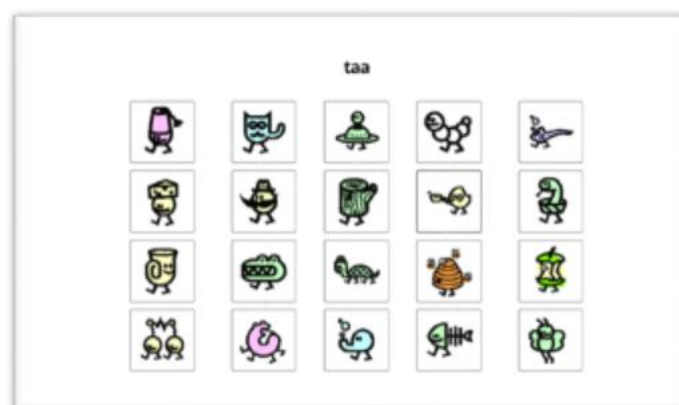
In the familiarization phase, participants have 2 minutes (i.e., 120,000 ms) to study and memorize the names associated with various objects. All objects and their corresponding names are displayed simultaneously on the screen. A countdown timer appears at the top of the screen during the last 30 seconds (30,000 ms) to signal the end of the familiarization phase.

Figure 6. *Familiarization phase of the aptitude task.*



During the testing phase, a single word is displayed at the top of the screen, and participants must select the correct object from the options before automatically advancing to the next screen. The test items were presented in a fixed order for all participants. There is no time limit for this part of the test, and once an answer is submitted, it cannot be revised.

Figure 7. *Example of a test trial from the aptitude task.*



At the beginning of the task, participants were informed that the aim was to explore their ability to learn new words in an unfamiliar language. They were told that the task involves an artificial language and unknown objects, and that it consists of two parts: a practice phase and a test phase. In the first phase, lasting two minutes, they were instructed to study and

memorize the associations between the words and the images. In the second phase, they were told they would be tested on their memory and that they must select the correct object corresponding to a given word.

3.3.2.3. Proficiency Tasks

To establish a baseline for participants' language knowledge in both languages, beyond their self-reported proficiency, a standardized proficiency test was employed. The DIALANG system (Alderson & Huhta, 2005) is an online language assessment tool available in 14 European languages, including German and Greek. For this study's purposes, only the DIALANG placement test was adapted and utilized in both languages. This test measures users' vocabulary size and operates similarly to a lexical decision task, where participants must determine whether each stimulus is a valid word in the given language.

Figure 8. *Example of a test trial from the L2 proficiency task.*



The proficiency task administered in this study comprises 75 items per language, consisting of 50 words and 25 nonwords. Participants are presented with one item at a time on the screen and must determine whether each item is a word in the target language. Responses are indicated by clicking a green button for words and a red button for nonwords. To ensure the reliability of responses and prevent the use of external resources (e.g., online searches), a time limit of 10,000 milliseconds (10 seconds) per item was set, with a countdown timer displayed to remind participants of the remaining time. Once a response is submitted, the test automatically transitions to the next item, and responses cannot be revised.

Before beginning the task, participants were informed of its purpose to assess their lexical knowledge in German and Greek. They were instructed to decide whether the presented

items were actual words in the respective language by clicking the appropriate response button, with 10 seconds per item.

3.3.2.4. Questionnaire

The purpose of the questionnaire is to gather general demographic information of the participants and to assess individual differences associated with first language attrition. Previous research highlights the importance of such contextual factors in studies of language attrition, as they contribute to a deeper understanding of how personal experience shapes language maintenance and decline. This questionnaire was specifically developed for this study, drawing inspiration from existing sociolinguistic instruments used in attrition research (Schmid & Dusseldorp, 2010; Keijzer, 2007), and was organized into four sections: demographic characteristics, linguistic background, frequency and type of use, and individual attitudes. A combination of open-ended questions, predefined closed-ended, numeric ranges and rating scales was used for data collection. Participants were informed that the questionnaire explored their personal background, language use, and attitudes. It consisted of 40 questions focused exclusively on their own experiences, with no right or wrong answers. Honest and spontaneous responses were encouraged, and participants were free to revisit their answers without any time restrictions.

In line with the study's aims, the questionnaire was organized into four structured sections, each targeting a specific aspect of the participants' linguistic profile and experience:

- **Category A - Demographic information:** The first section gathered demographic and biographical data relevant to the study's inclusion criteria and research aims. Participants reported their gender, age, native language, self-reported proficiency in the second language, use of additional languages in daily life, place of birth, current place of residence, highest level of education, occupation, and any diagnosed speech or learning disorders. It consisted of 12 items, and data were collected using a combination of predetermined nominal scales (e.g., gender, native language), ordinal scales (e.g., education level, self-rated proficiency), and open-ended questions (e.g., occupation, other languages).
- **Category B - Acquisition and exposure:** This category investigated participants' language learning trajectory. It comprised three items focused on the age of acquisition, the amount of time spent in relevant countries, and of the formal instruction received in

each language. Responses were collected using a predefined ratio scale with values ranging from 0 to 18+ years.

- **Category C - Language use:** The third section assessed participants' frequency and type of language use across a range of contexts for their first (L1), second (L2), and, optionally, a regularly used third language (L3). It was divided into three thematic sets. Set 1 focused on social and professional interactions, including use with family, a partner or at work. For L1, two additional items were included to distinguish between usage in the home country and in the country of residence, resulting in 7 items for L1, compared with 5 for the other languages. Set 2 contained 4 items targeting internal cognitive use, such as counting, dreaming, cursing, and self-talk, domains often associated with subconscious or automatic language processing. Set 3 also consisted of 4 items, addressing receptive and productive language skills across modalities, such as listening to music, reading books and using social media. Responses were collected on a 10-point rating scale (1 = never, 10 = always).
- **Category D - Motivation and attitude:** The final section explored participants' attitudes toward their languages, cultural identity, and future intentions. It focused on their desire to sound native-like, views on the importance of language for future generations, and attitudes toward foreign languages in general. Participants rated their agreement with 9 statements on a 10-point scale (1 = strongly disagree, 10 = strongly agree).

3.4. Data Preparation

3.4.1. *Aptitude, Proficiency, and Questionnaire Data*

Raw scores for both the aptitude and proficiency measures were automatically extracted from the Gorilla platform. Each trial was binary-coded as correct (1 point) or incorrect (0 points), resulting in a total maximum score of 20 points for the aptitude test and 75 points for each proficiency test. The data contained no missing responses. To facilitate comparability across participants and tasks, all aggregate scores were converted to percentages prior to statistical analysis.

The questionnaire had no missing values, as it was designed to prevent submission unless all mandatory items were completed. Data from non-mandatory questions, such as those

on the potential use of an L3, were documented descriptively but excluded from the statistical analysis, since they were considered peripheral to the study's core research questions. As the questionnaire focused on personal experiences and opinions, no automatic internal-consistency checks were implemented, and no statistical outliers were identified.

The questionnaire elicited nominal, ordinal, and ratio data. To manage the dataset effectively, two processing strategies were applied based on the nature of the questions. For Categories A and B, raw ratio-scale data (e.g., age) were processed directly, whereas open-ended responses (e.g., occupation, other languages spoken) were reported verbatim for descriptive purposes only. Nominal data were organized categorically and coded for statistical analysis. For example, the host countries were encoded as 0 = Austria, 1 = Germany, 2 = Switzerland (see Appendix C2). For Categories C and D, a different approach was taken to support interpretability. In some cases, aggregating responses across related items created composite scores. Seven constructs were identified: frequency, orientation, judgment, affinity, retrieval, migration, and repatriation. For each construct, thematically similar items were grouped, and their sums were converted to percentages to produce a single interpretable metric. Separate metrics were calculated for each language where applicable. This approach assumes equal item weighting within each group and enables consistent comparisons across participants and languages, as well as uniformity across measures.

Table 1. *Questionnaire composite metrics.*

Construct	Source	Description	Metrics
Frequency	C (all items)	Language use across diverse contexts	2 (L1/ L2)
Orientation	D (3 items)	Cultural identity and intergenerational transmission	2 (L1/ L2)
Judgement	D (1 item)	Perceived change in language proficiency	2 (L1/ L2)
Affinity	D (2 items)	Fondness of languages and self-perceived talent	1
Retrieval	D (1 item)	Frustration over lexical retrieval difficulty in L1	1
Migration	D (1 item)	Evaluation of migration decision	1
Repatriation	D (1 item)	Intention to return to home country	1

3.4.2. Attrition Data

To ensure data integrity for the picture naming task, specific criteria for handling missing data and excluding trials were applied. Trials with missing audio files or missing

responses were removed from both accuracy and reaction time. Trials with incorrect responses, including ‘pass’ responses, were excluded from RT analyses since their latencies do not reflect cognitive processes and would otherwise distort interpretation. These trials were nevertheless retained for accuracy scoring to capture overall retrieval success.

To improve stimulus reliability, an item-level inspection was conducted to identify non-canonical or poorly performing stimuli. Accuracy served as the primary exclusion criterion, as it directly indicates stimulus malfunction, whereas slow RTs typically reflect genuine processing difficulty. Six items (i.e., 15, 22, 33, 37, 42, 56) showed consistently low accuracy across participants and were removed to ensure data quality. Excluding 10 percent of the original 60 items struck a balance between maintaining data quality and preserving statistical power, resulting in a sufficiently large and reliable item set for analysis.

After data cleaning of the picture naming task outputs, its two metrics were prepared for analysis using standard procedures. Reaction times were recorded in milliseconds, and the mean RTs of all valid trials for each participant were calculated to provide an interpretable estimate of processing speed. The mean was chosen because it preserves natural variability in RTs and is compatible with parametric analysis. Retaining RT outliers was appropriate in this context, since extreme latencies can be highly informative rather than spurious errors in attrition research. For accuracy, the maximum attainable raw score was 180 points (60 items $\times$ 3 points). Accuracy was computed as the proportion of the obtained score relative to the maximum possible score for non-missing trials only, and subsequently converted to percentages to ensure comparability across variables. This process yielded the RT and accuracy measures used in the statistical analysis.

To derive a unified measure of lexical retrieval efficiency, accuracy and speed were integrated into a composite labeled Maintenance Index. This index was created by standardizing both metrics using *z*-scores, with RT *z*-scores inverted so that higher values consistently reflect better performance. This approach is particularly relevant in L1-attrition research, where accuracy ceiling effects often obscure subtle processing differences, as it captures the characteristic speed–accuracy dynamics of bilingual lexical access. The resulting composite, therefore, provides a more comprehensive indicator of L1 retrieval efficiency than either metric alone.

4. Results

This chapter reports the study's statistical results examining individual differences in first language maintenance among late bilinguals, with a primary focus on the role of language aptitude. The statistical analyses draw on behavioral performance in linguistic tasks and questionnaire-based measures capturing linguistic experience and sociolinguistic background. All analyses were conducted using JASP (version 0.95.3), an open-source statistical software package that provides an accessible interface for conducting classical statistical analyses (JASP Team, 2025). Results are presented descriptively and inferentially, without interpretive discussion, and statistical assumptions and effect sizes are reported where appropriate.

Section 4.1 presents descriptive statistics and distributional characteristics for the sample and key variables, including demographic characteristics and questionnaire-based measures, in order to provide an overview of the data and assess their suitability for inferential analyses. Section 4.2 reports the inferential analyses. First, confirmatory analyses examine the primary hypothesis that language aptitude predicts L1 lexical maintenance using regression-based approaches. Second, exploratory analyses examine correlational associations with sociolinguistic and extralinguistic variables to identify patterns of individual variability in L1 retention outcomes.

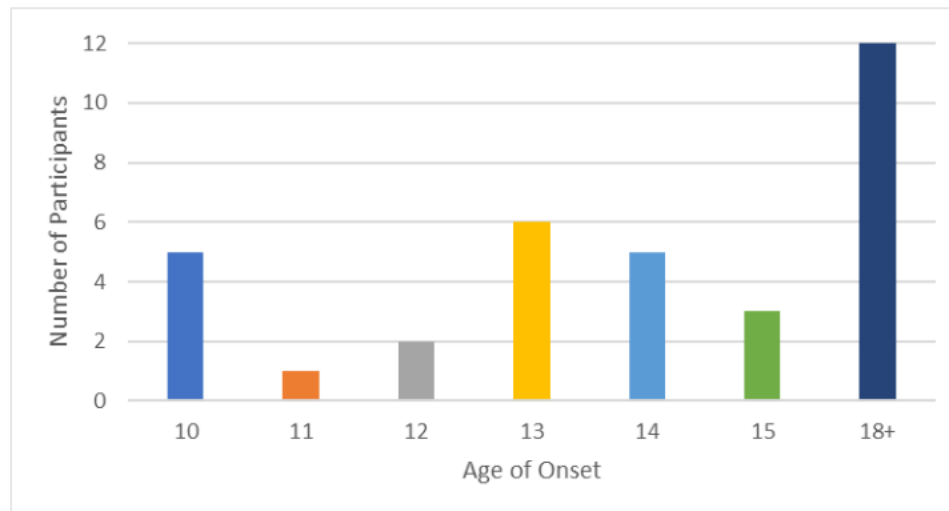
4.1. Descriptive Statistics

4.1.1. Demographic Data

Thirty-four ($N = 34$; 22 female) Greek sequential bilinguals residing in German-speaking countries (see Appendix Table C2.1) participated in the study. Participants' ages ranged from 21,4 to 61,10 ($M = 36,07$, $SD = 10.08$). All participants were born in Greece and received their primary education there, being raised monolingually in Greek. They began learning German as a foreign language at the earliest age of 10, while many did not start until adulthood ($N = 12$)¹.

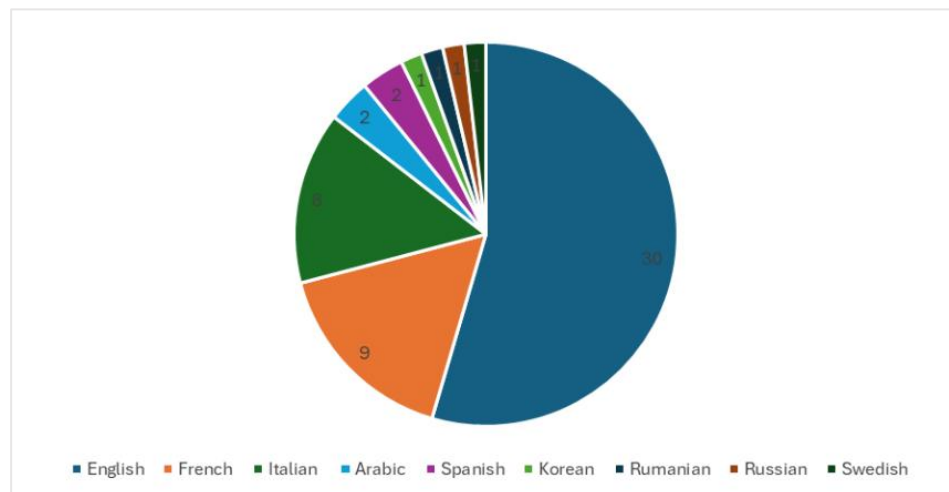
¹ Age of Acquisition (AoA) was collected as a categorical variable (0–18+) rather than as a continuous measure, which prevents the calculation of descriptive statistics such as mean or standard deviation. Full details are provided in the appendix.

Figure 9. *Age of onset of L2 learning.*



At the time of testing, participants self-reported their German proficiency as at least level B2 according to the CEFR (Council of Europe, 2001), with the majority ($N = 19$) claiming a C1 level. Additionally, nearly half of the participants ($N = 20$) reported using a third language daily (see Appendix Table C2.2), while even more ($N = 24$) reported speaking an additional language (see Figure 10). However, their level of proficiency in these additional languages was not verified.

