Beyond Modalities: Robust Neural representation of Language in the Brain

UNDERGRADUATE THESIS

Submitted in partial fulfillment of the requirements of
BITS F421T Thesis

By

Shreyas K
ID No. 2020B4AA0714G

Under the supervision of:

Dr. Gabriel KREIMAN

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, GOA CAMPUS
December 2024

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, GOA CAMPUS

Abstract

MSc Mathematics and Bachelor of Engineering in Electronics and Communication

Beyond Modalities: Robust Neural representation of Language in the Brain

by Shreyas K

This thesis investigates the neural mechanisms underlying cross-modal language processing through intracranial recordings in human participants. Using a novel experimental paradigm combining auditory and visual sentence presentation, we demonstrate distinct neural signatures for semantic and grammatical processing across sensory modalities. Our findings reveal both modality-specific and modality-independent neural representations in language processing, contributing to our understanding of how the brain integrates linguistic information across different sensory inputs.

Acknowledgements

I extend my heartfelt gratitude to Dr. Gabriel Kreiman, my thesis supervisor, for his unwavering guidance and support throughout this research journey. Special appreciation goes to my lab mates Elisa Pavarino, Dianna Hidalgo, Morgan Talbot and Bastien le lan for their invaluable insights during every lab meeting that enriched the depth of my work.

I would like to express my sincere gratitude to Dr. Amalin Prince, Head of the Department of Electrical and Electronics and my on-campus supervisor, for granting me the opportunity to embark on this project.

My deepest appreciation goes to my parents and friends for their constant encouragement and understanding during this challenging endeavour.

I am deeply indebted to Dr. Pranav Misra, a PhD student at the Kreiman Lab, for providing me with the opportunity to work on this thesis and for the valuable resources and support from the lab. The Kreiman Lab gave me the opportunity to study language in the brain with high spatiotemporal resolution via sEEG recordings, enabling me to explore new horizons in the realm of Computational Neuroscience and Machine Learning

Contents

Declaration of Authorship **i**

Certificate **ii**

Abstract **iv**

Acknowledgements **v**

Contents **vi**

List of Figures **viii**

Abbreviations **x**

1 Introduction **1**
1.1 Background 1
1.2 Significance 2
1.3 Research Objectives 3
1.4 Approach 3

2 Literature Review **5**
2.1 Foundations of Language Processing 5
2.2 From Localization to Distributed Networks 6
2.3 Modality Effects in Language Comprehension 6
2.4 Cross-Modal Integration and Abstract Representations 7
2.5 Temporal Dynamics of Language Processing 7
2.6 Bridging Neuroscience, Linguistics, and Technology 8
2.7 Gaps and Future Directions 8

3 Methods **10**
3.1 Task Design 10
3.1.1 Stimulus Categories 10
3.1.2 Presentation Parameters 11
3.2 Data Analysis 12
3.2.1 Signal Processing 12
3.2.2 Classification Analysis 12

Classification Code Example . . . 13
3.3 Statistical Methods . . . 15

4 Results **16**
4.1 Individual Electrode Analysis . . . 16
4.1.1 Grammar Processing . . . 16
4.1.2 Semantic Processing . . . 18
4.2 Neural Decoding Results . . . 20
4.2.1 Normalized Regional Analysis . . . 21
4.2.2 Cross-Modal Generalization . . . 21
4.3 Behavioral Performance . . . 22
4.4 Time-Resolved Analysis . . . 23
4.4.1 Temporal Dynamics . . . 23
4.5 Cross-Modal Integration . . . 23
4.5.1 Grammar Processing . . . 23
4.5.2 Semantic Processing . . . 25
4.6 Modality-Specific Processing . . . 25

5 Discussion **28**
5.1 Key Findings . . . 28
5.2 Theoretical Implications . . . 29
5.3 Limitations . . . 30
5.4 Future Directions . . . 30

6 Conclusion **31**
6.1 Summary . . . 31
6.2 Impact . . . 32
6.3 Final Remarks . . . 32

List of Figures

3.1 Electrode locations for the sentence task. The location of all electrodes is overlaid on the Desikan-Killiany Atlas, presented from multiple perspectives. Each white circle denotes the position of a single electrode. Panel (a) shows the left lateral view (n=730), (b) the left medial view (n=730), (c) the superior, whole-brain view (n=1563), (d) the inferior, whole-brain view (n=1563), (e) the right lateral view (n=833), and (f) the right medial view (n=833). 11

3.2 Sentence Task Design. In this paradigm, each trial presented participants with four-word sentences delivered either visually or auditorily, followed by a corresponding image. The sentences were constructed to be either semantically correct (GS), semantically implausible but grammatically correct (NS), or grammatically violated while preserving lexical content (NG). Participants were instructed to indicate via a button press whether the meaning conveyed by the sentence accurately described the subsequent image. 11

4.1 Neural signals distinguish between grammatically correct and incorrect sentences. An example electrode in the left pars opercularis (see location in panel g) demonstrates how normalized gamma-band power responses vary across semantically correct (GS: blue), semantically correct but grammatically intact and anomalous (NS: red), and syntactically violated (NG: black) sentences. Panels (a) and (b) show these distinctions for auditory (n=438 trials) and visual (n=432 trials) modalities, respectively. The vertical dashed line indicates word onset, and shaded areas represent standard error of the mean. Colored horizontal lines denote statistically significant differences in neural responses. Panels (c-f) display the Z-scored β coefficients for the GLM analysis, focusing on the area under the curve from 200 ms to 800 ms post-word onset. Panels (g) and (h) map all electrodes exhibiting audiovisual differences specifically linked to grammatical processing, with GvsNG as the only significant predictor. 17

4.2 Neural signals distinguish between semantically correct and incorrect sentences. An example electrode in the left lateral orbitofrontal cortex demonstrates differentiated gamma-band responses to semantically coherent versus incoherent sentences in both the auditory and visual domains. Panels (a) and (b) show the trial-averaged normalized power for each condition, illustrating modality-consistent patterns of semantic sensitivity. Panels (c-f) present the Z-scored β coefficients from the GLM, highlighting semantic factors as the sole significant predictors of neural responses. Panels (g-i) depict electrode locations exhibiting these semantic effects across modalities, confirming the presence of a stable neural substrate for semantic interpretation. 19

4.3 Neural signals distinguish between grammatically/semantically correct and incorrect sentences. By examining trial-averaged normalized gamma-band power responses and aggregating across multiple electrodes, distinct neural patterns emerged that were associated with either grammatical or semantic violations. These patterns were consistently observed across hemispheres, providing a stable signature of linguistic anomaly detection at a population level. 20
4.4 Decoding performance with an equal number of electrodes across all examined regions preserves the observed hemispheric asymmetry. This approach confirms that the lateralization of linguistic anomaly detection cannot be attributed solely to sampling differences and instead reflects genuine neurocognitive organization. 21
4.5 Auditory and visual crosstraining results highlight the modality-independent character of neural representations for grammatical and semantic processing. Classifiers trained on auditory data and tested on visual data (and vice versa) maintained above-chance decoding performance, indicating that abstract linguistic representations are shared across sensory modalities. 22
4.6 Temporal evolution of grammar processing within modalities. These timecourse results reveal how neural signals track syntactic information from word onset to later integrative phases, illuminating the dynamic interplay of syntactic parsing as sentences progress. 24
4.7 Cross-modal generalization of grammar processing. Neural signatures of grammatical structure learned from one modality were successfully applied to another, underscoring the abstract and time-resilient character of syntactic representations in the human brain. 24
4.8 Temporal evolution of semantic processing within modalities. As sentences unfold, semantic representations emerge and strengthen, reflecting the brain's ongoing efforts to assemble coherent meaning. 25
4.9 Cross-modal generalization of semantic processing. Similar to grammar, semantic representations exhibit temporal transferability across sensory modalities, reinforcing the view that neural meaning extraction operates at an abstract level, independent of the input channel. 26
4.10 Comparison of processing within individual modalities shows distinct temporal patterns for auditory and visual inputs. Although the underlying representations of grammar and meaning remain abstract, the temporal envelopes of processing differ slightly, reflecting the unique characteristics of auditory and visual language signals. 27

Abbreviations

GS Semantically & Grammatically Correct

NS Semantically Incorrect

NG Grammatically Incorrect

GLM Generalized Linear Model

FDR False Discovery Rate

PET positron emission tomography

EEG electroencephalogram

fmri functional magnetic resonance imaging

ERP Event related potential

Chapter 1

Introduction

1.1 Background

Understanding how the human brain processes language across different sensory modalities remains one of the fundamental challenges in neuroscience. While traditional language research has often focused on single modalities, natural language comprehension frequently involves integrating information across multiple sensory channels. The complexity of this integration process, coupled with the brain's remarkable ability to seamlessly process linguistic information from diverse sensory inputs, presents a fascinating area of investigation. This thesis delves deep into the neural mechanisms underlying both modality-specific and cross-modal aspects of language processing through direct intracranial recordings.

