Robust and Multimodal Signals for Language in the Brain

a dissertation presented
by
Pranav Misra
to
The Department of Biophysics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biophysics

Harvard University
Cambridge, Massachusetts
August 2024

Thesis advisor: Professor Gabriel Kreiman Pranav Misra

Robust and Multimodal Signals for Language in the Brain

Abstract

An impressive aspect of human language is its ability to maintain consistent meaning and grammatical correctness across different sensory modalities. Elucidating the modality-invariant internal representation of language in the brain has major implications for cognitive science, brain disorders, and artificial intelligence. A pillar of linguistic studies is the notion that words have defined functions, often referred to as parts of speech, such as nouns and adjectives. This dissertation tries to answer two questions: how can we find a multimodal and generalizable representation for language processing? Is there a representation for parts of speech in the brain?

To address these questions, I recorded neural responses from 1,801 electrodes in 20 participants with epilepsy while they were presented with two-word minimal phrases consisting of an adjective and a noun, in both auditory and visual presentations. I observed neural signals that distinguished between these two parts of speech (POS), localized within a small region in the left lateral orbitofrontal cortex. The representation of POS showed invariance across several criteria: visual and auditory presentation modalities, and robustness to word properties like length, order, frequency, and semantics. The results also generalized across two different languages in a bilingual participant. I found that these selective signals provide key elements for the compositional processes of language, highlighting a localized and invariant representation of POS. Furthermore, I extended these ideas by evaluating how parts of speech are processed within full sentences. Recording activity from 1,593 electrodes in 17 participants, I found neural signals that separated nouns from verbs in sentences. This selective, invariant, and localized representation of parts of speech provides key elements for the representation of language. Going beyond POS, I also found multimodal representations for grammar and syntax processing in sentences, a first with neurophysiological signals.

In conclusion, I collected and analyzed a dataset of over 60 hours of stimuli, 40,000 two-word phrases, and 15,000 sentences by recording neural responses to audiovisual stimuli from 3,394 stereo-electroencephalography electrode contacts across 37 participants. My findings report neural representations of language that are both robust and adaptable, contributing to a deeper understanding of core linguistic processes within the human brain. This work lays a foundation for more nuanced studies of language processing. To the best of my knowledge, this is the first study to rigorously and systematically evaluate the invariant representation of parts of speech and multimodal representations for semantic and grammatical processing of sentences at the invasive neurophysiological level from the human brain.

Contents

Listing of figures

List of Tables

Love, kindness and knowledge can change someone’s life forever,
Dedicated to those who showed me how abundantly we are surrounded by them,
Like fish are surrounded by the ocean.

Acknowledgments

Meeting Prof. Gabriel Kreiman, my advisor, for weekly meetings has been one of the most positive and liberating thing over the last 6 years. I would like to express my deepest gratitude towards him, for his unwavering support, inspiration, and rigorous guidance throughout this journey. His mentorship has been invaluable in shaping this work and my life.

I also want to thank all the patients who participated in my experiments from different hospitals in Boston, Cleveland, and Taipei, and the doctors who made these experiments possible and inspired me with their unwavering service to healing: Drs. Joseph R. Madsen, Scellig Stone, Daniel Weisholtz, Demitre Serletis, Bill Bingaman, Juan Bulacio, YenCheng Shih, and Yu Hsiang-Yu.

To Dr. Kavita Singh, who has been my life line for research and life.

I am immensely grateful to my lab mates for their intellectual and moral support. A special thanks to Jie Zheng for her guidance and support when my experiments weren't working. I would also like to thank Spandan Madam, Elisa Pavarino, Paula Sanchez-Lopez, Yuchen Xiao, Jiarui Wang, Morgan Talbot, Dianna Hidalgo, Georgia Dellaferrera, Katarina Bendtz, and David Mazumder for creating an enriching and supportive lab environment. Spandan, Paula, Elisa and Jie have always been there to vent and brainstorm. I feel truly lucky to have them in my life.

