

Japanese Speakers' Second Language Sentence Processing: Behavioral, ERP, and MRI Evidence

シャルボウシュ, ジャナネ

<https://hdl.handle.net/2324/7443210>

出版情報 : Kyushu University, 2025, 博士 (文学), 課程博士
バージョン :
権利関係 :



PhD thesis

**Japanese Speakers' Second Language
Sentence Processing: Behavioral, ERP,
and MRI Evidence**

Jananeh Shalpoush

ジャナネ シャルポウシュ

Graduate School of Humanities
Kyushu University

Supervisor: Professor Shinri Ohta

January 2026

Declaration

I hereby declare that this dissertation entitled “Japanese Speakers’ Second Language Sentence Processing: Behavioral, ERP, and MRI Evidence” is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University.

Those parts of this thesis which are based on previously published or publicly available work are as follows:

- 1) Chapter 2 and the corresponding sections of the General Discussion are based on the following preprint:

Shalpoush, J., Gallagher, D. C., Yamada, E., & Ohta, S. (2025).

Non-Canonical ERP Patterns in Second-Language Sentence Processing: Evidence from Syntactic, Semantic, and Phonological Violations in Japanese Learners of English. bioRxiv. <https://doi.org/10.64898/2026.01.05.697613>

All other parts of this dissertation have not been previously published or submitted for any other degree or qualification.

Jananeh Shalpoush

Kyushu University

January 2026

Table of Contents

Chapter 1 Background

1.1 Introduction.....	8
1.2 Research Areas of Interest and Purpose of This Study.....	12
1.3 Structure of the Thesis.....	15
1.4 Significance and Contribution to Linguistics.....	16
1.5 Novelty and Originality of the Present Study.....	21

Chapter 2 ERP–Behavioral Experiment

2.1 Introduction.....	23
2.2 Methods.....	23
2.2.1 Participants.....	23
2.2.2 Stimuli.....	24
2.2.3 Procedure.....	28
2.2.4 EEG Recording and Preprocessing.....	29
2.2.5 Behavioral and ERP Statistical Analysis	30
2.3 Results.....	32
2.3.1 Behavioral Results.....	32
2.3.2 ERP Results.....	35
2.4 ERP–Behavioral Correlations.....	42
2.5 Discussion.....	46

Chapter 3 MRI Analysis of Cortical Thickness

3.1 Introduction.....	51
-----------------------	----

3.2 Methods.....	53
3.2.1 Participants.....	55
3.2.2 MRI Data Acquisition.....	55
3.2.3 Cortical Thickness Estimation and ERP–MRI Correlation Analysis.....	56
3.3 Results.....	57
3.3.1 ERP–MRI Structure–Function Associations	57
3.4 Discussion.....	63

Chapter 4 General Discussion

4.1 Introduction.....	67
4.2 Interpretation of Behavioral and ERP Results.....	67
4.3 Structure–Function Relationships from ERP–MRI Correlations.....	72
4.4 Summary.....	75

Chapter 5 Conclusions

5.1 Introduction.....	77
5.2 Limitations.....	79
5.3 Future Directions.....	80
5.4 Final Remarks.....	82

Acknowledgements.....	84
------------------------------	-----------

References.....	85
------------------------	-----------

Appendix A.....	90
------------------------	-----------

Appendix B.....	106
------------------------	------------

Appendix C.....	111
------------------------	------------

List of Figures

Figure 1. Overview of the experimental sentence design	25
Figure 2. Trial sequence of a single experimental trial	29
Figure 3. Mean accuracy by linguistic condition	33
Figure 4. Mean accuracy by cue condition	34
Figure 5. Mean reaction times by cue condition	35
Figure 6. Grand-average ERP difference waves by condition and ROI	36
Figure 7. Scalp topographies of ERP difference waves (N400, P600)	41
Figure 8. Correlation between phonological accuracy and syntactic ERP	44
Figure 9. Correlation between syntactic accuracy and syntactic ERP	45
Figure 10. Correlation between semantic ERP (Parietal) and right posterior cingulate	58
Figure 11. Correlation between semantic ERP (Centroparietal) and right posterior cingulate	59
Figure 12. Correlation between semantic ERP and left rostral middle frontal	60
Figure 13. Correlation between phonological ERP and right pars opercularis	61
Figure 14. Correlation between phonological ERP and left fusiform gyrus	62
Figure 15. Correlation between syntactic ERP and right rostral ACC	63

Abstract

Despite substantial progress in the neurocognitive study of second-language (L2) processing, few studies have systematically compared syntactic, semantic, and phonological processing within a single experimental framework while linking real-time neural dynamics to brain structure. This thesis examines L2 sentence comprehension in native Japanese speakers with intermediate English proficiency by integrating behavioral measures, event-related potentials (ERPs), and structural MRI. Participants completed a cued auditory sentence judgment task involving well-formed English sentences and sentences containing syntactic, semantic, or phonological violations. Behavioral results showed highest accuracy for control and semantic conditions, with significantly lower performance for syntactic and phonological violations. ERP analyses revealed non-canonical neural response patterns: phonological violations elicited the most robust and sustained negativities across both 300–500 ms and 500–800 ms time windows, whereas syntactic and semantic anomalies evoked weaker or less consistent effects. These findings suggest that phonological mismatches disrupt lexical activation and downstream integration processes, producing broad negativities that overlap with, but are not equivalent to, the canonical N400. Structural MRI analyses further demonstrated systematic relationships between ERP amplitudes and cortical thickness in perisylvian and executive-control regions, indicating that individual neuroanatomical variability shapes functional processing in L2 comprehension. Together, the results support a view of L2 sentence processing in native Japanese speakers as relying on adapted and partially redistributed neural networks shaped by

L1–L2 typological differences, processing demands, and individual brain structure. This multimodal approach provides a more comprehensive account of variability in L2 sentence comprehension.

Keywords:

Second language (L2) acquisition; auditory sentence processing; electroencephalography (EEG); event-related potentials (ERP); speech perception; Japanese; syntax; semantics; phonology; structural magnetic resonance imaging (MRI); cortical thickness; neurocognitive processing

Chapter 1

Background

1.1 Introduction

The human brain is a highly dynamic and interconnected system in which cognitive functions arise from the coordinated activity of distributed neural networks rather than isolated cortical regions [1]. Language, one of the most complex human cognitive abilities, is supported by a widespread left-lateralized fronto-temporo-parietal network underpinning both comprehension and production [2,3]. Within this network, the left inferior frontal gyrus (LIFG), commonly known as Broca's area, is especially important for morphosyntactic processing, working memory, and controlled aspects of language production [4–6]. In contrast, phonological decoding and lexical-semantic mapping are more centrally supported by the left superior temporal gyrus (LSTG) and adjacent temporal areas including the superior temporal sulcus (STS) [7,8]. Regions within this larger network function in synergy, communicating through temporo-frontal pathways including the arcuate fasciculus that enable real-time integration of form and meaning and sound [9,10].

Importantly, this network interacts with the parietal and subcortical regions, forming a distributed yet functionally specialized language system capable of adapting to varying

linguistic contexts and demands [2,3]. While this framework is well established for native (L1) language processing, considerably less is known about how these neural systems adapt to a second language (L2), particularly when the L1 and L2 strongly differ in typology, phonology, and morphosyntax [11–14]. This issue is especially important because it can provide valuable insight into neural plasticity and the stability of linguistic representations across different processing contexts.

L2 acquisition generally occurs after the period of maximal neural plasticity for language; therefore, it relies on at least partially distinct neural strategies from those used in native language processing. Although both languages similarly engage a common set of cortical regions, L2 processing is often associated with more extensive and bilateral networks, reflecting greater cognitive effort, working memory load, and control demands. In this way, L2 processing may not be a simple reuse of L1 networks but rather an adapted configuration shaped through proficiency, exposure, and L1–L2 structural differences [11–15].

Thus, understanding how multiple linguistic subsystems are processed in real-time is essential for the characterization of the nature and limits of neural plasticity in adult bilinguals. One of the most fundamental questions in L2 speech perception is how the brain integrates syntactic, semantic, and phonological information during the comprehension of spoken sentences in a L2. These domains interact dynamically, and their interplay reflects the underlying mechanisms of linguistic representation and processing efficiency.

Event-related potentials (ERPs) provide key insights into the temporal dynamics of these processes. The N400, peaking around 400 ms, indexes semantic expectancy and contextual integration [16,17]. The P600, a late positivity often emerging around 500–800 ms,

classically has been linked to syntactic reanalysis and repair, although its functional interpretation is nowadays understood to be broader and to encompass conflict monitoring, controlled retrieval, and integration difficulty [10,18–20].

Various ERP studies on second-language learners revealed a non-canonical response pattern where ERP components differed in amplitude, latency, or scalp distribution from those observed in native speakers [21–25]. For instance, some L2 learners may show delayed or reduced N400 effects to semantic violations, or exhibit P600-like responses to semantic rather than syntactic anomalies, a phenomenon termed as "semantic P600" [19,21,22]. Such atypical patterns are thought to reflect differences in automaticity and reliance on conscious reanalysis strategies, and the influence of the L1 system on L2 processing [26,27].

These findings suggest that L2 sentence processing is characterized by reduced automaticity, increased reliance on controlled processes, and greater susceptibility to L1 transfer, especially when L1 and L2 diverge substantially [26–28].

Despite notable advances in the neural mechanisms of L2 processing, few studies have systematically compared multiple linguistic categories within one and the same experimental design [21,24,25,29,30]. There is a striking absence of comparable data, in particular for auditory sentence comprehension in late, unbalanced Japanese L2 learners. To date, most studies have examined these linguistic domains independently, which limits our understanding of how the brain concurrently processes and integrates multiple linguistic subsystems during real-time comprehension.

Moreover, the neurocognitive mechanisms driving the way that Japanese learners of English perceive and process linguistic information remain largely unexplored. Taking into consideration the considerable structural and typological differences between Japanese and English, a comprehensive multimodal approach is indispensable [31–33].

Beyond filling a descriptive gap, the current work also engages directly with fundamental theoretical debates in second-language sentence processing. One influential proposal is Clahsen and Felser’s “Shallow Structure Hypothesis” [26], which argues that adult L2 learners often rely on lexical and semantic cues and build only partial or underspecified syntactic representations during real-time comprehension. In this light, morphosyntactic information such as agreement morphology or function words is not fully utilized online, particularly under time pressure, and learners may succeed in offline grammaticality judgments without constructing the same depth of syntactic structure as native speakers. The present thesis directly tests this hypothesis by targeting exactly those grammatical domains that are weakly encoded in Japanese (agreement, articles, countability, prepositions) and by comparing syntactic processing with semantic and phonological processing within one experimental paradigm. Consequently, if syntactic cues are indeed less integrated than semantic or phonological information, this should be reflected not only in behavioral accuracy but also in the presence, absence, or modulation of canonical ERP markers of structural reanalysis.

Another important line of research concerns predictive and probabilistic accounts of sentence comprehension in L2. Contemporary models generally propose that native speakers use contextual information to generate expectations about upcoming words and structures, and that ERP components such as the N400 are sensitive to violations of these

expectations [16,17,27]. In L2 learners, however, it remains unclear to what extent predictive mechanisms are engaged, and whether such prediction relies primarily on lexical-semantic associations, grammatical structure, or surface phonological cues [27]. By manipulating syntactic, semantic, and phonological well-formedness in spoken sentences while recording time-locked ERPs, the present study is well suited to address this issue. As shown in later chapters, the behavioral and neural response patterns observed in Japanese learners of English speak directly to the question of whether L2 sentence comprehension involves deeply structured prediction or a more conservative, cue-based processing strategy that privileges certain linguistic domains over others.

1.2 Research Areas of Interest and Purpose of This Study

The purpose of this study is to investigate how Japanese L2 learners of English process different linguistic information such as syntax, semantics, and phonology during auditory sentence comprehension. While these domains are often studied in isolation, natural language comprehension requires their real-time integration. Yet, few studies have systematically compared these domains within a single experimental design.

This discrepancy is particularly pronounced in late, unbalanced bilinguals, where the typological distance between the L1 (Japanese) and the target L2 (English) is vast, specifically in terms of grammar, word order, phonology, and morphosyntactic structures [31–33]. Understanding how Japanese learners of English navigate these differences offers valuable insights into the limits of neural plasticity, the role of L1 transfer, and the developmental trajectories of L2 processing strategies. Additionally, Japanese lacks certain grammatical markers that are obligatory in English (e.g., pluralization, article usage), and its syllabic structure differs markedly from English phonotactics, which may affect L2

phonological and syntactic acquisition [31–33].

To fill these gaps, the present study explored how late Japanese learners of English process syntactic, semantic, and phonological anomalies during auditory sentence comprehension.

In particular, this study sought to answer three central questions:

1. How do different types of linguistic information (grammatical, phonological, and semantic) influence sentence judgment accuracy in Japanese L2 learners of English?
2. What are the temporal and spatial characteristics of ERP responses (N400, P600) to syntactic, semantic, and phonological violations in L2 learners, and do these responses resemble the canonical components typically observed in native speakers?
3. How do behavioral accuracy, neural responses, and anatomical brain structure relate across linguistic conditions?

To address these questions, a controlled auditory sentence judgment task was implemented. Participants heard sentences that were either well-formed or contained syntactic, semantic, or phonological violations. Before each trial, a Japanese cue directed participants to evaluate the upcoming sentence based on the cued linguistic dimension. Their task was to judge whether each sentence was acceptable with respect to the given cue.

Behavioral accuracy data were collected alongside EEG recordings to capture neural responses with millisecond-level temporal precision. ERP components of interest included the N400 (300–500 ms), associated with semantic processing and lexical integration [16,17], and the P600 (500–800 ms), linked to syntactic reanalysis and conflict monitoring

[10,18–20]. To further explore structure-function associations, structural MRI data were collected to investigate how cortical thickness in language-related brain regions correlates with functional ERP components.

Based on existing L2 ERP research, we hypothesized that semantic violations would elicit N400 effects, though potentially delayed or reduced in amplitude relative to native-like responses [21,22,24,25,36]. Syntactic violations were expected to evoke late positivities (P600), reflecting controlled reanalysis processes. Phonological violations were expected to elicit early-to-mid latency neural modulations, potentially cascading into semantic processing as reflected in the N400 time window. Overall, these predictions reflect canonical modulations consistent with adaptive neural reorganization in L2 processing, indicative of increased cognitive effort and reduced automaticity. However, given the typological distance between Japanese and English, we also expected ERP responses to deviate from canonical L1 patterns. By integrating behavioral accuracy with time-resolved neural measures, this study aims to provide a multidimensional characterization of how native Japanese L2 learners process multiple linguistic domains during auditory sentence comprehension. These findings provide empirical support for the notion that L2 sentence comprehension in late Japanese bilinguals engages a qualitatively distinct neural configuration, shaped by the structural characteristics of Japanese and the processing demands of English.

Further, we hypothesized that individual differences in cortical thickness within core language-related regions (e.g., left inferior frontal gyrus, bilateral superior temporal gyri, and temporoparietal areas) would relate to functional ERP responses and behavioral performance across linguistic domains. For example, larger cortical thickness in the left

IFG may be associated with higher P600 amplitudes in response to syntactic violations, due to greater morphosyntactic processing capacity, whereas temporal or parietal regions may indicate associations with semantic or phonological processing demands. These structure-function relationships were expected to support the view that L2 processing in late bilinguals is scaffolded by both functional adaptation and underlying anatomical variability [11–15].

1.3 The Structure of the Thesis

This thesis is organized into five chapters, each contributing to a comprehensive understanding of second-language sentence processing in Japanese learners of English through behavioral, electrophysiological, and neuroanatomical methods.

Chapter 1 (Background) introduces the theoretical and empirical foundations of the study, including key concepts in second-language acquisition, sentence comprehension, neural plasticity, and event-related potentials. It also outlines the research questions and hypotheses and highlights the significance of adopting a multimodal approach to L2 processing.

Chapter 2 (ERP–Behavioral Experiment) details the experimental design, materials, and procedures used to investigate how syntactic, semantic, and phonological violations are processed during L2 auditory sentence comprehension. It presents both behavioral results and ERP findings across defined time windows (N400, P600) and scalp regions of interest. It also highlights the correlation between ERP responses and sentence judgment accuracy by identifying functional neural markers which are associated with individual differences in L2 processing performance.

Chapter 3 (MRI Analysis of Cortical Thickness) focuses on structural brain data, providing a region-by-region analysis of cortical thickness and exploring how these anatomical features may support language processing in bilinguals. It also describes the extraction of cortical thickness measures and the analytical approach to linking them with ERP amplitudes. Moreover, it presents cross-modal associations between ERP components and cortical thickness across brain regions, revealing how anatomical features may support or constrain functional processing in L2 learners.

Chapter 4 (General Discussion) integrates the behavioral, ERP, and MRI findings to discuss the cognitive and neural mechanisms underlying L2 sentence processing. This chapter reflects on the implications of domain-specific processing patterns, non-canonical ERP profiles, and structural variability in bilingual brains, and situates the findings within the broader literature on L2 neurocognition and bilingualism.

Finally, Chapter 5 (Conclusions) summarizes the main findings and theoretical contributions of the thesis, discusses limitations, and proposes directions for future research in the neuroscience of second-language processing. It concludes with final remarks on the significance of the present multimodal approach for understanding L2 sentence comprehension.

1.4 Significance and Contribution to Linguistics

The present thesis lies at the intersection of linguistics, psycholinguistics, and cognitive neuroscience, but its primary goal is to address questions that are central to linguistics: how syntax, semantics, and phonology are represented and engaged during sentence comprehension, and how these components are shaped by the learner's first language as well as by the structure of the brain. By examining Japanese speakers' comprehension of English sentences through

behavioral performance, electrophysiological responses, and cortical structure, this work contributes to a more precise understanding of how linguistic knowledge is instantiated and deployed in a L2.

From the perspective of theoretical and descriptive linguistics, L2 sentence comprehension is informative because it shows which aspects of the grammar and phonological system are robust and which parts remain fragile under processing pressure. In this project, syntactic, semantic, and phonological information are manipulated in a controlled way but always within full sentences. This design allows us to examine one essential and fundamental linguistic issue: when Japanese speakers listen to English sentences, which linguistic cues do they actually rely on to construct an interpretation? Are grammatical morphemes (agreement, articles, prepositions) actively used online during processing, or do learners depend more heavily on lexical semantics and surface phonological patterns? The behavioral and ERP findings in this thesis indicate that, even when explicit cues direct attention to grammatical or phonological correctness, these domains are less reliably employed than semantic information. This pattern has direct implications for models of grammatical representation in L2, as it implies that certain morphosyntactic features might be represented more weakly or accessed less efficiently than lexical meaning, particularly when the L1 does not encode the same differences.

More broadly, this study can be considered as a part of a long-standing and ongoing question in linguistics: the relationship between grammatical competence and real-time processing. Traditional linguistic theories often describe a fixed system of rules and representations; on the other hand, psycholinguistic and neurolinguistic experiments emphasize the dynamics of using those representations in real-time. By highlighting the neurophysiological underpinnings of sentence comprehension in L2, this thesis provides

concrete evidence about how “grammar,” “meaning,” and “sound” are implemented in the brain during actual comprehension. The non-canonical ERP patterns which are observed in this study and will be discussed in the following chapters, show that having some explicit knowledge of grammar does not guarantee native-like processing routines. For linguistics, this means that theories of representation should be compatible and consistent with the idea that access to those representations can be greatly modulated by proficiency level, typological distance, and processing demands.