Figure 10. *Languages spoken by the participants.*

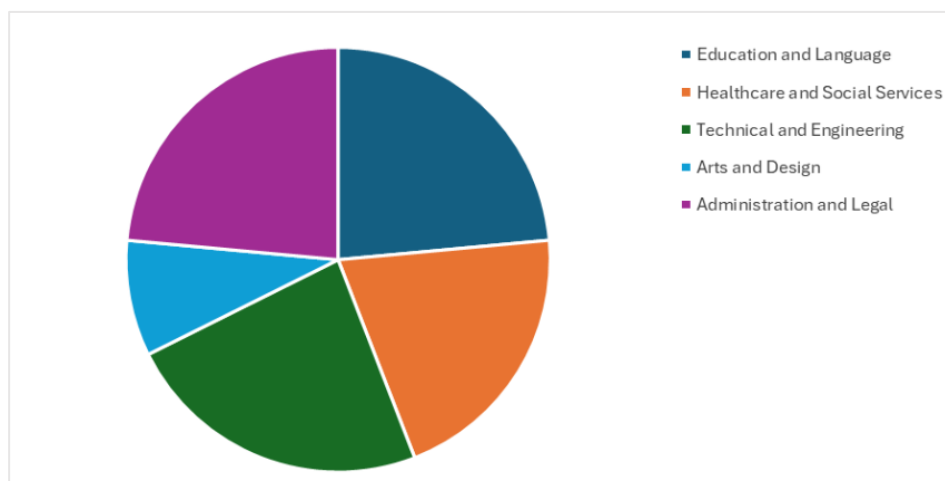


The participants reside in Austria ($N = 17$), Germany ($N = 16$), and Switzerland ($N = 1$), while some ($N = 6$) reported having lived in another country for at least 6 months in the past. Length of residence (LoR)² ranged from 5 to over 18 years, with most participants having lived

² Length of Residence (LoR) was collected as a categorical variable (0–18+) rather than as a continuous measure, which prevents the calculation of descriptive statistics such as mean or standard deviation. Full details are provided in the appendix.

in a German-speaking context for between 7 and 12 years. Furthermore, the participants were highly educated, with the vast majority ($N = 29$) having completed or currently attending higher education, almost half of whom were at the master's degree level.

Figure 11. *Distribution of participants by profession or field of study.*



All in all, the data reflect a relatively homogeneous and highly educated sample of late Greek–German bilinguals with advanced L2 proficiency and substantial L2 immersion in a German-speaking environment, making them suitable for investigating L1 maintenance outcomes.

4.1.2. *Variable Overview*

This section provides an overview of the key linguistic and sociolinguistic variables collected in the study. Measures are presented together with their descriptive statistics (mean, standard deviation, median, range) and corresponding boxplots to illustrate distributional patterns across participants. These summaries serve to contextualize the subsequent inferential analyses.

a) Main Variables

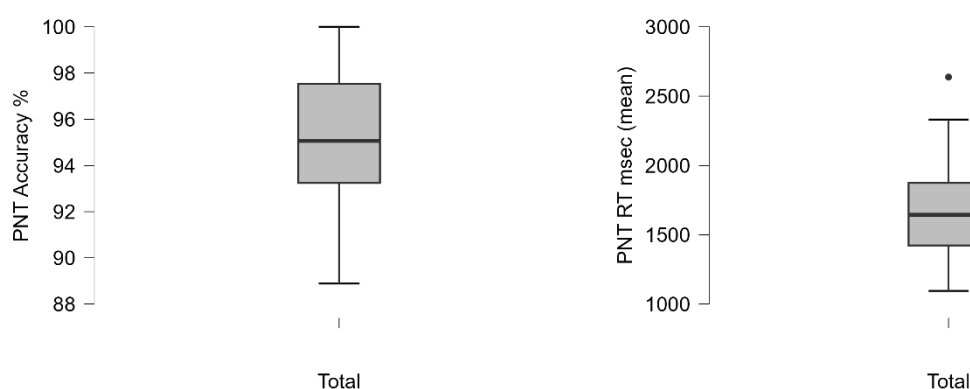
Descriptive analyses were performed for the picture naming task accuracy in percentages and reaction times (RT) in milliseconds. Participants demonstrated high accuracy overall ($M = 95.40\%$, $SD = 2.99$), indicating consistently correct performance across items.

Reaction times were more variable ($M = 1660$ ms, $SD = 367$), reflecting individual differences in response speed. Both measures met the assumption of normality, as indicated by the Shapiro–Wilk tests (Acc: $p = .225$; RT: $p = .337$).

Table 2. Descriptive statistics for accuracy (%) and reaction times (ms) ($N = 34$).

Variable	M	SD	Median	Range
Accuracy	95.40	2.98	95.06	88.89–100.00
Reaction Times	1660	367.0	1644	1095–2638

Figure 12. Boxplots of (a) accuracy (%) and (b) reaction times (ms).

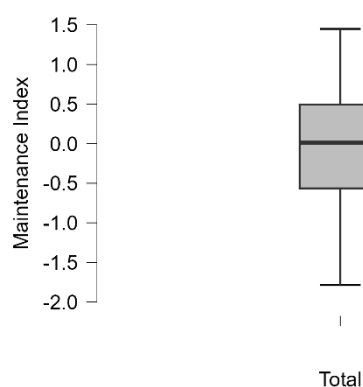


The composite maintenance score, combining standardized accuracy and RT values, showed good distributional properties ($M = 0$, $SD = 0.78$). Both skewness (0.07) and kurtosis (−0.20) indicate approximate normality, confirmed by the Shapiro–Wilk test ($p = .751$).

Table 3. Descriptive statistics for maintenance index (z-score) ($N = 34$).

Variable	M	SD	Median	Range
Maintenance	0.00	0.78	0.01	−1.78–1.45

Figure 13. Boxplot of maintenance index.

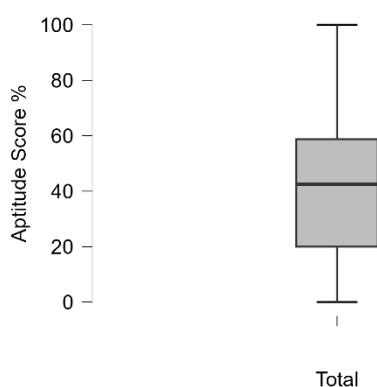


The aptitude scores ranged widely (0–100%), with a mean of 43.53% ($SD = 26.53$), indicating substantial variability in participants' aptitude levels. The distribution was approximately normal ($skewness = 0.38$, $kurtosis = -0.50$), supported by the non-significant Shapiro–Wilk test ($p = .256$). Overall, the data suggest a broad spread of performance with no major deviations from normality, suitable for parametric analyses.

Table 4. *Descriptive statistics for aptitude ($N = 34$).*

Variable	M	SD	Median	Range
Aptitude	43.53	26.53	42.50	0.00–100.00

Figure 14. *Boxplot of aptitude scores.*



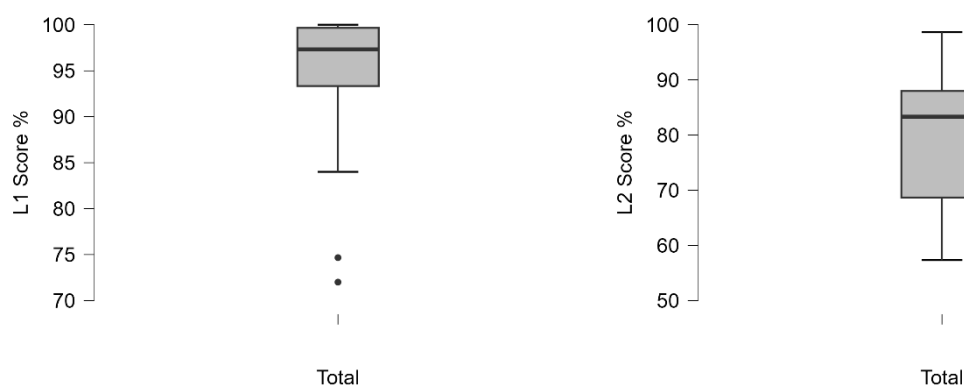
b) Control Variables

In the L1 proficiency test, very high accuracy was consistently achieved ($M = 94.67\%$, $SD = 7.07$), whereas the L2 performance was significantly lower and showed greater variability ($M = 79.80\%$, $SD = 11.36$), reflecting substantial individual differences in L2 proficiency.

Table 5. *Descriptive statistics for L1 and L2 accuracy ($N = 34$).*

Variable	M	SD	Median	Range
L1 Accuracy	94.67	7.07	97.33	72.00–100.00
L2 Accuracy	79.80	11.36	83.33	57.33–98.67

Figure 15. Boxplots of (a) L1 and (b) L2 accuracy scores.

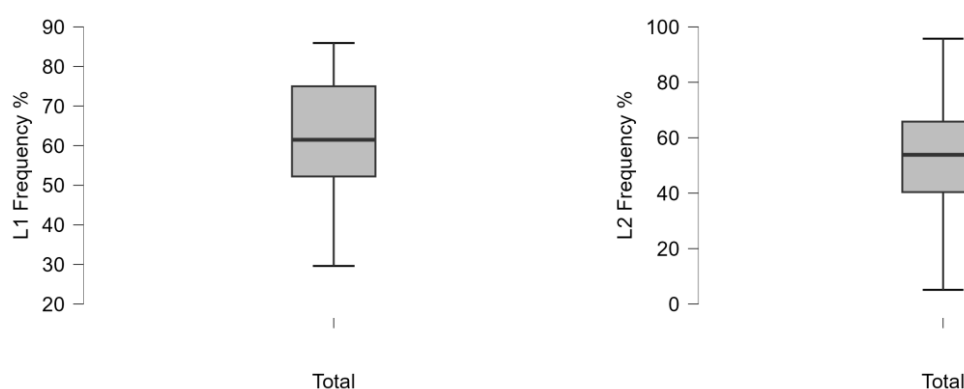


Participants reported more frequent use of L1 ($M = 62.46\%$, $SD = 14.28$) than L2 ($M = 52.69\%$, $SD = 20.48$). L1 use was relatively stable across participants, whereas L2 use varied more widely, suggesting differing degrees of engagement with the second language in daily life.

Table 6. Descriptive statistics for L1 and L2 frequency of use ($N = 34$).

Variable	M	SD	Median	Range
L1 Frequency	62.46	14.28	61.48	29.62–85.92
L2 Frequency	52.69	20.48	53.58	5.12–95.73

Figure 16. Boxplots of (a) L1 and (b) L2 frequency of use.

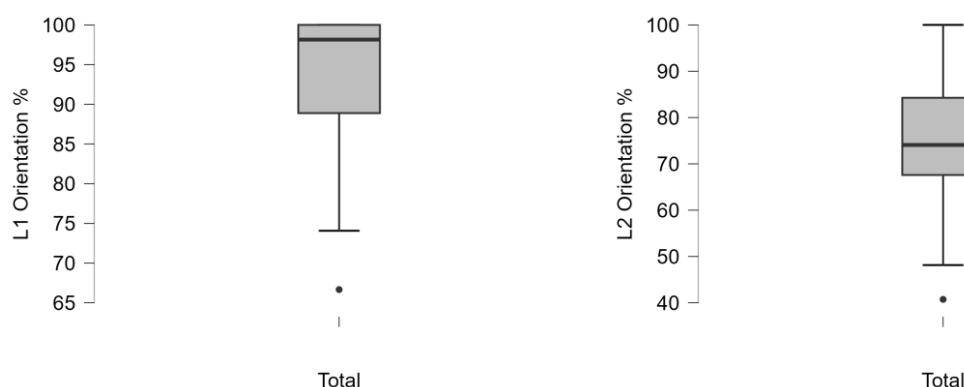


A stronger sense of belonging and identification with their L1 community ($M = 92.92\%$, $SD = 9.18$) than with the L2 community ($M = 74.18\%$, $SD = 14.02$) was reported. While most participants expressed high L1 orientation, L2 orientation varied more widely, indicating individual differences in cultural affiliation with the L2 context.

Table 7. Descriptive statistics for L1 and L2 orientation ($N = 34$).

Variable	M	SD	Median	Range
L1 Orientation	92.92	9.18	98.15	66.67–100.00
L2 Orientation	74.18	14.02	74.07	40.74–100.00

Figure 17. Boxplots of (a) L1 and (b) L2 orientation.

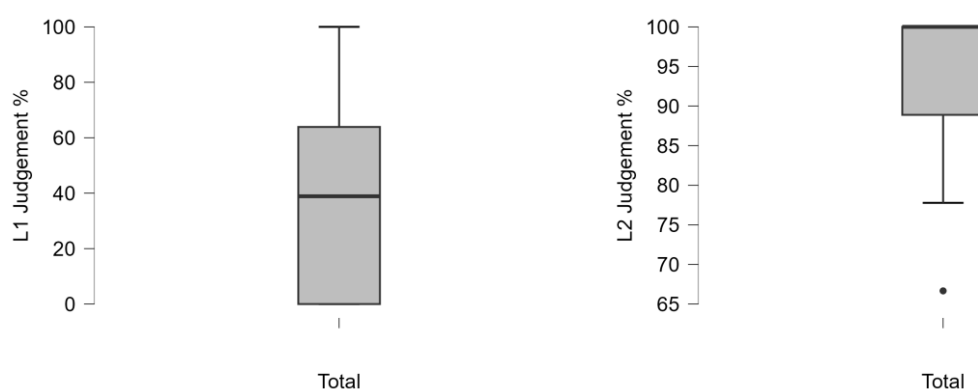


Participants overwhelmingly judged their L2 proficiency to have increased since relocation ($M = 92.48\%$, $SD = 12.15$). In contrast, the mean for L1 judgment was very low ($M = 36.60\%$, $SD = 34.66$), indicating that many participants may implicitly acknowledge some decline.

Table 8. Descriptive statistics for L1 and L2 judgement ($N = 34$).

Variable	M	SD	Median	Range
L1 Judgement	36.60	34.66	38.89	0.00–100.00
L2 Judgement	92.48	12.15	100.00	66.67–100.00

Figure 18. Boxplots of (a) L1 and (b) L2 judgement.

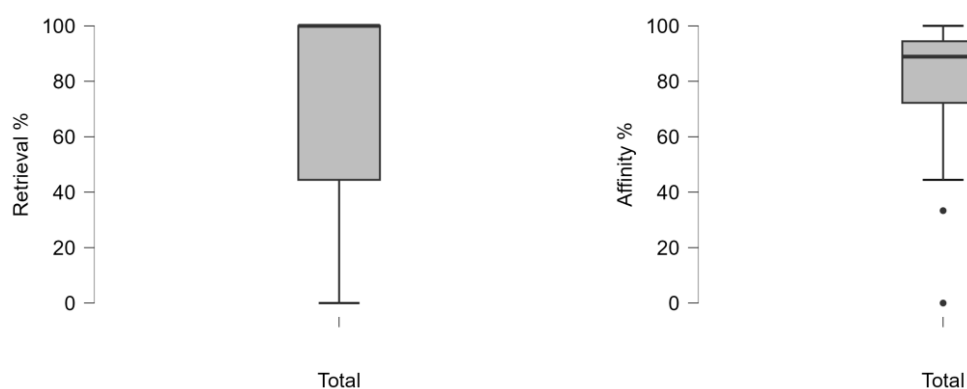


Participants largely admitted feeling frustrated when unable to recall a word in their L1 ($M = 77.12\%$, $SD = 31.66$), though responses showed considerable variability. Reported language affinity was similarly high ($M = 79.90\%$, $SD = 22.31$), reflecting generally positive attitudes toward languages and confidence in their own language learning ability.