Recent advances in human electrophysiology have revolutionized our ability to access and analyze neural activity during language processing. These technological breakthroughs have provided unprecedented insights into the intricate patterns of brain activity associated with language comprehension. However, despite these advances, there remains a significant gap in our understanding of how the brain processes and integrates linguistic information across different sensory modalities. This limitation in our knowledge is particularly striking given the inherently multimodal nature of natural language processing, where humans regularly and effortlessly integrate auditory and visual linguistic inputs in everyday communication.

The challenge of understanding cross-modal language processing is further complicated by the distributed nature of language networks in the brain. Different aspects of language processing, from phonological analysis to semantic interpretation, involve multiple brain regions working in concert. These regions must rapidly coordinate their activities to integrate information from different sensory modalities, all while maintaining the speed and accuracy necessary for effective communication.

1.2 Significance

The ability to process language through multiple sensory channels represents a fundamental aspect of human communication, and understanding this capability has far-reaching implications across multiple domains. The significance of this research extends beyond basic neuroscience, touching upon clinical applications, technological development, and our fundamental understanding of human cognition.

From a theoretical perspective, this research provides crucial insights into the neural architecture supporting language processing. By examining how the brain processes linguistic information across different modalities, we can better understand the fundamental organizing principles of the language system. This includes investigating how sensory processing interacts with higher-level linguistic computation, and how the brain maintains abstract linguistic representations that can be accessed through different sensory channels.

The clinical implications of this research are particularly noteworthy. Understanding how the brain processes language across different modalities can inform the development of more effective therapeutic approaches for language disorders. This knowledge is especially valuable for conditions affecting specific sensory modalities or the integration of cross-modal information. Furthermore, insights gained from this research could lead to improved rehabilitation strategies for patients with language impairments resulting from stroke, traumatic brain injury, or neurodegenerative diseases.

In the context of modern technology, this research gains additional relevance with the increasing prevalence of multimodal communication technologies. As our interaction with technology becomes increasingly multimodal, understanding how the brain integrates information across different sensory channels becomes crucial for developing more intuitive and effective human-machine interfaces.

1.3 Research Objectives

This research aims to address several fundamental questions about the neural basis of language processing through a comprehensive set of interconnected objectives. Our primary goal is to characterize the neural signatures of semantic and grammatical processing across auditory and visual modalities. This involves detailed analysis of neural activity patterns during language processing tasks, with particular attention to how these patterns differ or remain consistent across modalities.

A critical aspect of our investigation focuses on identifying brain regions that support modality-specific versus modality-independent language processing. This distinction is crucial for understanding how the brain achieves both specialized processing for different input modalities and abstract linguistic representations that transcend sensory inputs. Through careful analysis of neural activity patterns, we aim to map out the network of brain regions involved in cross-modal language processing and understand their specific contributions to language comprehension.

The temporal dynamics of cross-modal language integration form another key focus of our research. By leveraging the high temporal resolution of our recording techniques, we can track the precise timing of neural events during language processing. This temporal information is crucial for understanding how different aspects of language processing unfold over time and how information from different sensory modalities is integrated.

Additionally, we seek to establish clear relationships between behavioral performance and neural decoding across modalities. This objective bridges the gap between neural activity patterns and observable behavior, providing insight into how neural processing differences manifest in language comprehension performance.

1.4 Approach

Our research employs a comprehensive and methodologically rigorous approach that combines multiple advanced techniques for studying neural activity during language processing. At the core of our methodology is the use of high-resolution intracranial recordings, utilizing an extensive network of 1,563 electrodes across 17 participants. This approach provides exceptional spatial and temporal resolution, allowing us to capture neural activity patterns with unprecedented precision.

The experimental design incorporates carefully controlled linguistic stimuli presented in both auditory and visual modalities. These stimuli are meticulously crafted to isolate specific aspects of language processing while maintaining ecological validity. We employ advanced machine learning techniques for neural decoding, allowing us to extract meaningful patterns from the complex neural data and identify the neural signatures associated with different aspects of language processing.

Our analysis pipeline includes sophisticated time-resolved analysis of neural signals during language processing, enabling us to track the evolution of neural activity patterns with millisecond precision. This temporal resolution is crucial for understanding the dynamic nature of language processing and how different brain regions coordinate their activity during language comprehension.

This multifaceted approach allows us to examine language processing from multiple perspectives, providing a comprehensive view of how the brain handles linguistic information across different sensory modalities. The combination of high spatial and temporal resolution, carefully controlled stimuli, and advanced analytical techniques positions us to make significant contributions to our understanding of neural language processing.

Chapter 2

Literature Review

2.1 Foundations of Language Processing

The study of how the human brain processes language has a long and multifaceted history, spanning the fields of linguistics, psychology, neuroscience, and computational modeling. Early theoretical frameworks primarily relied on the analysis of linguistic structure, focusing on syntax, semantics, and phonology as discrete and somewhat independent levels of representation. Pioneering figures such as Chomsky argued for innate linguistic principles, suggesting that universal grammar provides a biologically determined scaffold upon which all human languages are built. This perspective led researchers to consider language processing as an internally driven mechanism, one that operates according to abstract syntactic rules largely independent of sensory modalities. Over time, however, advances in psycholinguistics and neurolinguistics broadened these views. Classic lesion studies in patients with aphasia, for example, highlighted that language abilities are supported by distinct brain regions—such as Broca's and Wernicke's areas—which selectively impact syntactic and semantic functions. These early findings motivated the development of more neurobiologically grounded models, aiming to understand language not only as an abstract symbolic system but also as a function emergent from specific cortical architectures and neural circuits.

2.2 From Localization to Distributed Networks

With the advent of neuroimaging technologies, including PET and fMRI, the field began to move beyond simple localization of function. A growing body of evidence suggested that language processing does not reside in a handful of dedicated cortical sites, but rather emerges from dynamic interactions among distributed brain networks. Studies using functional connectivity and network analyses revealed that the classical language areas interact with auditory and visual regions, the motor system, and higher-order association cortices. This shift from localizationist to network-based models was accompanied by a conceptual reorientation: language processing came to be viewed as an integrated function, with syntactic and semantic computations unfolding through spatiotemporally distributed neural activity. Research on the time course of language comprehension, facilitated by techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), further supported this distributed view. Event-related potential (ERP) components, notably the N400 and P600 responses, provided temporal markers for semantic and syntactic processing respectively, linking distinct neural signatures to particular aspects of linguistic computation. This evidence underscored the complexity of language as a cognitive function and pointed toward the need for models that account for both time and space, capturing how information flows through interconnected networks as linguistic input is parsed and understood.

2.3 Modality Effects in Language Comprehension

While early research tended to focus on audition as the primary modality for language, it soon became clear that visual language input—through reading or sign languages—activates overlapping yet not fully identical neural circuits. Comparisons of spoken and written language comprehension revealed both common neural substrates and modality-specific processing stages. For example, while both forms of input ultimately engage regions associated with lexical-semantic access and syntactic integration, the initial stages of processing often differ. Auditory inputs require the extraction of phonological information over time, whereas reading involves the parallel recognition of orthographic patterns. Studies combining neuroimaging with psychophysical tasks demonstrated that comprehension is influenced by the temporal characteristics of the input: spoken language unfolds sequentially, encouraging incremental parsing, while written language can be scanned and revisited at will, potentially altering how semantic and syntactic structures are assembled. Research on signed languages added further complexity, showing that the human language system can flexibly co-opt visual and motoric circuits to process linguistic information conveyed through hand shapes, facial expressions, and body movements. These findings collectively suggest that while language relies on common representational levels—syntax, semantics, pragmatics—the neural routes to these representations can vary depending on sensory modality. The idea of modality-independent "amodal" or "supramodal" linguistic representations emerged from this literature, positing that despite differences in input format, the brain constructs meaning and structure via shared conceptual and grammatical frameworks.

2.4 Cross-Modal Integration and Abstract Representations

As researchers delved deeper, the concept of cross-modal integration came to the fore. Rather than treating auditory and visual language as separate streams that converge only at the level of meaning, recent studies show that the brain dynamically integrates cues from multiple sources. Behavioral experiments demonstrated that comprehension can be enhanced when corresponding speech and text are presented together, and neuroimaging studies identified multisensory convergence zones where auditory and visual information is merged. At the same time, computational modeling approaches, including connectionist networks and deep neural networks, provided algorithmic frameworks for understanding how different modalities might be reconciled. These models often rely on shared representational layers that do not encode modality-specific features but instead capture abstract linguistic regularities. Such abstraction allows the system to generalize beyond the specifics of one input channel, supporting robust comprehension even when sensory conditions vary. For example, cross-situational learning paradigms have shown that listeners and readers can leverage patterns from one modality to inform processing in another, underscoring the brain's remarkable adaptability and capacity for multimodal language learning.