The choice of Japanese–English as an L1–L2 pairing is also of theoretical interest. Japanese and English have systematic differences in morphosyntactic and phonology, whereby Japanese does not obligatorily mark plurality or definiteness in the same way as English. Moreover, Japanese employs a different syllabic structure and segmental contrasts. By focusing on agreement, article usage, countability, and segmental phonology, this thesis examines exactly those domains where the two grammars diverge. The findings, therefore contribute to typologically informed linguistics by showing how specific L1–L2 differences shape the online use of grammatical and phonological cues. For instance, the persistent difficulties with phonological and morphosyntactic violations in English can be interpreted as evidence that certain categories and distinctions are not easily mapped to pre-existing Japanese representations, and that has implications for how we think about cross-linguistic influence.

Another important contribution of this thesis is both methodological and conceptual. It treats syntax, semantics, and phonology not as isolated modules, but as interacting subsystems that must be coordinated moment by moment during sentence comprehension. Rather than using separate experiments for each domain, the study employed closely matched sentence materials that were systematically manipulated to create syntactic,

semantic, and phonological violations. This unified design allows more direct cross-domain comparison while reducing confounds introduced by using entirely different materials across conditions. From a linguistic perspective, this approach provides a clearer view of how different levels of representation interact when listeners interpret spoken language.

The multimodal approach of this thesis, which integrates EEG and MRI in the same group of participants, further contributes to linguistic theories by linking abstract constructions such as syntactic processing or semantic integration to concrete properties of brain structure. For Experiment 1 (the ERP-behavioral experiment), the design of the study aims to address questions about the relative weight of various linguistic cues. However, ERP data alone cannot explain why some learners show stronger or more canonical responses than others. Experiment 1 reveals how the sentence is processed in real-time, but it does not tell us why individuals differ in their linguistic processing profiles.

Hence, Experiment 2 (the cortical thickness analysis) was conducted to complement and extend the ERP findings. Although, structural MRI does not capture moment-to-moment processing, it provides information about relatively stable, individual differences in the neuroanatomical substrate that supports language. By correlating cortical thickness with ERP amplitudes in different conditions, Experiment 2 addresses a different but linguistically relevant perspective.

The combination of Experiments 1 and 2 therefore has a clear logic. Experiment 1 shows the canonical or non-canonical patterns in Japanese learners of English when processing syntactic, semantic, and phonological information, revealing domain-specific vulnerabilities and adaptations. However, ERP data alone could not reveal whether these patterns are based purely on functional processing strategies or are influenced and

constrained by individual neuroanatomical profiles. Experiment 2 fills this gap by demonstrating systematic structure–function relationships in both perisylvian and executive-control regions, thus showing that individual anatomical variability is not just a general cognitive factor but is meaningfully related to how different components of the linguistic system are implemented during sentence comprehension. Together, the two experiments allow this thesis to argue that L2 sentence processing is supported by adapted, partially redistributed neural networks whose configuration is shaped both by linguistic experience and by individual brain structure.

Taken together, these findings contribute to linguistics in several important ways. First, they provide a specific empirical investigation of how different types of linguistic information interact in an adult L2 system under real-time constraints, which is directly relevant to theories of sentence processing, feature checking, and form–meaning mapping. Second, they show that phonological and morphosyntactic processing in L2 can remain non-automatized even when basic comprehension succeeds, reinforcing the need for linguistic theories that distinguish between competence, access, and processing routes. Third, by relating ERP patterns to cortical thickness in language-related regions, the thesis maps abstract linguistic categories to specific neural substrates, helping to ground linguistic theory in the biology of the brain without reducing it to purely anatomical descriptions.

Lastly, our work will have implications in applied and educational linguistics. Understanding which aspects of the L2 (for example, segmental contrasts, plural marking, or article use) are particularly vulnerable in online comprehension can help the design of teaching materials and assessment tasks that focus more directly on those areas. Furthermore, highlighting the individual differences in brain structure and neural sensitivity and their possible connection to L2 processing outcomes might suggest that

learners may benefit from tailored approaches that not only take their explicit knowledge into account, but also consider their processing profiles and cognitive strengths. Therefore, this interdisciplinary, multimodal approach is not only compatible with the goals of linguistics but also enriches them by grounding linguistic theory in the dynamics and structure of the brain, and providing a complete account of what it means to know and process a L2.

1.5 Novelty and Originality of the Present Study

In addition to its theoretical and methodological implications, the present thesis offers several novel contributions to the study of second-language sentence processing and linguistics more broadly. To our knowledge, it is among the few studies to examine syntax, semantics, and phonology together within a single auditory sentence judgment paradigm. Most previous ERP studies of L2 processing have focused on only one linguistic domain at a time. By using closely matched sentence sets across conditions, the present study enables a more direct comparison of how grammatical, semantic, and phonological information interact during real-time sentence comprehension.

Second, the research targets late and unbalanced Japanese learners of English, a group that has been relatively under-represented in multimodal neurolinguistic work. Also, there is considerable typological divergence in the morphosyntactic and phonological profiles of Japanese and English. So by targeting areas where the two languages vary, the study offers a typologically informed test case for how L1 structure shapes L2 sentence processing.

Third, the use of trial-by-trial linguistic cues in the L1 (Japanese) is also a novel aspect of the

study. By explicitly cueing “grammar”, “meaning”, or “phonology” before each sentence, it is possible to dissociate domain-specific attention from domain-general comprehension, and reveal what happens when learners are explicitly asked to judge a certain type of information.

A fourth source of originality lies in the combination of behavioral, ERP, and structural MRI data from the same participant group. Although electrophysiological studies have frequently examined ERP responses to L2 anomalies, and structural research has related language skills to cortical thickness, and it is still rare to link online sentence-level ERPs to cortical thickness within a single experimental setting that also manipulates multiple linguistic subsystems.

Fifth, a few studies have highlighted cross-domain and cross-measure relationships in L2 research. For instance, researchers have examined the correlation between ERP amplitudes for a given condition and their corresponding behavioral accuracy scores to assess processing effort. Similarly, studies have explored the correlations between ERP amplitudes and cortical thickness to link functional neural responses with structural brain properties.

Overall, from a linguistic perspective, this thesis is original in its connection of abstract linguistic categories to their neurobiological implementation. It offers a novel and interdisciplinary contribution to the field, demonstrating how linguistically informed experimental design can illuminate the neurocognitive basis of second-language sentence comprehension in native Japanese speakers.

Chapter 2

ERP–Behavioral Experiment

2.1 Introduction

Based on the theoretical and empirical foundation established in Chapter 1, this chapter explores the experimental methods employed to examine how Japanese learners of English process syntactic, semantic, and phonological information in the auditory sentence comprehension process. A controlled sentence judgment task was developed in which participants listened to sentences that were either well-formed or contained a specific linguistic manipulation, and made acceptability judgments based on a preceding linguistic cue. Both behavioral accuracy and event-related potentials (ERPs) were recorded to assess real-time processing across linguistic domains. The chapter first describes participants, materials, and procedures, followed by EEG recording and preprocessing protocols. Finally, it outlines the statistical approaches used to analyze both behavioral and ERP data in relation to the research questions advanced in the study.

2.2 Methods

2.2.1 Participants

Twenty-seven native speakers of Japanese (14 males, 13 females; age range = 18–25 years, $M \pm SD = 21.0 \pm 2.0$) participated in the study. All participants were undergraduate or graduate students at Kyushu University (Fukuoka, Japan), had normal or corrected-to-normal vision, and were self-identified as right-handed. None reported a history of neurological, speech, or hearing disorders, nor had any participant lived in an English-speaking country for more than

one year. All participants had successfully passed the mandatory, credit-bearing English language course requirement administered by Kyushu University, which typically corresponds to an intermediate level of English proficiency required for graduation. Completion of this course served as an index of participants' formal English learning experience. Additionally, participants provided self-assessed ratings of their English listening proficiency on a 7-point scale ($M = 4.6$, $SD = 1.0$), offering an individualized measure of their perceived auditory comprehension ability.

Informed consent was obtained from each participant after a full explanation of the experiments, in accordance with the World Medical Association's Declaration of Helsinki. The experimental procedures were approved by the Ethics Committee of the Faculty of Humanities at Kyushu University.

2.2.2 Stimuli

The stimuli consisted of 360 English sentences, including 240 main and 120 filler sentences. The main sentences were organized into four linguistic conditions targeting common sources of difficulty for Japanese learners of English: subject-verb agreement, prepositional usage (on vs. of), countability distinctions in nouns, and definiteness and article choice (a/an/the). Within each category, sentences were distributed across four sub-conditions: syntactically correct, syntactic violation, phonological violation, and semantic violation (Fig. 1).

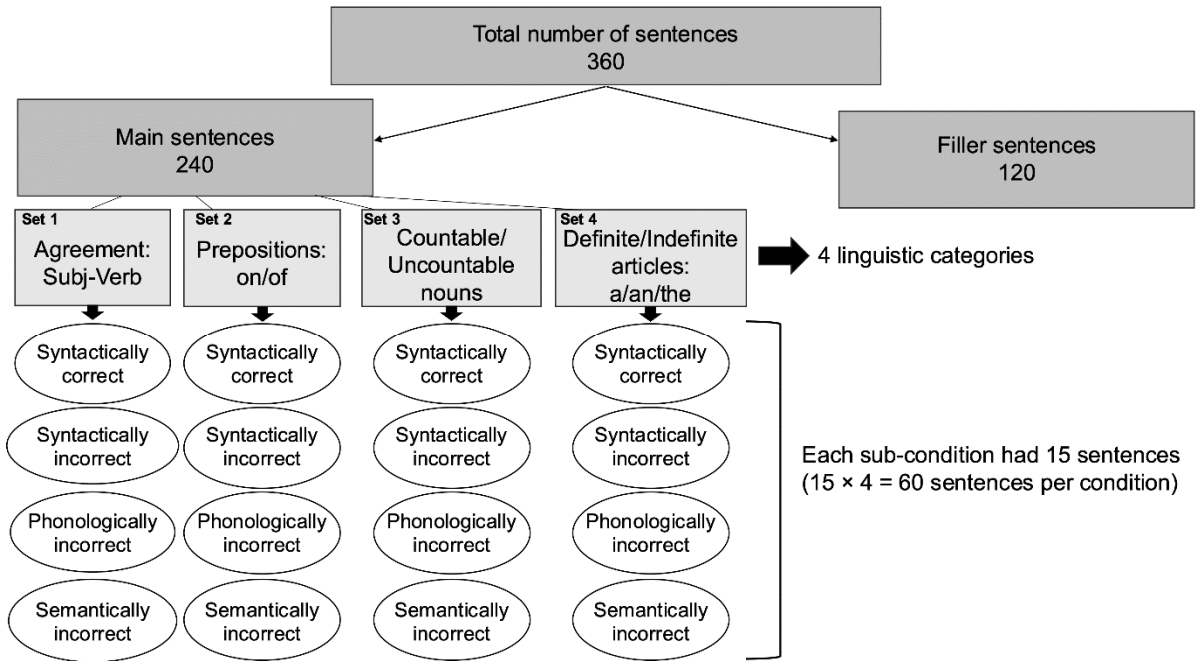


Figure 1. Overview of the experimental sentence design. Each of the four linguistic categories included four sub-conditions, with 15 sentences per condition (60 per category), totaling 240 main sentences plus 120 fillers.

The syntactic violations encompassed subtypes belonging to the target domain (e.g., subject-verb agreement mismatches, omission of required prepositions, ungrammatical count/non-count usage, and omission of articles). For instance, in the subject-verb agreement set, a representative violation involved incorrect verb inflection (e.g., *The cat chase the mouse*), while in the prepositional set, violations omitted required prepositions (e.g., *She ran out time*). Countability violations involved omission of plural marking (e.g., *She saw four bird*), and definiteness violations omitted obligatory articles (e.g., *She visits friend*). This heterogeneity was intentional, as the aim was to broadly probe L2 learners' sensitivity to different types of syntactic anomalies and to examine whether a unified ERP response pattern would emerge across structurally distinct violations. By including diverse violation types, we hoped to explore whether the neural response to syntactic anomalies in L2 learners is generalized or dissociable

across subtypes. For statistical analyses, these items were collapsed into a single syntactic condition to maintain statistical power and comparability across domains. We discuss the implications of this decision in the Discussion section.

Semantic violations followed a consistent structure across all categories, replacing a critical element with an implausible yet grammatically well-formed alternative (e.g., *The cat studies the mouse*; *She thunders beautifully*; *She bought some happiness*; *She visits a carpet*). Phonological violations were also consistent across sets, created by making a controlled mispronunciation (e.g., *The cat shases the mouse*).

From a psycholinguistic perspective, the semantic violations in this study are best characterized as lexical–semantic anomalies and selectional restriction violations. In each case, the critical word was grammatically well-formed but semantically inappropriate given the preceding context. Such anomalies violate stored knowledge about typical event structures and argument–predicate combinations, and therefore target the lexical–semantic system and world knowledge rather than syntactic form. This design follows the classic N400 literature in which semantically anomalous but syntactically well-formed words are used to probe contextual integration and prediction during comprehension.

The phonological violations were designed as segmental-level mispronunciations of otherwise familiar words. In most items, a single consonant was substituted, yielding nonwords or low-likelihood forms while preserving the original syllable count and overall prosodic structure. In psycholinguistic terms, these stimuli target sublexical phonological encoding and phoneme-to-lexicon mapping, making a mismatch at the level of individual segments rather than at the level of stress pattern or sentence-level prosody. Because the resulting strings are phonotactically

plausible English-like forms, they approximate the kind of mispronunciations that L2 listeners may encounter in real communicative settings.

The syntactic violations can be classified as local morphosyntactic anomalies involving agreement and function word usage. All of the syntactic manipulations preserved lexical identity and basic word order but disrupted inflectional morphology or required function words. They therefore test learners' sensitivity to the grammatical scaffolding of the sentence instead of the lexical plausibility or phonetic well-formedness.

Each condition contained 15 sentences, yielding a total of 60 target sentences per category and 240 target sentences overall. The 120 filler sentences served as baseline control items and were grammatically, semantically, and phonologically well-formed and natural. All sentences were constructed by the authors to ensure precise control over lexical content, grammatical structure, and sentence length, and were designed to be easily interpretable by native Japanese speakers with intermediate English proficiency.

For auditory stimulus presentation, the sentences were synthesized using the text-to-speech software Ondoku3.com, ensuring a uniform speech rate, intonation, and pronunciation across all sentences, thereby minimizing acoustic variability that could confound ERP measurements. Half of the sentences ($n = 180$) were synthesized with a male voice, and the remaining half ($n = 180$) with a female voice, providing balanced gender representation across the auditory stimuli.

A complete list of sentences, including main sentences and the filler sentences, is provided in Appendices A and B.

2.2.3 Procedure

The experiment was performed using Presentation software (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), which controlled stimulus presentation, timing, and response collection.

Prior to the main experimental session, participants received task instructions in Japanese and completed a practice phase designed to familiarize them with the procedure. Feedback was provided exclusively during this practice phase. The main experimental phase commenced after confirming, through verbal interaction, that each participant fully understood the task requirements. This approach was used to ensure that participants did not proceed to the main task until they demonstrated a solid understanding of it.

Participants then completed an auditory sentence judgment task while their brain activity was continuously recorded using electroencephalography (EEG). The task was specifically designed to assess how Japanese learners of English process distinct linguistic domains (syntactic, semantic, and phonological information) during real-time sentence comprehension.

Each trial began with fixation crosses (“+++”) presented at the center of the screen for 500 ms, during which participants were instructed to maintain fixation and minimize blinking (Fig. 2). The fixation was followed by a blank screen interval of 5000 ms preceding the onset of the stimulus sequence. To indicate the type of linguistic information on which participants needed to evaluate each trial, a written cue was displayed at the beginning of every single trial. The cues used in this study consisted of a single Japanese word centered on the screen: “文法” (grammar), “意味” (meaning), and “音韻” (phonology). All these cues were presented in standard-font, white-colored text without any accompanying images, shapes, or color-coding,

and remained on the screen for 5000 ms before the sentence was played. Immediately following the cue, participants heard a sentence and were instructed to judge its acceptability with respect to the cued dimension by pressing one of two buttons on a response pad: the blue button to indicate “correct” and the red button to indicate “incorrect”. The auditory sentences had an average duration of approximately 3000 ms. An inter-trial interval of 200 ms was included between trials to ensure clear separation of successive events. The experiment was organized into 15 blocks of 24 trials (360 total). After each block, participants were provided with a brief rest period, during which the laboratory door was opened to reduce fatigue and maintain alertness. The entire experimental session, including task instructions, practice trials, the main task during which EEG data were recorded, and short breaks, lasted approximately 70 min.

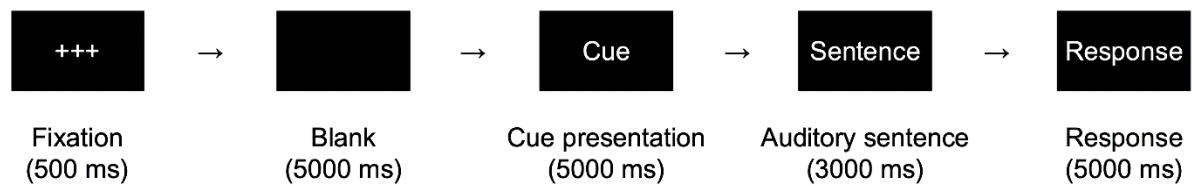


Figure 2. Trial sequence of a single experimental trial.

2.2.4 EEG Recording and Preprocessing

EEG data were recorded continuously using 32 Ag/AgCl electrodes positioned according to the International 10-10 system. The signal was acquired using a Neurofax EEG-1200 system (Nihon Kohden), in combination with an actiCAP electrode cap (Brain Products). Vertical and horizontal electrooculographic (EOG) activity was monitored using electrodes placed below the left eye and at the outer canthus. The electrode impedance was consistently maintained below 20 k Ω throughout the experiment. During acquisition, the EEG was sampled at 1000 Hz. The signals were recorded relative to a specific online reference (FCz).

EEG data preprocessing and analysis were performed using MNE-Python (version 1.5.1) and custom Python scripts. All procedures were performed in a Python 3.11 (64-bit) environment using the Spyder IDE (Anaconda distribution) [34]. First, event lists were created to identify and label trigger events in the continuous EEG recordings. Following this, bad channels were visually inspected, manually rejected, and subsequently interpolated to preserve the overall channel structure. The continuous EEG data were then subjected to standard preprocessing steps, including high-pass filtering and re-referencing. Ocular artifacts were corrected using independent component analysis (ICA), with components corresponding to blink activity identified and removed. The cleaned data were then segmented into epochs spanning from –200 ms to 2000 ms relative to the stimulus onset. Baseline correction was applied using the pre-stimulus interval (–200 to 0 ms). Automated bad epoch rejection was performed using the autoreject algorithm to identify and exclude excessively noisy trials [35].

Grand average Event-Related Potentials (ERPs) were computed by averaging the cleaned epochs across trials within each experimental condition for each participant. The resulting averaged data were low-pass filtered at 30.0 Hz prior to grand averaging for visualization and subsequent statistical analysis.

For ERP analyses, the N400 and P600 components were extracted. Mean amplitudes were calculated for each component within predefined regions of interest (Central, Parietal, and Centroparietal) within the established temporal windows of 300–500 ms for the N400 and 500–800 ms for the P600 [16–18].

2.2.5 Behavioral data and ERP statistical analysis

Behavioral accuracy was calculated for each participant and condition by computing the

proportion of correct responses. Mean accuracy rates were then averaged across participants and visualized using bar graphs with standard error bars. To assess whether accuracy differed statistically across conditions, a repeated-measures ANOVA was conducted with Condition (Control, Syntactic, Semantic, Phonological) as a within-subjects factor. In addition to the overall condition-based accuracy analysis, cue-based behavioral accuracy and reaction times were also summarized descriptively in order to assess how effectively participants used the three trial-by-trial cue types (Semantic, Phonological, and Grammar) during sentence judgment. ERP mean amplitudes (N400 and P600) were analyzed using repeated-measures ANOVAs with factors Condition (Control, Syntactic, Semantic, Phonological) and Region (Central, Parietal, Centroparietal) for the N400 (300–500 ms) and P600 (500–800 ms) time windows. All statistical analyses were conducted in JASP (version 0.95.3, <https://jasp-stats.org/>). Effect sizes are reported as ω^2 , and post hoc comparisons were adjusted using the Holm–Bonferroni method.