Table 9. *Descriptive statistics for retrieval and affinity ($N = 34$).*

Variable	M	SD	Median	Range
Retrieval	77.12	31.66	100.00	0.00–100.00
Affinity	79.90	22.31	88.89	0.00–100.00

Figure 19. *Boxplots of (a) retrieval and (b) affinity.*

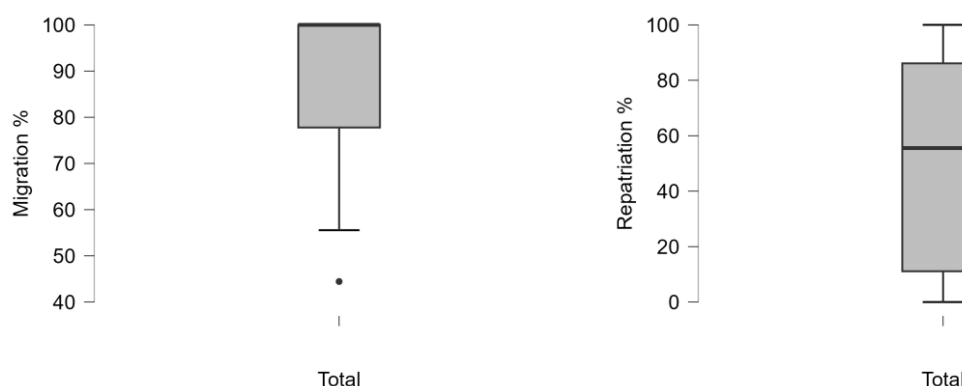


The decision to migrate was almost unanimously supported ($M = 90.20\%$, $SD = 14.41$), indicating overall satisfaction. In contrast, responses regarding repatriation were highly variable ($M = 50.98\%$, $SD = 39.18$), reflecting diverse personal attitudes toward returning to the home country.

Table 10. *Descriptive statistics for migration and repatriation ($N = 34$).*

Variable	M	SD	Median	Range
Migration	90.20	14.41	100.00	44.44–100.00
Repatriation	50.98	39.18	55.56	0.00–100.00

Figure 20. *Boxplots of (a) migration and (b) repatriation.*



Overall, the descriptive results demonstrated sufficient variability and acceptable distributional properties across the key variables, thereby justifying the use of parametric and non-parametric analyses. The following section examines whether language aptitude predicts L1 maintenance, followed by exploratory analyses involving additional sociolinguistic factors.

4.2. Inferential Statistics

4.2.1. Confirmatory Analysis

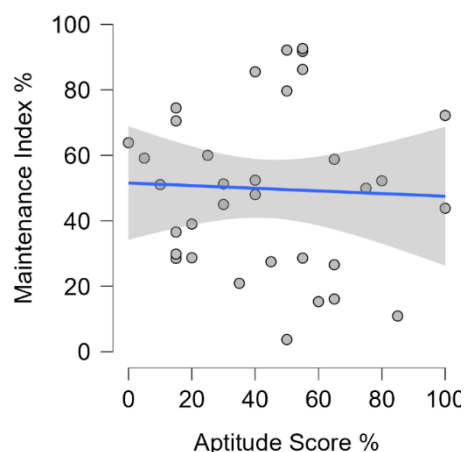
This section presents the confirmatory analysis, testing the central hypothesis of this study (H1), namely that individual language aptitude predicts the degree of first-language (L1) maintenance in late bilinguals. To evaluate this, a series of regression models was conducted using the composite Maintenance Index and its component measures. Statistical significance was assessed at an alpha level of $p < .05$.

A simple linear regression tested whether aptitude predicted overall L1 maintenance. The model was not significant, $F(1, 32) = 0.06$, $p = .811$, explaining negligible variance ($R^2 = .002$). The regression coefficient was negative ($\beta = -0.04$), indicating no meaningful relationship between aptitude and maintenance.

Table 11. *Linear regression predicting L1 maintenance index from aptitude score (%).*

Variable	B	SE	t	p	R ²
Maintenance	-0.04	0.17	-0.24	.811	.002

Figure 21. Scatterplot of (a) aptitude score and (b) maintenance index (%).



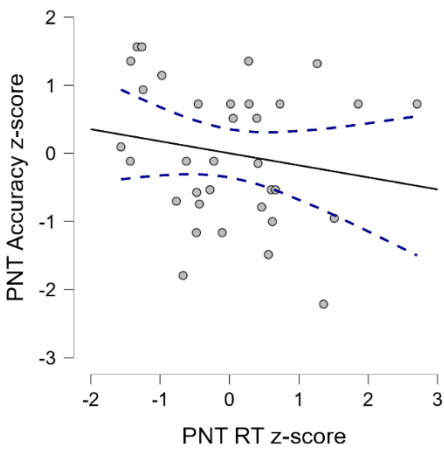
Since the composite L1 Maintenance Index did not yield a significant relationship with aptitude, a post hoc Pearson correlation was conducted to assess whether the composite itself was conceptually justified. This correlation tested the association between PNT standardized accuracy and reaction times (z-scores) to determine whether both measures reflected related underlying processes of L1 lexical access.

The correlation was weak and nonsignificant ($r = -.18, p = .32$), indicating that accuracy and response speed did not co-vary meaningfully across participants. This suggests that, although the index was statistically well-behaved, its interpretive validity as a combined maintenance measure is limited and likely captures distinct aspects of lexical access performance.

Table 12. Correlation between accuracy and reaction times (z-scores).

Variables	Pearson's r	p
Accuracy - RT	– 0.18	.318

Figure 22. Scatterplot of (a) accuracy and (b) reaction time (z-scores).



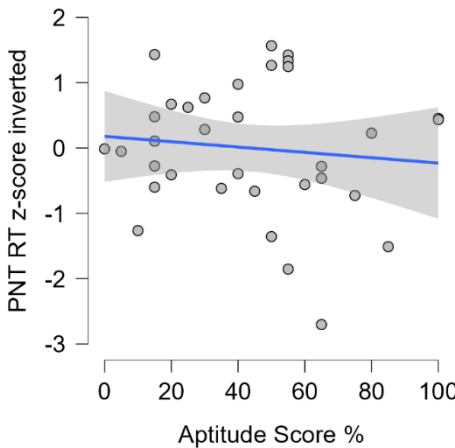
Follow-up analyses were conducted on the component measures separately, i.e., reaction times and accuracy, to explore whether aptitude relates differentially to speed- and accuracy-based aspects of L1 performance.

A simple linear regression tested whether aptitude predicted reaction times, which were standardized and inverted so that higher values indicate faster responses, aligning interpretability with accuracy and aptitude scores. The model was not significant, $F(1, 32) = 0.37, p = .547$, explaining only 1.1% of the variance ($R^2 = .011$). The regression coefficient for aptitude was small and nonsignificant ($B = -0.00, SE = 0.01, p = .547$), indicating no meaningful relationship between aptitude and response speed.

Table 13. Linear regression predicting reaction times (z-score) from aptitude score (%).

Variable	B	SE	t	p	R ²
RT	-0.00	0.01	-0.61	.547	.011

Figure 23. Scatterplot of (a) aptitude score (%) and (b) reaction times (z-score).

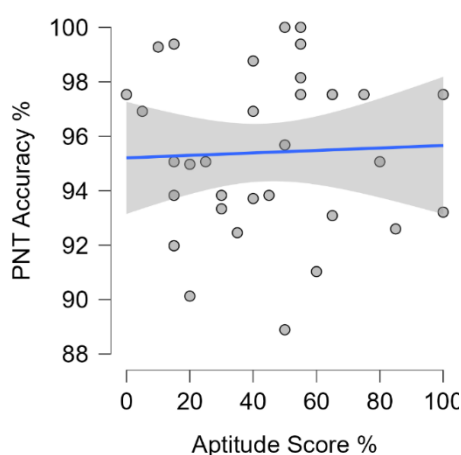


Finally, a regression analysis tested whether aptitude predicted accuracy. The result was not significant, $F(1, 32) = 0.05$, $p = .821$, explaining only 0.2% of the variance ($R^2 = .002$). The regression coefficient was close to zero and positive ($B = 0.01$, $SE = 0.02$, $p = .821$), indicating a negligible and nonsignificant trend in the expected direction, namely higher aptitude associated with slightly higher accuracy.

Table 14. *Linear regression predicting accuracy from aptitude score (%).*

Variable	B	SE	t	p	R ²
Accuracy	0.01	0.02	-0.23	.821	.002

Figure 24. *Scatterplot of (a) aptitude score and (b) accuracy (%).*



Across all three regression models (maintenance, reaction time, accuracy), aptitude did not significantly predict any aspect of L1 performance. As no confirmatory effects were observed, exploratory analyses were subsequently conducted to identify potential patterns that may inform future hypotheses or account for the observed variability.

4.2.2. *Exploratory Analysis*

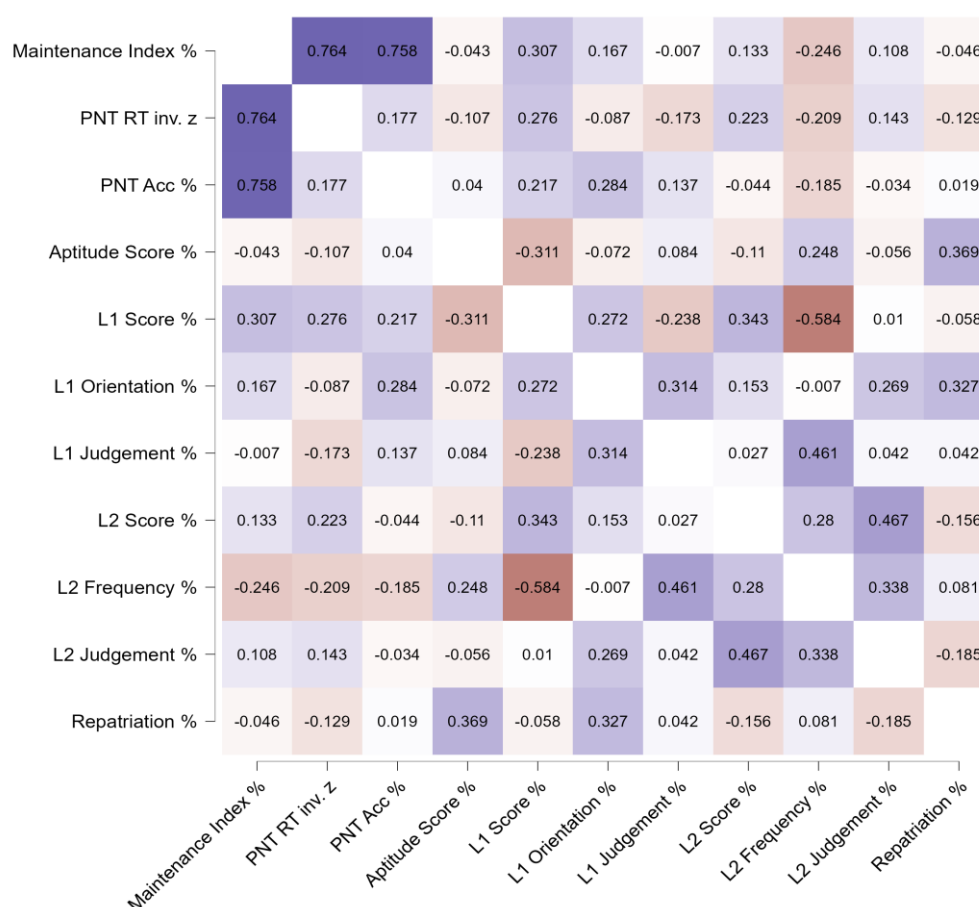
a) Cross-Variable Relationships

As part of the exploratory analyses, a full correlation matrix across variables served as a diagnostic tool to screen for potential relationships and patterns (see Appendix Figure C3.1). To improve interpretability, a reduced heatmap (see Figure 25) was then generated, displaying only correlations of approximately $|r| \geq .30$, a common threshold for highlighting moderate and potentially meaningful associations while suppressing visual noise (Gignac & Szodorai, 2016).

The Maintenance Index showed the expected strong associations with its components, correlating positively with PNT Accuracy ($r = .76, p < .001$) and with PNT RT (inverted; $r = .76, p < .001$), confirming internal consistency. Beyond these task-internal relationships, several moderate cross-variable correlations emerged.

The strongest correlations involved interactions between L1 proficiency and L2 usage. L1 Score showed a substantial negative association with L2 Frequency ($r = -.58, p < .001$), indicating that participants who used their L2 more frequently tended to obtain lower scores on the L1 proficiency measure. A related pattern emerged for self-rated L1 ability, with L1 Judgement correlating positively with L2 Frequency ($r = .46, p = .006$). Within the L2 measures themselves, objective and self-reported proficiency aligned well, as reflected in the correlation between L2 Score and L2 Judgement ($r = .47, p = .005$).

Figure 25. Heatmap of significant correlations.



A number of moderate trends were also observed. L2 Judgement was positively associated with L2 Frequency ($r = .34, p = .051$), and L1 Score showed a comparable association with L2 Score ($r = .34, p = .047$). Aptitude Score correlated moderately with

Repatriation attitude ($r = .37, p = .032$). L1 Orientation displayed two borderline associations: one with L1 Judgement ($r = .31, p = .071$) and one with Repatriation ($r = .33, p = .059$). Finally, several smaller but still notable trends appeared just below the moderate range. Some tendencies were visible between L1 Score and PNT RT inverted ($r = .28, p = .115$) and between PNT Accuracy and L1 Orientation ($r = .28, p = .104$). L2 Score showed a weak positive correlation with L2 Frequency ($r = .28, p = .109$), and L1 Orientation showed a similar trend with L2 Judgement ($r = .27, p = .124$). Aptitude Score displayed a small positive association with L2 Frequency ($r = .25, p = .157$). All remaining correlations were below $|r| = .25$ and thus statistically nonsignificant, suggesting no further consistent relationships among the variables.

b) Background Factors

To explore whether general demographic variables contributed to individual differences in L1 maintenance, several background factors were examined using non-parametric analyses. A Spearman correlation was used to test the association between age and the maintenance index. The result showed no meaningful relationship ($\rho = .12, p = .51$), indicating that age did not predict L1 maintenance in this sample.

Table 15. *Spearman's correlations between age and maintenance index (%).*

Variable	Spearman's rho	p
Age	0.118	.506

To facilitate non-parametric analysis, the AoA and LoR variables were converted into two conceptually distinct, dichotomous groups: 'early onset' (≤ 14) vs. 'late onset' (≥ 15) for AoA, and 'shorter residence' (≤ 9) vs. 'longer residence' (≥ 10) for LoR. A Mann–Whitney U test comparing the two AoA groups showed no difference in maintenance scores ($U = 117.0, p = .39; r = .18$). Likewise, the comparison between the two LoR groups yielded no effect ($U = 140.0, p = .95; r = .02$). Both results indicate that neither AoA nor LoR contributed to L1 maintenance outcomes.

Table 16. *Comparison of maintenance index across AoA and LoR groups.*

Variable	U	p	r
AoA	117.0	.391	0.18
LoR	140.0	.945	0.02

A Kruskal–Wallis test was used to assess whether maintenance scores differed across education levels. The result was non-significant ($H = 0.61$, $p = .90$), suggesting that education did not contribute to variation in L1 maintenance.

Table 17. *Kruskal–Wallis test for education and maintenance index (%).*

Variable	Statistic	df	p
Education	0.605	3	.895

c) Group Comparisons

To further investigate individual variability in L1 maintenance, additional exploratory analyses were conducted focusing on performance-based exploratory group comparisons. The aim was to examine whether potential non-linear patterns might emerge when isolating and comparing individuals with extreme performances, rather than across the entire sample. This approach reduces noise from mid-range performers and allows clearer inspection of possible group differences.

Participants were first grouped according to their lexical retrieval performance. A bivariate median-split was applied to standardized scores for accuracy and inverse reaction time to identify profiles with consistently strong or weak retrieval performance. Participants who scored above the sample median on both measures formed the High Maintenance group ($N = 7$), whereas those who scored below both medians were classified as Low Maintenance ($N = 8$). A Mann–Whitney U test comparing aptitude scores between these two groups showed a directional trend, with slightly higher aptitude scores on average in the High Maintenance group. This was a small, non-significant difference ($U = 21.00$, $p = .449$; $r = .25$), indicating no meaningful group effect.

Table 18. *Mann–Whitney U test for aptitude scores (high vs. low maintenance).*

Variable	U	p	r
Aptitude	21.00	.449	0.25

A second exploratory grouping focused on participants' aptitude performance. Here, a fixed-threshold method was used, defining Low aptitude ($N = 14$) as scores of 35 or below and High aptitude ($N = 8$) as scores of 65 or above. A Mann–Whitney U test comparing the Maintenance Index between the aptitude groups showed lower maintenance scores on average

for the High aptitude group. This difference was not statistically significant ($U = 67.00$, $p = .482$) with a small effect size ($r = -0.20$), indicating a negligible directional trend.