2.5 Temporal Dynamics of Language Processing

Parallel to the investigation of cross-modal processes, there has been growing interest in the temporal structure of language comprehension. Technological innovations, particularly intracranial recordings and advanced signal processing techniques, have enabled researchers to probe the millisecond-scale dynamics of neural activity. Such time-resolved analyses revealed that syntax and semantics may be processed in partially overlapping windows but also maintain distinct temporal profiles. Some studies reported that semantic integration occurs relatively early, closely tied to the unfolding of the speech or text stream, whereas syntactic computation might proceed more gradually, building hierarchical structures over longer intervals. Additionally, the temporal dimension is crucial for understanding how contextual information accumulates. As words are processed over time, the brain updates its interpretation of sentence meaning and anticipates upcoming linguistic material, a phenomenon reflected in predictive coding frameworks. This temporal lens thus provides a richer picture of how modality-specific and modality-independent representations interact dynamically, aligning incrementally parsed inputs with higher-level grammatical and conceptual knowledge.

2.6 Bridging Neuroscience, Linguistics, and Technology

Modern research on language comprehension stands at the intersection of multiple disciplines. Insights from neuroscience have begun to inform linguistic theories, offering empirical grounding for abstract concepts such as syntactic structures and semantic features. Meanwhile, computational linguistics and artificial intelligence research leverage findings from neurocognitive studies to design more human-like language models. Functional neuroimaging data can validate or challenge algorithmic assumptions in natural language processing, and machine learning techniques can help decode complex neural patterns that correspond to different linguistic conditions. Furthermore, the growing field of neurorehabilitation has capitalized on these interdisciplinary connections. By understanding the neural basis of language processing, clinicians and engineers can develop brain-computer interfaces and rehabilitative protocols that aid patients with language disorders. Such applications highlight the practical importance of identifying modality-independent neural signatures and understanding how different sensory inputs are integrated within the language system.

2.7 Gaps and Future Directions

Despite considerable progress, several key questions remain. One pressing issue is to delineate more precisely the boundaries between modality-specific and modality-independent processing stages. While current evidence supports the existence of abstract representations, it remains unclear how these representations emerge developmentally, how they adapt to bilingual and multilingual contexts, and how they are influenced by socio-cultural factors. Additionally, the neural mechanisms that enable effortless cross-modal transitions in naturalistic settings—where language is embedded in rich auditory, visual, and situational contexts—are not fully understood. Future research will likely integrate more diverse methodologies, from high-density electrophysiological recordings and intracranial studies to immersive virtual reality paradigms, aiming to capture language comprehension as it naturally occurs. Such interdisciplinary efforts promise not only to resolve existing debates in the literature but also to open new avenues for understanding how the human brain negotiates the delicate balance between specialization and abstraction in the service of language.

‘‘‘

Chapter 3

Methods

3.1 Task Design

3.1.1 Stimulus Categories

In this study, participants were presented with three distinct types of sentences, each designed to probe different aspects of language processing. The first category consisted of semantically and grammatically correct sentences (GS), such as "the girls ate cakes," which served as a control condition. These well-structured and meaningful sentences established a baseline for normal comprehension processes. The second category introduced a violation of semantic expectations while preserving grammatical structure (NS). An example sentence from this category would be "the cakes ate girls," which appears syntactically sound but is semantically implausible, thereby isolating the role of meaning in language comprehension. The final category introduced a grammatical violation while employing the same lexical items used in the previous categories (NG). A representative example is "the ate girls cakes," which disrupts the syntactic order of the sentence, allowing us to examine the impact of morphosyntactic errors on comprehension independently of lexical or semantic properties. By interspersing these three sentence types, each trial offered a nuanced platform to investigate how human language processing mechanisms adapt to and resolve different forms of linguistic irregularities.

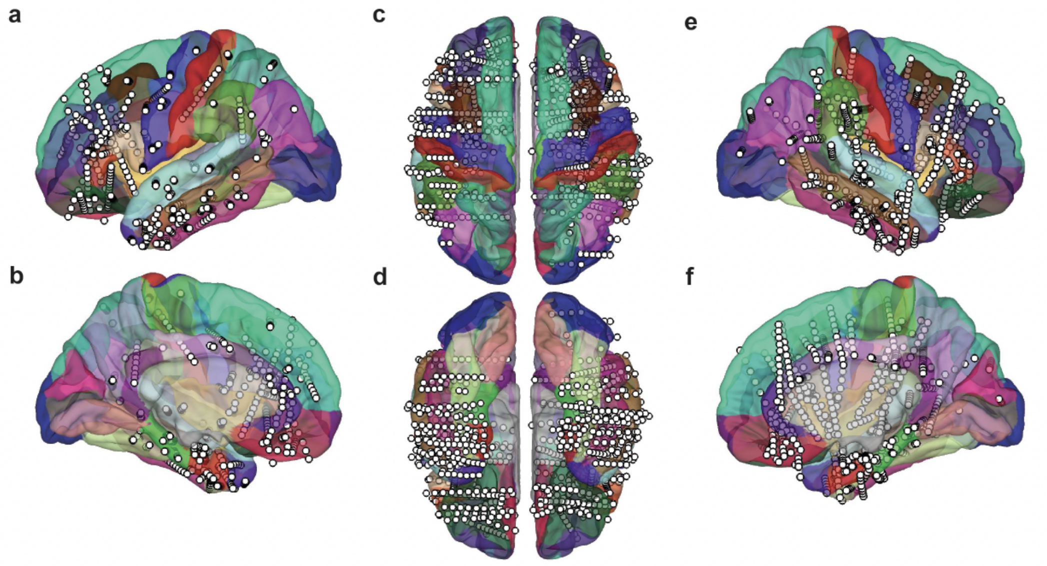

FIGURE 3.1: Electrode locations for the sentence task. The location of all electrodes is overlaid on the Desikan-Killiany Atlas, presented from multiple perspectives. Each white circle denotes the position of a single electrode. Panel (a) shows the left lateral view (n=730), (b) the left medial view (n=730), (c) the superior, whole-brain view (n=1563), (d) the inferior, whole-brain view (n=1563), (e) the right lateral view (n=833), and (f) the right medial view (n=833).

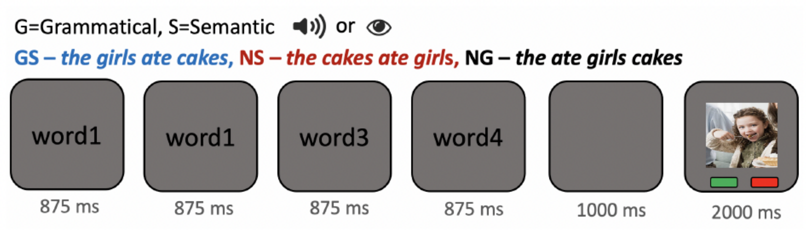

FIGURE 3.2: Sentence Task Design. In this paradigm, each trial presented participants with four-word sentences delivered either visually or auditorily, followed by a corresponding image. The sentences were constructed to be either semantically correct (GS), semantically implausible but grammatically correct (NS), or grammatically violated while preserving lexical content (NG). Participants were instructed to indicate via a button press whether the meaning conveyed by the sentence accurately described the subsequent image.

3.1.2 Presentation Parameters

Each trial began with an initial fixation cross that remained visible for 600 ms to orient the participant's attention. After this brief fixation period, the four words of the sentence were presented one after the other, each displayed or played for a duration of 875 ms. Between the final word and the subsequent image presentation, there was an inter-stimulus interval of 1 second during which a neutral gray screen appeared. This timing structure ensured that participants had a momentary pause to process the sentence before verifying its meaning against the ensuing image. The image presentation phase required participants to respond by pressing a button, indicating whether the sentence they had just perceived accurately matched the depicted scenario. These timing parameters were carefully chosen to balance the need for adequate comprehension and processing of each word against the practical constraints of maintaining participant engagement and minimizing fatigue.

3.2 Data Analysis

3.2.1 Signal Processing

The neural signals recorded during the sentence comprehension task underwent a rigorous series of processing steps designed to isolate the physiologically meaningful high-frequency neural responses associated with language comprehension. High gamma activity, defined as power fluctuations in the 30-150 Hz range, was extracted from the raw signals using multi-taper spectral analysis. This approach provided a robust and stable estimation of spectral power while mitigating the influence of noise. To preserve temporal resolution and capture the dynamic evolution of neural responses over the course of sentence processing, a sliding window procedure was employed. Windows of 200 ms, incremented in 50 ms steps, allowed for a time-resolved characterization of neural activity. Throughout this process, careful artifact rejection and stringent quality control checks were implemented. These measures ensured that the final data set was free from contaminations—such as electrical interference or muscle artifacts—allowing for a more accurate interpretation of the underlying neural processes supporting language comprehension.

3.2.2 Classification Analysis

A subsequent analytical phase focused on the classification of sentence types using machine learning techniques. Linear Support Vector Machine (SVM) classifiers were employed to distinguish among the three categories of sentences and to probe whether the neural activity patterns elicited by these stimuli could be reliably decoded. To robustly assess the generalizability and stability of the classification models, a five-fold cross-validation procedure was applied. This involved systematically partitioning the data into training and testing sets, ensuring that the decoding performance was not biased by any particular subset of data. Analyses were conducted both within the same modality to establish modality-specific decoding accuracy and across modalities to evaluate the potential for cross-modal generalization. Moreover, these classification procedures were performed in a time-resolved manner, examining how the discriminability of the different sentence types evolved throughout the sentence presentation. By elucidating not only the peak decoding performance but also its temporal trajectory, this approach yielded insights into the time course of information processing underlying sentence comprehension.