A special shoutout to Dr. Akshay Sharma whose exceptional support and proactive efforts were instrumental in establishing a new collaboration at the Cleveland Clinic.

I am thankful to the Biophysics Program for providing a stimulating academic environment and for their support. I extend my appreciation to Michele R. Eva, Prof. Venkatesh Murthy (Venki), Prof. Rachelle Gaudet, Prof. Martha Bulyk, Daniel Barabasi, and Jonathan Jackson for their insights and encouragement. Venki has been a unique source of support and inspiration. His curiosity and openness for science and scientific culture is a force in itself.

I also want to thank my Committee Members, Prof. Boris Katz, Prof. John Assad, Prof. Alfonso Caramazza and Prof. Ziv Williams. They really went into all the details of my work and helped iron out the loose ends before peer review.

To my friend Eshaan Matthew, who has been a constant since the first day of grad school, thank you for the late night airport drop-offs and pick-ups. You made leaving Boston hard and coming back easier. To my friends and housemates, Jonas Herman, Felicia & Arvid Anderesen and Rafael Marengoni, thank you for being my anchors and for always being there. To Eeshit Vaishnav, a true friend and inspiration. To Gargi Dwivedi, for showing me my own light and re-teaching me friendship and trust.

I am forever grateful to my parents for giving me the freedom to pursue my dreams, especially my mother, who has given me freedom in the most complete sense of the word. To my grandmothers (Dadi: Prof. P.K. Misra, Nani: Mrs. G.G. Shukla and Manglik Nani: Prof. Mohini Manglik), thank you for being my first teachers. To my sister, Radhieka, who showed immense growth for herself and has become a pillar of support and inspiration. To my aunts, uncles and cousins who are an ocean of love so big and so deep, that I can venture away from home without worrying about my family - Kiki & Tanu Mausi, Peeku Mama, Bhuma Nani and Wakhlu Family. Also to my dogs, Oscar, Waldi, Kaiser, and Biggie - you simplify things and bring so much love.

To those writers and scientists who saved me a tonne of time! Thank you for thinking stuff that helped me make sense of life. Einstein, Heisenberg, Sri Aurobindo, Melanie Klein, John Dewey, John Hopfield, Donald Hebb, Leo Tolstoy, Anuradha Roy, Alfred Adler, Kahneman, and Godel.

Lastly, to my friends at Harvard Ed School: Arundhati Sural, Muskaan Maharia, Mehaak Maharia, Kavya Krishna, Amrita Raghavendran, Deepanjali Rao and Kruti Goyal. They say there is light at the end of the tunnel, but I didn't know there was so much of it! Life's been different since I ran into everyone since November 2023. Special thanks to Arundhati and Muskaan for the wholesome conversations. Anything is doable after a hot cup of chai.

0
Introduction

An undergraduate class on the "Theory of Computation" introduced me to formal proofs for algorithms and compositional computations. The course material overlapped with frameworks from linguistic theory that classify speech patterns based on their complexity, like the famous Chomsky Hierarchy, and naturally brought me to the question, "Can human language be described by an algorithm?"

After extensive review and deep contemplation, I realized that human language was too complex and undefined to be called an algorithm*. Unlike computer algorithms, which are designed for specific outcomes, language emerges from the human brain within social interactions, serving many functions like communication, reasoning, and planning. While we might create algorithms for different parts of language and solve the individual puzzle pieces, understanding the whole system remains a larger challenge. The success of Large Language Models (LLMs) in predicting the next word in a sentence demonstrates that modeling even one salient aspect of language can be very effective, leading to the creation of "foundation models" repurposable for various downstream tasks (OpenAI, 2023).