Table 19. *Mann–Whitney U test for maintenance scores (high vs. low aptitude).*

Variable	U	p	r
Maintenance	67.00	.482	−0.20

Overall, the exploratory analyses did not reveal any statistically significant relationships between the variables of interest, and effect sizes across all comparisons were small or negligible. Given the limited sample sizes and the resulting constraints on statistical power, extending the analyses to additional variables or more complex models was not methodologically justified. Accordingly, the exploratory analyses were limited to correlational screening and did not allow for formal testing of interaction effects as specified in the second hypothesis. Further testing under these conditions would risk producing unstable or spurious patterns rather than meaningful insights. For these reasons, the exploratory component was intentionally kept focused and limited to the most theoretically relevant contrasts.

5. Discussion

This chapter interprets the findings of the present study in relation to the research questions and hypotheses concerning the role of language aptitude in first-language lexical maintenance among late bilinguals. The results are situated within established theoretical and empirical accounts of bilingual language processing and language attrition, with particular emphasis on the implications of the observed null effects for aptitude. Methodological considerations delimit the scope of the conclusions and clarify the conditions under which the findings can be generalized.

Section 5.1 interprets the main findings, first addressing the absence of a predictive relationship between language aptitude and L1 lexical maintenance and subsequently examining the contribution of experiential and sociolinguistic factors. Section 5.2 outlines the study's limitations, including constraints related to design, sample characteristics, measurement, and statistical power. Section 5.3 proposes directions for future research, highlighting the potential of neurocognitive methods, longitudinal designs, refined operationalizations, and broader sampling strategies. Section 5.4 concludes the chapter by summarizing the central insights and situating them within the broader literature on language attrition and individual differences.

5.1. Interpretation of Findings

5.1.1. The Role of Aptitude in L1 Lexical Maintenance

The primary hypothesis, positing a protective role for language aptitude in first-language lexical attrition, was not supported. Across the confirmatory analyses, aptitude failed to significantly predict any dimension of L1 lexical maintenance. The composite Maintenance Index provided the most direct test of this relationship, yielding a negligible and non-significant effect in a simple linear regression ($F(1, 32) = 0.06, p = .811, R^2 = .002, \beta = -0.04$). Although accuracy and reaction times were combined into a composite measure for analytical consistency, this index is a statistical combination rather than a psychometrically validated construct. This is evident in the weak association between its components ($r = -.18, p = .318$), suggesting that they capture partially distinct cognitive aspects of lexical performance, namely

lexical knowledge and retrieval efficiency. Follow-up analyses confirmed that this null effect persisted when the components were examined separately: aptitude did not predict picture-naming accuracy ($F(1, 32) = 0.05, p = .821, R^2 = .002$) nor response speed ($F(1, 32) = 0.37, p = .547, R^2 = .011$).

Importantly, the absence of aptitude effects cannot be attributed to restricted variance or insensitivity of the measures. Substantial inter-individual variability was observed in both aptitude scores ($M = 43.53\%, SD = 26.53\%$) and picture-naming reaction times ($M = 1660$ ms, $SD = 367.0$ ms), indicating meaningful differences in learning ability and processing efficiency across participants. As expected for late bilinguals, picture-naming accuracy reached near-ceiling levels ($M = 95.40\%, SD = 2.98\%$), reflecting largely intact lexical knowledge. Exploratory group comparisons further corroborated the robustness of the null finding, revealing no significant differences between high- and low-maintenance performers ($U = 21.00, p = .449$) or between high- and low-aptitude groups ($U = 67.00, p = .482$). Together, these patterns indicate that the null results reflect a genuine absence of association rather than a methodological artifact.

These findings suggest that individual differences in language aptitude, as operationalized in the present study as vocabulary learning ability, do not systematically relate to the maintenance of well-established L1 lexical representations in late bilinguals immersed in an L2 environment. While aptitude is a robust predictor of success in second-language learning, its contribution to first-language lexical maintenance appears limited.

5.1.2. The Influence of Other Factors on L1 Lexical Maintenance

Beyond language aptitude, exploratory analyses examined whether sociolinguistic and experiential factors were associated with L1 lexical maintenance. The majority of the tested factors, including demographic and attitudinal factors, failed to show statistically significant predictive relationships with L1 maintenance. Among the parameters considered, language use emerged as the most robust correlate of L1 performance.

In particular, a highly significant negative correlation was observed between objective L1 proficiency and L2 frequency of use ($r = -.58, p < .001$), indicating that participants who reported using their L2 more frequently tended to perform worse on the L1 test. This finding aligns with usage-based accounts of attrition, which hold that reduced L1 exposure and increased competition from the dominant L2 contribute to subtle L1 decline (Paradis, 1993). A

related pattern was evident in participants' subjective evaluations of their L1. Self-rated L1 judgement correlated positively with L2 frequency ($r = .46, p = .006$), suggesting that increased engagement with the L2 was associated with a more negative perception of L1 ability, despite objective performance remaining intact. This indicates that usage patterns shape the speakers' subjective awareness of L1 retrieval difficulty.

Additional patterns were consistent with this interpretation. Objective and self-rated L2 proficiency were positively correlated ($r = .47, p = .005$), supporting the internal consistency of proficiency and usage measures. Several further trends pointed in theoretically expected directions, such as positive associations between L2 judgement and L2 frequency ($r = .47, p = .005$), between L1 orientation and L1 judgement ($r = .31, p = .071$), and L1 orientation and repatriation attitude ($r = .33, p = .059$), while aptitude showed a small positive association with L2 Frequency ($r = .25, p = .157$). These findings, however, did not reach conventional levels of significance and should thus be interpreted cautiously.

5.1.3. Theoretical and Methodological Implications

The present findings have several implications for theoretical models of first language attrition and for the methodological study of individual differences in late bilinguals. Most notably, the absence of a robust association between language aptitude and L1 lexical maintenance suggests that, in highly proficient, immersed late bilinguals, L1 outcomes are not primarily governed by cognitive traits linked to language learning ability. Instead, the observed pattern points to a greater role of external and dynamic factors, particularly frequency of language use, in shaping lexical performance in adulthood.

From a theoretical perspective, the present findings are consistent with the well-established role of language aptitude in L2 attainment, whereas no corresponding effect on first language lexical maintenance was observed in this cohort. Additionally, the absence of aptitude effects herein may reflect the role of developmental timing. For these late bilinguals, whose L1 was fully consolidated before L2 immersion, aptitude was not a reliable predictor of L1 maintenance. This outcome is broadly consistent with previous research on late bilinguals reporting null (Bylund & Ramírez-Galán, 2014) or weak (Opitz, 2011) aptitude effects, in contrast to significant effects for early bilinguals (Bylund et al., 2009). Therefore, these patterns suggest that aptitude appears to be primarily relevant to the acquisition of new linguistic systems in late bilinguals and to the retention of an existing one in early bilinguals, whereas its

influence on the retention of an established L1 in adulthood may be indirect, attenuated, or difficult to detect under the experimental conditions, as illustrated by the present study.

These results are also compatible with accounts that conceptualize attrition as a processing-based phenomenon driven by reduced activation and increased cross-linguistic competition. Despite near-ceiling accuracy, substantial individual variability in processing efficiency was observed, indicating that attrition in late bilinguals primarily affects the automaticity of lexical retrieval rather than the integrity of underlying lexical representations (Chamorro & Sorace, 2019). Notably, this interpretation is supported by participants' self-reports, which indicate metacognitive awareness of these subtle L1 changes associated with increased L2 use (Lazaridou-Chatzigoga & Karatsareas, 2022).

The absence of effects for background factors can plausibly be attributed to the characteristics of the present population, which was deliberately selected to minimize extraneous variability and isolate attrition-related effects. Age at testing ranged narrowly within adulthood, limiting the likelihood of age-related influences (Higby & Obler, 2017), while the L2 was acquired post-puberty in predominantly explicit instructional contexts (Li, 2019), where AoA effects are known to be attenuated. Similarly, length of residence yielded null effects, likely reflecting a stabilization phase (Schmid, 2009), since attrition effects are already detected within months of immersion (Baus et al., 2013). Education showed no association, probably due to restricted variance, as all participants were highly educated (Jarvis, 2019). In contrast, the lack of effects for attitudinal factors, including cultural identification and motivation, was less expected, given their frequent association with attrition in the literature. However, this pattern is consistent with evidence suggesting that such factors tend to shape subjective experience more than objective linguistic performance outcomes (Köpke & Schmid, 2004).

Methodologically, the present study underscores the value of lexical retrieval tasks for investigating L1 attrition in late bilinguals. Time-sensitive measures, such as reaction times in picture naming, proved informative in capturing subtle processing differences that accuracy measures alone are unlikely to detect under ceiling conditions. The weak association between speed and accuracy further underscores the importance of treating these dimensions separately. In addition, the emergence of meaningful associations among usage-related variables supports the use of within-group correlational designs as a valid approach for studying attrition processes in the absence of monolingual control groups. Importantly, the consistency of null effects across multiple analytic approaches, combined with the satisfactory distributional properties of all key measures, indicates that these findings are unlikely to be attributable to data quality or violations of statistical assumptions.

Taken together, these findings indicate that language aptitude did not predict L1 lexical maintenance in late bilinguals. Instead, L1 outcomes appear to be primarily processing-related and shaped by dynamic factors associated with language use. Although the hypotheses were not supported, this result should not be interpreted as evidence against a role of aptitude in L1 retention more broadly. Rather, it suggests that under the specific methodological conditions and population profile of the present study, any such influence was not detectable, despite substantial evidence linking aptitude to variation in L1 knowledge.

5.2. Limitations

The present study is subject to limitations that should be considered when interpreting its findings. Given the novelty of this topic and the limited prior empirical work examining the role of language aptitude in first language attrition, certain methodological and analytical trade-offs were unavoidable. The following limitations range from design, sample, and material characteristics to statistical constraints and measurement issues.

The study employed a quantitative, cross-sectional, correlational design focusing on within-group comparisons to investigate the relationship between language aptitude and first language attrition. While this approach is effective for identifying associations and patterns of variability in L1 maintenance, it inherently precludes causal inference. Consequently, the findings cannot establish whether language aptitude plays a causal role in L1 maintenance, nor can they determine the directionality of the observed relationships.

The generalizability of the findings is limited by the characteristics of the participant sample. Although recruitment was conducted via widely accessible online platforms, participation was self-selected, resulting in a sample predominantly composed of highly educated, linguistically skilled late bilinguals. While this profile strengthens internal consistency, it restricts the applicability of the findings to more diverse migrant populations with different educational backgrounds or lower levels of linguistic engagement.

The overall sample size, while comparable to that of prior L1 attrition studies, restricted statistical power. This limitation is particularly relevant given that the statistical analyses did not reveal significant relationships among the variables of interest. It also reduced the study's sensitivity to small or negligible effects that may have been present, especially given that reaction time effects in attrition research are often subtle. Very small group sizes in exploratory

analyses further limited the robustness of group-based comparisons, while dichotomization is associated with well-known methodological limitations (MacCallum, 2002).

These power constraints directly restricted the range and complexity of statistical models that can be reliably employed. Given the ratio of sample size to the number of measured variables, the use of multivariate or mixed-effects modeling was not methodologically justified, as it would risk producing unstable or spurious patterns rather than meaningful insights. As a result, analyses were limited to simpler statistical approaches (e.g., regression), which may not fully capture complex interrelations among cognitive, linguistic, and experiential factors.

Certain data collection decisions further constrained analytical sensitivity. In the questionnaire, key background variables such as age of onset and length of residence were collected using bounded response options (e.g., 0–18+). This constraint required collapsing these variables into categorical groupings to facilitate analysis and interpretability. While pragmatic, this approach reduced data granularity and may have obscured more subtle gradient effects associated with these experiential factors.

The picture naming task employed stimuli systematically controlled for various factors to ensure item-level validity. The images were drawn from a database containing materials standardized in multiple languages; however, at the time of this study, the dataset had not been specifically validated for Greek. Despite extensive item control and preliminary quality assurance checks, this reliance may have introduced some item-level variability, potentially affecting the precision of the lexical retrieval measures.

Language aptitude was measured using the LLAMA test battery, which is not a standardized instrument. While LLAMA_B is widely used to assess individual differences in lexical learning ability, the lack of standardization limits the interpretability and comparability of the aptitude scores (Singleton, 2017; Meara, 2005). In addition, the task involves memorizing novel word–image pairings, raising the question of whether test performance reflects the narrowly defined construct of lexical aptitude or broader memory abilities. From this perspective, the aptitude findings should be interpreted with caution.

Finally, the study was conducted entirely in an unsupervised online testing environment, which was necessary for accessing the target population of expatriates. Although extensive quality-control measures were implemented, remote testing inevitably introduces technical and contextual variability beyond the experimenter’s control, potentially compromising the integrity and consistency of behavioral measures. For example, inattentiveness and distractions (Maniaci & Rogge, 2014), or varying RT due to technical reasons (Anwyl-Irvine et al., 2020).

5.3. Future Directions

Building on the findings of the present study, future research should further investigate the potential role of language aptitude in first-language lexical attrition by adopting methodological and conceptual approaches that are better suited to capturing subtle processing-based effects in late bilinguals. Four directions appear particularly relevant.

A particularly promising line of inquiry for future research lies in integrating neurocognitive methods, which are widely regarded as central to advancing the study of language attrition (Gallo et al., 2021; Keijzer, 2014; Rossi et al., 2019). Yet, to date, research on L1 attrition has relied predominantly on behavioral paradigms (Rossi et al., 2019; Keijzer, 2014), leaving covert neural reorganization that precedes overt performance changes largely unobserved. Techniques with high temporal resolution (e.g., EEG/MEG) can reveal altered processing dynamics, whereas those with high spatial resolution (e.g., fMRI) can map functional and structural correlates of L1 attrition. For example, existing neurocognitive evidence indicates that attriters exhibit qualitatively different neural responses despite accurate performance, suggesting altered processing routines rather than loss of representation (Gallo et al., 2021). Employing imaging and electrophysiological approaches will therefore capture these neural patterns more comprehensively, thereby filling a critical methodological gap.

A second key direction concerns the adoption of longitudinal designs. This approach allows within-speaker performance baselines to be defined, making any subsequent shifts attributable to attrition processes rather than individual differences. This is particularly relevant for lexical retrieval, a domain characterized by substantial inter-individual variability even among monolingual speakers (Fuhrmeister et al., 2024). Since language attrition is not a linear process (Schmid et al., 2013), longitudinal data are necessary to capture its temporal trajectory, including the emergence, stabilization, and potential attenuation of attrition-related effects following renewed L1 immersion, as well as to test whether language aptitude modulates the rate of change. Such designs also enable stronger inferences about directionality, namely about whether changes in language use precede, follow, or co-occur with changes in lexical access.

The field would also benefit from more refined operationalizations of both language aptitude and lexical access to improve construct validity and interpretability. With regard to aptitude, reliance on single-task measures may obscure the contribution of distinct cognitive components. Vocabulary-learning tasks might not be optimal for predicting production efficiency; complementary measures targeting memory, analytic ability, or language-specific processing would clarify which aspects of aptitude, if any, are relevant to L1 maintenance. With

respect to lexical access, task triangulation represents a valuable extension. Complementing picture naming with tasks such as verbal fluency, receptive lexical decision, and language-switching paradigms enables a more comprehensive assessment of lexical processing under varying task demands and reduces task-specific bias.

Finally, future studies should aim for larger and more diverse populations. A larger sample size would permit more sophisticated modeling of individual variability and enhance sensitivity to subtle effects. Broader sampling across educational backgrounds, migration histories, usage profiles, and language pairs would further improve generalizability. Where theoretically appropriate, inclusion of comparison groups, such as non-expatriate bilinguals, could help establish baselines of native-language variability and contextualize attritional outcomes.