Let N_{correct} be the number of correctly classified trials and N_{total} the total number of trials. Then the decoding accuracy A is given by:

$$A = \frac{N_{\text{correct}}}{N_{\text{total}}}.$$

Classification Code Example

The following MATLAB code illustrates the decoding procedure described in this thesis:

```
function [accuracy, std_dev, shuffled_accuracies] = perform_decoding(X,
   Y)
    % Ensure inputs are proper numeric arrays
    X = double(X);
    Y = double(Y);

    % Remove any trials with NaN or Inf values
    valid_trials = all(isfinite(X), 2);
    X = X(valid_trials, :);
    Y = Y(valid_trials);

    % Determine number of folds based on data size
    n_samples = size(X, 1);
    k = min(5, n_samples);  % Use 5-fold CV or less if fewer samples

    % Perform decoding using SVM
    cv = cvpartition(Y, 'KFold', k);
    accuracies = zeros(cv.NumTestSets, 1);
    shuffled_accuracies = zeros(cv.NumTestSets, 100);

    for i = 1:cv.NumTestSets
        train_idx = cv.training(i);
```

```
        test_idx = cv.test(i);

        % Train and test SVM
        try
            mdl = fitcsvm(X(train_idx, :), Y(train_idx), ...
                'KernelFunction', 'linear', ...
                'Standardize', true);
            accuracies(i) = sum(predict(mdl, X(test_idx, :)) == Y(
test_idx)) / sum(test_idx);

            % Shuffled labels
            for j = 1:100
                Y_shuffled = Y(randperm(length(Y)));
                mdl_shuffled = fitcsvm(X(train_idx, :), Y_shuffled(
train_idx), ...
                    'KernelFunction', 'linear', ...
                    'Standardize', true);
                shuffled_accuracies(i, j) = sum(predict(mdl_shuffled, X(
test_idx, :)) == ...
                    Y_shuffled(test_idx)) / sum(test_idx);
            end
        catch ME
            fprintf('Error in fold %d: %s\n', i, ME.message);
            accuracies(i) = NaN;
            shuffled_accuracies(i, :) = NaN;
        end
    end

    % Remove any NaN results
    accuracies = accuracies(~isnan(accuracies));
    shuffled_accuracies = shuffled_accuracies(~any(isnan(
shuffled_accuracies), 2), :);

    if isempty(accuracies)
        error('Decoding failed for all folds');
    end

    accuracy = mean(accuracies);
```

```
    std_dev = std(accuracies);
    shuffled_accuracies = mean(shuffled_accuracies, 1);
end
```

3.3 Statistical Methods

All statistical assessments were carried out with stringent controls to maintain the integrity of the inferences drawn from the data. Permutation tests were utilized, running 100 random shuffles of condition labels to generate empirical null distributions. By comparing observed classification accuracies, spectral power differences, or other relevant metrics to these null distributions, it was possible to determine whether the results exceeded what could be expected by chance alone. To correct for multiple comparisons—an issue that arises when conducting multiple tests across time points, frequency bands, or anatomical regions—false discovery rate (FDR) correction procedures were employed, reducing the likelihood of inflating Type I error rates. In addition, bootstrap confidence intervals were calculated to provide robust estimates of variability and enhance the reliability of reported effects. Where meaningful differences were detected, effect sizes were computed to contextualize the magnitude of these differences, ensuring that statistically significant findings were also interpretable in terms of their practical relevance.

Chapter 4

Results

4.1 Individual Electrode Analysis

4.1.1 Grammar Processing

The analyses of individual electrodes revealed distinct neural responses that differentiated grammatically correct from incorrect sentences. In these recordings, neural signals were measured as normalized gamma-band power, which showed clear modulation when participants processed violations in the grammatical structure of sentences. As illustrated in Figure 4.1, the responses of an example electrode located in the left pars opercularis demonstrated discernible patterns of activity in response to semantically correct sentences compared to those that were semantically intact yet grammatically anomalous, and those that were semantically implausible but grammatically intact. Notably, this electrode's activity patterns were evident across both auditory and visual modalities, as indicated in panels (a) and (b) of the figure. The timing of these responses was aligned to the onset of each word and the waiting period preceding image onset, providing a detailed temporal profile of how the brain detects and processes violations in syntactic structure. Statistical analyses, employing Generalized Linear Models (GLMs) with Z-scored β coefficients, further supported these observations. Only the grammatical violation predictor emerged as consistently significant in a subset of electrodes, including those highlighted in panels (c-f), indicating that syntactic processing could be isolated as a distinct neural signal separate from semantic interpretation. The spatial distribution of these electrodes, depicted in panels (g) and (h), further confirmed that certain cortical regions, including the left pars opercularis, are sensitive to syntactic aberrations and can track these violations across different sensory input modalities.

If we let A_{obs} be the observed accuracy and $A_{\text{null}}^{(p)}$ be the accuracy of the p-th permutation, the p-value is given by:

$$p\text{-value} = \frac{\sum_{p=1}^{P} \mathbb{I}(A_{\text{null}}^{(p)} \geq A_{\text{obs}})}{P}$$

where $\mathbb{I}$ is an indicator function and P is the total number of permutations.

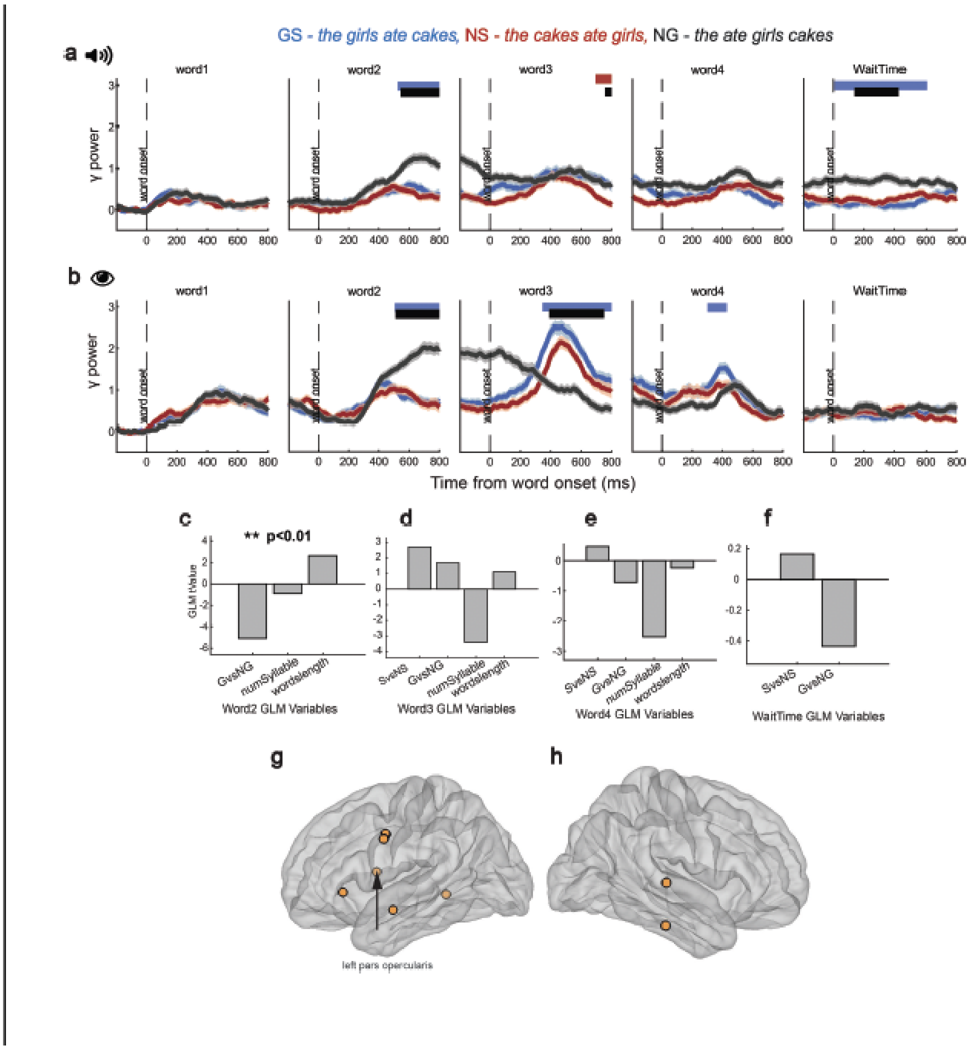

FIGURE 4.1: Neural signals distinguish between grammatically correct and incorrect sentences. An example electrode in the left pars opercularis (see location in panel g) demonstrates how normalized gamma-band power responses vary across semantically correct (GS: blue), semantically correct but grammatically intact and anomalous (NS: red), and syntactically violated (NG: black) sentences. Panels (a) and (b) show these distinctions for auditory (n=438 trials) and visual (n=432 trials) modalities, respectively. The vertical dashed line indicates word onset, and shaded areas represent standard error of the mean. Colored horizontal lines denote statistically significant differences in neural responses. Panels (c-f) display the Z-scored β coefficients for the GLM analysis, focusing on the area under the curve from 200 ms to 800 ms post-word onset. Panels (g) and (h) map all electrodes exhibiting audiovisual differences specifically linked to grammatical processing, with GvsNG as the only significant predictor.