But for my thesis, I took a different more experimental approach and want to focus on how language processes might be implemented in the brain. After I started my Ph.D. in Biophysics in 2018, I developed an appreciation for techniques within the neurobiological and cognitive neuroscience community. The experimental rigor and the gap that was needed to understand language mechanisms for clinically relevant outcomes drew me in. I found a lack of neural evidence supporting linguistic theories and computational models. Additionally, studies showed that animals lack complex language-like behaviors, suggesting human uniqueness and a sense of urgency, given the overarching role of language in our lives †. This shifted my interest toward under-

*An algorithm is a finite set of instructions carried out in a specific order to perform a particular task.

†Many studies have tried to explore this phenomena of computability and grammars in animals (Stobbe et al., 2012, Saffran et al., 2008, Malassis et al., 2020, Reber, 1967, Abe and Watanabe, 2011, Gentner et al., 2006, Fitch and Hauser, 2004, Wilson et al., 2020, Heimbauer et al., 2018, Jiang et al., 2018). While some showed similarities to humans, most revealed significant differences in generalization, especially when simple cues like color were removed (van Heijningen et al., 2009). Additionally, animals could not self-report their actions like humans, who could verify their logic.

standing language processes in the brain rather than entirely creating language-like algorithms[‡] that were divorced from their organ of origin. At this point, I faced the fundamental experimental questions: what to measure and how to measure it.

0.1 What do I measure for?

Core functions of the linguistic circuit, can be conceptualized along the following three axes (Chomsky, 1995) [§]:

1. Superficial: production and perception of speech and writing

2. Syntax: grammatical roles and correctness, and phrase composition

3. Semantics: word meaning, and phrase meaning

Many neurophysiological experiments have begun to investigate neural signals associated with the presentation of individual words or short phrases (Woolnough, 2021, Nourski, 2022, Forseth, 2018, Forseth et al., 2021, Murphy, 2022, Ding et al., 2016, Cometa, 2023, Artoni, 2020, Keshishian, 2023, Khanna et al., 2024). Research has examined orthographic features of real versus pseudowords (Woolnough, 2021, Vigliocco et al., 2011, Castellucci et al., 2022), phonetic features of word comprehension (Forseth, 2018, Woolnough, 2021, Yi et al., 2021, Gwilliams et al., 2022) and production (Forseth et al., 2021, 2020, Bhaya-Grossman and Chang, 2022, Hamilton et al., 2021) and the retrieval of semantic information in audio-visual naming-

[‡]For algorithms modeling language in the brain, one would require them to predict neural data and share similarities with neural responses and perhaps neural architectures, mechanisms, etc.

[§]These axes are not necessarily orthogonal to each other in a strict sense, and often overlap and intersect. E.g., the tone (*superficial*) of a particular phrase, (such as "Did you eat your lunch?") can modulate whether it was said sarcastically or not (*semantic*).

to-definition tasks (Forseth, 2018). These studies have shed light on the early processes associated with detecting, comprehending, and producing words. However, beyond individual words, the heart of linguistic structures lies in the notion that words serve specific functions within a sentence, including articles, nouns, adjectives, and verbs. These parts of speech (POS) are building blocks for generative syntax mechanisms, facilitating the creation of infinitely many phrases and sentences. POS are widely shared across languages, combined according to defined grammatical rules, and play critical roles in natural language processing (NLP) algorithms (Chomsky, 1995, Chomsky et al., 2019, Murphy, 2022, OpenAI, 2023, Hagoort and Indefrey, 2014, Calinescu et al., 2023, Goldstein et al., 2022, Jamali et al., 2024, Cai et al., 2023). Despite the emphasis from NLP algorithms and linguistic theory on POS, their neural representations remain ambiguous. We did not know if there is a POS representation in the brain.