5.4. Conclusion

This thesis was motivated by a strong interest in first language attrition as a distinctive consequence of bilingualism, particularly as it reflects change in a fully developed linguistic system that was long assumed to be largely stable. Against this background, the role of language aptitude has emerged as an open question in attrition research. Lexical access was targeted due to its sensitivity to attritional change, and late bilinguals were examined to ensure that observed variability reflected attrition rather than incomplete acquisition.

The findings of the present study indicate that language aptitude did not emerge as a detectable predictor of first language lexical maintenance in highly proficient, long-term immersed late bilinguals. Instead, variability in L1 outcomes was more closely associated with patterns of language use and dominance. In parallel, the results are consistent with the view that attrition in late bilinguals primarily affects performance rather than linguistic competence *per se*, underscoring the importance of investigating covert processing changes.

This outcome also helps explain why the relationship between language aptitude and language attrition has remained difficult to establish, as reflected in the scarcity and heterogeneity of existing empirical findings. Both constructs are independently complex and theoretically evolving, and examining their interaction further compounds this complexity. From this perspective, null findings are informative, as they constrain theoretical expectations and underscore the limitations of current behavioral approaches in isolating subtle relationships between cognitive traits and attritional change.

The absence of detectable aptitude effects in this study must not be interpreted as evidence against a role for aptitude in language attrition more generally. Rather, it suggests that behavioral approaches may be insufficiently sensitive to capture such effects, if they exist, in late bilingual populations. Advancing this line of research will therefore require renewed empirical attention and more intricate methodological frameworks, such as longitudinal designs with neurocognitive measures, capable of tracing change over time and assessing underlying processing dynamics beyond overt performance.

References

- Alario, F.-Xavier, et al. (2004). *Behavior Research Methods, Instruments, & Computers* article is: Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36(1), 140–155
- Alderson, J.C., & Huhta, A. (2005). The Development of a Suite of Computer-Based Diagnostic Tests Based on the Common European Framework. *Language Testing*, 22(3), 301-320.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407.
- Artieda, G. & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences* 50. 42–48.
- Baddeley, A. D. (1986). *Working memory*. Clarendon Press; Oxford University Press.
- Banks, B., & Connell, L. (2022). *Category production norms for 117 concrete and abstract categories*. *Behavior Research Methods*, 55(2), 1292–1313.
- Barkat-Defradas, M., Gayraud, F., Köpke, B., & Lefebvre, L. (2019). Linguistic regression in bilingual patients with Alzheimer’s disease. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 136–145). Oxford University Press.
- Bates, E., D’Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotchewa, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344–380.
- Baus, C., Costa, A., Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, 142 (3), 402–409.
- Bergmann, J., Kroll, J. F., & Bialystok, E. (2015). The role of language exposure in lexical attrition. *Bilingualism: Language and Cognition*, 18(3), 345–359.
- Berinsky, A.J., Margolis, M.F., & Sances, M.W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, 58(3), 739–753.

- Brooks, T.L., Cid de Garcia, D. (2015) Evidence for morphological composition in compound words using MEG. *Frontiers in Human Neuroscience*. 9:215.
- Bylund, E. & Ramirez-Galan, P. (2016). Language Aptitude in First Language Attrition: A Study on Late Spanish-Swedish Bilinguals. *Applied Linguistics*, 37(5), 621–638.
- Bylund, E., N. Abrahamsson, & K. Hyltenstam. (2012). Does L1 maintenance hamper L2 nativelikeness? A study of L2 ultimate attainment in early Spanish-Swedish bilinguals,' *Studies in Second Language Acquisition* 34: 215–41.
- Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House, 83–118.
- Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test (MLAT)*. New York: Psychology Corporation.
- Chalari, A. (2024). "Cultural identities among Greek Diaspora in the UK," GreeSE – Hellenic Observatory Papers on Greece and Southeast Europe 203, Hellenic Observatory, LSE.
- Chamorro, G., Sorace, A. (2019). The Interface Hypothesis as a framework for studying L1 attrition. In: Schmid, M. S. and Köpke, B., eds. *The Oxford Handbook of Language Attrition*. Oxford University Press.
- Chamorro, G., Sorace, A., & Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition*, 19(3), 520–532.
- Chang, C. B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 191–203). Oxford, UK: Oxford University Press.
- Chen, O., Paas, F. & Sweller, J. (2023) A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review* 35, 63.
- Connor, L. T., Spiro, A., 3rd, Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *The journals of gerontology. Series B, Psychological sciences and social sciences*, 59(5), 203–209
- Cook, V. J. (1991a). The poverty of the stimulus argument and multicompetence. *Second Language Research*, 7(2), 103–117.
- Council of Europe. (2001). *Common European Framework of Reference for language learning and teaching*. Cambridge, UK: Cambridge University Press.
- Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* 178. 222–235.

- de Bot, K. (1997). Language attrition. In G. R. Tucker, & D. Corson (Eds.), *Encyclopedia of language and education*. (Vol. 1, pp. 51 – 62). Dordrecht: Kluwer Academic Publishers.
- de Bot, K., & Bülow, L. (2021). Cognitive factors of language contact. In E. Adamou & Y. Matras (Eds.), *Routledge handbook of language contact* (pp. 168-184). Routledge.
- De Graaff, R. (1997). The *Experanto* experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, 19, 249–276.
- de Leeuw, E. (2014). 2. Maturational Constraints in Bilingual Speech. In E. Thomas & I. Mennen (Ed.), *Advances in the Study of Bilingualism* (pp. 25–37). Bristol, Blue Ridge Summit: Multilingual Matters.
- de Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, 13(1), 33–40
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- Dimitropoulou, M., Duñabeitia, J., Avilés, A., Corral, J. & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behaviour: the case of Greek. *Frontiers in Psychology*, 1:218, 1-12.
- Ding Jinfeng , Liu Wenjuan , Yang Yufang. (2017). The Influence of Concreteness of Concepts on the Integration of Novel Words into the Semantic Network. *Frontiers in Psychology*. Volume 8 – 2017.
- Dorian, N. C. (1981). *Language death: The life cycle of a Scottish Gaelic dialect*. University of Pennsylvania Press.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford: Blackwell.
- Doussias et al. (2021). Language contact in the lab. In E. Thomas & I. Mennen (Ed.), *Advances in the Study of Bilingualism* (pp. 28–45). Bristol, Blue Ridge Summit: Multilingual Matters.
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808-816.
- Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *PLoS ONE*, 16.

- Fuhrmeister, P., Elbuy, S. and Bürki, A. (2024) ‘Are Faster Participants Always Faster? Assessing Reliability of Participants’ Mean Response Speed in Picture Naming’, *Journal of Cognition*, 7(1), p. 12.
- Gallo, F., Bermudez-Margaretto, B., Shtyrov, Y., Abutalebi, J., Kreiner, H., Chitaya, T., Petrova, A., & Myachykov, A. (2021). First Language Attrition: What It Is, What It Isn't, and What It Can Be. *Frontiers in human neuroscience*, 15.
- Gignac, G. E., Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers, *Personality and Individual Differences*, Volume 102, 74-78.
- Goral, M. (2022). What can aphasia tell us about how the first-acquired language is instantiated in the brain? *Languages*, 7(4), 283.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3–15.
- Higby, E., Lerman, A., Korytkowska, M., Malcolm, T., & Obler, L. K. (2019). *Ageing as a Confound in Language Attrition Research: Lexical retrieval, language use, and cognitive and neural changes*. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 121–135). Oxford University Press.
- Higby, E., Obler, L. (2017). Chapter 4. Length of residence: Does it make a difference in older bilinguals?. In E. Bialystok & M. Sullivan (Ed.), *Growing Old with Two Languages: Effects of Bilingualism on Cognitive Aging* (pp. 55-75). John Benjamins Publishing Company.
- Hu, X., Hermann Ackermann, J. M., Michael Erb, S. W., & Susanne Reiterer. (2012). Language aptitude for pronunciation in advanced second language (L2) Learners: Behavioural predictors and neural substrates. *Brain and Language* 127(3). 366–376.
- Jarvis, S. (2009). 5. Lexical Transfer. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 99-124). Bristol, Blue Ridge Summit: Multilingual Matters.
- Jarvis, S. (2019). Lexical Attrition, in Monika S. Schmid, and Barbara Köpke (eds), *The Oxford Handbook of Language Attrition*, Oxford Handbooks
- JASP Team (2025). JASP (Version 0.95.3)[Computer software]
- Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321–340.
- Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. PhD dissertation, Vrije Universiteit Amsterdam.

- Keijzer, M. (2014). Changes in Neural Activation Patterns and Brain Anatomy as a Function of Non-Pathological First Language Attrition. *Journal of Neurological Disorders*, 2(04).
- Konrad, E., Ebner, V., Kremmel, B. (2024). Language learning policies and opportunities for migrants and their practical implementation in Austria. *Slovenščina 2.0*, 12(2): 35–62.
- Köpke, B. & Genevska-Hanke, D. (2018). First Language Attrition and Dominance: Same or Different? *Frontiers in Psychology*.
- Köpke, B. & M. S. Schmid (2004). Language Attrition: The Next Phase. *Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. M. S. Schmid, B. Köpke, M. Keijzer and L. Weilemar. Amsterdam, John Benjamins. 28: 1-43.
- Köpke, B. (2004). “Attrition is not a unitary phenomenon: on different possible outcomes of language contact situations,” in *Bilingual Socialization and Bilingual Language Acquisition: Proceedings from the Second International Symposium on Bilingualism*, eds A. M. L. Suarez, F. Ramallo, and X.-P. Rodriguez-Yanez (Vigo: Servizo de Publicacions da Universitdade de Vigo), 1331-1347.
- Köpke, B. (2007). Language attrition at the crossroads of brain, mind, and society. In Köpke, B., Schmid, M. S., Keijzer, M., Dostert, S. (eds.), *Language attrition: Theoretical perspectives* (pp. 9–38). Amsterdam/Philadelphia: John Benjamins.
- Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena, & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131-152). (Language Learning and Language Teaching; Vol. 35). John Benjamins.
- Kupisch, T. (2019). 2L1 simultaneous bilinguals as heritage speakers. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford University Press.
- Lang, J. (2022). Employment effects of language training for unemployed immigrants. *J Popul Econ* **35**, 719–754.
- Lazaridou-Chatzigoga, D., & Karatsareas, P. (2022). Language Attrition and Lived Experiences of Attrition among Greek Speakers in London. *Languages*, 7(4), 307.
- Lehner, A. (2018). The Role of Language Aptitude in Second Language Attrition. In Reiterer, M. Susanne (eds.), *Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience*, 305–321. Berlin: Springer.
- Lenneberg, E.H. (1967). *Biological foundations of language*. New York: Wiley.
- Leonard L. B. (2014). Children with specific language impairment and their contribution to the study of language development. *Journal of child language*, 41 Suppl 1(0 1), 38–47.

- Li, S. (2019). Six decades of language aptitude research: A comprehensive and critical review. In Z. Wen, P. Skehan, A. Biedron, S. Li, & R. Sparks (Eds.), *Language aptitude: Advancing theory, testing, research, and practice*. New York: Routledge.
- Liu, Y., Dong, B., Yang, J., Ejima, Y., Wu, J., Wu, Q., & Zhang, M. (2021). Stimulus Intervals Modulate the Balance of Brain Activity in the Human Primary Somatosensory Cortex: An ERP Study. *Frontiers in neuroinformatics*, 14,
- Llompart, M. & Dąbrowska, E. (2023). “Foreign” language aptitude predicts individual differences in native grammatical proficiency. *Linguistics*, 61(5), 1165-1193.
- Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): a meta-analysis of serial reaction time task performance. *Cortex; a journal devoted to the study of the nervous system and behavior*, 51(100), 1–10.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40.
- MacWhinney, B. (2019). Language attrition and the competition model. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 30–48). Oxford University Press.
- Malikouti-Drachman, A. (2001). Greek Phonology: A Contemporary Perspective. *Journal of Greek Linguistics*, 2(1), 187–243.
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83.
- Meara, P. M., & Rogers, V. E. (2020). *The LLAMA Tests v3*. Lognostics.
- Meara, P. (2005). LLAMA language aptitude tests: The manual (Tech. Rep.). Lognostics.
- Montrul, S., & Polinsky, M. (2019). Introduction to heritage language development. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford University Press.
- Montrul, S., & Yoon, J. (2019). Morphology and Language Attrition. In *Oxford Research Encyclopedia of Linguistics* Oxford University Press.
- Opitz, C. (2011). First language attrition and second language acquisition in a second language environment. PhD thesis. Centre for Language and Communication Studies. Trinity College Dublin. Dublin.
- Oppenheimer, D., Meyvis, T. & Davidenko, N. (2009). Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology*. 45. 867-872.

- Paradis, J. (2007). Bilingual children with specific language impairment: Theoretical and applied issues. *Applied Psycholinguistics*, 28, 551-564.
- Paradis, M. (1993). Linguistic, psycholinguistic, and neurolinguistic aspects of "interference" in bilingual speakers: The activation threshold hypothesis. *International Journal of Psycholinguistics*, 9(2)[26], 133–145.
- Paul Boersma & David Weenink (1992–2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.09, retrieved 2 March 2023 from <https://www.praat.org>.
- Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics*, 11 (2), 175–205.
- Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery*. New York, NY: Harcourt Brace Jovanovich.
- Ramoser, C., Fischer, A., Caspers, J. *et al.* Language aptitude is related to the anatomy of the transverse temporal gyri. *Brain Struct Funct* **230**, 14 (2025).
- Regel, S., Opitz, A., Müller, G., & Friederici, A. D. (2019). Processing inflectional morphology: ERP evidence for decomposition of complex words according to the affix structure. *Cortex*, 116, 143–153.
- Rimfeld, K., Philip S. Dale & Robert Plomin. (2015). How specific is second language learning ability? A twin study exploring the contributions of first language achievement and intelligence to second language achievement. *Translational Psychiatry* 5. 1–6.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17(4), 368-392.
- Rodd, J., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 256-271.
- Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, 1(1), 49-60.
- Rossi, Eleonora, Yanina Prystauka, and Michèle T. Diaz, 'Neuroimaging Perspectives on L1 Attrition and Language Change', in Monika S. Schmid, and Barbara Köpcke (eds), *The Oxford Handbook of Language Attrition*, Oxford Handbooks
- Roux, Frédéric & Armstrong, Blair & Carreiras, Manuel. (2016). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*. 49(5), 1864–1881.
- Schmid, M. (2011). *Language Attrition (Key Topics in Sociolinguistics)*. Cambridge: Cambridge University Press.

- Schmid, M. S. (2009). On L1 attrition and the linguistic system. *EUROSLA Yearbook*, 9(1), 212–244.
- Schmid, M. S. (2011). What is attrition? In *Language Attrition* (pp. 11–17). chapter, Cambridge: Cambridge University Press.
- Schmid, M. S., & Jarvis, S. (2014). *Lexical access and lexical diversity in first-language attrition*. *Bilingualism: Language and Cognition*, 17(4), 729-748.
- Schmid, M. S., & Yilmaz, G. (2018). Predictors of Language Dominance: An Integrated Analysis of First Language Attrition and Second Language Acquisition in Late Bilinguals. *Frontiers in psychology*, 9, 1306.
- Schmid, M. S., & Yilmaz, G. (2018). *Predictors of language dominance: An integrated analysis of first-language attrition and second-language acquisition in late bilinguals*. *Frontiers in Psychology*, 9.
- Schmid, M. S., Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research* 26:125–60.
- Schmid, M. S., Köpke B. (2009). L1 attrition and the mental lexicon. Pavlenko, Aneta. *The Bilingual Mental Lexicon. Interdisciplinary approaches, Multilingual Matters*, pp.209-238, 2009, *Bilingual Education & Bilingualism*
- Schmid, M.S., & Köpke, B (2017). *The relevance of first language attrition to theories of bilingual development*. *Linguistic Approaches to Bilingualism*, 7 (6). pp. 637–667.
- Schmid, M.S., Barbara K., De Bot, K. (2013). Language attrition as a complex, non-linear development. *International Journal of Bilingualism*, SAGE Publications, 17 (6), pp.675-682.
- Schmid, M.S., Cherciov, M. (2019). Introduction to Extralinguistic Factors in Language Attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 266–276). Oxford University Press.
- Schmid, M. S. (2013). First language attrition. *WIREs Cognitive Science*, 4(2), 117–123.
- Schmitt, E. & Sorokina, A. (2024) Guest editorial: Language attrition – a comprehensive introduction, *The Language Learning Journal*, 52:2, 133-144
- Schmitt, E. (2019). Morphological attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 228–240). Oxford University Press.
- Shao, Z., Meyer, A.S. & Roelofs, A. Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition* **41**, 1200–1211 (2013).

- Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching*, 7, 89–103.
- Skehan, P. & L. Duroquet. (1988). A Comparison of First and Foreign Language Learning Ability. Working Documents 8, ESOL Department. London University.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33.
- Sparks, R. L., Patton, J., & Luebbers, J. (2019). Individual differences in the L2 achievement mirror individual differences in the L1 skills and L2 aptitude: Crosslinguistic transfer of L1 to L2 skills. *Foreign Language Annals*, 52, 2, 1–29.
- Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257-277.
- Turker, S., Seither-Preisler, A., & Reiterer, S. M. (2021). Examining Individual Differences in Language Learning: A Neurocognitive Model of Language Aptitude. *Neurobiology of language (Cambridge, Mass.)*, 2(3), 389–415.
- Ullman M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231–270.
- Ullman, M. and Pullman, M. (2015) A Compensatory Role for Declarative Memory in Neurodevelopmental Disorders. *Neuroscience and Biobehavioral Reviews*, 51, 205-222.
- van Hell, J. G., & Dijkstra, T. (2002). Models of the bilingual lexical representation. *Psycholinguistic Bulletin*, 33(2), 215-232.
- Vogt, A., Hauber, R., Kuhlen, A. K., & Rahman, R. A. (2022). Internet-based language production research with overt articulation: Proof of concept, challenges, and practical advice. *Behavior Research Methods*, 54(4),
- Wen, Z. (2016). *Working memory and second language learning: Towards an integrated approach*. Bristol: Multilingual Matters.
- Wen, Z., Adriana Biedroń & Peter Skehan. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching* 50. 1–31.
- Yılmaz, G., & Schmid, M. S. (2012). L1 lexical accessibility among Turkish–Dutch bilinguals. *The Mental Lexicon*, 7, 249–274.

Appendices

Appendix A: Recruitment and Information Materials

A1. Recruitment poster



**universität
wien**
Institut für
Sprachwissenschaft

TeilnehmerInnen für Online-Studie gesucht!



20€
Gutschein
zu gewinnen

Sprachgebrauch griechischer Muttersprachler im deutschen Sprachraum

Teilnahmebedingungen
Alter: 25-55
Sprachen:
○ Muttersprache: Griechisch
○ Fremdsprache (B2-C2): Deutsch (ab 15. Lebensjahr)
Herkunftsort: Griechenland
Wohnort: Deutschland, Österreich, Schweiz (7+ Jahre, ab 15. Lebensjahr)
Keine Sprach- bzw. Lernstörungen

Ablauf
Dauer: 25-30'
Computer/Laptop erforderlich

Aufgaben
1. Bildbenennung mit Tonaufnahme
2. Multiple-Choice
3. Richtig/Falsch
4. Fragebogen

Kontakt
[Redacted contact information]

A2. Participant information sheet (translated)



universität
wien

Institut für Sprachwissenschaft

Masterstudium Allgemeine Linguistik: Grammatiktheorie und kognitive Sprachwissenschaft

Participation Instructions

Thank you for your interest!

The purpose of this study is to analyze the linguistic abilities of Greeks living in German-speaking countries. Please confirm that you meet the participation criteria and agree to the experimental procedure.

A) Requirements

Native language:	Greek
Age:	18+
Place of birth and residence:	Greece, until the 15 th year of life
Current place of residence:	Germany, Austria or Switzerland
Years abroad:	At least 7 consecutive years
German knowledge:	B2+
Age of German acquisition:	15+

B) Procedure

Mode	On-line	You will receive a link to the study. A computer (desktop or laptop) and basic computer knowledge are required. You may start at your own time. A quiet environment is required.
Parts	Part 1:	Test on word production readiness (Greek): Voice recording of responses
	Part 2:	General language ability test: Multiple choice
	Part 3:	2 language tests (Greek, German): True/False
	Part 4:	Questionnaire: Demographic and linguistic profile
Duration	25-30 minutes	It is not possible to pause the test and continue later. Breaks are allowed between parts. The duration may vary accordingly.
Data protection		Personal data and voice recordings shall remain confidential.
Reward	20€	By completing the study, you enter a draw for 3 € 20 Amazon vouchers.
Warning		This study requires concentration, and a lack thereof can compromise the results. Each participation comes with costs for the experimenter. Thus, please consider carefully the above before signing up. There are no penalties for withdrawing at any time, as participation is voluntary.

Contact information

E-mail:



Appendix B: Experimental Tasks and Materials

B1. Within-experiment instructions - excl. Questionnaire (translated)

Note: Section numbering was added to the appendix for clarity, but it did not appear in the live experiment.

1. Introduction

1.1. Welcome!

Thank you for your interest in participating in the present study!

The study is conducted by **student's name**, a student in the Department of Linguistics at the University of Vienna, as part of her master's thesis, under the supervision of **professor's name**. The purpose of this study is to analyze the lexical abilities of Greek native speakers living in German-speaking countries.

1.2. Consent form

I hereby declare that I

- have read the information carefully
- have been adequately informed by the experimenter
- comprehend the content and the processes of the study
- understand that my participation is voluntary
- understand that I can withdraw at any time without penalty
- understand that my data is treated with strict confidentiality
- understand that audio of my responses is recorded
- understand that the researcher will handle my data

I accept the above and consent to participate in the present study.

1.3. General Instructions

The following study requires concentration.

- Read the instructions carefully. Each task is unique.
- Start only when you are certain about the process of each task.
- It is not possible to return to a previous page.
- Set your mobile phone to silent mode.
- Quite a space is required (e.g., no music, chatting with other people)
- Close all the other open tabs on your browser.

Remember that despite the absence of supervision, you are participating in a formal scientific study!

2. Part 1

2.1. Introduction

This part tests your **speed and accuracy** in **recognizing and producing** words in **Greek**. Your oral responses are being **recorded**.

2.2. Technical check

Make sure that your microphone is working!

You will see an image on the next screen. Click *Start Recording* to begin. Click *Accept* in the pop-up notification requesting access to your microphone. Say aloud what is depicted. Click *Stop Recording* once you are done. Provided the microphone is operational, you will hear your recording automatically. If you wish to make a new recording, click *Re-record*.

Should you encounter technical issues, discontinue the study and promptly inform the researcher.

If no issues arise, click Next when you are ready.

2.3. Description

Please name the following **60 images** in **Greek**.

Before each image, a cross will appear on your screen. The image appears for **10 seconds**. **Recording starts automatically** once the image appears. Say loud and clear the word that corresponds to the image. To proceed to the next image, you can either a) wait for the time limit to expire and automatically advance to the next image, or b) click *Next*. Attention: do not pass any images by pressing *Next* when the time limit is reached!

The images depict nouns such as animals, professions, foods, body parts, tools, etc. The words are either simple (e.g., paper) or closed compounds (e.g., newspaper), but not open compounds (e.g., paper cutter). One-word answers are required.

Tips:

- Recording starts automatically once the image appears. There is no start button.
- Reply spontaneously and rapidly.
- One-word answers are required.
- If you do not know an answer, say 'pass'.

- Avoid ‘uhms’ and ‘ehms’.
- Avoid noises (e.g., do not play with the headphone cable).
- You can change your answer by simply saying a new word.

2.4. Instructions

- Automatic audio recording.
- Say aloud what you see on the screen.
- One-word answers.
- Respond clearly and spontaneously.
- If you are uncertain, say ‘pass’.
- You have 10 seconds for each image.
- After responding, wait or click Next to continue.

2.5. Attention question 1

When does the voice recording start?

- Automatically, once the image appears.
- Once I press the start button.

Feedback:

- Correct: You seem concentrated!
- Incorrect: You probably need a recap... (2.4. Instructions are repeated)

2.6. Practice

Before the testing session begins, familiarize yourself by completing 5 **examples**. In the end, the correct answers are provided so that you can confirm you understood the task.

Click Next to start practicing.

2.7. Test

You may proceed to the main task, as long as you comprehend the procedure.

Reminder:

- Say aloud what you see on the screen.
- One-word answers.
- Respond clearly and spontaneously.
- If you are uncertain, say ‘pass’.
- You have 10 seconds for each image.
- After responding, wait or click Next to continue.

Attention: Once you click Next, you will be automatically transferred to the first image, and the study commences immediately!

When you are ready, click Next.

2.8. Attention Question 2

What do you do in case you do not know the answer?

- I do not reply.
- I say 'pass'.

Feedback:

- Correct: You seem concentrated!
- Incorrect: You probably need a recap... (2.4. Instructions are repeated)

2.9. End

Wonderful!

You may take a short break before continuing to Part 2!

When you are ready, click Next.

3. Part 2

3.1. Introduction

This part tests your ability to learn new words in foreign languages.

Try to learn 20 **words** in an **artificial** (non-real) **language** within a limited time. These words are the names of imaginary (non-real) objects. This part consists of two phases.

3.2. Phase 1: Learning

Memorize the words. 20 images and their names underneath appear on the screen. Learn the images and their respective names. You have **2 minutes** to complete this phase. A countdown appears in the last 30 seconds. Page transition occurs automatically.

Click Next when you are ready to learn the words.

3.3. Phase 2: Testing

Remember the words? A **single word** from the ones you have just learned appears each time at the top of the screen. Choose which object the word stands for by clicking on that image. You cannot change your response. Once you click on the image, you are automatically transferred to the next word. There is no time limit.

Click Next to continue to the test.

3.4. End

Perfect!

You may take a short break before continuing to Part 3!

When you are ready, click Next.

4. Part 3

4.1. German

This part tests your lexical abilities in German and Greek.

75 lexical items (verbs) for each language are included. Decide whether each item is a **real word or not**. If an item is a real word, click the green button on the screen, whereas if it is not, click the red button. You have **10 seconds** for each item!

Click Next to start with **German**.

4.2. Greek

Now, continue with **Greek**. If the item is a real word, click the green button. If it is not, click the red button. You have **10 seconds** for each item!

When you are ready, click Next.

4.3. End

Great!

You may take a short break before continuing to Part 4!

When you are ready, click Next.

B2. Background Questionnaire (translated)

Note: Section numbering was added to the appendix for clarity and did not appear in the live experiment.

1. Introduction

This questionnaire investigates the participants' personal background, language use and attitude. It consists of 40 questions, all of which concern exclusively your own linguistic experience. Answer truthfully and spontaneously. There are no right or wrong answers! You may revisit your answers. There is no time limit.

2. Category A

1. Date of birth (month/year)	
2. Gender	• Female • Male • Other
3. Native language	• Greek • Greek & German • Greek & other • Other (Please specify)
4. Level in German	• Upper intermediate (B2) • Advanced (C1) • Proficient (C2)
5. Do you use other languages daily? Which?	
6. Have you been diagnosed with any speech/learning disorders?	• No • Yes (Please specify)
7. Place of birth	• Greece • Other (Please specify)
8. Place of current residence	• Germany • Austria • Switzerland • Other (Please specify)
9. Have you lived in another country for longer than 6 months?	• No • Yes (Please specify country and duration, e.g., England 2 years)
10. Highest level of education	• Primary education • Secondary education • Vocational education • Undergraduate education • Postgraduate education
11. Occupation (pensioners, unemployed: previous occupation, students: field of studies)	
12. Do you speak other languages? Which?	

3. Category B

1. At what age did you start learning the following languages?	
• Greek • German • Other (regularly used)	0 - 18+

2. How many years have you lived in a country where the following languages are spoken?	
• Greek	0 - 18+
• German	
• Other (regularly used)	
3. How many years did you receive schooling (e.g., school, university, foreign language courses) in the following languages?	
• Greek	0 - 18+
• German	
• Other (regularly used)	

4. Category C

In this part, you are requested to rate from 1 (never) to 10 (always) how often you use each language in each of the following situations.

a) Greek	
1. Family	1 (never) - 10 (always)
2. Partner	
3. Friends	
4. Work	
5. Hobbies	
6. People in Greece	
7. People in the country of residence	
b) German	
1. Family	1 (never) - 10 (always)
2. Partner	
3. Friends	
4. Work	
5. Hobbies	
c) Other (regularly used)	
1. Family	1 (never) - 10 (always)
2. Partner	
3. Friends	
4. Work	
5. Hobbies	
a) Greek	
1. Count	1 (never) - 10 (always)
2. Dream	
3. Cuss	
4. Self-talk	
b) German	
1. Count	1 (never) - 10 (always)
2. Dream	
3. Cuss	
4. Self-talk	
c) Other (regularly used)	
1. Count	1 (never) - 10 (always)
2. Dream	
3. Cuss	
4. Self-talk	
a) Greek	

1. Listen to music, watch movies	1 (never) - 10 (always)
2. Read books, magazines	
3. Write e-mails, diary	
4. Use social media	
b) German	
5. Listen to music, watch movies	1 (never) - 10 (always)
6. Read books, magazines	
7. Write e-mails, diary	
8. Use social media	
c) Other (regularly used)	
9. Listen to music, watch movies	1 (never) - 10 (always)
10. Read books, magazines	
11. Write e-mails, diary	
12. Use social media	

5. Category D

In this part, you are requested to rate from 1 (disagree) to 10 (agree) how much you agree with each of the following statements.

1. I want to sound like a native speaker.	
• Greek	1 (disagree) - 10 (agree)
• German	
• Other (regularly used)	
2. It is important that my children learn...	
• Greek	1 (disagree) - 10 (agree)
• German	
• Other (regularly used)	
3. I identify with this culture.	
• Greek	1 (disagree) - 10 (agree)
• German	
• Other (regularly used)	
4. My language skills have improved ever since I migrated.	
• Greek	1 (disagree) - 10 (agree)
• German	
• Other (regularly used)	
1. I intend to repatriate to Greece.	
1 (disagree) - 10 (agree)	
2. My decision to move here was right.	
1 (disagree) - 10 (agree)	
3. It bothers me when I cannot remember a word in Greek.	
1 (disagree) - 10 (agree)	
4. I enjoy learning foreign languages.	
1 (disagree) - 10 (agree)	
5. I acquire foreign languages with ease.	
1 (disagree) - 10 (agree)	

6. End

Congratulations!

You have reached the end of the study.

7. Voucher

Fill in your e-mail address to participate in the raffle for the vouchers:

Increase your chances of winning! Describe the technical characteristics of your device.

Microphone (e.g., internal, headset)

Mouse (e.g., wireless, touchpad)

8. Thanks

Thank you for participating!

To complete the study, click Next.