In addition to the repeated-measures ANOVAs, complementary linear mixed-effects (LME) analyses were conducted for the 300–500 ms and 500–800 ms time windows in order to evaluate condition effects at the trial level while accounting for within-participant variability. These analyses were intended to complement the ROI-based ANOVA approach and to provide a more sensitive assessment of subtle ERP differences across conditions. As a supplementary step, temporal cluster-based permutation analyses were also conducted over the post-stimulus interval after baseline correction. Clusters were defined on the basis of temporal adjacency and a cluster-forming threshold, and no explicit minimum duration criterion (e.g., 50 ms) was imposed. Instead, cluster significance was determined by the permutation procedure itself. Because these temporal clustering analyses were supplementary to the primary ROI-based and LME analyses, the results are reported in the Appendix C.

2.3 Results

2.3.1 Behavioral Results

Mean accuracy rates were computed for each linguistic condition to verify task engagement. Participants exhibited the highest accuracy in the Control condition ($M = 76.7 \pm 2.46\%$), followed by the Semantic condition ($M = 64.7 \pm 2.39\%$). In contrast, performance declined substantially in the Phonological ($M = 39.9 \pm 3.31\%$) and Syntactic ($M = 36.9 \pm 2.52\%$) conditions. Values reflect mean $\pm$ standard error of the mean (SEM) (Fig. 3). For this analysis, accuracy was first computed separately for each of the four syntactic violation types and then averaged to yield a single Syntactic condition accuracy score for each participant

To statistically evaluate the significant effect of conditions, a one-way repeated-measures ANOVA was conducted with Condition (Control, Syntactic, Semantic, and Phonological) as a within-subjects factor. The analysis revealed a significant main effect of Condition ($F(3, 78) = 90.22, p < 0.001$), indicating that task performance varied reliably across the linguistic conditions. Additionally, post-hoc comparisons confirmed that both Phonological and Syntactic conditions differed significantly from the Control and Semantic conditions, but not from each other ($p > 0.10$), indicating comparably low accuracy for these two violation types.

These results suggest that sentences containing syntactic and phonological violations posed the greatest difficulty, whereas semantic violations were more readily detected. This pattern of performance aligns with previous evidence that native-language experience shapes speech perception, which can make phonological processing particularly challenging for L2 users [32].

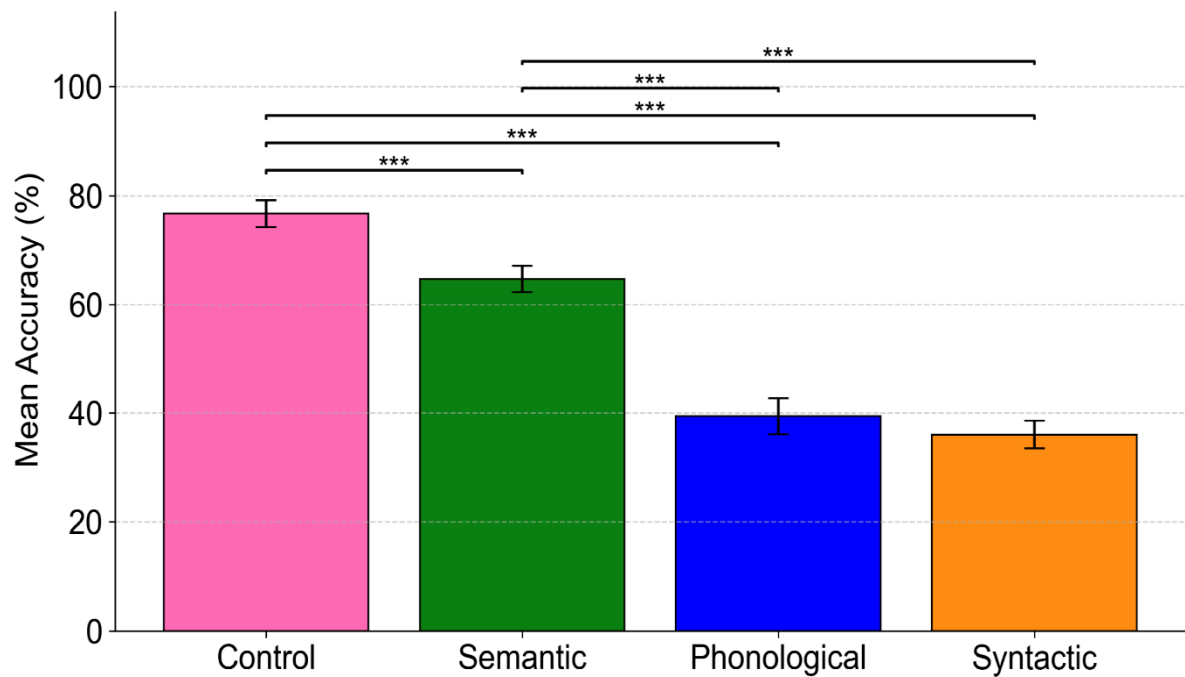


Figure 3. Mean accuracy rates for each linguistic condition. Error bars represent ± 1 standard error of the mean (SEM). Asterisks indicate statistically significant pairwise comparisons ($p < 0.05$, Bonferroni-corrected); horizontal lines denote the tested contrasts.

In addition to overall condition-based performance, cue-based behavioral accuracy was examined descriptively to assess how effectively participants used the three cue types presented before each sentence. Performance was highest for the Semantic cue (意味; $M = 70\%$, $SD = 9\%$, $SEM = 1.7\%$), followed by the Phonological cue (音韻; $M = 58\%$, $SD = 11\%$, $SEM = 2.1\%$) and the Grammar cue (文法; $M = 56\%$, $SD = 9\%$, $SEM = 1.8\%$). Thus, although participants were able to use all three cue types above chance, judgments based on semantic information were clearly the most accurate, whereas phonological and grammatical cues yielded weaker and less reliable performance. This pattern suggests that meaning-based information was relatively more accessible than grammatical or phonological information for explicit sentence judgments in these late L2 learners.

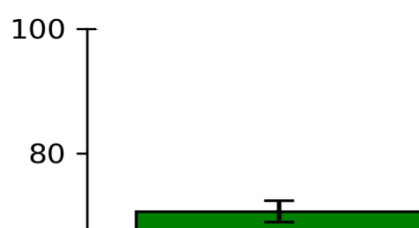


Figure 4. Mean accuracy by cue condition (Semantic, Phonological, and Grammar). Error bars represent ± 1 standard error of the mean (SEM).

Reaction times (RTs) were also examined across cue conditions to assess whether the lower accuracy observed for some cue types could be attributed to speeded responding. Mean RTs were broadly similar across the three cue conditions: Grammar cue (文法), $M = 1525$ ms, $SD = 655$ ms, $SEM = 124$ ms; Semantic cue (意味), $M = 1583$ ms, $SD = 664$ ms, $SEM = 126$ ms; and Phonological cue (音韻), $M = 1421$ ms, $SD = 593$ ms, $SEM = 112$ ms. Overall, participants took comparable amounts of time to respond across cue conditions, with mean response latencies falling within a relatively narrow range of approximately 1400–1600 ms. This suggests that the lower accuracy observed for grammar and phonology cues was not simply due to fast guessing or reduced task engagement, but more likely reflects the greater difficulty of making explicit L2 judgments in these domains.

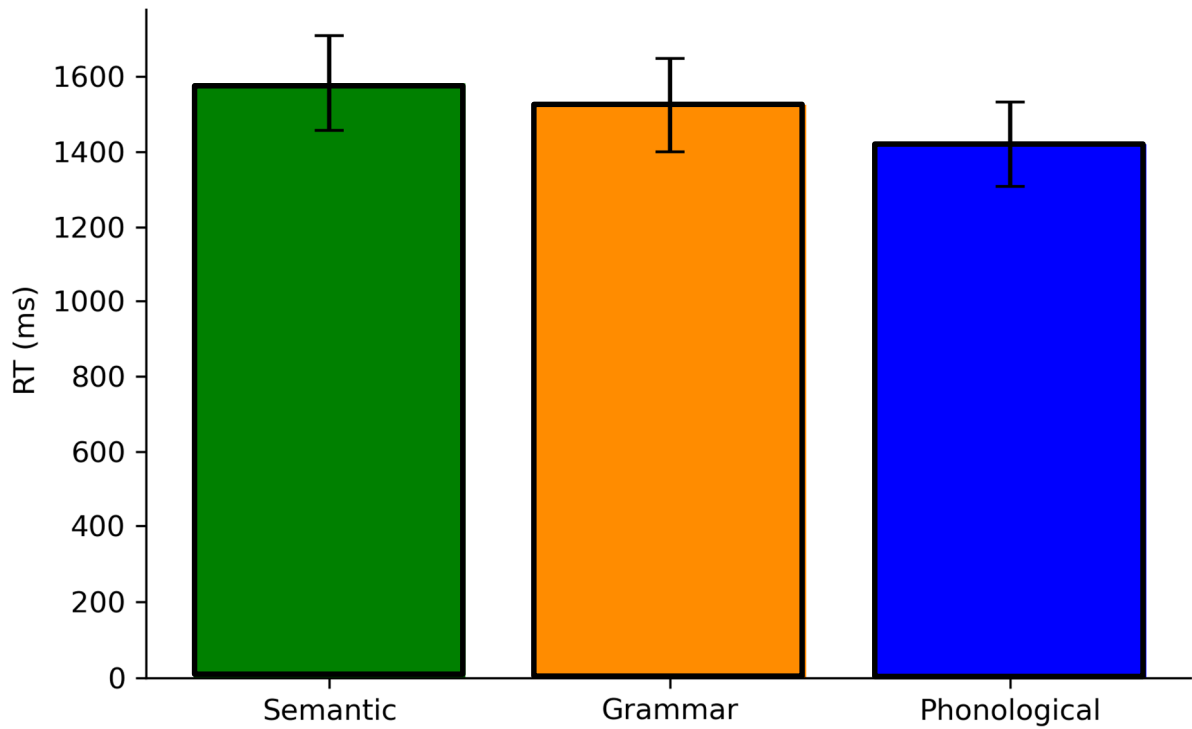


Figure 5. Mean reaction times (ms) by cue condition (Grammar, Semantic, and Phonological). Error bars represent ± 1 standard error of the mean (SEM).

2.3.2 ERP Results

ERP results revealed distinct temporal and regional patterns across linguistic conditions. ERP responses were analyzed for syntactic, semantic, and phonological violations relative to control sentences in two time windows: N400 (300–500 ms) and P600 (500–800 ms) (Fig. 6).

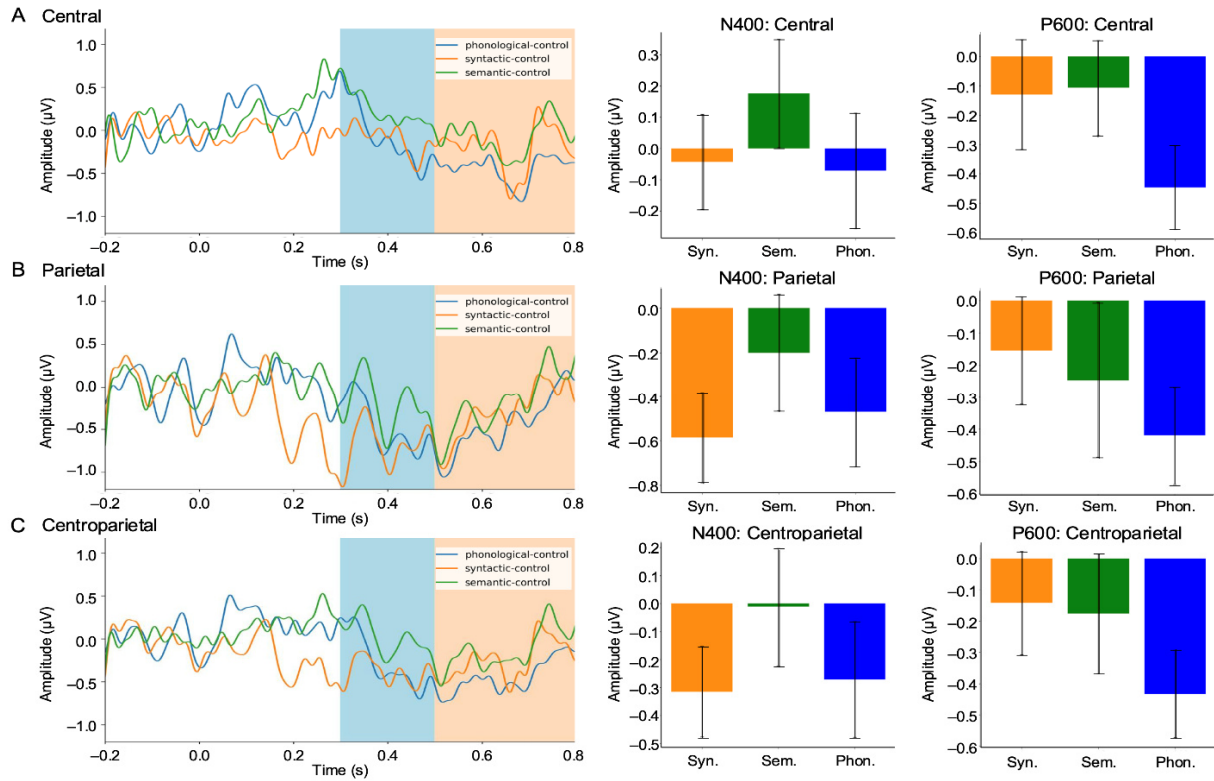


Figure 6. Grand-average ERP difference waves (violation minus control) for syntactic (orange), semantic (green), and phonological (blue) conditions at Central (A), Parietal (B), and Centroparietal (C) regions of interest. Shaded areas indicate the N400 (300–500 ms, blue) and P600 (500–800 ms, orange) analysis windows. Amplitude is shown in microvolts (μV), time in seconds, with negative plotted downwards. Error bars represent ± 1 standard error of the mean (SEM).

As shown in Figure 6, grand-average ERP difference waves (violation minus control) are plotted for the phonological, syntactic, and semantic conditions across the Central, Parietal, and Centroparietal regions of interest. Because the difference wave represents the subtraction of the control condition from each violation type, deviations from zero indicate differential processing relative to well-formed sentences. In the 300–500 ms (N400) window, the phonological condition shows a sustained negative deflection across all regions, with the largest numerical

negativity in the parietal region (Fig. 6B). This negativity may reflect increased processing demands associated with phonological anomalies. The syntactic condition also shows a small negative shift, but this effect is weaker and less consistent across regions, with mean amplitudes generally closer to zero than in the phonological condition. By contrast, the semantic condition remains close to baseline across all regions and shows no clear deviation from the control condition.

To statistically assess these observations, a repeated-measures ANOVA was conducted on mean N400 amplitudes during the 300–500 ms time window with factors Condition (Control, Syntax, Semantic, and Phonological) and Region (Central, Parietal, and Centroparietal) as within-subjects factors. This analysis revealed a significant main effect of Region ($F(2,52) = 32.21, p < 0.001, \omega^2 = 0.222$), indicating that ERP amplitudes varied across electrode sites. No significant main effect of Condition was observed ($F(3,78) = 1.44, p = 0.237$), but the Condition $\times$ Region interaction was significant ($F(6,156) = 3.10, p = 0.007, \omega^2 = 0.032$), suggesting that the effect of condition differed depending on scalp location.

Follow-up pairwise comparisons in the 300–500 ms time window showed a non-significant trend for the phonological condition to elicit more negative amplitudes than the control condition at parietal ($p = .091$) and centroparietal ($p = .153$) regions, with numerically larger differences over posterior sites. Although the waveforms showed a broad negativity consistent with phonological sensitivity, these effects did not reach statistical significance in the ROI-based comparisons. For the syntactic condition, a significant difference from the control condition was observed only at the parietal region ($p = .012$), although the waveform did not show the typical distribution, amplitude, or morphology of a canonical N400. As discussed later, this may reflect the heterogeneity of the syntactic violation types and the learners' reduced

engagement of syntactic reanalysis processes. The semantic condition did not differ significantly from the control condition at any region (all p s > 0.10), consistent with the visual pattern of minimal divergence from baseline. This absence of a canonical semantic N400 is particularly notable, given that semantic anomalies typically elicit robust centroparietal N400 effects in native speakers. Overall, these findings suggest that phonological anomalies elicited the clearest ERP modulation in the early time window, particularly over posterior scalp regions, whereas syntactic and semantic violations produced weaker or less consistent effects.

To further clarify the condition effects in the 300–500 ms window, complementary linear mixed-effects analyses were conducted. These analyses showed that both the phonological condition ($p = 0.017$) and the syntactic condition ($p = 0.005$) elicited significantly more negative amplitudes than the control condition, whereas the semantic condition did not differ significantly from control ($p = 0.892$). Thus, although the omnibus ANOVA indicated a more regionally distributed pattern, the LME results provided clearer evidence that phonological and syntactic anomalies were associated with enhanced early negativities relative to well-formed sentences.

In the later 500–800 ms (P600) window, there is limited evidence for a classic P600-like late positivity in response to the syntactic condition. Instead, the phonological condition continued to show negative-going activity across all regions, with the strongest effect observed over the parietal region. These patterns suggest that among the three conditions, the phonological condition elicited the most distinct ERP response relative to the control, whereas syntactic and semantic conditions elicited comparatively more subtle effects.

To statistically evaluate these observations, a repeated-measures ANOVA was conducted on

mean P600 amplitudes during the 500–800 ms time window with Condition (Control, Syntactic, Semantic, and Phonological) and Region (Central, Parietal, and Centroparietal) as within-subjects factors.

A significant main effect of Condition ($F(3,78) = 2.98$, $p = 0.036$, $\omega^2 = .056$) was found, indicating that the ERP responses in this window differed significantly across the four conditions. However, there was no main effect of Region ($p = 0.846$), and the Condition $\times$ Region interaction was also non-significant ($p = 0.776$), indicating that these condition effects were not strongly dependent on scalp distribution

Follow-up pairwise comparisons indicated that the phonological condition differed significantly from the control condition at all three regions: central ($p = 0.009$), parietal ($p = 0.030$), and centroparietal ($p = 0.016$). However, these differences were driven by continued negative-going activity, not a canonical P600 activity, suggesting prolonged processing difficulty or mismatch detection rather than syntactic reanalysis.

Furthermore, the syntactic condition did not differ significantly from control conditions at any region (all $ps > 0.25$), although the waveform showed a small positive shift centrally. Similarly, Semantic conditions did not differ significantly from control conditions in any region (all $ps > 0.30$).

Complementary LME analyses for the 500–800 ms window again revealed a robust effect for the phonological condition, which showed significantly more negative amplitudes than the control condition ($p < 0.001$). In addition, the semantic condition also differed significantly from control in this later window ($p = 0.033$), whereas the syntactic condition showed only a

non-significant trend ($p = 0.080$). These findings suggest that phonological anomalies continued to impose strong processing costs into the later time range, while semantic anomalies also showed a later emerging effect not fully captured by the ROI-based ANOVA alone.

Taken together, these patterns suggest that phonological violations elicited the most robust ERP effects across both time 300–500 ms and 500–800 ms time windows, while syntactic and semantic violations evoked less consistent neural responses in this L2 auditory sentence comprehension task.