Many studies in patients with brain lesions have focused on deficits in the retrieval of individual nouns and verbs (Vigliocco et al., 2011, Rapp and Caramazza, 2002, Caramazza and Hillis, 1991, Woolnough, 2021, Aflalo et al., 2020, Damasio and Tranel, 1993, Crepaldi et al., 2011); however, previous studies, including neuroimaging and non-invasive magnetic and electrophysiology experiments, could not resolve the explicit neural processes in POS processing due to insufficient signal-to-noise ratio, spatial or temporal resolution[¶]. Furthermore, recent work has suggested that parts-of-speech may be implicitly learned and represented in modern large language models

[¶]Note: In word retrieval studies, the brain can possibly make use of semantic memories associated with a given word, in addition or in the extreme case entirely separate from its POS and grammatical roles. While these approaches remain important for benchmarking neuro-rehabilitation for speech and language production, they obfuscate the study of grammatical processes for language perception.

(Tenney et al., 2019, Elazar et al., 2021).

A remarkable hallmark of language is its universality. We can interpret the word 'cat' when uttering it, writing it, listening to it, reading it, and even when examining a photograph of a cat. This universality tempts us to speculate that there may be an invariant representation of language concepts in the brain. Several studies have examined potential correlates of language processing using only unimodal signals (Murphy, 2022, Keshishian, 2023, Cai et al., 2023, Jamali et al., 2024, Sinai, 2005, Ding et al., 2016, Cometa, 2023, Woolnough, 2021, Forseth et al., 2020, Chomsky et al., 2019, Goldstein et al., 2022). Noticing this gap, I decided to study language using an intersectional approach, incorporating both auditory and visual stimuli. Within this paradigm, I made the foundational elements of grammar, namely POS, the focus of my initial study.

Following initial successes from the first study, I aimed to build a more complex understanding of language processing and designed an experiment for full sentences. To the best of my knowledge, there is no existing dataset with human intracranial signals for language in both auditory and visual modalities with full sentences. In this study design, I incorporated elements of both syntax and meaning processing.

During my rotation before joining the Kreiman Lab, I also conducted extensive experiments in participants with intracranial electrodes with naturalistic stimuli from movies, eventually collecting a dataset of 43 hours of movies with 36,000 sentences (205,000 words) across 10 participants. This is the largest dataset of intracranial recordings featuring grounded naturalistic language, one of the largest English universal dependencies (UD) treebanks in general, and one of only a few UD treebanks aligned to multimodal vision-language features.

Ultimately, I created unique datasets of over 2 terabytes from 5,064 electrode contacts implanted within 47 participants to study the neural representation of language, and conducted extensive analysis, leading to foundational results. I will expand upon the methods and challenges in the next subsection.

0.2 How I chose to do it - my approach and challenges

Experimental design, statistical power, reported effect sizes, and time-to-conclude-study are widely dependent upon the choice of the measurement technique (Waldert, 2016). There are three main axes along which current techniques of measuring brain function fall into :

1. invasive versus non-invasive,
2. spatial resolution, and
3. temporal resolution.

The most salient among these axes is the *invasive versus non-invasive* because it translates to a number of advantages and challenges. The invasive methods, namely single-unit recording (32 Khz, gives action potentials from single neurons) (Dubey and Ray, 2019, Johnson and Knight, 2015, Mukamel and Fried, 2012a), stereoelectroencephalography (sEEG, up to 2 KHz), and Electrocorticography (ECoG, up to 2 KHz), exhibit the highest spatiotemporal resolution and signal-to-noise ratio that are the keys to the study of human cognition. SEEG and ECoG were originally developed for localizing epileptogenic foci for pharmacologically-intractable epilepsy patients and are currently widely used for addressing neuroscience questions. The challenges with invasive methods are the low rate of participant recruitment, which is solely guided by clinical needs, high costs, high expertise and a high risk clinical environment that is required to facilitate its use.

Non-invasive methods, namely magnetoencephalography (MEG), electroencephalogram (EEG), positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), either lack the spatial resolution, temporal resolution or both depending on the choice of measurement. Additionally, all the non-invasive techniques have minimal signal-to-noise ratio. The advantage with these techniques is a higher rate of participant recruitment, and the possibility of using these methods in non-clinical and lower stakes settings.