B3. Picture naming task stimuli

Table B3.1. *Stimuli list.*

Item	Original	Target	Synonym
1	chin	πιγούνι	σαγόνι
2	mermaid	γοργόνα	
3	prisoner	φυλακισμένος	
4	frame	κάδρο	κορνίζα
5	clover	τριφύλλι	
6	screwdriver	κατσαβίδι	
7	fireman	πυροσβέστης	
8	saw	πριόνι	
9	belly	κοιλιά	
10	button	κουμπί	
11	calendar	ημερολόγιο	
12	snake	φίδι	
13	fridge	ψυγείο	
14	bin	κάδος	σκουπιδοτενεκές
15	hanger	κρεμάστρα	
16	goalkeeper	τερματοφύλακας	
17	computer	υπολογιστής	
18	printer	εκτυπωτής	
19	seagull	γλάρος	
20	dummy	πιπίλα	
21	knee	γόνατο	
22	salt	αλάτι	
23	postman	ταχυδρόμος	
24	torch	φακός	
25	heel	τακούνι	
26	knot	κόμπος	
27	compass	πυξίδα	
28	plumper	υδραυλικός	
29	thread	κλωστή	
30	letterbox	γραμματοκιβώτιο	
31	dice	ζάρι	
32	razor	ξυράφι	ξυραφάκι
33	spaceship	διαστημόπλοιο	
34	soldier	στρατιώτης	
35	ruler	χάρακας	
36	fan	ανεμιστήρας	
37	strawberry	φράουλα	
38	stapler	συρραπτικό	
39	mountain	βουνό	
40	ballot box	κάλη	
41	scissors	ψαλίδι	
42	peg	μανταλάκι	
43	apron	ποδιά	

44	rose	τριαντάφυλλο	
45	nail	καρφί	πρόκα
46	swan	κύκνος	
47	fence	φράχτης	
48	watermelon	καρπούζι	
49	ant	μυρμήγκι	
50	iron	σίδηρο	
51	corn	καλαμπόκι	
52	pencil case	κασετίνα	
53	coat	παλτό	
54	windmill	ανεμόμυλος	
55	runner	δρομέας	
56	fire extinguisher	πυροσβεστήρας	
57	cards	τράπουλα	
58	washing machine	πλυντήριο	
59	scarecrow	σκιάχτρο	
60	wallet	πορτοφόλι	

Table B3.2. *Item-level linguistic and visual properties.*

Item	Image	Length	Lg10CD	WF	Correct	Unknown	Visual
1	PICTURE_6	3	1.2304	0.8203	99	0	1,87
2	PICTURE_12	3	1.5563	1.6838	97	1	2,72
3	PICTURE_27	5	2.0719	6.1309	98	0	2,77
4	PICTURE_33	2	1.2304	0.8203	94	3	2,23
5	PICTURE_40	3	1.1761	1.6838	96	2	2,21
6	PICTURE_46	4	1.7482	3.0223	98	2	2,34
7	PICTURE_48	4	1.7324	5.0515	98	1	2,86
8	PICTURE_52	3	1.8388	4.1017	95	2	1,98
9	PICTURE_94	2	2.5670	20.1198	98	0	1,96
10	PICTURE_64	2	2.6284	29.8343	98	1	1,80
11	PICTURE_108	6	2.5276	28.9276	99	0	2,47
12	PICTURE_133	2	2.4624	23.0557	99	0	3,38
13	PICTURE_138	3	2.6902	28.7117	100	0	2,68
14	PICTURE_145	2	1.2788	0.8203	100	0	2,16
15	PICTURE_147	3	1.2788	0.9499	100	0	2,00
16	PICTURE_171	6	1.2788	1.1657	99	0	2,99
17	PICTURE_196	5	2.2480	10.8802	99	0	2,58
18	PICTURE_204	4	0.9542	0.3886	99	0	2,99
19	PICTURE_220	2	1.2041	0.8203	98	0	2,86
20	PICTURE_221	3	1.0000	0.4318	97	1	2,36
21	PICTURE_222	3	2.2330	8.5919	99	0	2,23
22	PICTURE_233	3	2.3032	13.6435	100	0	2,26
23	PICTURE_242	4	1.8633	5.5696	97	0	2,86
24	PICTURE_262	2	1.8921	3.8426	99	0	2,31
25	PICTURE_266	3	1.6128	2.0724	99	0	1,95
26	PICTURE_267	2	1.6128	2.0292	96	3	2,07
27	PICTURE_292	3	1.9395	7.2103	90	6	2,65
28	PICTURE_294	4	1.6435	2.5474	93	3	2,89
29	PICTURE_300	2	1.9345	4.1017	93	3	2,58
30	PICTURE_305	6	1.6335	2.2020	94	4	2,38
31	PICTURE_308	2	0.9031	0.3886	100	0	1,92
32	PICTURE_311	3	2.0374	5.9582	95	3	2,42
33	PICTURE_315	5	2.0645	10.0167	96	0	2,83
34	PICTURE_317	3	2.6180	33.6769	98	0	2,97
35	PICTURE_336	3	0.6990	0.1727	97	2	1,89
36	PICTURE_363	5	1.2553	0.8635	99	1	2,64
37	PICTURE_381	3	1.7160	3.2813	100	0	2,60
38	PICTURE_387	4	1.3802	1.3816	99	1	2,52
39	PICTURE_407	2	2.6335	34.8858	96	1	3,05
40	PICTURE_415	2	0.8451	0.4318	96	3	2,50
41	PICTURE_453	3	2.0899	8.0738	98	2	1,90
42	PICTURE_467	4	0.6990	0.1727	93	4	2,11
43	PICTURE_476	2	1.9590	5.3969	94	2	2,56
44	PICTURE_480	5	1.9138	5.5265	100	0	2,80
45	PICTURE_485	2	2.4082	16.1045	94	4	1,88

46	PICTURE_528	2	1.4472	1.8134	97	2	2,94
47	PICTURE_530	2	1.6232	2.6769	99	1	2,10
48	PICTURE_560	3	1.4624	1.4248	99	1	2,72
49	PICTURE_585	3	1.6435	3.7994	96	4	2,89
50	PICTURE_586	3	2.0294	5.8719	100	0	2,44
51	PICTURE_587	4	2.0569	7.7716	98	0	2,94
52	PICTURE_609	4	0.7782	0.2159	99	0	2,59
53	PICTURE_644	2	2.6730	32.8134	99	1	2,49
54	PICTURE_675	5	0.7782	0.2591	93	5	2,96
55	PICTURE_688	3	1.3617	1.4680	100	0	2,92
56	PICTURE_702	5	0.9542	0.4749	97	3	2,50
57	PICTURE_723	3	1.8129	4.4903	98	0	2,10
58	PICTURE_734	4	2.2304	10.4485	99	0	2,89
59	PICTURE_743	2	1.5682	3.2813	96	3	3,00
60	PICTURE_382	4	2.6684	31.0432	99	0	2,33

Label	Description
Image, Correct, Unknown, Visual	MultiPic norms (label, correctness, unknown responses, visual complexity)
Length	Word length in syllables
Lg10CD, WF	Subtlex-GR word frequency measures

Figure B3.3. *Distribution of stimuli by frequency and syllable length.*

N=60		FREQUENCY (Lg10CD)		
		low <1.5	medium 1.5–2.0	high >2.0
LENGTH (syllables)	short 2	Condition 1 6 items	Condition 4 6 items	Condition 7 6 items
	medium 3	Condition 2 7 items	Condition 5 7 items	Condition 8 7 items
	long 4-6	Condition 3 7 items	Condition 6 7 items	Condition 9 7 items

B4. Stimuli of other tests

Table B4.1. *Proficiency task items (L1 and L2).*

Item	L1-Word	Type	Item	L2-Word	Type
1	προσφέρω	word	1	schwören	word
2	αριστεύω	word	2	zurückstecken	word
3	παρατηρώ	word	3	schuttern	nonword
4	εζητείνω	nonword	4	umstellen	word
5	προηγούμαι	word	5	aggressieren	nonword
6	οδεύω	word	6	probieren	word
7	λυσσομανώ	word	7	stillen	word
8	καραδοκώ	word	8	fliehen	word
9	επιστρέφω	word	9	schmieren	word
10	παστεριώνω	word	10	zerdrücken	word
11	σιχαίνομαι	word	11	eindräuen	nonword
12	θαλασσοδέρνω	word	12	dreien	nonword
13	δυσαρεστώ	word	13	binden	word
14	ναυλώνω	word	14	verderben	word
15	εγκαταλείπω	word	15	erinnern	word
16	ατενασσάνω	nonword	16	schweiken	nonword
17	νικοδημίζω	nonword	17	ausblintern	nonword
18	θορυβώ	word	18	chiffrieren	word
19	μουχλιάζω	word	19	steuern	word
20	αξιοποιώ	word	20	hinstürzen	word
21	αυξάνω	word	21	umhaupten	nonword
22	κρεπώνω	nonword	22	umrahmen	word
23	ωραιοπλαίω	nonword	23	gemaunen	nonword
24	μεθοδεύω	word	24	hinhauen	word
25	πατώ	word	25	niederkämpfen	word
26	αισθάνομαι	word	26	abkratzen	word
27	γελοστίζω	nonword	27	erdulden	word
28	θεσπίζω	word	28	schildieren	nonword
29	εκτοξεύω	word	29	halbieren	word
30	επαινώ	word	30	bauen	word
31	καμαρώνω	word	31	bremsen	word
32	καρβαίνω	nonword	32	hineinbekommen	word
33	αστοιχολώ	nonword	33	zermalmen	word
34	εκτυπώνω	word	34	verhutzeln	nonword
35	υπολογίζω	word	35	herauspasten	nonword
36	παρανοικλίνω	nonword	36	erkrellen	nonword
37	τεσσάνομαι	nonword	37	hochjagen	word
38	ρούχομαι	nonword	38	herausfinden	word
39	πληροφορώ	word	39	krönen	word
40	ανατινάζω	word	40	abstitzen	nonword
41	παρακολουθώ	word	41	schlopfen	nonword
42	υστερώ	word	42	entfremden	word
43	υπονοώ	word	43	lecken	word

44	φρεσκάρω	word
45	φθίνω	word
46	βανατρυλώνω	nonword
47	καπνίζω	word
48	δωστολογώ	nonword
49	περιλαμβάνω	word
50	απαρτερώ	nonword
51	κεριστάζομαι	nonword
52	νερολλάζω	nonword
53	απολύω	word
54	δευτεριλώνω	nonword
55	γευματώνω	nonword
56	φεροκρατώ	nonword
57	προάγω	word
58	νοτοχαρίζω	nonword
59	χρωστώ	word
60	ζώομαι	nonword
61	ανθώ	word
62	μαγειρεύω	word
63	περιστρέφω	word
64	συμβαδίζω	word
65	ψηλοδηγώ	nonword
66	τρυγώ	word
67	υπερισχύω	word
68	βαλοιδώ	nonword
69	οργανώνω	word
70	συγκινώ	word
71	περιφρονώ	word
72	ισοπλίζω	nonword
73	τρεναρίζω	nonword
74	μουσκεύω	word
75	στενοχωρώ	word

44	leihen	word
45	beherrschen	word
46	zählen	word
47	entklupfern	nonword
48	mindern	word
49	ansprechen	word
50	mitlaufen	word
51	vereinlichen	nonword
52	seitern	nonword
53	festlegen	word
54	sträubern	nonword
55	kleben	word
56	glasieren	word
57	pöhlen	nonword
58	bemollen	nonword
59	einarbeiten	word
60	verfeinern	word
61	entlaben	nonword
62	wählen	word
63	herstellen	word
64	einpfauen	nonword
65	vertrenken	nonword
66	leisten	word
67	geschehen	word
68	orientieren	word
69	mögen	word
70	heraufsetzen	word
71	hineinstopfen	word
72	plaschen	nonword
73	vergessen	word
74	fordern	word
75	verballen	nonword

Table B4.2. *Aptitude task items.*

Item	Word	Item	Word
1	taa	11	iva
2	lulu	12	yo.o
3	chaka	13	suma
4	saa	14	tuu
5	ina	15	ixa
6	chuku	16	so.o
7	ko.o	17	yua
8	yutun	18	xika
9	yuka	19	suku
10	ita	20	tachi

Appendix C: Supplementary Data and Results

C1. Task Scores

Table C1.1. *Scores in the aptitude and proficiency tasks in percentages.*

ID	Aptitude	L1	L2
p1	15,00	96,00	85,33
p2	40,00	100,00	93,33
p3	5,00	98,67	62,67
p4	65,00	97,33	81,33
p5	50,00	86,67	82,67
p6	100,00	97,33	94,67
p7	15,00	98,67	93,33
p8	55,00	100,00	65,33
p9	20,00	100,00	98,67
p10	40,00	97,33	73,33
p11	15,00	98,67	88,00
p12	80,00	93,33	68,00
p13	40,00	72,00	74,67
p14	30,00	90,67	74,67
p15	50,00	98,67	78,67
p16	45,00	94,67	84,00
p17	65,00	97,33	65,33
p18	55,00	98,67	85,33
p19	35,00	89,33	84,00
p20	100,00	74,67	66,67
p21	20,00	97,33	64,00
p22	65,00	84,00	57,33
p23	15,00	100,00	65,33
p24	55,00	93,33	89,33
p25	75,00	94,67	84,00
p26	30,00	100,00	90,67
p27	25,00	100,00	90,67
p29	85,00	98,67	96,00
p30	50,00	100,00	88,00
p31	55,00	100,00	66,67
p32	10,00	94,67	86,67
p34	0,00	100,00	88,00
p35	60,00	84,00	70,67
p36	15,00	92,00	76,00

Table C1.2. *Item-level performance and items excluded (in grey).*

Item	Acc (sum)	RT (ms)	Item	Acc (sum)	RT (ms)
It. 1	85	2311,31	It. 31	90	1907,67
It. 2	92	1961,38	It. 32	96	1935,09
It. 3	98	1434,83	It. 33	69	1976,00
It. 4	100	1798,62	It. 34	96	2193,00
It. 5	99	1161,83	It. 35	102	1327,38
It. 6	86	2038,38	It. 36	100	1448,62
It. 7	87	1509,10	It. 37	71	3228,58
It. 8	102	1519,85	It. 38	98	1540,17
It. 9	99	1276,08	It. 39	88	2149,67
It. 10	91	2220,08	It. 40	92	2977,67
It. 11	91	1768,00	It. 41	102	1986,46
It. 12	94	1478,50	It. 42	78	2316,17
It. 13	85	1639,00	It. 43	85	2576,75
It. 14	93	2235,00	It. 44	98	1930,42
It. 15	83	2096,33	It. 45	99	1189,17
It. 16	96	1872,18	It. 46	98	1502,50
It. 17	93	1481,45	It. 47	99	1027,83
It. 18	92	2887,67	It. 48	90	1463,83
It. 19	89	2496,92	It. 49	93	1222,91
It. 20	96	1547,00	It. 50	99	1838,42
It. 21	99	1578,77	It. 51	99	1484,92
It. 22	81	2181,58	It. 52	96	1921,85
It. 23	96	2396,15	It. 53	99	1209,17
It. 24	98	1389,17	It. 54	97	1126,50
It. 25	98	1392,67	It. 55	84	1399,25
It. 26	99	1455,00	It. 56	84	1733,36
It. 27	90	1872,67	It. 57	86	1569,17
It. 28	97	1293,75	It. 58	99	1724,42
It. 29	97	1242,83	It. 59	96	1220,50
It. 30	96	1564,42	It. 60	99	1478,75

C2. Questionnaire Results

Table C2.1. *Demographic background.*

ID	Gender	Age	Education	Host	LoR	Other host
p1	1	40,6	3	1	9	1
p2	0	32,1	3	1	7	0
p3	0	38,2	3	1	5	0
p4	1	26,11	0	0	8	0
p5	0	20,9	2	1	7	0
p6	0	47,2	1	2	18	0
p7	1	48,2	1	1	18	0
p8	1	55,5	2	0	8	0
p9	0	61,1	2	1	18	1
p10	1	34	3	0	10	1
p11	1	29,11	2	0	12	0
p12	1	29	3	0	9	0
p13	1	26	3	0	9	1
p14	1	25,7	3	0	7	0
p15	0	30,7	3	0	10	0
p16	1	21,4	2	0	9	0
p17	0	38,8	2	1	11	0
p18	1	48,3	2	1	16	0
p19	1	36,3	3	0	10	0
p20	1	30,8	2	1	5	0
p21	1	42,9	1	1	5	0
p22	0	36,1	2	1	15	0
p23	0	53,4	3	0	6	1
p24	0	37,5	3	1	10	0
p25	1	31,11	2	0	5	0
p26	1	37,8	3	0	9	0
p27	1	52,6	2	0	18	0
p29	1	36,3	3	1	18	0
p30	0	27,2	3	1	5	0
p31	1	35,4	2	1	10	0
p32	1	34,6	2	1	11	0
p34	0	26,9	2	0	9	1
p35	1	28,3	2	0	5	0
p36	1	26,5	0	0	7	0

Variable	Coding
Gender	0 = male; 1 = female
Age	year, month
Education	0 = secondary; 1 = vocational; 2 = bachelor; 3 = master
Host country	0 = Austria; 1 = Germany; 2 = Switzerland
LoR (years)	18 = 18+
Other host	0 = no; 1 = yes