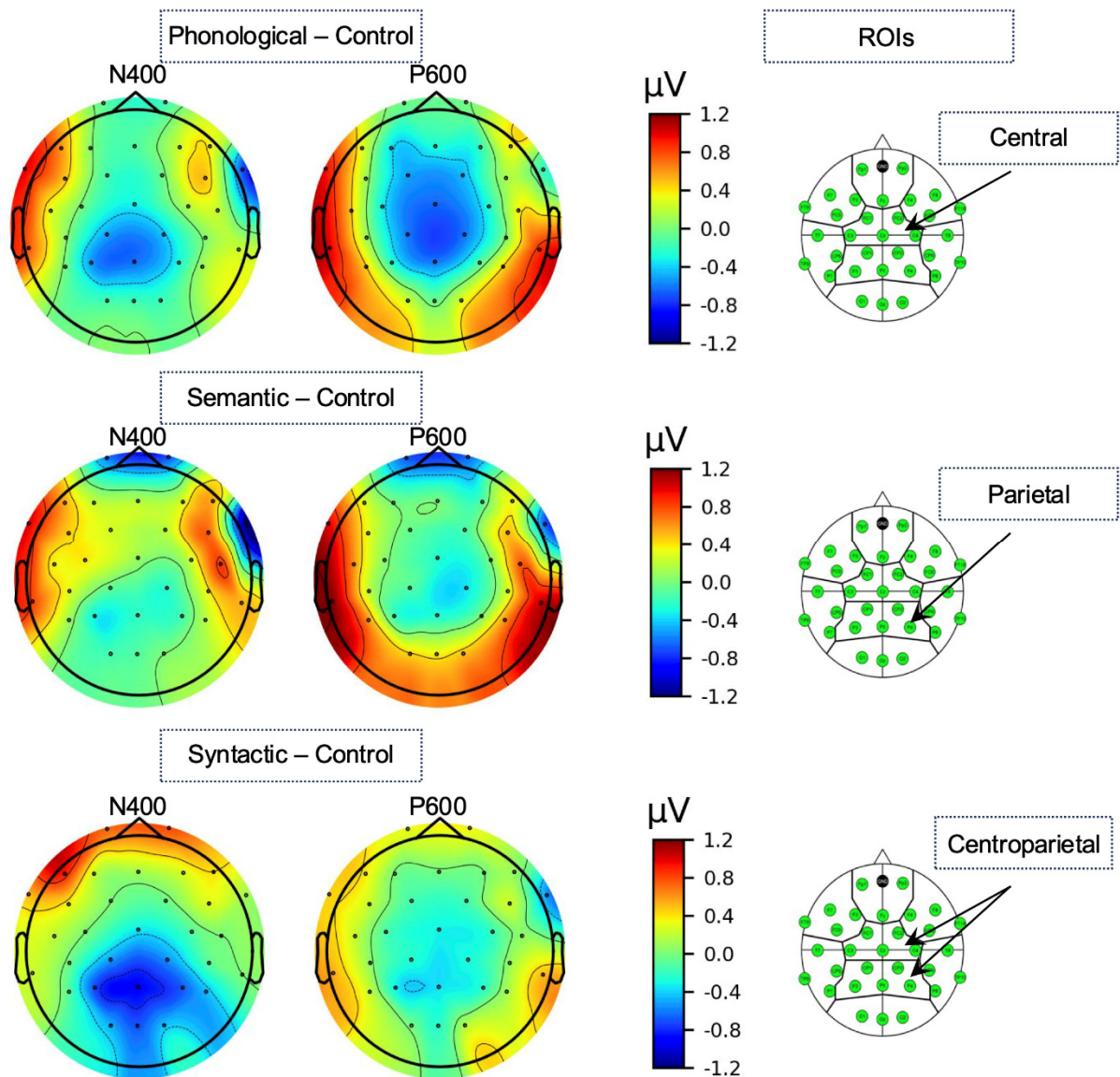


Figure 7. Scalp topographies of ERP difference waves (violation minus control) for phonological, semantic, and syntactic conditions during the N400 (300–500 ms) and P600 (500–800 ms) time windows in the Central, Parietal, and Centroparietal regions of interest. These topographies support the scalp distribution described in the waveform plots, particularly highlighting the posterior negativity elicited by phonological anomalies.

Figure 7 displays the scalp topographies of the difference waves (violation minus control) for phonological, semantic, and syntactic conditions across the Central, Parietal, and Centroparietal

regions of interest during the N400 (300–500 ms) and P600 (500–800 ms) time windows. In the N400 window, the topography map shows that phonological conditions elicited a sustained posterior negativity, particularly in the parietal region, which extends into the later time window. This would suggest increased processing effort or detection difficulty for phonological anomalies. Moreover, syntactic conditions produced potentially weaker or diffused effects across both time windows, with no clear evidence of a canonical P600. Semantic conditions also showed minimal topographic modulation, with only a subtle posterior negativity in the N400 window. Overall, the topographic maps (Fig. 7) support the ERP wave forms and statistical results highlight that phonological anomalies produced the most robust ERP effects, while syntactic and semantic conditions yielded relatively inconsistent neural responses (Fig. 6).

2.4 ERP–Behavioral Correlations

While the group-level ERP and behavioral accuracy results provided valuable insights into second-language sentence processing, they cannot capture important individual differences that may modulate neural and behavioral responses. To address this, we examined correlations between ERP amplitudes and sentence judgment accuracy to determine how variability in neural activity relates to performance in L2 sentence comprehension. Specifically, we tested whether the amplitudes of the N400- and P600-range components were linked to sentence judgment accuracy in the different linguistic conditions. Identifying such relationships provides insight into the functional significance of ERP modulations and clarifies whether stronger or more canonical ERP responses are associated with more efficient or accurate processing in L2 learners. This analysis complements the group-level findings by relating participant-level performance to individual patterns of neural activation.

To investigate the relationship between neural activity and behavioral performance, a series of Pearson correlation analyses was conducted between mean ERP amplitudes and sentence judgment accuracy. For each participant, mean ERP amplitudes were extracted in the 300–500 ms (N400) and 500–800 ms (P600) time windows at the central, parietal, and centroparietal ROIs for each condition. These values were then correlated with accuracy in the four experimental conditions (control, syntactic, semantic, phonological). The analyses revealed two significant correlations in the early (300–500 ms) time window at the parietal ROI.

As shown in Figure 8, there was a significant negative correlation between behavioral accuracy in the phonological condition and ERP amplitudes elicited by syntactic violations at the parietal ROI ($r = -0.421$, $p = 0.0286$). Participants who showed more negative-going ERP amplitudes to syntactic anomalies tended to achieve higher accuracy when judging phonological correctness. This cross-condition relationship suggests that individuals with greater early neural sensitivity to syntactic structure also perform better when evaluating phonological aspects of the sentences, perhaps reflecting more efficient engagement with the overall morphosyntactic frame that supports phonological mapping.

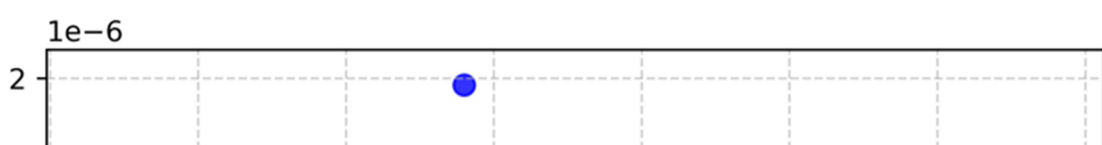


Figure 8. Scatterplot showing the correlation between behavioral accuracy in the phonological condition and ERP amplitude for syntactic violations at the parietal ROI during the 300–500 ms window. Greater negativities (μV) were associated with higher performance in phonological judgments.

Similarly, for the syntactic condition, a significant negative correlation was observed between sentence judgment accuracy and ERP amplitudes for syntactic violations at the parietal ROI in the same 300–500 ms window ($r = -0.407$, $p = 0.0349$; Fig. 9). Participants with more negative ERP responses in this early time window tended to perform more accurately in detecting syntactic anomalies. This indicates that stronger early neural responses to syntactic violations are associated with better behavioral sensitivity to grammatical errors in L2 sentence processing.

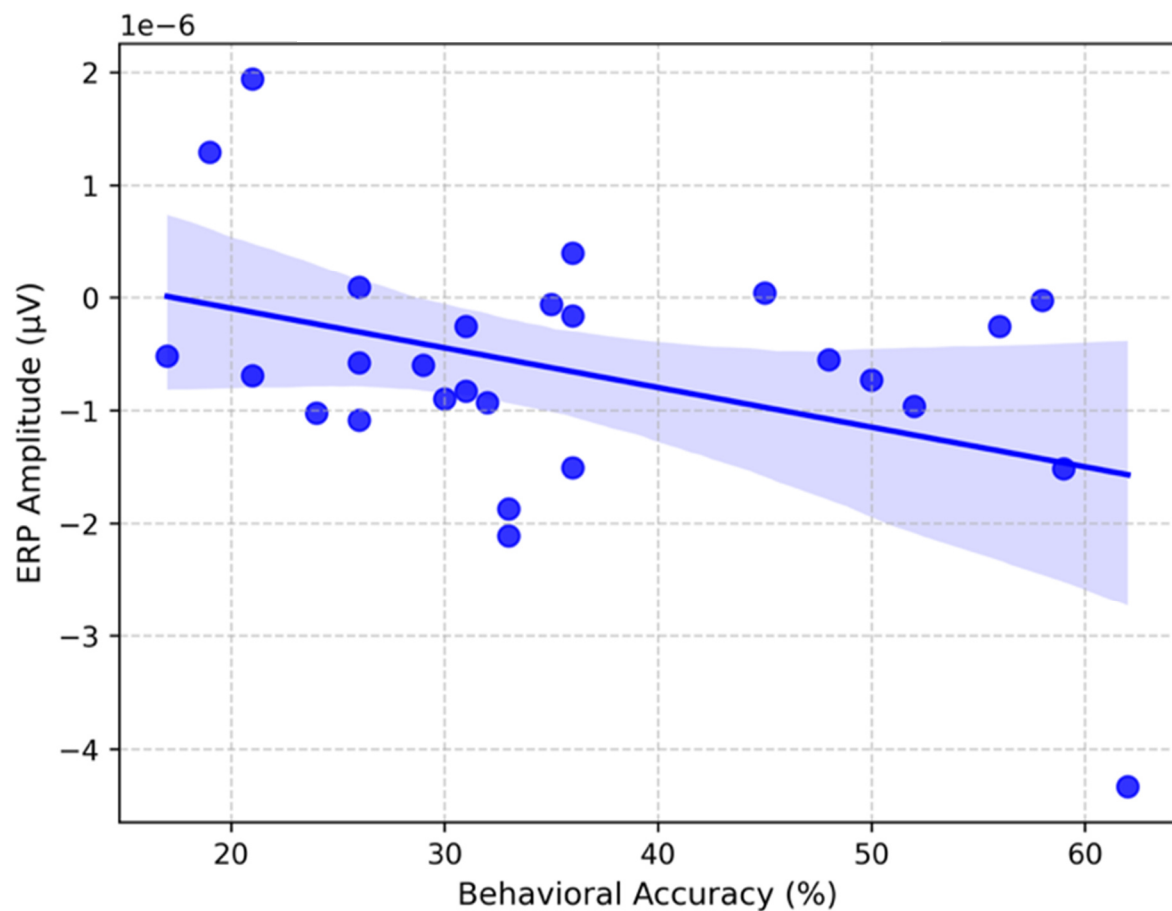


Figure 9. Scatterplot showing the correlation between behavioral accuracy in the syntactic condition and ERP amplitude for syntactic violations at the parietal ROI during the 300–500 ms window. Participants with more negative ERP responses tended to perform more accurately in detecting syntactic violations.

Together, these correlations suggest that early syntactic-related neural activity may serve as a marker of individual differences in second-language sentence processing. Greater ERP sensitivity in the 300–500 ms window appears to reflect more robust engagement with morphosyntactic structure, which in turn is linked to improved performance in both phonological and syntactic tasks. The cross-condition association between phonological

accuracy and syntactic ERP activity highlights shared or interacting mechanisms, whereby learners who engage more effectively with syntactic structure may also be better equipped to detect phonological anomalies under increased processing demands.

These findings are in line with previous work underlining the role of variability in ERP profiles as neurocognitive markers of language proficiency [24,28,36]. Prior studies have shown that inter-individual differences in ERP amplitudes, especially in components associated with morphosyntactic processing such as the N400 and P600, reflect meaningful differences in learners' grammatical sensitivity and allocation of cognitive resources. In the present study, the observed correlations further support the idea that early ERP dynamics capture functionally relevant aspects of L2 processing success and difficulty across linguistic domains.

2.5 Discussion

The ERP-behavioral experiment provides a detailed picture of how Japanese learners of English process multiple linguistic domains during auditory sentence comprehension. At the behavioral level, participants showed relatively high accuracy for control and semantic trials, but markedly reduced performance in the syntactic and phonological conditions. This asymmetry suggests that, for this learner group, grammatical and phonological cues are less reliably exploited in real-time than semantic information, even when explicit instructions direct attention to a specific linguistic dimension. The low accuracy in syntactic and phonological conditions is consistent with prior work indicating that late L2 learners often struggle to automatize morphosyntactic parsing and to segment non-native phonological forms, particularly when L1 and L2 differ substantially in prosodic structure, phonotactics, and grammatical marking [26,27,31–33]. In the present context, the lack of obligatory plural marker and article marking in Japanese, as well as differences in segmental contrasts and syllabic

structure, likely contribute to persistent difficulty in mapping English input onto stable morphosyntactic and phonological representations.

The additional cue-based behavioral analyses further clarified this pattern. Participants were most accurate when their judgments were guided by semantic cues, whereas grammar and phonology cues yielded weaker and less reliable performance. Importantly, reaction times were broadly comparable across the three cue conditions, indicating that the lower accuracy for grammar and phonology was not driven by speeded responding or lack of effort. Rather, these results suggest that explicit judgments based on grammatical and phonological information were genuinely more difficult for these Japanese learners of English, whereas meaning-based information was relatively more accessible.

The ERP results qualify this behavioral pattern in important ways. Across the N400 (300–500 ms) and P600 (500–800 ms) windows, phonological violations elicited the most consistent neural modulation, characterized by sustained negative-going activity over posterior regions. Although these negativities do not correspond to a canonical, sharply peaked N400, their timing and topography suggest increased processing demands when phonological forms deviate from learners' expectations. One interpretation might be that mispronunciations disrupt early phonological encoding and, in turn, hinder lexical access, which would lead to downstream integration difficulties indexed by broad negativities in the N400 time range. The fact that these effects continue into the later window further supports the idea that phonological anomalies impose prolonged processing costs in L2 comprehension, consistent with accounts emphasizing the fragility of phonological mapping in late bilinguals [30–33].

On the contrary, semantic violations failed to elicit robust N400 effects relative to the control

condition, despite being designed as clear plausibility violations while in native speakers, comparable anomalies reliably produce large centro-parietal N400 responses [16,17]. The absence of a canonical semantic N400 in the present study might suggest that L2 semantic integration was not fully automatic or that semantic conflict was not consistently detected at the neural level. Task-related factors likely contributed to this matter too, for instance the use of explicit trial-by-trial cues, and an emphasis on overt acceptability judgments may have encouraged a strategy in which participants prioritized anomaly detection and surface-level cues over deeper semantic integration. Nevertheless, as similar paradigms still elicit robust N400s in native listeners [16,17,39], the current findings indicate that the combination of L2 proficiency level, typological distance, and task demands leads to qualitatively different patterns of semantic processing in L2 comprehension rather than simply delayed native-like responses.

Moreover, syntactic violations produced only weak and inconsistent ERP effects. Although a parietal difference from the control condition was observed in the 300–500 ms window, the waveform did not exhibit the typical morphology or distribution of a canonical N400 or P600, and no clear late positivity emerged in the 500–800 ms window. Several factors might have contributed to this pattern. First, the syntactic condition intentionally combined different violation types (subject–verb agreement, prepositional usage, countability, and article choice) to probe general sensitivity to morphosyntactic anomalies. While this design increased ecological validity and allowed a broader coverage of typical L2 difficulties, collapsing across heterogeneous subtypes may have increased within-condition variability and diluted subtype-specific ERP effects. Second, the low behavioral accuracy in the syntactic condition implies that many violations were not reliably detected or were processed only shallowly, reducing the likelihood of robust reanalysis-related positivities. This may suggest that late L2 learners often

rely on lexical and local cues rather than constructing fully specified syntactic representations that would trigger a strong P600 [22–24,37,38].

The complementary LME analyses also sharpened the interpretation of the ERP findings. In the early 300–500 ms window, both phonological and syntactic conditions showed significantly more negative amplitudes than the control condition, whereas the semantic condition did not. In the later 500–800 ms window, phonological anomalies again showed the strongest negative-going effect, while semantic violations also differed significantly from control and syntactic violations showed only a marginal trend. Taken together, these results suggest that the ROI-based ANOVAs may have underestimated some condition effects that became clearer in the trial-level mixed-effects framework.

The ERP–behavioral correlations further explain the functional importance of these neural patterns. More negative amplitudes to syntactic violations in the 300–500 ms window at parietal sites were linked to higher accuracy in both the syntactic and phonological conditions. This cross-condition relationship shows that early neural sensitivity to structural anomalies supports better performance not only when grammatical correctness is explicitly targeted, but also when phonological correctness is evaluated. One interpretation is that learners who engage more efficiently with the morphosyntactic frame of the sentence can more effectively map phonological input onto lexical and structural expectations, leading to improved detection of both syntactic and phonological violations. The correlation between syntactic ERP amplitudes and syntactic accuracy reinforces the view that individual differences in early neural responses capture meaningful variability in L2 grammatical sensitivity, as suggested by previous work on ERP-based proficiency markers [24,28,36].

It should be noted that several methodological considerations should be taken into consideration when interpreting this experiment. The sample consisted of intermediate Japanese learners with relatively homogeneous educational backgrounds, limiting generalizability to beginners or highly advanced speakers. The reliance on text-to-speech synthesis ensured uniform prosody and timing but may have reduced some natural variability that listeners exploit in everyday communication. The cued design, while critical for isolating domain-specific processing, may have shifted attention away from spontaneous integration toward a more strategic, feature-based mode of listening as we are making some forms of awareness. Finally, collapsing syntactic violation subtypes into a single condition was necessary for adequate statistical power but may have masked subtype-specific ERP signatures.

Despite these constraints, the ERP-behavioral experiment demonstrates that, in late Japanese learners of English, phonological violations are the most salient at both behavioral and neural levels, while syntactic and semantic anomalies evoke weaker or non-canonical ERP responses. The observed correlations between early ERP amplitudes and performance indicate that individual differences in neural sensitivity are closely tied to L2 processing success. These findings motivate the multimodal approach adopted in this thesis.

Chapter 3

MRI Analysis of Cortical Thickness

3.1 Introduction

Studies on L2 sentence processing need to explore not only the dynamic brain activity associated with language comprehension but also the neuroanatomical architecture associated with such brain functions. Although electrophysiological measures like event-related potentials (ERPs), which have good temporal resolution, have successfully identified the characteristic components associated with L2 processing, like the N400 and P600 (see Chapter 2), they lack information about the variability associated with the human neuroanatomical architecture.

Recently, cortical thickness, an index of offline brain anatomy that is quantified using high-resolution T1-weighted MRI, has emerged as a candidate biomarker for cognitive domains such as language acquisition abilities. Cortical thickness is defined as the distance from the pial surface to the white matter boundary (WM) and is typically variable across individuals because of genetic and experiential differences. It is quite sensitive to experience-driven plasticity in adulthood [11–15].