4.1.2 Semantic Processing

Parallel analyses of individual electrodes also revealed neural correlates that distinguished semantically correct sentences from those containing semantic violations, regardless of whether the sentences were grammatically sound. As shown in Figure 4.2, an example electrode located in the left lateral orbitofrontal cortex exhibited distinct gamma-band power responses when encountering sentences that conveyed meaning consistent with expected semantic structures, compared to sentences that were semantically implausible. These neural distinctions were evident during both auditory and visual presentations, thus underscoring that semantic comprehension—and its disruption—can be mapped onto spatially localized and modality-invariant brain responses. Statistical models comparing the GLM predictors identified semantic processing as the only significant factor in a subset of electrodes. This provides compelling evidence that semantic anomalies evoke a unique neural signature, enabling the differentiation of lexical-meaningful combinations from those that challenge comprehension. By examining electrode locations (g-i), it became apparent that certain cortical areas are consistently engaged in tracking semantic coherence across modalities, highlighting a system that is robust and flexible enough to support meaningful interpretation of linguistic input, irrespective of the sensory channel through which it is delivered.

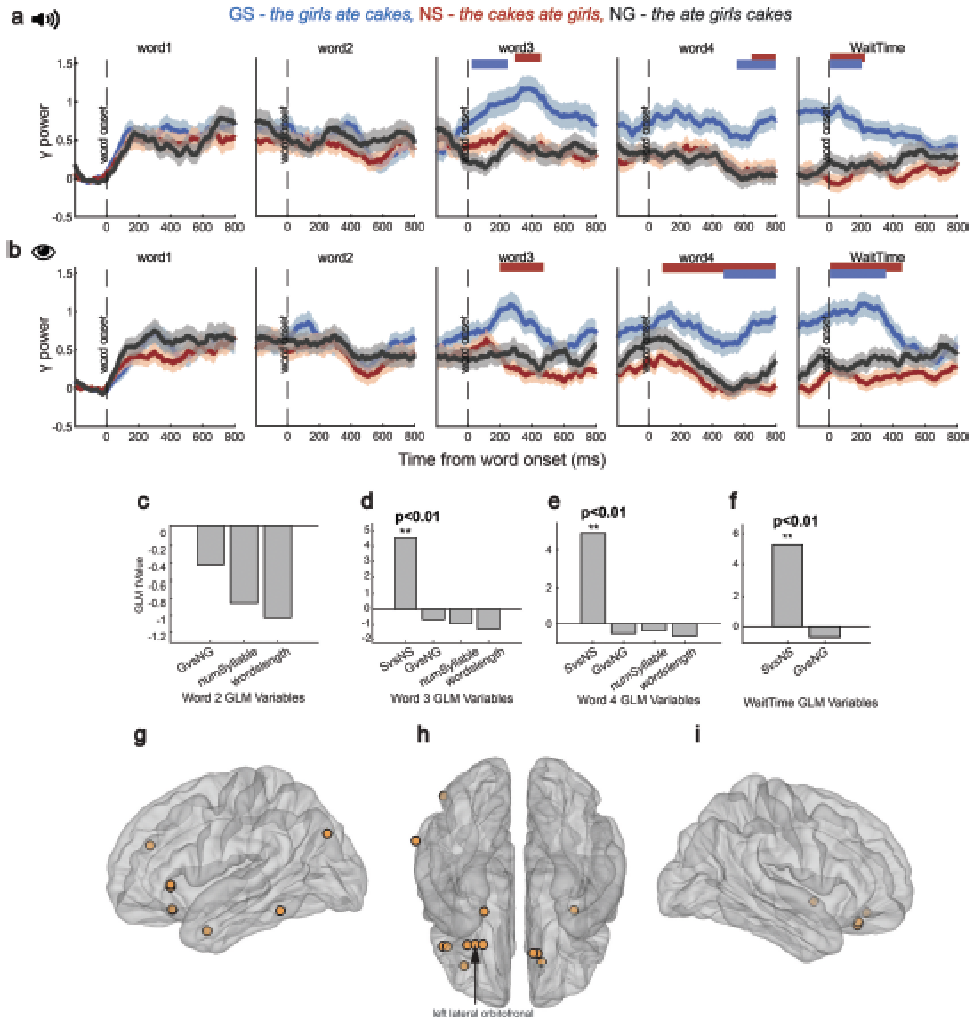

FIGURE 4.2: Neural signals distinguish between semantically correct and incorrect sentences. An example electrode in the left lateral orbitofrontal cortex demonstrates differentiated gamma-band responses to semantically coherent versus incoherent sentences in both the auditory and visual domains. Panels (a) and (b) show the trial-averaged normalized power for each condition, illustrating modality-consistent patterns of semantic sensitivity. Panels (c-f) present the Z-scored β coefficients from the GLM, highlighting semantic factors as the sole significant predictors of neural responses. Panels (g-i) depict electrode locations exhibiting these semantic effects across modalities, confirming the presence of a stable neural substrate for semantic interpretation.

4.2 Neural Decoding Results

Building on the electrode-level analyses, ensemble decoding approaches were applied to assess whether patterns of neural activity could reliably distinguish between different sentence conditions at a population level. As illustrated in Figure 4.3, neural signals displayed distinct patterns of gamma-band power for grammatically and semantically correct versus incorrect sentences. These differences emerged across both hemispheres, reflecting widespread involvement of cortical networks in handling linguistic anomalies. By pooling activity from multiple electrodes, it became possible to identify robust, condition-specific signatures that allowed a classifier to decode the presence of grammatical or semantic violations with above-chance accuracy. Thus, the decoding results support the notion that the brain maintains separable neural representations for syntactic and semantic information, which can be quantified and tracked in real-time as linguistic input unfolds.

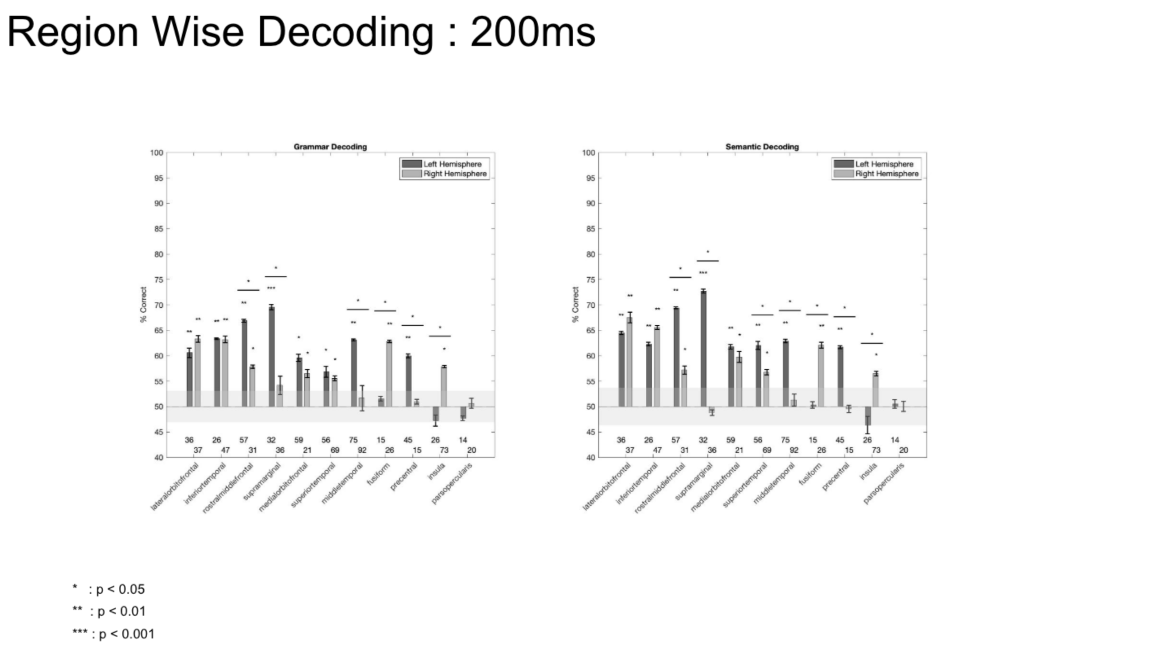

FIGURE 4.3: Neural signals distinguish between grammatically/semantically correct and incorrect sentences. By examining trial-averaged normalized gamma-band power responses and aggregating across multiple electrodes, distinct neural patterns emerged that were associated with either grammatical or semantic violations. These patterns were consistently observed across hemispheres, providing a stable signature of linguistic anomaly detection at a population level.