Table C2.2. *Linguistic background.*

ID	L2 reported	L3 daily use	L4 knowledge	L2 AoA
p1	1	1	0	10
p2	2	1	1	13
p3	1	0	1	18
p4	0	1	0	17
p5	1	1	0	14
p6	1	1	0	14
p7	1	0	1	18
p8	0	0	1	18
p9	2	1	1	12
p10	1	1	0	18
p11	1	0	1	11
p12	1	1	1	14
p13	1	1	0	13
p14	1	1	1	14
p15	0	0	1	15
p16	1	1	1	13
p17	0	0	1	18
p18	0	0	1	18
p19	1	0	1	18
p20	1	1	0	13
p21	0	0	1	18
p22	2	0	0	13
p23	0	0	1	12
p24	1	0	1	18
p25	0	0	0	18
p26	1	1	1	13
p27	2	1	1	15
p29	2	1	1	14
p30	1	1	1	10
p31	0	1	1	10
p32	1	1	1	10
p34	2	1	1	18
p35	1	0	0	15
p36	1	1	1	10
Variable	Coding			
L2 proficiency	0 = B2; 1 = C1; 2 = C2			
L3 daily use	0 = no; 1 = yes			
L4 knowledge	0 = no; 1 = yes			
L2 AoA	18 = 18+			

Table C2.3. *Language- and culture-related reports (composite %).*

ID	Frequency		Orientation		Judgement	
	L1	L2	L1	L2	L1	L2
p1	60,74	52,99	92,59	74,07	44,44	100,00
p2	64,44	21,37	100,00	51,85	0,00	100,00
p3	85,92	29,06	100,00	40,74	0,00	100,00
p4	76,29	57,26	88,89	70,37	66,67	88,89
p5	59,25	52,14	100,00	55,56	0,00	100,00
p6	80,00	87,18	100,00	74,07	100,00	100,00
p7	38,51	62,39	100,00	48,15	0,00	100,00
p8	71,11	36,75	88,89	74,07	0,00	66,67
p9	37,77	66,67	77,78	70,37	22,22	100,00
p10	53,33	41,03	100,00	77,78	77,78	100,00
p11	54,07	63,25	100,00	59,26	66,67	100,00
p12	53,33	54,70	100,00	74,07	44,44	77,78
p13	81,48	95,73	100,00	88,89	100,00	100,00
p14	51,85	46,15	74,07	51,85	44,44	77,78
p15	71,11	34,19	88,89	88,89	0,00	100,00
p16	76,29	76,92	100,00	77,78	33,33	100,00
p17	58,51	33,33	100,00	81,48	44,44	100,00
p18	51,85	44,44	100,00	66,67	0,00	100,00
p19	56,29	72,65	81,48	77,78	0,00	100,00
p20	29,62	88,03	66,67	70,37	0,00	100,00
p21	66,66	40,17	92,59	88,89	44,44	88,89
p22	48,14	70,09	100,00	100,00	88,89	88,89
p23	78,51	19,66	96,30	77,78	0,00	66,67
p24	66,66	55,56	100,00	85,19	55,56	77,78
p25	49,62	55,56	100,00	62,96	77,78	100,00
p26	85,92	44,44	100,00	74,07	44,44	100,00
p27	46,66	59,83	100,00	96,30	100,00	100,00
p29	62,22	70,94	96,30	92,59	0,00	100,00
p30	71,11	26,50	85,19	88,89	0,00	100,00
p31	76,29	5,13	81,48	62,96	0,00	66,67
p32	78,51	71,79	92,59	85,19	66,67	77,78
p34	60,00	47,01	88,89	74,07	33,33	100,00
p35	70,37	51,28	77,78	77,78	55,56	66,67
p36	51,11	57,26	88,89	81,48	33,33	100,00

Table C2.4. *Attitudinal and migration-related reports (composite %).*

ID	Attrition	Affinity	Migration	Repatriation
p1	44,44	55,56	100,00	0,00
p2	100,00	94,44	88,89	77,78
p3	100,00	100,00	77,78	100,00
p4	100,00	88,89	88,89	100,00
p5	100,00	88,89	100,00	100,00
p6	44,44	50,00	100,00	100,00
p7	44,44	77,78	100,00	0,00
p8	100,00	72,22	88,89	77,78
p9	100,00	94,44	88,89	0,00
p10	100,00	88,89	100,00	77,78
p11	100,00	100,00	77,78	11,11
p12	100,00	100,00	66,67	66,67
p13	88,89	94,44	100,00	22,22
p14	77,78	72,22	77,78	11,11
p15	0,00	33,33	100,00	44,44
p16	100,00	94,44	100,00	88,89
p17	100,00	0,00	100,00	44,44
p18	33,33	88,89	77,78	77,78
p19	100,00	100,00	100,00	0,00
p20	88,89	72,22	100,00	66,67
p21	88,89	61,11	100,00	22,22
p22	33,33	72,22	77,78	100,00
p23	0,00	83,33	100,00	88,89
p24	100,00	44,44	77,78	100,00
p25	44,44	77,78	100,00	0,00
p26	100,00	77,78	100,00	22,22
p27	88,89	94,44	77,78	44,44
p29	100,00	100,00	100,00	100,00
p30	100,00	88,89	100,00	0,00
p31	44,44	100,00	100,00	0,00
p32	100,00	88,89	100,00	77,78
p34	100,00	88,89	44,44	44,44
p35	66,67	72,22	55,56	66,67
p36	33,33	100,00	100,00	0,00

C3. Correlation Heatmaps (Pearson's r)

Figure C3.1. Heatmap of all parameters.

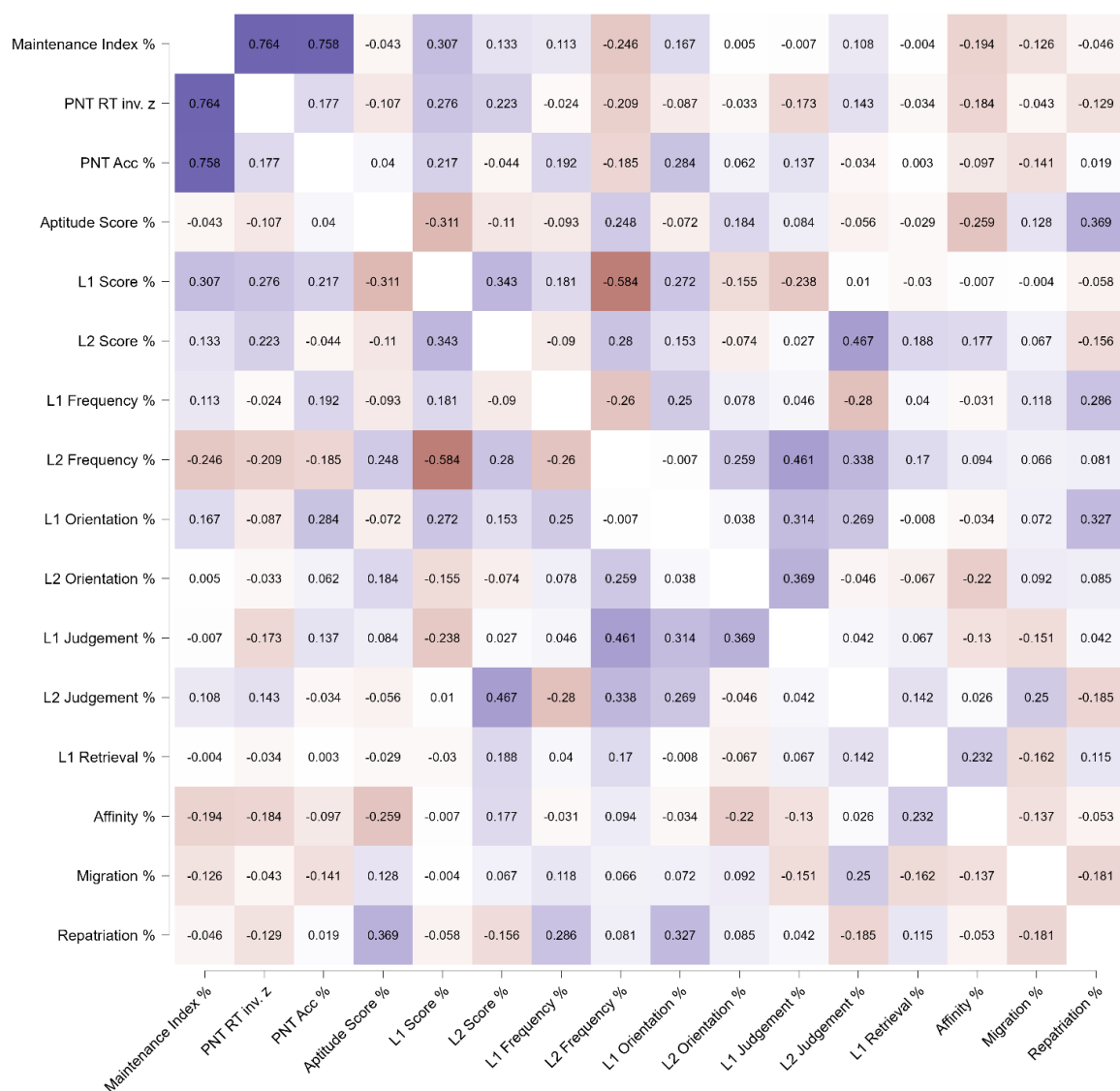


Figure C3.2. *Heatmap of all L1 parameters.*

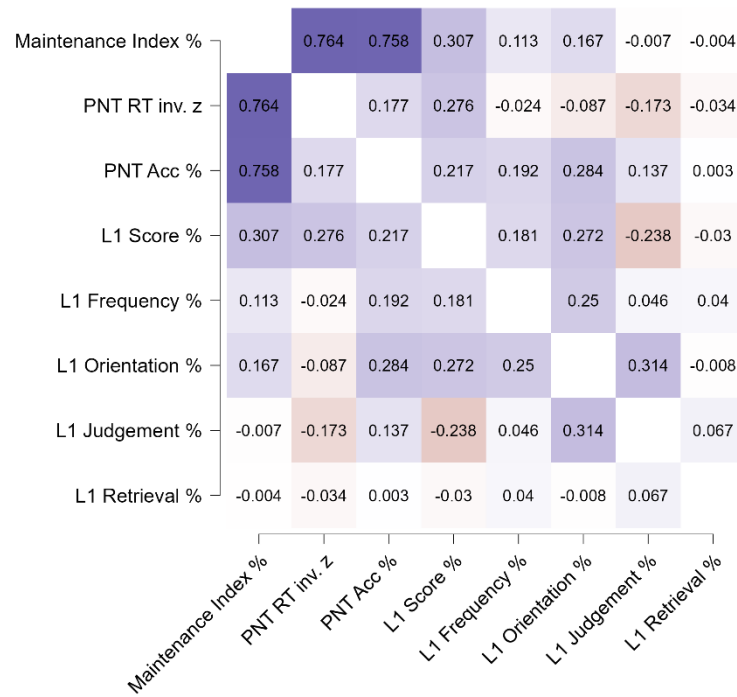
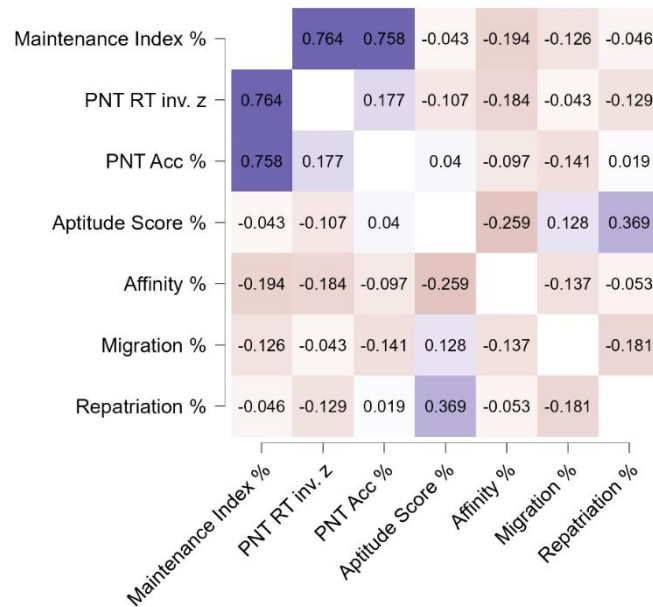


Figure C3.3. *Heatmap of all attitudinal parameters.*



C5. Technical Data

Table C5.1. *Breakdown of the trial configuration of the picture naming task.*

Configuration	Description	Duration
Fixation Cross	Signals the start of the trial.	250 ms
Task Execution	Image display and participant response.	10 s
Trial Progression	Advances automatically or manually.	10 s

Table C5.2 *Technical setup of participants' devices.*

OS	Browser	Monitor	Viewport	Mouse	Microphone
Mac OS 10.15	Firefox 102.0	1440x900	1439x731	Touchpad	Headphones
Windows 10	Chrome 103.0.0.0	1280x720	1280x569	NA	NA
Windows 10	Chrome 103.0.0.0	1366x768	1366x625	NA	NA
Mac OS X 10.5	Chrome 103.0.0.1	1440x900	1440x789	Touchpad	Internal
Windows 10	Chrome 103.0.5060.114	1920x1080	1440x730	Wired mouse	Internal
Windows 10	Firefox 102.0	1536x864	1536x731	Touchpad	Internal
Windows 10	Edge 103.0.1264.71	1280x720	1280x609	Touchpad	Internal
Windows 10	Edge 103.0.1264.71	1920x1080	1872x969	Wireless mouse	Internal
NA	Chrome 103.0.0.0	1440x900	1440x693	NA	NA
Windows 10	Edge 104.0.1293.47	1536x864	1536x714	Touchpad	Internal
Windows 10	Chrome 105.0.0.0	1536x864	1536x714	Wireless mouse	Internal
Windows 10	Chrome 106.0.0.0	1368x912	1368x769	Wireless mouse	Internal
NA	Safari 16.0	1440x900	1315x725	Touchpad	Internal
Windows 10	Edge 106.0.1370.47	1536x864	1536x722	NA	NA
Windows 10	Chrome 107.0.0.0	1536x864	1536x754	Wireless mouse	Internal
NA	Safari 16.0	1440x900	1432x721	Wireless mouse	Internal
Windows 10	Chrome 107.0.0.0	1536x864	1536x754	Touchpad	Internal
Windows 10	Firefox 106.0	1600x900	1284x722	Wired mouse	Internal
NA	Safari 16.0	1680x1050	1680x970	NA	NA
Windows 10	Chrome 107.0.0.0	1061x707	1061x567	Touchpad	Internal
Windows 10	Chrome 108.0.0.0	1680x1050	1680x907	Wired mouse	Internal
NA	Safari 16.1	2240x1260	2240x1180	NA	NA
NA	Chrome 108.0.0.0	1280x800	1280x689	Touchpad	Internal
Windows 10	Firefox 93.0	1366x768	1366x607	NA	NA
Windows 10	Edge 107.0.1418.62	1366x768	1319x657	NA	NA
Windows 10	Firefox 107.0	1366x768	1366x643	NA	NA
Windows 10	Edge 107.0.1418.56	1536x864	1490x754	Wireless mouse	Internal
Windows 10	Chrome 108.0.0.0	1280x720	1280x573	Touchpad	Internal
Windows 10	Edge 108.0.1462.54	1600x900	1600x757	Touchpad	Internal
Windows 10	Firefox 108.0	1366x768	1366x643	Touchpad	Internal
Windows 10	Chrome 108.0.0.0	1536x864	1536x754	Wired mouse	Internal
Windows 10	Chrome 108.0.0.0	1280x720	1280x520	Touchpad	Internal
Windows 10	Chrome 109.0.0.0	1536x864	1366x762	NA	NA
Windows 10	Chrome 111.0.0.0	1280x720	1280x569	Touchpad	Internal
Windows 10	Chrome 113.0.0.0	1536x864	1519x754	NA	NA
Windows 10	Chrome 118.0.0.0	1280x720	1280x595	NA	NA

NA = not reported by the system/participant.