Within the field of linguistics, previous studies have also found a relationship between cortical thickness measurements and linguistic functions, especially in perisylvian areas of the left hemisphere of the brain [9–11,15]. For instance, heightened cortical thickness in

the left inferior frontal gyrus has been correlated with enhanced syntactic analysis and verbal fluency abilities [4,10,11], while areas including STG and MTG have also been identified as being responsible for phonological and semantic decoding [7,8]. In bilingual individuals, cortical changes have also been identified in areas within the homologous region of the right brain hemisphere and the executive control areas, including ACC, suggesting that bilingualism and the acquisition of an L2 are linked to cortical differences even outside the conventional language areas of the brain [11–13].

Moreover, it has been shown that L2 experience can cause plastic changes in the thickness of the cortex as well as gray matter density [11–15]. For example, a prior study revealed the positive effect of bilingualism, which leads to higher gray matter density of the LIFG compared to that of monolingual individuals, especially among early and highly proficient bilinguals. Similarly, another study showed that the positive correlation between the level of L2 proficiency and the thickness of the cortex exists, particularly for the left anterior cingulate cortex and the right inferior parietal lobule. The mentioned studies demonstrated that the experience of L2 learning has a positive effect on the modulation of structural anatomy, which, as a consequence, affects the level of processing efficiency for the respective language system(s). At the same time, a limited number of studies have explored the positive correlation between the thickness of the cortex and the amplitude of the corresponding ERP responses among L2 learners. Also another study demonstrated that the level of amplitude of the ERP responses depends positively upon the structural properties of the LIFG during the semantic processing of the sentence judgment task among native English speakers. However, this work was conducted in native speakers, and no comparable studies to date have examined ERP–MRI correlations in late L2 learners, particularly within a single experimental design that compares multiple linguistic

subsystems (syntax, semantics, and phonology).

However, a gap in the literature remains, as there is no systematic study that specifically explores the amplitude of ERP responses among native Japanese speakers and the corresponding structural MRI data.

Therefore, the current chapter aims to fill this gap by examining structure-function associations in Japanese learners of English through a multimodal analysis integrating ERPs and structural MRI. By correlating mean ERP amplitudes from key time windows (N400: 300–500 ms; P600: 500–800 ms) with cortical thickness across a wide set of brain regions, we aim to determine the anatomical predictors of functional neural responses to various types of linguistic violations in a L2. This approach addresses questions regarding which language-related brain regions exhibit structural associations with ERP responses to syntactic, semantic, and phonological violations in L2 sentence comprehension; whether these associations reflect modality-specific (e.g., semantic) or domain-general (e.g., executive control) processes; and whether the structure-function relationships are lateralized or distributed. By combining cortical thickness with neural data, this chapter contributes to a more complete neurocognitive model of second-language processing.

3.2 Methods

Building on the separate ERP and MRI findings reported in previous sections, the next step is to bring these two approaches together and ask how individual differences in brain structure relate to the atypical ERP patterns observed in Japanese learners of English as a L2. At the group level, the ERP results showed that phonological violations elicited the strongest neural responses, whereas syntactic and semantic anomalies produced weaker or

less typical effects. However, these averaged patterns may mask important variability across learners in how neural resources are engaged during sentence processing.

The cortical thickness analysis likewise revealed considerable variation across individuals, both in core language regions (such as the left inferior frontal gyrus, superior and middle temporal gyri, and temporo-parietal areas) and in regions associated with executive control (including the anterior cingulate cortex and inferior parietal lobule). This variability suggests that L2 learners differ in the structural neural architecture that supports language processing.

By correlating regional cortical thickness with mean N400 and P600 amplitudes, we can test whether greater thickness in specific brain areas is associated with stronger, more canonical ERP responses, or whether structural differences instead support compensatory, non-canonical processing strategies. Linking these structure-function relationships to behavioral accuracy further allows us to assess whether particular combinations of brain anatomy and electrophysiological responses are tied to more successful L2 sentence comprehension.

Accordingly, the following section presents a set of ERP–MRI correlation analyses aimed at identifying (1) which brain regions show reliable structure-function relationships for phonological, syntactic, and semantic violations, (2) whether these relationships are lateralized or bilateral, and (3) how they relate to individual differences in L2 performance.

3.2.1 Participants

We scanned 27 native speakers of Japanese (14 males, 13 females; age range = 18–25 years, $M \pm SD = 21.0 \pm 2.0$), who participated in the ERP-Behavioral experiment. All

participants were undergraduate or graduate students at Kyushu University (Fukuoka, Japan), had normal or corrected-to-normal vision, and were self-identified as right-handed. None reported a history of neurological, speech, or hearing disorders, nor had any participant lived in an English-speaking country for more than one year. All participants had successfully passed the mandatory, credit-bearing English language course requirement administered by Kyushu University, which typically corresponds to an intermediate level of English proficiency required for graduation. Completion of this course served as an index of participants' formal English learning experience.

Informed consent was obtained from each participant after a full explanation of the experiments, in accordance with the World Medical Association's Declaration of Helsinki. The experimental procedures were approved by the Ethics Committee of the Faculty of Humanities at Kyushu University.

3.2.2 MRI Data Acquisition

For structural MRI data acquisition, a participant was in a supine position, and his or her head was immobilized inside the radio-frequency coil. The MRI scans were conducted on a 3.0 T MRI system equipped with a 32-channel phased-array head coil (MAGNETOM Lumina; Siemens Healthineers, Erlangen, Germany). High-resolution T1-weighted images of the whole brain (320 axial slices, $0.8 \times 0.8 \times 0.8 \text{ mm}^3$) were acquired from all participants with a magnetization-prepared rapid gradient echo (MPRAGE) sequence (TR = 2500 ms, TE = 1.94 ms, FA = 8° , FOV = $256 \times 256 \text{ mm}^2$, slice thickness = 0.8 mm, without gap).

3.2.3 Cortical Thickness Estimation and ERP–MRI Correlation Analysis

Cortical thickness values were estimated from T1-weighted structural images using the CAT12 toolbox implemented in SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK). For each participant, mean cortical thickness was extracted across bilateral cortical regions as implemented in CAT12.

To explore the association of brain structural differences with individual differences in neural activity patterns, Pearson correlation analyses were performed between the mean amplitudes of the ERPs and measures of cortical thickness. The analysis was carried out separately for each participant, collecting the ERP data from four experimental conditions (syntactic, semantic, phonological, and control), the two time windows related to the N400 (300–500 ms) and the P600 (500–800 ms), and the three predefined scalp regions: central, parietal, and centroparietal.

In the first step, the analysis identified multiple statistically significant correlations ($p < .05$) across all conditions and ROIs. However, because a high number of correlations were found, it was necessary to concentrate on regions that were previously shown to be significantly involved in language processing in general. In the next step, we restricted our attention to regions known to play key roles in language processing. This filtering process reduced the set to 37 language-relevant correlations. From these, eight correlations were selected for detailed examination based on their theoretical relevance, effect sizes, and consistency with the ERP patterns observed at the group level. Considering the large number of correlations examined, these ERP–MRI analyses are exploratory in nature and the reported associations should be interpreted with caution and validated in future studies. These “top” correlations were primarily linked to the phonological and semantic conditions, which also showed the most robust ERP effects in this study.

3.3 Results

3.3.1 ERP–MRI Structure–Function Associations

Among the most prominent findings, the amplitudes of semantic ERP components between 300 and 500 ms in the parietal area were negatively related to the cortical thickness in the right posterior cingulate area ($r = -0.64$, $p = 0.0006$) (Fig. 8). The region has previously been related to the processes of semantic monitoring and integration and the default mode network and might suggest participants with a thinner cortex in this area may rely on more efficient or specialized semantic networks during L2 sentence processing.

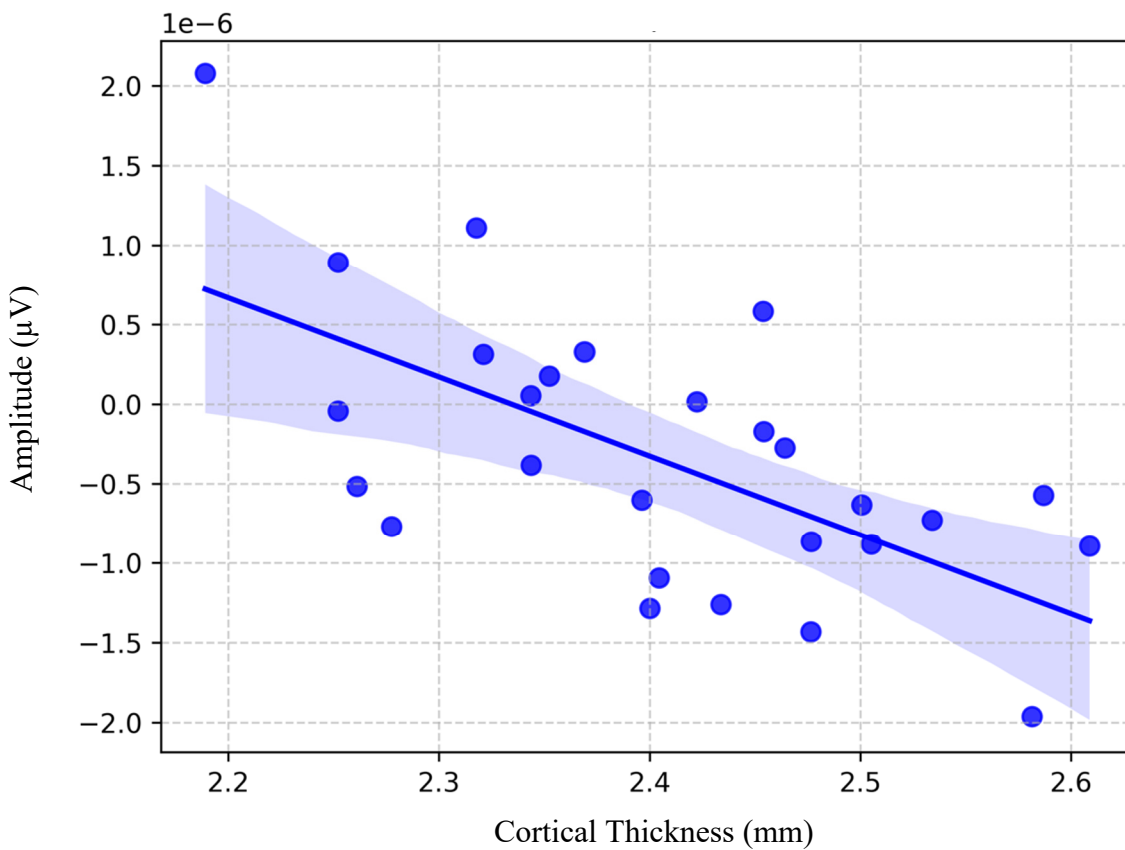


Figure 10. Correlation between semantic ERP amplitudes (300–500 ms, parietal ROI) and cortical thickness in the right posterior cingulate cortex ($r = -0.64$, $p = 0.0006$). Each dot represents one participant; the regression line and 95% confidence band are shown.

Amplitude (μV)

The similar relation between the amplitudes of the semantic ERP components and the cortical thickness was revealed for the centroparietal area and the area of the posterior cingulate of the right hemisphere ($r = -0.52$, $p = 0.0076$) (Fig. 10). Together, these results underscore the involvement of posterior midline structures in supporting semantic comprehension in L2 learners.

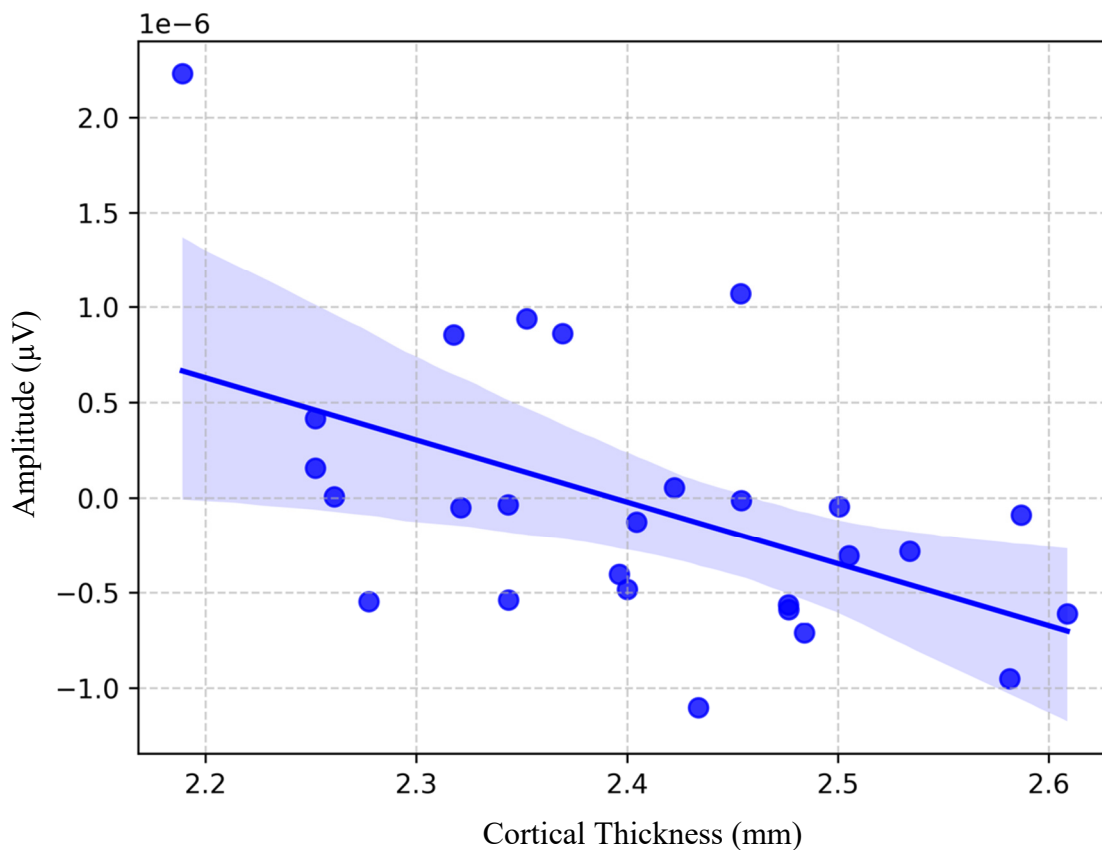


Figure 11. Correlation between semantic ERP amplitudes (300–500 ms, centroparietal ROI) and cortical thickness in the right posterior cingulate cortex ($r = -0.52$, $p = 0.0076$). Each dot represents one participant; the regression line and 95% confidence band are shown.

Late semantic activity in the 500–800 ms time window was also found to be negatively correlated with the cortical thickness in the left rostral middle frontal region ($r = -0.51$, p

= 0.0095) (Fig. 11). The region is associated with working memory and the integration of executive functions. This finding suggests that individuals with thinner cortex in the prefrontal region may have streamlined or optimized processes for semantic integration over time.

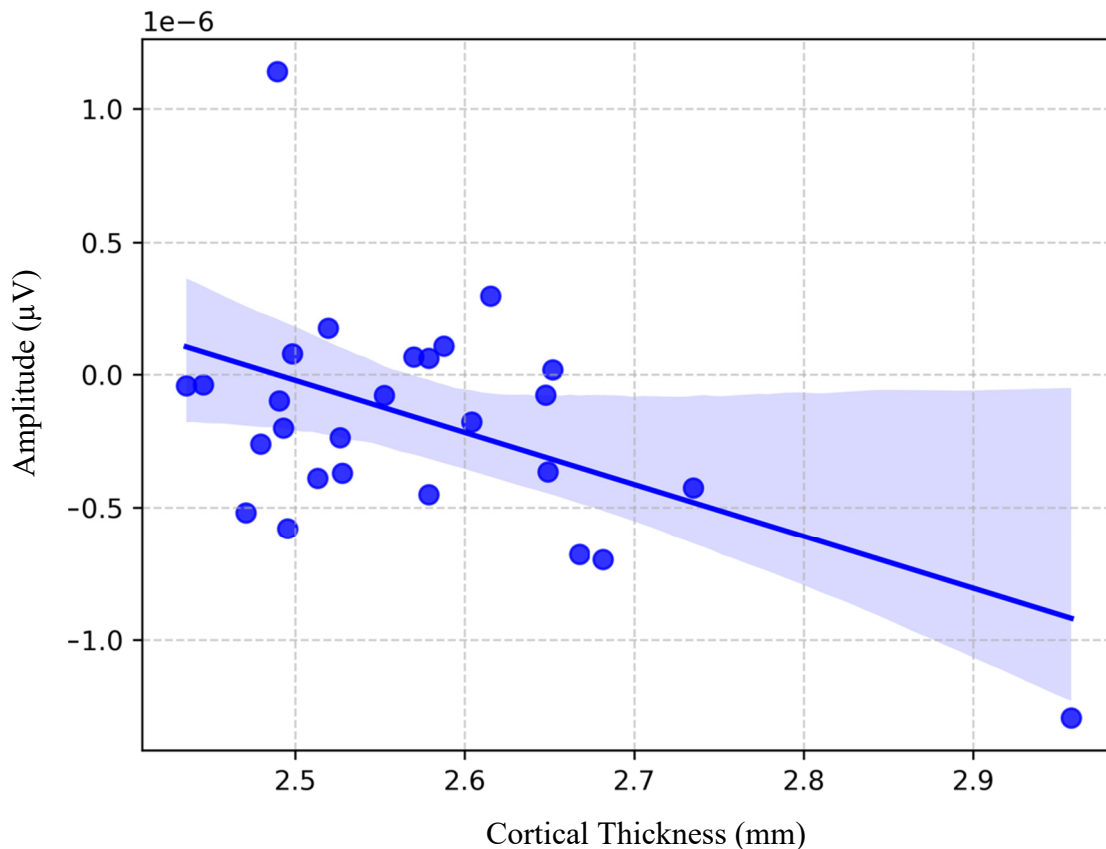


Figure 12. Correlation between semantic ERP amplitudes (500–800 ms, centroparietal ROI) and cortical thickness in the left rostral middle frontal gyrus ($r = -0.51$, $p = 0.0095$). Each dot represents one participant; the regression line and 95% confidence band are shown.

Apart from the semantic condition, the control condition also showed significant correlates for the right pars opercularis within the time window of the P600 for the parietal and centroparietal areas ($r = 0.60$ and $r = 0.53$, respectively). These correlations point to the engagement of Broca’s area and adjacent frontal regions even when participants process grammatically correct sentences, possibly reflecting baseline syntactic fluency, articulatory

planning, or both.

Additionally, in the phonological conditions, early ERP components (central, 300–500 ms) were negatively correlated with cortical thickness in the right pars opercularis ($r = -0.42$, $p = 0.034$) (Fig. 12). This finding is in line with the involvement of this region in phonological error detection, articulatory planning, and prosodic sensitivity in L2 contexts.