4.2.1 Normalized Regional Analysis

To further evaluate the stability and generality of decoding performance, an analysis was conducted in which the number of electrodes contributing to classification was normalized across distinct cortical regions. As shown in Figure 4.4, when equalizing electrode counts across different areas, the hemispheric asymmetry that characterized neural decoding of linguistic anomalies remained intact. This finding highlights that the detected differences are not merely artifacts of uneven sampling density but rather reflect underlying neurobiological asymmetries in language processing networks. It suggests that the left hemisphere, traditionally implicated in language function, may still hold a relative advantage in detecting and resolving linguistic violations, even under conditions of balanced sampling.

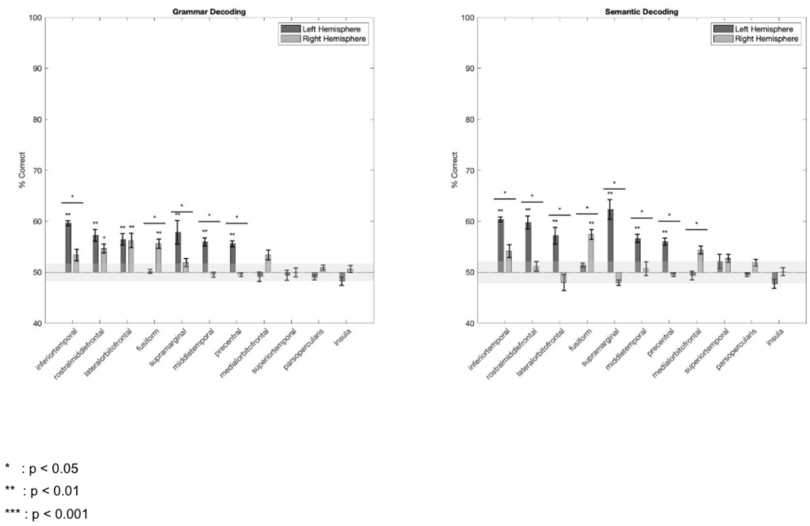

FIGURE 4.4: Decoding performance with an equal number of electrodes across all examined regions preserves the observed hemispheric asymmetry. This approach confirms that the lateralization of linguistic anomaly detection cannot be attributed solely to sampling differences and instead reflects genuine neurocognitive organization.

4.2.2 Cross-Modal Generalization

Additional decoding analyses assessed whether neural representations underlying grammatical and semantic processes could generalize across sensory modalities. By training classifiers on data from one modality (e.g., auditory) and testing them on another (e.g., visual), the results, depicted in Figure 4.5, demonstrate a remarkable degree of cross-modal generalization. The brain's encoding of grammatical and semantic information appears to be sufficiently abstract and invariant that it transcends the particular perceptual channel through which language is delivered. This cross-modal resilience underlines the high-level nature of these linguistic representations and suggests a common neural code for language comprehension that is not strictly tied to modality-specific input features.

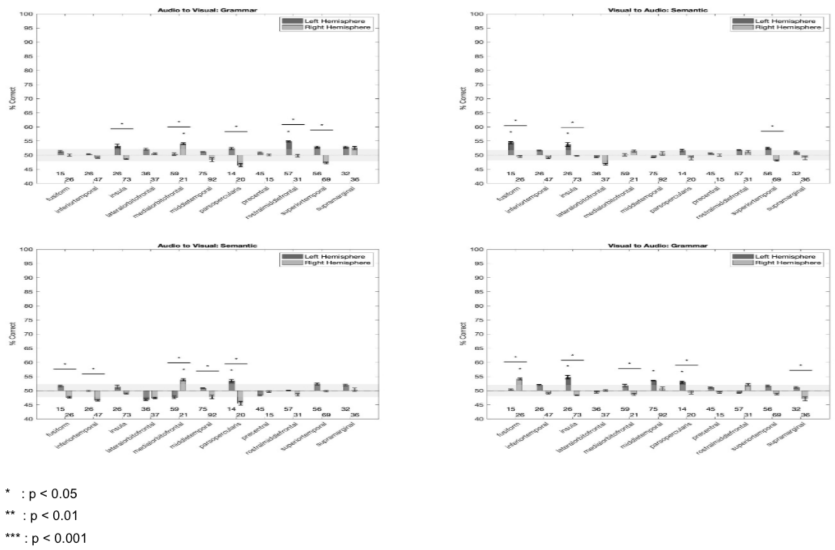

FIGURE 4.5: Auditory and visual crosstraining results highlight the modality-independent character of neural representations for grammatical and semantic processing. Classifiers trained on auditory data and tested on visual data (and vice versa) maintained above-chance decoding performance, indicating that abstract linguistic representations are shared across sensory modalities.

4.3 Behavioral Performance

Examination of behavioral measures revealed that participants maintained robust comprehension accuracy across the various experimental conditions. On average, accuracy approached 86%, with only moderate variability across subjects. Although comparable performance levels were observed for auditory and visual modalities, participants tended to be slightly more accurate in detecting grammatical violations than semantic ones. This suggests that subtle morphosyntactic disruptions are salient cues for listeners and viewers, allowing them to more readily reject mismatched images when sentences fail to adhere to expected syntactic structures. Overall, these behavioral results align with the neural findings, implying that the observed neural patterns associated with grammatical and semantic anomalies are functionally relevant and support participants' comprehension at a behavioral level.

4.4 Time-Resolved Analysis

4.4.1 Temporal Dynamics

Analyses incorporating a time-resolved perspective revealed that different brain regions exhibited distinct temporal profiles of linguistic processing. Frontal areas, for instance, tended to show peak decoding performance at approximately 300 ms after word onset, indicating a relatively delayed but sustained involvement in higher-level integrative functions of language. By contrast, temporal regions displayed earlier and more transient peaks, around 100 ms post word onset, suggesting a swifter and possibly more feedforward-driven response to incoming linguistic information. These dynamic patterns were not confined to isolated time windows; instead, some sustained effects persisted throughout the entire sentence processing interval, reflecting the continuous re-analysis and integration of linguistic input as sentences unfold. Such fine-grained temporal assessments provide a rich view of how different cortical territories coordinate in time to decode the syntactic and semantic structure of language.

4.5 Cross-Modal Integration

4.5.1 Grammar Processing

Temporal analyses of grammatical processing within and across modalities, exemplified in Figures 4.6 and 4.7, further corroborated the notion that these neural representations are not only distributed across cortical space but also exhibit a complex temporal organization. Within modalities, grammar-related signatures emerged relatively early and persisted over the subsequent processing windows, indicating an incremental accumulation of syntactic information as the sentence unfolds. When cross-modal training was introduced, the results showed that these temporally structured representations of grammatical rules could be transferred from one modality to another. Such findings underscore the temporal robustness of grammatical representations and highlight their modality-independent encoding, suggesting a common temporal framework that the brain exploits to integrate and evaluate structural cues in language.

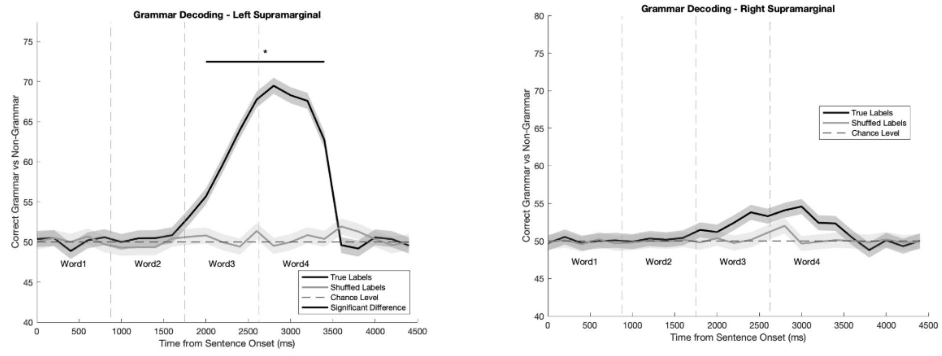

Figure 4.6: Temporal evolution of grammar processing within modalities. These timecourse results reveal how neural signals track syntactic information from word onset to later integrative phases, illuminating the dynamic interplay of syntactic parsing as sentences progress.

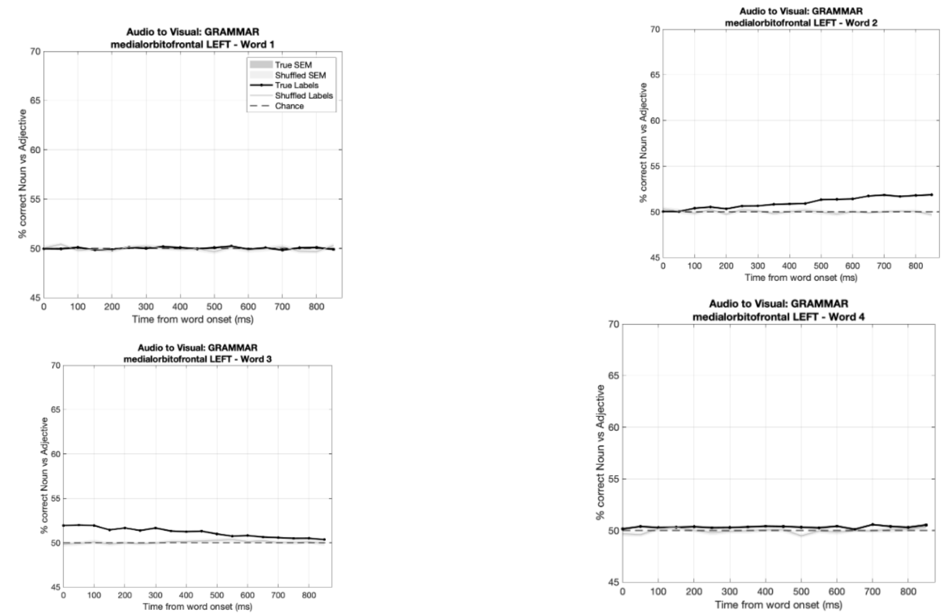

Figure 4.7: Cross-modal generalization of grammar processing. Neural signatures of grammatical structure learned from one modality were successfully applied to another, underscoring the abstract and time-resilient character of syntactic representations in the human brain.

4.5.2 Semantic Processing

Parallel temporal analyses of semantic processing (Figures 4.8 and 4.9) offered similar insights into the evolving and cross-modal nature of meaning extraction. Within each modality, semantic contrasts emerged and developed over time, indicating that the brain incrementally refines its interpretation of word meanings in context. Cross-modal crosstraining results revealed that the temporal patterns of semantic comprehension, once established in one modality, could be redeployed in another. This temporal transferability aligns closely with the evidence for modality-invariant semantic representations and suggests that the neural code for meaning is temporally structured and highly flexible.

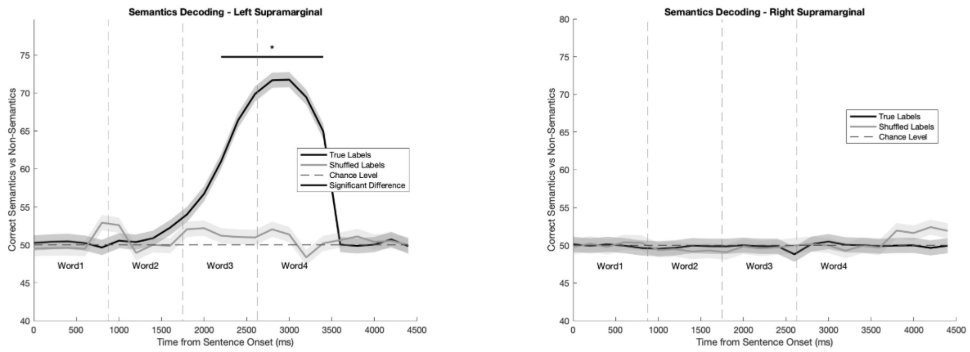

FIGURE 4.8: Temporal evolution of semantic processing within modalities. As sentences unfold, semantic representations emerge and strengthen, reflecting the brain's ongoing efforts to assemble coherent meaning.

4.6 Modality-Specific Processing

Finally, the examination of modality-specific effects (Figure 4.10) confirmed that while the underlying grammatical and semantic representations are abstract and cross-modally generalizable, some differences in temporal dynamics between auditory and visual inputs do arise. The processing timelines for spoken sentences may be slightly earlier or more sustained in certain cortical regions compared to written input, reflecting the intrinsic differences in how auditory and visual language signals are sampled and transformed. These modality-specific

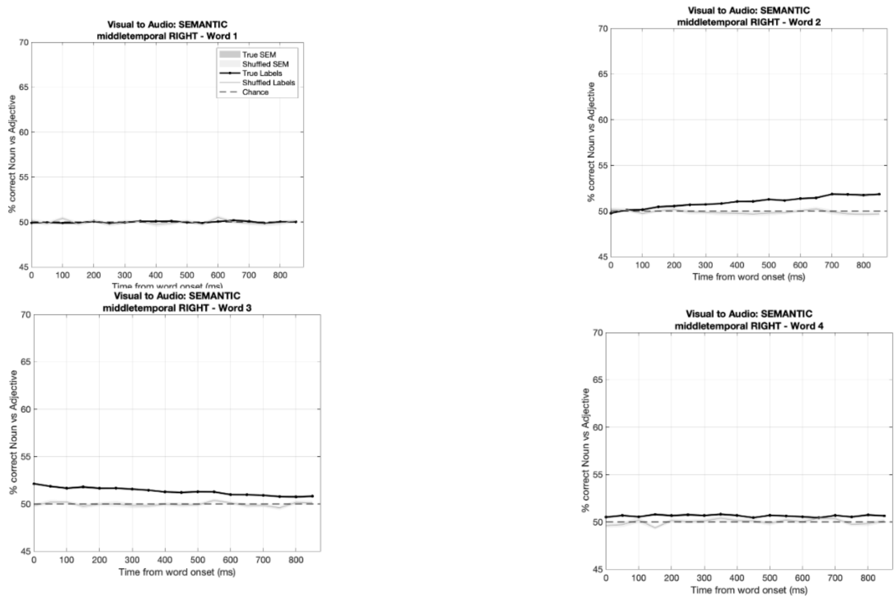

Figure 4.9: Cross-modal generalization of semantic processing. Similar to grammar, semantic representations exhibit temporal transferability across sensory modalities, reinforcing the view that neural meaning extraction operates at an abstract level, independent of the input channel.

temporal nuances do not undermine the broader claim of modality-invariant conceptual and structural representations; rather, they highlight that the human language system is equipped to handle the unique challenges posed by each sensory channel. The brain thus exploits a common representational substrate for grammar and semantics, while still tailoring aspects of the processing chronology to the sensory format of the input.

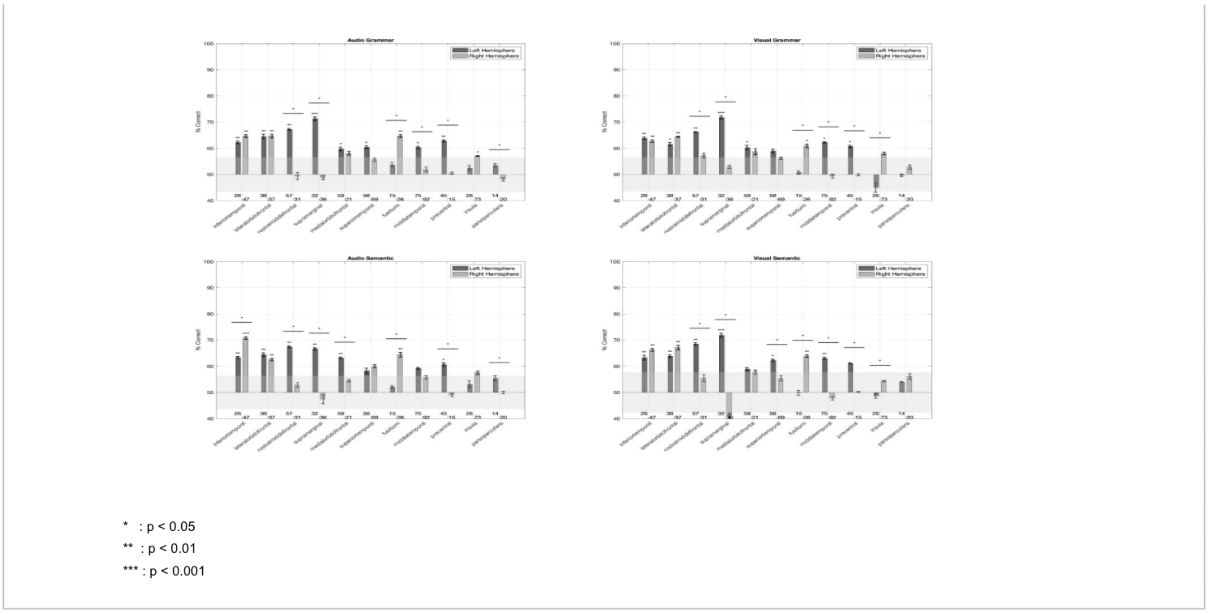

FIGURE 4.10: Comparison of processing within individual modalities shows distinct temporal patterns for auditory and visual inputs. Although the underlying representations of grammar and meaning remain abstract, the temporal envelopes of processing differ slightly, reflecting the unique characteristics of auditory and visual language signals.

Chapter 5

Discussion

5.1 Key Findings

In this study, we investigated how the human brain processes grammatical and semantic information in sentences presented through both auditory and visual modalities. The findings reveal that language comprehension engages an intricate interplay between modality-specific and modality-independent neural mechanisms. Although the underlying representations of linguistic structure and meaning can be abstracted away from any single sensory channel, the temporal profiles of neural responses were not identical across modalities. Instead, a combination of shared linguistic representations and modality-tuned processing stages worked in tandem to achieve robust comprehension. This duality is evidenced by successful cross-modal decoding, which indicates that certain neural signatures of language comprehension are preserved across sensory inputs, while differences in timing and processing sequences suggest that each modality makes use of partially distinct computational pathways. Moreover, hemispheric specialization persisted even when the sensory modality changed, reinforcing the view that certain cortical networks—particularly those in the left hemisphere—are preferentially engaged in aspects of linguistic analysis, regardless of whether the input is spoken or written.