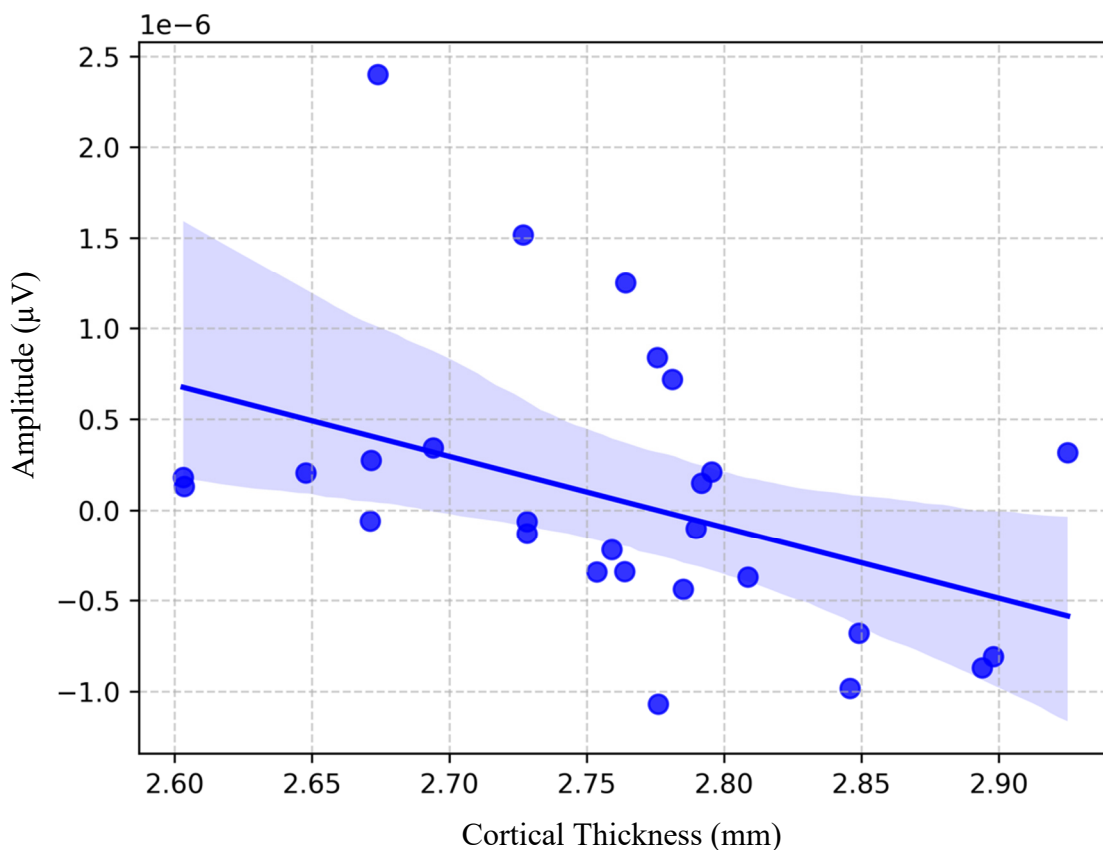


Figure 13. Correlation between phonological ERP amplitudes (300–500 ms, central ROI) and cortical thickness in the right pars opercularis ($r = -0.42$, $p = 0.034$). Each dot represents one participant; the regression line and 95% confidence band are shown.

Furthermore, the late phonological ERP component (central, 500–800 ms) was negatively associated with cortical thickness in the left fusiform gyrus ($r = -0.41$, $p = 0.044$) (Fig. 13),

a region known to be involved in the visual-phonological mapping and lexical processing. This might be indicative of a compensatory recruitment of lexical processing mechanisms when phonological violations occur.

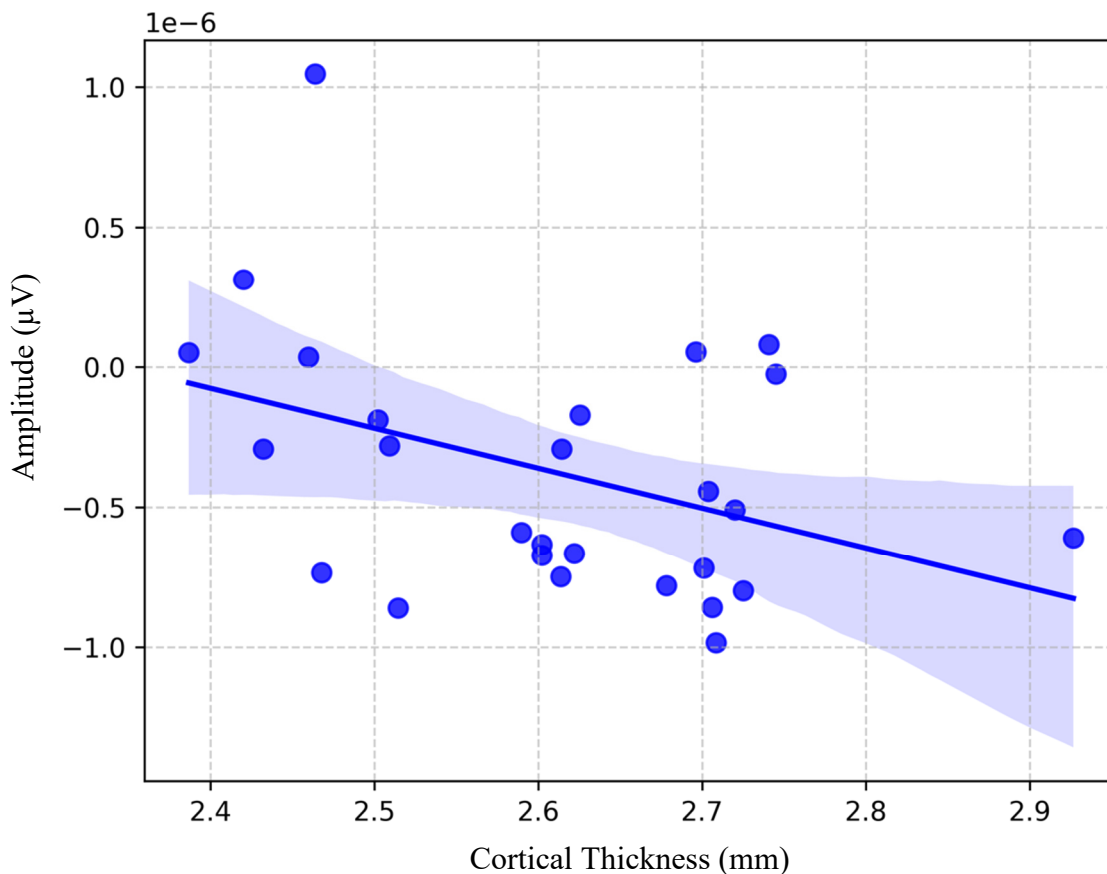


Figure 14. Correlation between phonological ERP amplitudes (500–800 ms, central ROI) and cortical thickness in the left fusiform gyrus ($r = -0.41$, $p = .044$). Each dot represents one participant; the regression line and 95% confidence band are shown.

Finally, the syntactic ERP component for the time window of 500–800 ms in centroparietal region was positively correlated with cortical thickness in the right rostral anterior cingulate cortex ($r = 0.43$, $p = 0.031$) (Fig. 14). This finding points to the involvement of domain-general monitoring and executive control processes for resolving syntactic anomalies in L2

sentence comprehension.

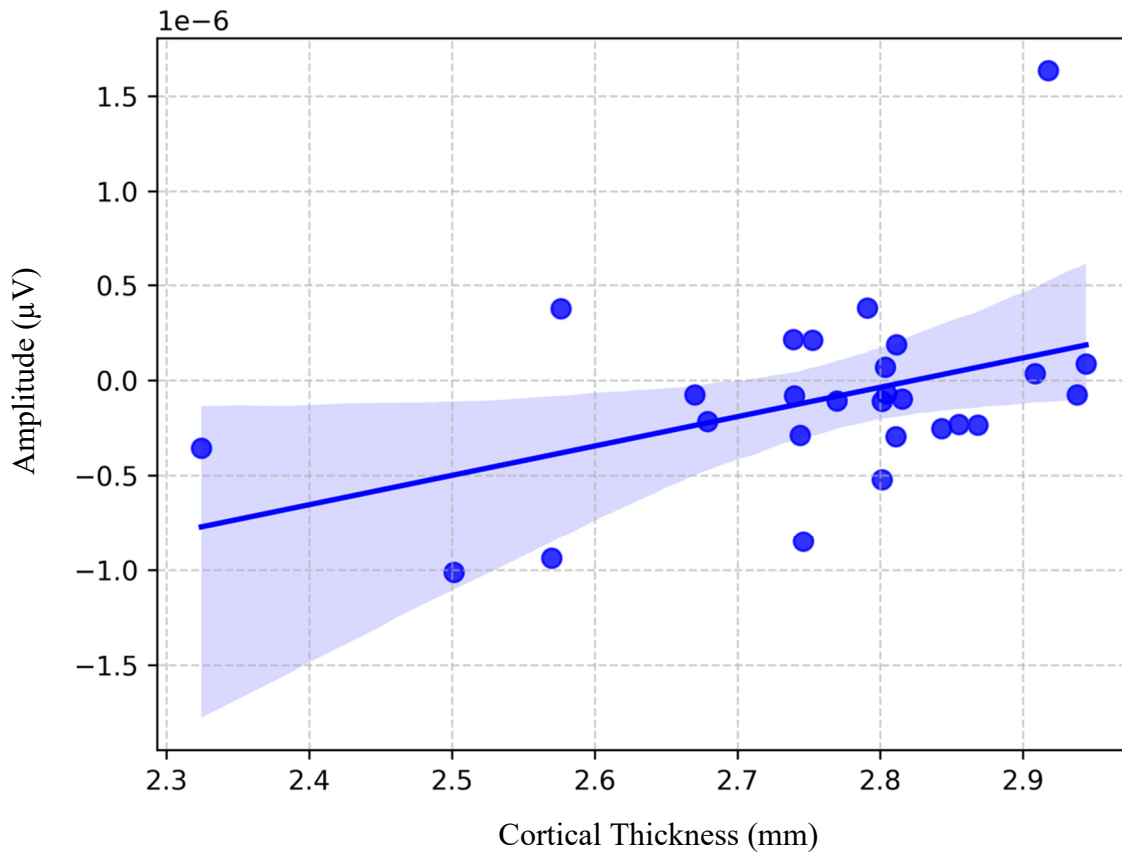


Figure 15. Correlation between syntactic ERP amplitudes (500–800 ms, centroparietal ROI) and cortical thickness in the right rostral anterior cingulate cortex ($r = 0.43$, $p = 0.031$). Each dot represents one participant; the regression line and 95% confidence band are shown.

3.4 Discussion

The cortical thickness analyses in this chapter were designed to identify structural substrates associated with the atypical ERP patterns observed in Japanese learners of English. Several consistent structure–function relationships emerged among syntactic, semantic, phonological, and control conditions, indicating that individual variability in L2

sentence processing is linked to variations in both language-related and executive-control regions.

First, the negative association between semantic ERP amplitudes in the 300–500 ms time window and cortical thickness in the right posterior cingulate cortex was observed for both parietal and centroparietal regions. Participants with thinner cortex in this posterior midline region showed larger (more negative) semantic ERP responses. Functionally, the posterior cingulate has been implicated in semantic monitoring, integration, and default-mode network dynamics. The present pattern suggests that streamlined or more specialized circuitry in this region may support more efficient engagement with semantic information during L2 sentence processing. So, reduced thickness here may index experience-dependent refinement of networks that support the integration of meaning in a L2.

A second major pattern involved late semantic ERP activity (500–800 ms) and cortical thickness in the left rostral middle frontal gyrus. Here again, a negative correlation emerged, and a thinner cortex in this prefrontal region was associated with larger semantic ERP effects. The rostral middle frontal cortex is usually linked to working memory, controlled retrieval, and executive aspects of semantic integration. The present findings are therefore consistent with the view that L2 comprehension relies on prefrontal control mechanisms to support the incremental construction of meaning. Individuals with reduced cortical thickness in this region may be able to deploy these resources more efficiently, resulting in clearer electrophysiological signatures of semantic processing, even in the context of overall non-canonical ERP profiles.

The phonological condition also revealed meaningful structure–function relationships.

Early phonological ERP components (central, 300–500 ms) were negatively correlated with cortical thickness in the right pars opercularis, while late phonological activity (central, 500–800 ms) showed a negative association with thickness in the left fusiform gyrus. The right pars opercularis is often implicated in prosodic analysis, articulatory planning, and error monitoring, particularly in demanding or non-native contexts. The observed negative correlation may suggest that thinner cortex in this region is linked to greater neural sensitivity to phonological anomalies. One possible interpretation is that this reflects experience-related specialization in circuitry supporting the detection of subtle deviations in L2 speech, although this explanation remains tentative. The association with the left fusiform gyrus, a region involved in visual–phonological and orthographic–lexical mapping, may indicate that learners recruit lexical-orthographic mechanisms when confronted with unexpected or not well-formed phonological input in the auditory modality. This is compatible with the idea that late L2 learners rely on a broad set of resources, including the regions associated with reading, to stabilize phonological representations in the L2.

In addition, for syntactic processing, we found a positive correlation between late ERP activity (500–800 ms) at centroparietal sites and cortical thickness in the right rostral anterior cingulate cortex. This underscores the role of monitoring and executive control in L2 syntactic processing. Unlike the negative correlations described above, this positive relationship shows that greater structural resources in the anterior cingulate are associated with stronger late neural responses to syntactic violations. One explanation might be that learners with thicker ACC are better able to detect and resolve conflicts arising from grammatical anomalies, engaging additional monitoring processes when the input conflicts with their developing L2 representations. This finding is consistent with the notion that L2

sentence processing in adulthood relies on an extended control network to compensate for less automatized structural parsing.

In sum, the set of correlations observed in this chapter points to a distributed architecture supporting L2 sentence comprehension. Language-related perisylvian regions (such as pars opercularis and fusiform gyrus) and posterior midline structures (posterior cingulate cortex) jointly contribute to the processing of semantic and phonological information, while executive-control regions (rostral middle frontal gyrus and anterior cingulate cortex) modulate the engagement of monitoring and integration processes. Importantly, several of the strongest correlations involved right-hemisphere homologues of classic left-hemisphere language areas, as well as control regions outside the traditional perisylvian language network. This pattern is consistent with the hypothesis that second-language processing in late learners entails a partial redistribution of processing load across both hemispheres and across language and control networks, rather than a simple reuse of the L1 system.

However, these structure–function correlations are to be considered in the context of the limitations of the methods used. The correlations are based on cross-sectional cortical thickness estimates and cannot establish causal direction. So, it remains unclear whether certain structural profiles support more efficient L2 processing, or whether sustained L2 experience leads to the observed anatomical differences. Finally, the sample consists of intermediate Japanese learners of English only. Therefore, it might be possible that different results might be obtained with more proficient or beginner learners of English or in pairs of languages with a more or less similar typological relation.

Chapter 4

General Discussion

4.1 Introduction

The present study investigated how Japanese L2 learners of English process syntactic, semantic, and phonological information in spoken L2 sentences, integrating behavioral and ERP measures to characterize both performance outcomes and underlying neural mechanisms.

4.2 Interpretation of Behavioral and ERP Results

Behavioral results indicated more accurate performance in the control and semantic conditions, with reduced accuracy for phonological and syntactic violations. These differences suggest that participants were less sensitive to structural and phonological anomalies, potentially reflecting reduced awareness or difficulty in detecting such violations during real-time sentence comprehension. This may indicate lower automaticity in parsing natural morphosyntax and in mapping L2 phonological forms onto lexical representations. A significant main effect of Condition confirmed these differences. We interpret this to suggest that syntactic and phonological features are less salient in L2 processing. This pattern is consistent with observations that L2 processing is characterized by persistent difficulty in automatizing grammatical features and segmenting non-native phonological input [26,27,31–33].

Additional behavioral analyses further refined this pattern. When performance was examined by cue type, participants were most accurate when judgments were guided by semantic cues, whereas phonological and grammar cues produced weaker and less reliable performance. Importantly, reaction times were broadly comparable across the three cue conditions, indicating that the lower accuracy for grammar and phonology was not driven by faster responding, guessing, or reduced effort. Rather, these findings suggest that explicit judgments based on grammatical and phonological information were genuinely more difficult for these Japanese learners of English, whereas meaning-based information remained relatively more accessible during spoken L2 sentence processing.

Unlike the behavioral data, ERP results revealed only subtle neural distinctions between linguistic conditions. In the N400 time window (300–500 ms), phonological violations elicited a small but consistent posterior negativity, particularly at the parietal region. While this negativity overlapped in timing and distribution with the canonical N400, it may not directly reflect phonological encoding. Instead, it may index downstream consequences of impaired lexical access due to earlier disruptions in phonological analysis. That is, mismatches in expected phonological forms may hinder lexical activation, triggering a negativity that resembles an N400 but reflects different underlying processes. This interpretation aligns with the notion that L2 learners, particularly at lower proficiency levels, rely more on surface-level phonological or lexical cues to process input, rather than robust form-meaning mappings [14,26,27].

In contrast, semantic violations did not elicit significant N400 effects relative to the control condition. The absence of a robust semantic N400, typically observed in native speakers,

implies that semantic integration processes in L2 learners may not be fully native-like. Rather than reflecting a delayed N400, this pattern may indicate qualitatively different integration mechanisms in L2 comprehension. Furthermore, task demands may have contributed to this effect. Previous work has demonstrated that the N400 amplitude is modulated by attentional allocation, such that reduced attention to semantic content results in attenuated or absent N400s [37–39]. In the current design, the use of explicit trial-by-trial cues may have directed attention toward more perceptually salient or task-relevant features, such as phonological anomalies, limiting deeper structural or semantic reanalysis. However, since similar tasks still reliably elicit the N400 in native speakers [16,17,39], the present findings indicate that L2 learners may prioritize attentional resources to task-relevant features, such as anomaly detection, rather than to semantic integration. This tradeoff reveals limits in the automaticity of semantic processing and highlights the role of attentional control in L2 learners.

Complementary linear mixed-effects analyses provided a clearer picture of these ERP patterns. In the 300–500 ms window, both the phonological and syntactic conditions showed significantly more negative amplitudes than the control condition, whereas the semantic condition did not differ significantly from control. In the later 500–800 ms window, phonological violations again showed the strongest effect, and semantic violations also differed significantly from control, while the syntactic condition showed only a non-significant trend. These results suggest that some condition effects were more clearly detected in the trial-level mixed-effects framework than in the omnibus ROI-based ANOVAs, reinforcing the conclusion that phonological anomalies produced the most robust and sustained disruption, while syntactic and semantic effects were weaker or less stable.

Although a significant parietal effect was observed for syntactic violations, the corresponding waveform lacked the characteristic shape, distribution, and amplitude of a canonical N400 or P600. This may reflect weak or distributed processing responses, possibly associated with general processing difficulty or low-level lexical disruption. One likely contributing factor is the heterogeneity of the syntactic condition. Our stimuli included subject-verb agreement errors, prepositional usage violations, countability distinctions in nouns, and definiteness and article choice errors. This diversity was a deliberate design feature intended to test whether L2 learners show converging ERP responses across different syntactic anomaly types. However, participants' relatively low accuracy in the syntactic condition resulted in an insufficient number of correct responses per subtype. To preserve statistical power and ensure comparability with other conditions, we therefore collapsed all syntactic violations into a single condition for analysis. While this approach maintained trial counts and analytic feasibility, it may have increased within-condition variability, potentially obscuring consistent ERP effects specific to individual violation types. This interpretation is also supported by the LME analysis, which detected a significant early negativity for the syntactic condition relative to control in the 300–500 ms window, suggesting that syntactic anomalies did elicit a measurable neural cost even though the effect was weaker and less canonical than those typically reported in native speakers.

Interestingly, prior research has shown that early-stage L2 learners often respond to syntactic anomalies with N400-like effects rather than P600s, suggesting reliance on lexically based detection mechanisms rather than structural reanalysis [22–24,37,38]. Our findings are consistent with this developmental account, where syntactic violations may be

processed through disrupted lexical expectations rather than syntactic repair processes. Thus, the absence of P600 effects for syntactic violations likely reflects the limited development of syntactic reanalysis mechanisms in this learner population. At the same time, it is also possible that structural violations were not fully detected or not processed at a depth sufficient to engage repair processes. Either account offers further evidence that the neural signatures of L2 processing differ qualitatively from native-like patterns, particularly for syntactic violations.

In the later P600 window (500–800 ms), no clear P600-like positivities were observed in any condition. Phonological violations continued to elicit a sustained posterior negativity, while semantic and syntactic conditions remained close to baseline. The ANOVA revealed a significant main effect of Condition in this time window, driven primarily by the phonological condition. The complementary LME analysis clarified this pattern further by showing that the phonological condition remained robustly different from control in the later window, while the semantic condition also showed a significant effect and the syntactic condition showed only a marginal trend. This reinforces the interpretation that later-stage L2 processing in the present task was dominated by sustained phonological processing costs rather than by canonical syntactic reanalysis.

Overall, the ERP results suggest that phonological violations were the most salient and consistently processed anomaly type in this study, while syntactic and semantic anomalies elicited weaker or non-canonical neural responses. These findings support the idea that L2 learners rely more heavily on phonological and lexical cues, particularly under time pressure and at lower proficiency levels. This domain-specific modulation of ERP responses is consistent with prior findings, which showed that ERP patterns in L2 learners

are less native-like during early stages of acquisition, and that the nativization process varies by both linguistic domain and modality [21,25,29,30]. In these ERP studies, phonological and semantic processing showed distinct developmental trajectories, suggesting that different neural mechanisms underlie the acquisition of structural versus surface-level processing skills. The current results extend this pattern by showing that phonological violations elicited clear neural effects even when syntactic and semantic anomalies did not, reinforcing the idea that multiple interacting factors, including proficiency, attention, task demands, and linguistic domain, shape L2 processing. At the same time, the additional mixed-effects analyses suggest that subtle early syntactic effects and later semantic effects were present, but were less stable and less canonical than the phonological response. Together, these results contribute to the growing body of evidence that L2 comprehension is not simply a delayed version of native processing, but reflects qualitatively different strategies shaped by both learner-internal and task-external constraints.

These findings underscore the importance of considering linguistic domain, task design, and learner proficiency when interpreting ERP results in L2 contexts and suggest that neural sensitivity to linguistic violations in L2 is highly selective and developmentally dynamic.

4.3 Structure–Function Relationships from ERP–MRI Correlations

In addition to the behavioral and ERP results, the current ERP–MRI correlation analyses provide a novel interpretation of the relationship between individual differences in the brain's structural properties and the processing of L2 sentences. Overall, the ERP–MRI

correlation results suggest that the relationship between structure and function during L2 comprehension is not localized to a single language center but is spread over a set of brain regions. The regions identified as important for the relationship between brain structure and ERP amplitude measures are the pars opercularis, the fusiform gyri, and the rostral middle frontal gyrus. Though most of the identified relationships are left-lateralized as is typical for the linguistic hemisphere of the brain, strong relationships were also observed for the right frontal and cingulate regions.

The nature of the correlations also offers more information about the association between cortical structure and function. The negative correlations for the semantic and phonological tasks indicate that reduced cortical thickness in certain areas is potentially a reflection of more specialized or optimal circuitry for more efficient processing of specific types of information. The positive correlations in the front regions potentially suggest an association between greater cortical resources being required in order to accommodate the task demands in syntactic and L2 sentence analysis. The current results collectively highlight a complex interaction in which both diminished and increased cortical thickness are functionally significant in specific areas.

These structural–function relationships also have important implications for theories of L2 acquisition and neuroplasticity. A substantial literature has indicated that there are often differences between L2 processing on the one hand and L1 processing on the other, both on behavioral and on neurobiological grounds [11–14,21–23,26,27]. The correlations reported here might help to find the possible structural explanation for some of these differences. The engagement of the right-hemisphere homologues of the classic left-hemisphere language areas (such as the right pars opercularis), as well as the anterior

cingulate areas involved in control, suggests a model according to which the processing of the L2 in the adult brain might be the result of additional or alternative routes, which is compatible with a partial redistribution of the processing load on the cortical surface, as constrained by individual anatomical profiles [11–13,23–25].

Furthermore, the correlations found between cortical thickness and ERP components are consistent with experience-dependent plasticity [11–15]. The use of a L2 has the potential to alter cortical thickness as well as patterns of functional activation. Although the current results are cross-sectional and do not determine causality, they offer a necessary stepping stone for future longitudinal research concerning alterations in cortical thickness and related modifications in ERP characteristics of sentence processing during increased proficiency in an L2.

Third, it is only by integrating findings from ERPs and MRIs that it becomes clear how beneficial it can be to have a multi-modal approach to the neurocognition of L2 comprehension. Instead of structure and function being considered in a more or less isolated manner, it is made clear that structure and function are in fact highly interdependent processes. While ERPs are sensitive to the milliseconds involved with comprehension, cortical thickness is clearly a stable individual characteristic that affects these processes in certain ways, reflecting individual differences that might, in fact, look identical from a behavioral perspective.

Overall, the results obtained in this investigation support the existence of a systematic relation between the electrophysiological indices of the on-line processing of sentences in a L2 and the structure of the key areas involved in language use and executive control. This

research not only provides new insights into the neural substrate of bilingual sentence understanding, but it also paves the way for further research related to the role of individual structural differences in shaping L2 acquisition.

4.4 Summary

In this chapter, the behavioral, ERP, and ERP–MRI correlation findings were integrated to provide a unified account of L2 sentence processing in Japanese learners of English. The behavioral results showed that participants performed most successfully when judgments were guided by semantic information, whereas phonological and grammatical cue-based judgments were less accurate, despite broadly comparable reaction times across cue conditions. This pattern suggests that the lower performance in grammar and phonology reflected genuine processing difficulty rather than speeded responding or reduced task engagement.

The ERP findings showed that phonological anomalies elicited the most robust and sustained neural responses, whereas syntactic and semantic violations produced weaker or non-canonical patterns, suggesting domain-specific limits in automaticity and structural reanalysis. Complementary mixed-effects analyses further clarified this pattern by revealing early negativities for both phonological and syntactic violations and later effects for phonological and semantic violations, even when these were less clearly captured by the omnibus ANOVA approach.

At the same time, systematic associations between cortical thickness and ERP amplitudes indicated that individual neuroanatomical profiles shape how L2 sentences are processed, particularly in perisylvian and executive-control regions. Together, these findings support a

view of L2 comprehension as relying on adapted, partially redistributed neural networks, setting the stage for the broader conclusions discussed in Chapter 5.

Chapter 5

Conclusions

5.1 Introduction

The findings of this study broadly support earlier work showing that L2 listeners often experience reduced ability to predict upcoming linguistic elements and weaker structural commitment during real-time comprehension, as reflected in reduced semantic N400 sensitivity and attenuated P600 responses to syntactic anomalies [14,26,27]. In the present study, these effects were especially pronounced for phonological anomalies, which elicited consistent neural modulation in both early and late time windows. The contrast between behavioral performance and relatively weak or non-canonical neural markers suggests a dissociation between explicit decision processes and automatic neural detection in L2 comprehension. In particular, for Japanese learners, phonological mapping difficulties related to L1–L2 phonological mismatch, including segmental contrasts and reduced cues to morphology, may impede form–meaning correspondence early in the processing stream, potentially decreasing the cognitive resources available for subsequent syntactic analysis [30,31–33]. These findings contribute to growing evidence that morphosyntactic and phonological processing remain less automatized in adult-acquired languages, particularly in time-constrained comprehension tasks [22,23,26,27,38].

Additional behavioral analyses refined this interpretation by showing that participants were most accurate when judgments were guided by semantic cues, whereas phonological and grammatical cue-based judgments were less accurate. Importantly, reaction times were broadly similar across cue conditions, indicating that the lower performance in the phonological and grammatical domains did not reflect fast guessing or reduced effort, but rather genuine difficulty in making explicit L2 judgments based on these types of information. These findings suggest that meaning-based information remained relatively more accessible to the learners than phonological or grammatical information during spoken sentence comprehension.

Complementary linear mixed-effects analyses also clarified the ERP findings beyond the omnibus ANOVA results. In the 300–500 ms window, both the phonological and syntactic conditions showed significantly more negative amplitudes than the control condition, whereas the semantic condition did not differ significantly from control. In the later 500–800 ms window, the phonological condition again showed the strongest effect, while the semantic condition also differed significantly from control and the syntactic condition showed only a non-significant trend. Taken together, these results suggest that the behavioral and neural profile of L2 sentence processing is more differentiated across linguistic domains than is apparent from region-based omnibus analyses alone.

Future work should continue to evaluate how these domain-specific processing profiles evolve as a function of proficiency, exposure, and cognitive resources, an issue that may be especially relevant for Japanese learners given the typological characteristics of their native language. At the structural level, the MRI and ERP–MRI correlation analyses showed that cortical thickness in key language-related and executive-control regions is

systematically linked to individual variability in ERP responses to L2 sentence processing [11–15,23–25]. These structure–function associations clarify how anatomical differences in frontal, temporal, and cingulate areas constrain the efficiency with which learners detect and resolve phonological, semantic, and syntactic anomalies. More broadly, the findings delineate the neural architecture supporting bilingual sentence comprehension and highlight individual anatomical variability as an important factor shaping second-language learning outcomes in adulthood.

5.2 Limitations

Nevertheless, several limitations of the present study should be acknowledged. First, the limited number of participants and their relatively restricted range of English proficiency constrain the generalizability of the results. In other words, the findings primarily reflect intermediate Japanese learners of English and may not apply in the same way to beginners, advanced speakers, or learners with different educational or linguistic backgrounds.

Second, the study focused exclusively on one L1–L2 pairing, Japanese–English, which is characterized by substantial typological differences in both phonology and morphosyntax. As a result, some of the observed patterns, particularly the prominence of phonological difficulties, may be specific to this language pair rather than universal features of L2 processing. In addition, the syntactic condition included several different subtypes of violations, such as subject–verb agreement errors, prepositional misuse, countability violations, and article-related errors. Although this design allowed the study to assess syntactic processing across a broader range of structures, it may also have increased

within-condition variability and reduced the likelihood of observing a single robust and canonical ERP response.

Third, although the additional cue-based behavioral analyses, mixed-effects analyses, and temporal clustering analyses provided useful complementary information, some of these analyses should be interpreted with caution. The cue-based analyses were intended to refine the behavioral interpretation rather than serve as the sole basis for inferential claims, and the temporal clustering results were supplementary to the primary ROI-based and mixed-effects ERP analyses. Accordingly, these additional analyses strengthen the interpretation of the findings but do not eliminate the need for replication in larger samples.

Finally, the MRI analyses relied on cross-sectional cortical thickness estimates and did not incorporate other structural or functional measures such as white-matter connectivity, subcortical volumes, or fMRI-based activation patterns. A broader multimodal imaging approach would provide a more comprehensive account of the neural architecture underlying L2 sentence processing and may reveal additional pathways linking anatomy and online language function.

5.3 Future Directions

These limitations also point to several promising directions for future research. Longitudinal studies following learners over extended periods of time would be especially useful for understanding how changes in exposure, proficiency, and training influence ERP

components and structural characteristics of the cerebral cortex. Such work would help clarify whether the weak or non-canonical ERP responses observed here reflect a transitional developmental stage, a stable feature of adult L2 processing, or an interaction between the two.

Future studies should also compare different L1/L2 pairings and include native speaker control groups in order to distinguish patterns that are specific to Japanese learners of English from those that reflect more general aspects of L2 cognition. In particular, direct comparisons across typologically different language pairs would help determine whether the strong phonological effects observed in the present study are driven primarily by Japanese–English mismatch or reflect a broader tendency for phonological processing to dominate under certain task conditions.

Another important direction would be to examine cue-based processing more directly by systematically manipulating the reliability, salience, or attentional relevance of semantic, phonological, and grammatical cues. Because the present findings suggest that learners performed best when judgments were guided by semantic information, future work could investigate how cueing interacts with linguistic domain, attentional control, and task demands in shaping both behavior and neural responses during L2 comprehension.

Methodologically, larger samples would also allow more systematic comparisons between repeated-measures ANOVA, linear mixed-effects modeling, and cluster-based approaches. The present results suggest that some subtle effects, particularly early syntactic negativities and later semantic effects, may become more visible when trial-level variability is modeled explicitly. Future studies should therefore explore the extent to which different analytical

frameworks capture distinct but complementary aspects of L2 sentence processing.

Finally, the multimodal approach adopted here could be extended through the inclusion of diffusion-weighted imaging, resting-state or task-based fMRI, and more detailed cognitive profiling, including working memory, attentional control, and executive functioning. Such approaches would further clarify how individual differences in domain-general abilities and network connectivity contribute to variability in L2 comprehension. Together, these lines of research would build on the present findings to develop a richer and more comprehensive account of how structural and functional brain properties interact with language experience to shape second-language learning across the lifespan.

5.4 Final Remarks

The present thesis combined behavioral, electrophysiological, and structural MRI measures to provide an integrated account of how Japanese learners of English process syntactic, semantic, and phonological information in a L2. The behavioral findings showed that participants performed most successfully when judgments were guided by semantic information, whereas phonological and grammatical cue-based judgments were less accurate despite comparable response times, indicating genuine domain-specific difficulty rather than reduced effort or impulsive responding. At the neural level, phonological anomalies elicited the most robust and sustained ERP effects, whereas syntactic and semantic violations produced weaker or non-canonical patterns. Complementary mixed-effects analyses further clarified that subtle early syntactic effects and later semantic effects were present, even when these were less clearly captured by omnibus ANOVA results.

At the same time, the ERP–MRI correlations demonstrated that individual neuroanatomical variation contributes meaningfully to how L2 sentences are processed, particularly in regions associated with language, executive control, and integrative processing. These findings underscore the importance of considering not only linguistic domain and learner proficiency, but also structural brain differences when interpreting variability in second-language comprehension. Overall, this thesis contributes to a more nuanced understanding of second-language neurocognition by showing that L2 sentence processing is shaped by the interaction of behavioral demands, neural sensitivity, and individual anatomical constraints. In doing so, it provides a foundation for future multimodal research on bilingual language processing and the neurobiological basis of adult language learning.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor Shinri Ohta, for his guidance and support throughout the writing of this thesis. I am also grateful to Dr. Daniel C. Gallagher and Professor Emi Yamada for their valuable support and encouragement. I would like to thank my friend and lab mate, Haruki Noda, for his support throughout my PhD journey, as well as the other members of the lab.

Above all, I would like to thank my parents, Niloufar and Mehran, for their endless love, patience, encouragement, and sacrifices, without which this journey would not have been possible. Maman, Baba, doosetun daram. I am also deeply grateful to my dear friend Kuniko for her kindness, support, and constant presence. She always took my hand during difficult times, gave me hope, stood by me, and loved me like a member of her family.

I dedicate this dissertation to my mother, my father, my family, and Kuniko-san.

Jananeh Shalpoush

January 2026

References

1. Cocchi, L.; Zalesky, A.; Fornito, A.; Mattingley, J.B. Dynamic Cooperation and Competition between Brain Systems during Cognitive Control. *Trends Cogn. Sci.* 2013, *17*, 493–501. doi:10.1016/j.tics.2013.08.006.
2. Fedorenko, E.; Ivanova, A.A.; Regev, T.I. The Language Network as a Natural Kind within the Broader Landscape of the Human Brain. *Nat. Rev. Neurosci.* 2024, *25*, 289–312. doi:10.1038/s41583-024-00802-4.
3. Hertrich, I.; Dietrich, S.; Ackermann, H. The Margins of the Language Network in the Brain. *Front. Commun.* 2020, *5*, 519955.
4. Hashimoto, R.; Sakai, K.L. Specialization in the Left Prefrontal Cortex for Sentence Comprehension. *Neuron* 2002, *35*, 589–597. doi:10.1016/S0896-6273(02)00788-2.
5. Musso, M.; Moro, A.; Glauche, V.; Rijntjes, M.; Reichenbach, J.; Büchel, C.; Weiller, C. Broca's Area and the Language Instinct. *Nat. Neurosci.* 2003, *6*, 774–781. doi:10.1038/nn1077.
6. Hagoort, P. MUC (Memory, Unification, Control) and Beyond. *Front. Psychol.* 2013, *4*, 416. doi:10.3389/fpsyg.2013.00416.
7. Mesgarani, N.; Cheung, C.; Johnson, K.; Chang, E.F. Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science* 2014, *343*, 1006–1010. doi:10.1126/science.1245994.
8. Kocagoncu, E.; Clarke, A.; Devereux, B.J.; Tyler, L.K. Decoding the Cortical Dynamics of Sound–Meaning Mapping. *J. Neurosci.* 2017, *37*, 1312–1319.

- doi:10.1523/JNEUROSCI.2858-16.2016.
9. Pallier, C.; Devauchelle, A.D.; Dehaene, S. Cortical Representation of the Constituent Structure of Sentences. *Proc. Natl. Acad. Sci. USA* 2011, *108*, 2522–2527.
doi:10.1073/pnas.1018711108.
 10. Friederici, A.D. Towards a Neural Basis of Auditory Sentence Processing. *Trends Cogn. Sci.* 2002, *6*, 78–84. doi:10.1016/S1364-6613(00)01839-8.
 11. Perani, D.; Abutalebi, J. The Neural Basis of First and Second Language Processing. *Curr. Opin. Neurobiol.* 2005, *15*, 202–206.
 12. Abutalebi, J.; Green, D.W. Neuroimaging of Language Control in Bilinguals: Neural Adaptation and Reserve. *Biling. Lang. Cogn.* 2016, *19*, 689–698.
doi:10.1017/S1366728916000225.
 13. Abutalebi, J. Neural Aspects of Second Language Representation and Language Control. *Acta Psychol. (Amst.)* 2008, *128*, 466–478. doi:10.1016/j.actpsy.2008.03.014.
 14. Indefrey, P. A Meta-Analysis of Hemodynamic Studies on First and Second Language Processing: Which Suggested Differences Can We Trust and What Do They Mean? *Lang. Learn.* 2006, *56*, 279–304.
 15. Berken, J.A.; Gracco, V.L.; Chen, J.-K.; Klein, D. The Timing of Language Learning Shapes Brain Structure Associated with Articulation. *Brain Struct. Funct.* 2016, *221*, 3591–3600. doi:10.1007/s00429-015-1121-9.
 16. Kutas, M.; Federmeier, K.D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* 2011, *62*, 621–647.
 17. Kutas, M.; Hillyard, S.A. Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science* 1980, *207*, 203–205. doi:10.1126/science.7350657.
 18. Osterhout, L.; Holcomb, P.J. Event-Related Brain Potentials Elicited by Syntactic

- Anomaly. *J. Mem. Lang.* 1992, 31, 785–806. doi:10.1016/0749-596X(92)90039-Z.
19. Kuperberg, G.R. Neural Mechanisms of Language Comprehension: Challenges to Syntax. *Brain Res.* 2007, 1146, 23–49.
 20. Sassenhagen, J.; Schlesewsky, M.; Bornkessel-Schlesewsky, I. The P600-as-P3 Hypothesis Revisited: Single-Trial Analyses Reveal That the Late EEG Positivity Following Linguistically Deviant Material Is Reaction Time Aligned. *Brain Lang.* 2014, 137, 29–39.
 21. Steinhauer, K.; White, E.J.; Drury, J.E. Temporal Dynamics of Late Second Language Acquisition: Evidence from Event-Related Brain Potentials. *Second Lang. Res.* 2009, 25, 13–41.
 22. Hahne, A.; Friederici, A.D. Processing a Second Language: Late Learners' Comprehension Mechanisms as Revealed by Event-Related Brain Potentials. *Biling. Lang. Cogn.* 2001, 4, 123–141.
 23. Weber-Fox, C.M.; Neville, H.J. Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *J. Cogn. Neurosci.* 1996, 8, 231–256.
 24. McLaughlin, J.; Tanner, D.; Pitkänen, I.; Frenck-Mestre, C.; Inoue, K.; Valentine, G.; Osterhout, L. Brain Potentials Reveal Discrete Stages of L2 Grammatical Learning. *Lang. Learn.* 2010, 60, 123–150.
 25. Bowden, H.W.; Steinhauer, K.; Sanz, C.; Ullman, M.T. Native-like Brain Processing of Syntax Can Be Attained by University Foreign Language Learners. *Neuropsychologia* 2013, 51, 2492–2511.
 26. Clahsen, H.; Felser, C. Grammatical Processing in Language Learners. *Appl. Psycholinguist.* 2006, 27, 3–42.
 27. Kaan, E. Predictive Sentence Processing in L2 and L1: What Is Different? *Linguist.*

- Approaches Biling.* 2014, 4, 257–282.
28. Tanner, D.; Van Hell, J.G. ERPs Reveal Individual Differences in Morphosyntactic Processing. *Neuropsychologia* 2014, 56, 289–301.
 29. Gallagher, D.; Yano, M.; Ohta, S. The Neurophysiological Modality Effect in Native and Second Language Processing: An ERP Study. *bioRxiv* 2022.
doi:10.1101/2022.12.17.520859.
 30. Lin, Y.; Fan, R.; Mo, L. Differences in Phonetic Discrimination Stem from Differences in Psychoacoustic Abilities in Learning the Sounds of a Second Language: Evidence from ERP Research. *PLoS ONE* 2017, 12, e0187135.
 31. Miyawaki, K.; Jenkins, J.J.; Strange, W.; Liberman, A.M.; Verbrugge, R.; Fujimura, O. An Effect of Linguistic Experience: The Discrimination of [r] and [l] by Native Speakers of Japanese and English. *Percept. Psychophys.* 1975, 18, 331–340.
 32. Cutler, A. *Native Listening: Language Experience and the Recognition of Spoken Words*; MIT Press: Cambridge, MA, USA, 2012.
 33. Tremblay, A.; Broersma, M.; Coughlin, C.E.; Choi, J. Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation. *Front. Psychol.* 2016, 7, 985.
 34. Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; et al. MEG and EEG Data Analysis with MNE-Python. *Front. Neurosci.* 2013, 7, 267.
 35. Jas, M.; Engemann, D.; Raimondo, F.; Bekhti, Y.; Gramfort, A. Automated Rejection and Repair of Bad Trials in MEG/EEG. In *Proceedings of the 2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*; 2016; pp. 1–4.
 36. Tanner, D.; Inoue, K.; Osterhout, L. Brain-Based Individual Differences in Online L2 Grammatical Comprehension. *Biling. Lang. Cogn.* 2014, 17, 277–293.
 37. Erlbeck, H.; Kübler, A.; Kotchoubey, B.; Veser, S. Task Instructions Modulate the

Attentional Mode Affecting the Auditory MMN and the Semantic N400. *Front. Hum. Neurosci.* 2014, 8, 786.