These results highlight that language comprehension is neither purely modality-specific nor entirely abstract. Instead, our findings point toward a nuanced view in which abstract linguistic representations are supported by modality-dependent processing streams. The temporal dynamics observed here confirm that comprehension unfolds in stages, with earlier cortical responses likely reflecting lower-level perceptual analyses and later stages reflecting the integration of semantic and syntactic information into coherent meaning. By capturing this progression across time, we can understand how different brain regions contribute in sequence, each playing a unique role in assembling the final interpretation of linguistic input. The evidence also suggests that, as the sentence unfolds, multiple integration windows open, allowing information from different modalities to converge, thereby granting the language system both flexibility and robustness.

5.2 Theoretical Implications

The implications of these findings extend to multiple domains of linguistic theory and cognitive neuroscience. From a theoretical perspective, the evidence that abstract, modality-independent representations are interwoven with modality-dependent processes challenges overly simplified models of language comprehension that assume a single, modality-agnostic pathway. Instead, the data advocate for more integrative frameworks that incorporate both shared and specialized mechanisms. Such models must account for the ability of the brain to decode linguistic structure and meaning from diverse forms of input while also accommodating the reality that different modalities evoke distinct temporal and neural signatures.

In the broader cognitive framework, these results contribute to our understanding of sensory integration in language processing. The ability of the brain to handle cross-modal transfer—training on one modality and successfully decoding in another—demonstrates that the representations underlying grammar and semantics reside at a sufficiently high level of abstraction to be accessible from multiple sensory inputs. Furthermore, the observed hemispheric specialization, which endures irrespective of modality, aligns with well-established notions that particular cortical regions retain stable functional roles in linguistic analysis.

Such insights carry significance not only for fundamental linguistic theory but also for our broader understanding of neural representation. The interplay between modality-dependent processing streams and modality-independent representations reveals how the brain balances specialization with flexibility, ensuring that core linguistic constructs remain accessible despite changes in input format. By mapping these neural principles, we can refine theories of how language and cognition are organized in the human brain, paving the way for models that better reflect the complexity of real-world language use.

5.3 Limitations

It is important to acknowledge certain limitations of the present study. First, the electrode coverage was constrained by clinical considerations, and not all cortical regions implicated in language processing could be equally sampled. This selective coverage might have left some relevant neural territories unexplored. Additionally, the tasks and stimuli were designed for experimental rigor, potentially differing from the complexity and variability of naturalistic language. Such controlled conditions, while useful for isolating linguistic variables, may influence the generalizability of these findings to everyday language comprehension. Statistical power also warrants careful consideration, as certain subgroup analyses may have been underpowered to detect subtle effects. Finally, while the results reveal robust cross-modal integration, future work will be needed to clarify whether these patterns hold consistently across a wider range of linguistic contexts and with more diverse participant populations.

5.4 Future Directions

Future research can build on these findings by extending paradigms to encompass more complex linguistic structures and more naturalistic communication scenarios, including dialogues and narratives. Such investigations could examine how these neural mechanisms scale up when confronted with longer and more syntactically rich sentences or when integrating extralinguistic cues such as gestures, prosody, and visual context. Another promising direction involves exploring individual differences in language processing. Differences in linguistic proficiency, bilingualism, or language disorders might illuminate how these neural dynamics adapt or deteriorate under varying cognitive conditions.

In addition, future investigations could strive to translate these insights into clinical applications. Understanding how the brain integrates grammar and meaning across modalities could aid in developing targeted rehabilitation strategies for patients with language impairments, such as aphasia. Interventions might leverage the brain's capacity for modality-independent processing to support recovery. Finally, integrating these electrophysiological approaches with other neuroimaging modalities, such as functional MRI or MEG, would help paint a more comprehensive picture of the spatiotemporal dynamics of language processing. By blending high temporal-resolution data with detailed spatial maps, future studies can create richer models that inform both basic science and clinical practice.

Chapter 6

Conclusion

6.1 Summary

In this thesis, we have presented a comprehensive exploration into how the human brain processes language across different sensory modalities. By examining neural signals recorded during the comprehension of sentences that varied along semantic and grammatical dimensions, we identified distinct neural signatures that underlie these fundamental linguistic operations. In both auditory and visual domains, we observed clear and reliable neural markers that differentiated meaningful, syntactically well-formed sentences from those containing either semantic or grammatical violations. Importantly, our findings showed that certain cortical regions were consistently involved in processing linguistic information regardless of the modality of presentation, suggesting that the human language system maintains modality-independent representations. At the same time, subtle differences in the temporal dynamics of neural responses emerged between auditory and visual conditions, indicating that while the underlying linguistic computations can be abstracted from sensory form, the brain still tailors certain aspects of processing to the specific input channel. Finally, we linked these neural patterns directly to behavior, demonstrating that the extent to which neural activity reflects grammatical and semantic structure correlates with participants' ability to accurately discern meaning and syntactic integrity. Collectively, these results clarify the interplay between abstract linguistic principles and modality-specific processing demands, and they highlight the richness and complexity of the temporal trajectories that support language integration in the human brain.

6.2 Impact

The implications of this research extend well beyond the immediate findings and have the potential to inform various domains and applications. From a basic neuroscience perspective, the demonstration that language comprehension arises from both shared and modality-tuned neural mechanisms challenges the idea of a singular, unified language pathway. Instead, the results support more nuanced models that envision linguistic knowledge as a flexible resource, accessible from multiple sensory channels, yet fine-tuned to their distinct temporal and perceptual characteristics. This refined understanding can also guide clinical applications, particularly in the assessment and rehabilitation of individuals with language impairments. By uncovering neural markers that reliably track grammatical and semantic processing across modalities, clinicians may be able to better target specific deficits, leveraging modality-independent networks to compensate for impaired sensory channels or damaged cortical regions. Additionally, these insights can inform computational models of language processing and artificial intelligence systems, helping engineers and researchers design algorithms that integrate multiple input forms more naturally and process linguistic information in a manner that resembles human cognition. In language technology development, this perspective could improve speech recognition, reading aids, and multimodal interfaces, enabling more robust and contextually aware systems capable of adapting to user preferences and environmental constraints.

6.3 Final Remarks

The work presented here represents a significant advance in our understanding of the neural architecture underlying cross-modal language processing. By bridging the gap between earlier studies focused primarily on single modalities and more recent, integrative frameworks that consider multimodal inputs, we have generated evidence that the language system is both versatile and highly organized. The identification of shared linguistic representations that transcend sensory boundaries, combined with evidence for subtle, modality-specific temporal dynamics, paints a picture of a language network that is at once stable and adaptable. These findings not only refine theoretical models of language comprehension but also offer practical insights for clinical interventions and the development of language-related technologies. In sum, the research described in this thesis underscores the complexity and resilience of the human language system and sets the stage for future investigations that further illuminate the intricate interplay between abstraction, modality, and timing in human linguistic cognition.

References

[1] Noam Chomsky. *The minimalist program.* MIT Press, 1995.

[2] Kiefer J. Forseth et al. "A lexical semantic hub for heteromodal naming in middle fusiform gyrus". In: *Brain* 141 (2018), pp. 2112–2126. DOI: 10.1093/brain/awy120.

[3] Kenneth M. Heilman and Edward Valenstein. *Clinical Neuropsychology.* Oxford University Press, 1993.

[4] Bradford Z. Mahon and Alfonso Caramazza. "Concepts and categories: a cognitive neuropsychological perspective". In: *Annual Review of Psychology* 60 (2009), pp. 27–51. DOI: 10.1146/annurev.psych.60.110707.163532.

[5] M. M. Mesulam et al. "The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia". In: *Brain* 138 (2015), pp. 2423–2437. DOI: 10.1093/brain/awv154.

[6] Pranav Misra et al. "Invariant neural representation of parts of speech in the human brain". In: *bioRxiv* (2024). DOI: 10.1101/2024.01.15.575788.

[7] Elliot Murphy et al. "Minimal Phrase Composition Revealed by Intracranial Recordings". In: *Journal of Neuroscience* 42 (2022), pp. 3216–3227. DOI: 10.1523/JNEUROSCI.1575-21.2022.

[8] Kirill V. Nourski et al. "Gamma Activation and Alpha Suppression within Human Auditory Cortex during a Speech Classification Task". In: *Journal of Neuroscience* 42 (2022), pp. 5034–5046. DOI: 10.1523/JNEUROSCI.2187-21.2022.

[9] George Ojemann et al. "Cortical language localization in left, dominant hemisphere". In: *Journal of Neurosurgery* 71 (1989), pp. 316–326. DOI: 10.3171/jns.1989.71.3.0316.

[10] Michael Petrides. *Neuroanatomy of language regions in the human brain.* Elsevier, 2014.

[11] Sophie K. Scott. “From speech and talkers to the social world: The neural processing of human spoken language”. In: *Science* 366 (2019), pp. 58–62. DOI: 10.1126/science.aax0288.

[12] Elizabeth Warrington and Tim Shallice. “Category specific semantic impairments”. In: *Brain* 107 (1984), pp. 829–854.