38. Hahne, A.; Friederici, A.D. Differential Task Effects on Semantic and Syntactic Processes as Revealed by ERPs. *Cogn. Brain Res.* 2002, 13, 339–356.
39. Van Petten, C.; Luka, B.J. Neural Localization of Semantic Context Effects in Electromagnetic and Hemodynamic Studies. *Brain Lang.* 2006, 97, 279–293.

Appendices

Appendix A: 240 main experimental sentences

Set one: Focusing on agreement (third person singular/proper noun and verb form)

(n=15).

1)

Syntactically Correct: The cat chases the mouse.

Syntactically Incorrect: The cat chase the mouse.

Phonological violation: The cat shases the mouse.

Semantically Incorrect: The cat studies the mouse.

2)

Syntactically Correct: She sings beautifully.

Syntactically Incorrect: She sing beautifully.

Phonological violation: She zings beautifully.

Semantically Incorrect: She thunders beautifully.

3)

Syntactically Correct: Sara reads novels every day.

Syntactically Incorrect: Sara read novels every day.

Phonological violation: Sara reats novels every day.

Semantically Incorrect: Sara snows novels every day.

4)

Syntactically Correct: We eat dinner together.

Syntactically Incorrect: We eats dinner together.

Phonological violation: We ead dinner together.

Semantically Incorrect: We rain dinner together.

5)

Syntactically Correct: The flowers bloom in spring.

Syntactically Incorrect: The flowers blooms in spring.

Phonological violation: The flowers floom in spring.

Semantically Incorrect: The flowers groom in spring.

6)

Syntactically Correct: He plays tennis every Sunday.

Syntactically Incorrect: He play tennis every Sunday.

Phonological violation: He blay tennis every Sunday.

Semantically Incorrect: He rusts tennis every Sunday.

7)

Syntactically Correct: The dog barks loudly.

Syntactically Incorrect: The dog bark loudly.

Phonological violation: The dog varks loudly.

Semantically Incorrect: The dog flies loudly.

8)

Syntactically Correct: She washes the dishes.

Syntactically Incorrect: She wash the dishes.

Phonological violation: She vashes the dishes.

Semantically Incorrect: She barks the dishes.

9)

Syntactically Correct: Mary drives a red car.

Syntactically Incorrect: Mary drive a red car.

Phonological violation: Mary trives a red car.

Semantically Incorrect: Mary floods a red car.

10)

Syntactically Correct: They study English diligently.

Syntactically Incorrect: They studies English diligently.

Phonological violation: They stody English diligently.

Semantically Incorrect: They rattle English diligently.

11)

Syntactically Correct: I work very hard.

Syntactically Incorrect: I works very hard.

Phonological violation: I verk very hard.

Semantically Incorrect: I frost very hard.

12)

Syntactically Correct: The birds chirp in the morning.

Syntactically Incorrect: The birds chirps in the morning.

Phonological violation: The birds shurp in the morning.

Semantically Incorrect: The birds paint in the morning.

13)

Syntactically Correct: We swim in the pool.

Syntactically Incorrect: We swims in the pool.

Phonological violation: We svim in the pool.

Semantically Incorrect: We drip in the pool.

14)

Syntactically Correct: The hikers enjoy the view.

Syntactically Incorrect: The hikers enjoys the view.

Phonological violation: The hikers emjoy the view.

Semantically Incorrect: The hikers fizzle the view.

15)

Syntactically Correct: You cook delicious meals.

Syntactically Incorrect: You cooks delicious meals.

Phonological violation: You gook delicious meals.

Semantically Incorrect: You seep delicious meals.

Set two: Focusing on prepositions “on” and “of” (n=15).

1)

Syntactically Correct: She ran out of time.

Syntactically Incorrect (targeting "of"): She ran out time.

Phonological violation: She ran out of tine.

Semantically incorrect: She ran out of snow.

2)

Syntactically Correct: They get rid of every item.

Syntactically Incorrect (targeting "of"): They get rid every item.

Phonological violation: They get rid of efery item.

Semantically incorrect: They get rid of happiness.

3)

Syntactically Correct: They focused on language skills.

Syntactically Incorrect (targeting "on"): They focused language skills.

Phonological violation: They focused on lankuage skills.

Semantically incorrect: They focused on dust.

4)

Syntactically Correct: She took advantage of opportunities.

Syntactically Incorrect (targeting "of"): She took advantage opportunities.

Phonological violation: She took advantage of oppordunities.

Semantically incorrect: She took advantage of noise.

5)

Syntactically Correct: He was accused of crime.

Syntactically Incorrect (targeting "of"): He was accused crime.

Phonological violation: He was accused of crine.

Semantically incorrect: He was accused of ice.

6)

Syntactically Correct: She relies on friends.

Syntactically Incorrect (targeting "on"): She relies friends.

Phonological violation: She relies on vriends.

Semantically incorrect: She relies on oceans.

7)

Syntactically Correct: We insisted on refunds.

Syntactically Incorrect (targeting "on"): We insisted refunds.

Phonological violation: We insisted on rebunds.

Semantically incorrect: We insisted on error.

8)

Syntactically Correct: They drew on knowledge.

Syntactically Incorrect (targeting "on"): They drew knowledge.

Phonological violation: They drew on knowredge.

Semantically incorrect: They drew on chairs.

9)

Syntactically Correct: The project depended on effort.

Syntactically Incorrect (targeting "on"): The project depended effort.

Phonological violation: The project depended on affort.

Semantically incorrect: The project depended on rocks.

10)

Syntactically Correct: He warned of risks.

Syntactically Incorrect (targeting "of"): He warned risks.

Phonological violation: He warned of lisks.

Semantically incorrect: He warned of kitchen.

11)

Syntactically Correct: He leaned on expertise.

Syntactically Incorrect (targeting "on"): He leaned expertise.

Phonological violation: He leaned on ixpertise.

Semantically incorrect: He leaned on space.

12)

Syntactically Correct: The team piled on pressure.

Syntactically Incorrect (targeting "on"): The team piled pressure.

Phonological violation: The team piled on prassure.

Semantically incorrect: The team piled on color.

13)

Syntactically Correct: They banked on success.

Syntactically Incorrect (targeting "on"): They banked success.

Phonological violation: They banked on guccess.

Semantically Incorrect: They banked on spaghetti.

14)

Syntactically Correct: We took on responsibilities.

Syntactically Incorrect (targeting "on"): We took responsibilities.

Phonological violation: We took on desposibilities.

Semantically incorrect: We took on unicorn.

15)

Syntactically Correct: The lecture touched on topics.

Syntactically Incorrect (Targeting "on"): The lecture touched topics.

Phonological violation: The lecture touched on kopics.

Semantically incorrect: The lecture touched on spoons.

Set three: Focusing on countable and uncountable nouns (n=15).

1)

Syntactically Correct: I bought three books.

Syntactically Incorrect (targeting Countable): I bought three book.

Phonological violation: I bought three pooks.

Semantically incorrect: I bought three dreams.

2)

Syntactically Correct: She drank a glass of water.

Syntactically Incorrect (targeting Uncountable): She drank a glass of waters.

Phonological violation: She drank a glass of vater.

Semantically incorrect: She drank a glass of pizza.

3)

Syntactically Correct: She bought some milk.

Syntactically Incorrect (targeting Uncountable): She bought some milks.

Phonological violation: She bought some milg.

Semantically incorrect: She bought some happiness.

4)

Syntactically Correct: She needs some advice.

Syntactically Incorrect (targeting Uncountable): She needs some advices.

Phonological violation: She needs some adbice.

Semantically incorrect: She needs some cloud.

5)

Syntactically Correct: The teacher gave us two pens.

Syntactically Incorrect (targeting Countable): The teacher gave us two pen.

Phonological violation: The teacher gave us two pems.

Semantically incorrect: The teacher gave us two planets.

6)

Syntactically Correct: He purchased a loaf of bread.

Syntactically Incorrect (Targeting Uncountable): He purchased a loaf of breads.

Phonological violation: He purchased a loaf of pread.

Semantically incorrect: He purchased a loaf of thread.

7)

Syntactically Correct: They adopted three cats.

Syntactically Incorrect (targeting Countable): They adopted three cat.

Phonological violation: They adopted three gats.

Semantically incorrect: They adopted three bikes.

8)

Syntactically Correct: Mother packed a bag of rice.

Syntactically Incorrect (targeting Uncountable): Mother packed a bag of rices.

Phonological violation: Mother packed a bag of roice.

Semantically incorrect: Mother packed a bag of mice.

9)

Syntactically Correct: We bought some new chairs.

Syntactically Incorrect (targeting Countable): We bought some new chair.

Phonological violation: We bought some new jaires.

Semantically incorrect: We bought some new laughter.

10)

Syntactically Correct: We drank a cup of coffee.

Syntactically Incorrect (targeting Uncountable): We drank a cup of coffees.

Phonological violation: We drank a cup of covee.

Semantically incorrect: We drank a cup of pasta.

11)

Syntactically Correct: She saw four birds.

Syntactically Incorrect (targeting Countable): She saw four bird.

Phonological violation: She saw four pirds.

Semantically incorrect: She saw four rains.

12)

Syntactically Correct: I made a bowl of soup.

Syntactically Incorrect (targeting Uncountable): I made a bowl of soups.

Phonological violation: I made a bowl of shoups.

Semantically incorrect: I made a bowl of traffic.

13)

Syntactically Correct: The mechanic fixed two cars.

Syntactically Incorrect (targeting Countable): The mechanic fixed two car.

Phonological violation: The mechanic fixed two gars.

Semantically incorrect: The mechanic fixed two homeworks.

14)

Syntactically Correct: She needs some information.

Syntactically Incorrect (targeting Uncountable): She needs some informations.

Phonological violation: She needs some invormation.

Semantically incorrect: She needs some thunders.

15)

Syntactically Correct: She baked several cupcakes.

Syntactically Incorrect (targeting Countable): She baked several cupcake.

Phonological violation: She baked several gupcakes.

Semantically incorrect: She baked several books.

Set four: Focusing on definite and indefinite articles (n=15).

1)

Syntactically Correct (Indefinite Article - Singular): She visits a friend.

Syntactically Incorrect: She visits friend.

Phonological violation: She visits a vriend.

Semantically incorrect: She visits a carpet.

2)

Syntactically Correct (Indefinite Article - Singular): An apple fell from the tree.

Syntactically Incorrect: An apple fell from tree.

Phonological violation: An apple fell from the jree.

Semantically incorrect: An apple fell from the room.

3)

Syntactically Correct (Definite Article - Singular): That book on the table is interesting.

Syntactically Incorrect: That book on table is interesting.

Phonological violation: That book on the dable is interesting.

Semantically incorrect: That book on the dream is interesting.

4)

Syntactically Correct (Indefinite Article - Singular): He bought a watch.

Syntactically Incorrect: He bought watch.

Phonological violation: He bought a vatch.

Semantically incorrect: He bought a respect.

5)

Syntactically Correct (Definite Article - Singular): The bus stopped at the light.

Syntactically Incorrect: The bus stopped at light.

Phonological violation: The bus stopped at the loight.

Semantically incorrect: The bus stopped at the table.

6)

Syntactically Correct (Indefinite Article - Singular): She watched a film.

Syntactically Incorrect: She watched film.

Phonological violation: She watched a vilm.

Semantically incorrect: She watched a thought.

7)

Syntactically Correct (Definite Article - Singular): The bird sang a song.

Syntactically Incorrect: The bird sang song.

Phonological violation: The bird sang a zong.

Semantically incorrect: The bird sang a decision.

8)

Syntactically Correct (Indefinite Article - Singular): He started a job.

Syntactically Incorrect: He started job.

Phonological violation: He started a chob.

Semantically incorrect: He started a paper.

9)

Syntactically Correct (Definite Article - Singular): The movie at the cinema was interesting.

Syntactically Incorrect: The movie at cinema was interesting.

Phonological violation: The movie at the zinema was interesting.

Semantically incorrect: The movie at the desk was interesting.

10)

Syntactically Correct (Definite Article - Singular): He passed the test.

Syntactically Incorrect: He passed test.

Phonological violation: He passed the tist.

Semantically incorrect: He passed the air.

11)

Syntactically Correct (Definite Article - Singular): The house on the corner looks stunning.

Syntactically Incorrect: The house on corner looks stunning.

Phonological violation: The house on the colner looks stunning.

Semantically incorrect: The house on the sofa looks stunning.

12)

Syntactically Correct (Definite Article - Singular): The sun sets in the west.

Syntactically Incorrect: The sun sets in west.

Phonological violation: The sun sets in the wist.

Semantically incorrect: The sun sets in the concept.

13)

Syntactically Correct (Definite Article - Singular): The flower in the garden bloomed today.

Syntactically Incorrect: The flower in garden bloomed today.

Phonological violation: The flower in the carden bloomed today.

Semantically incorrect: The flower in the curtain bloomed today.

14)

Syntactically Correct (Indefinite Article - Singular): She lives in an apartment.

Syntactically Incorrect: She lives in apartment.

Phonological violation: She lives in an abartment.

Semantically incorrect: She lives in an atom.

15)

Syntactically Correct (Definite Article - Singular): Everything at the gallery was fantastic.

Syntactically Incorrect: Everything at gallery was fantastic.

Phonological violation: Everything at the garery was fantastic.

Semantically incorrect: Everything at the shirt was fantastic.

Appendix B: 120 filler experimental sentences

1. I cleaned my room.
2. We will study math.
3. He stopped the car.
4. I don't like cooking.
5. I prefer tea to coffee.
6. I will arrange the plan.
7. I made a new routine.
8. This house is very old.
9. Paris is very crowded.
10. It will rain tonight.
11. She chose a red pen.
12. He wears glasses.
13. She owns a nice flat.
14. He won't go to work.
15. You should accept her offer.
16. The weather is cloudy.
17. Festivals will be held.
18. My camera is on.
19. I need a highlighter.
20. I want an umbrella.
21. She confirms her order.

22. She is back to work.
23. He ordered a drink.
24. I will prepare a file.
25. We took a picture.
26. The show will start soon.
27. They will watch a drama.
28. This task is hard.
29. She cut her hair.
30. He wears a shirt.
31. We missed the flight.
32. He is very busy.
33. He has a son.
34. My meeting has started.
35. They play board games.
36. We watch scary movies.
37. The sun rises early.
38. She likes pop music.
39. She paints her nails.
40. He cooks tasty meals.
41. Kids ride new bikes.
42. The baby cries loudly.
43. She loves fresh flowers.
44. He repairs old cars.
45. Birds find worms easily.
46. We build strong houses.

47. The rain falls gently.
48. She wears red dresses.
49. He teaches math lessons.
50. Kids play in the park.
51. The dog runs fast.
52. She drinks hot coffee.
53. He catches fish.
54. Birds fly over trees.
55. They plant small seeds.
56. We paint white walls.
57. She buys new shoes.
58. Kids laugh at jokes.
59. He draws funny cartoons.
60. They watch videos.
61. She performs well.
62. We chat every day.
63. I will review the notes.
64. He plays guitar weekly.
65. It seems very cold.
66. They trained perfectly.
67. The wind blows fiercely.
68. We cycle in the park.
69. The dancers move quickly.
70. You bake tasty pies.
71. He went on a hike.

72. The jar is full of cookies.
73. Your hat is on the rug.
74. She is fond of singing.
75. My dog lies on the floor.
76. Their shed is on fire.
77. He is on the phone.
78. The leaves turn colors.
79. She is an artist.
80. The waves crash loudly.
81. You cook delicious soups.
82. The mechanic fixed two bikes.
83. The movie was thrilling.
84. A teacher gave a lecture.
85. I walk to school.
86. She makes jewelry.
87. He collects stamps.
88. She wears gloves.
89. She is sick.
90. I take English lessons.
91. He always wakes up early.
92. She drives trucks.
93. They iron their clothes.
94. They type emails.
95. She decorates new houses.
96. He polished his shoes.

97. She plants herbs.
98. I exercise every day.
99. We packed our luggage.
100. She organizes her closet.
101. Kids invent fun games.
102. He repairs old clocks.
103. She observed my classes.
104. The rain soaks the garden.
105. He slices fresh bread.
106. She waves hands.
107. The wind shakes trees.
108. They set the table.
109. We capture amazing photos.
110. They recycle everything.
111. We discuss the plan.
112. He adjusts his glasses.
113. She updates the software.
114. I assembled the furniture.
115. They complete the puzzle.
116. He monitors the situation.
117. She downloads new apps.
118. The sun dries clothes.
119. They exchange old books.
120. She records her voice.

Appendix C: Temporal Cluster-Based Analysis

In addition to the primary ROI-based ANOVAs and complementary linear mixed-effects analyses, an exploratory temporal cluster-based analysis was conducted on the epoched ERP data to examine whether the observed condition effects formed broad sustained temporal clusters across the post-stimulus interval. The analysis did not reveal broad or continuous clusters corresponding to the main ERP patterns reported in the thesis. Instead, it yielded only brief and fragmented effects, including a short syntactic effect around 190 ms and several brief phonological effects between approximately 387 and 539 ms. Because these effects were highly short-lived and did not form robust continuous clusters, they were interpreted cautiously and were not treated as primary evidence for the main conclusions of the study. The main interpretation of the ERP findings therefore relies on the ROI-based analyses and the complementary linear mixed-effects models, which provided more stable and interpretable results for the N400 and P600 time windows.