Finally, it is worth mentioning cognitive studies of patients with lesions or neurodegenerative disorders, that fall into a separate kind of non-invasive technique. These methods give unique insights into what and how brain functions are lost (aphasia) and recovered over time due loss of or injury to brain tissue (Caramazza and Hillis, 1991, Mesulam et al., 2015, 2014, 2022). However for these methods, the participant throughput and brain area coverage is unreliable, requiring long range research planning that is often focused on rehabilitation. Moreover, the spatial and temporal resolution are severely lacking compared to other methods. One noteworthy area of research with this approach is the study of Primary Progressive Aphasia (PPA), a neurodegerative disease, that has sometimes been called "the Alzheimer's of the Language Circuit". In the upcoming sections, I will discuss how my findings compare with research on PPA.

Given the limitations of non-invasive methods, which lack the signal quality and resolution to confirm or deny hypotheses derived from decades of linguistic theory, I realized the critical position of invasive methods in providing answers about core language processes. The opportunity at the Kreiman Lab at Boston Children's Hospital and several other collaborating hospitals had the opportunity to study language in the brain with high spatiotemporal resolution via sEEG recordings, providing the hope of filling the knowledge gaps mentioned earlier. After quick and focused review of language representations in the brain, I found two major gaps which helped me design a more focused and exciting research plan :

1. data from non-invasive methods lacked the resolution to confirm or deny the hypothesis derived from decades of linguistic theory (Crepaldi et al., 2011).

2. the existing literature from invasive methods mostly focused on speech production, detection or word orthography. There were a few notable invasive studies of language that relied on unimodal stimuli to study the phrase composition, but lacked the conclusive evidence for multimodal language processes (Woolnough, 2021, Nourski, 2022, Forseth, 2018, Forseth et al., 2021, Murphy, 2022, Ding et al., 2016, Cometa, 2023, Artoni, 2020, Keshishian, 2023, Khanna et al., 2024), and audiovisual semantic retrieval (Forseth, 2018).

Cognitive research using SEEG and ECoG requires long-range planning and overcoming high stakes challenges due to

Semantic processing is considered orthogonal and complementary to syntax processing and POS representations (Chomsky et al., 2019).

1. significant costs: $ 120,000 per procedure (Salehi et al., 2022) **,

2. limited participant availability: one participant every two months, 2-3 years of data collection cycles, and

3. strict clinical environment constraints: I typically had a narrow 2-hour window with no guarantees for repetition.

These challenges required me to meticulously plan an experiment, design and tailor it with sufficient statistical power to support the key research questions I wanted to tackle. Eventually, I designed an audiovisual minimal phrase task that encapsulated the ability to answer hypotheses for POS processing for my first study[††]. In parallel, I also collected data for BrainTreeBank (Wang et al., 2024), which is a large-scale dataset of electrophysiological neural responses, recorded from intracranial probes while 10 subjects watched one or more Hollywood movies. Subjects watched on average 2.6 Hollywood movies, for an average viewing time of 4.3 hours, and a total of 43 hours.

After these two studies, I designed an audiovisual sentences-picture matching task to reveal interactions between grammar and meaning. All three experiments combined, I collected and analyzed over **2 terabytes (Tb)** of neural data from **5,064 electrode contacts** across **47 participants** during my graduate program[‡‡].

**Adding monthly salaries of all the staff facilitating sEEG surgeries, for a 1 patient per month rate the cost of supporting these experiments go up to $ 250,000 per procedure.

[††]Both tasks detailed in the upcoming sections

[‡‡]I started working on BrainTreeBank as a rotation project and soon realized both the advantages and challenges of using naturalistic, movie-like stimuli to study language, especially with invasive methods and limited time per participant. The advantage is the engaging nature of movie-like stimuli, which can facilitate large-scale recordings. The challenge is that many other cues, such as visual elements, narrative, tone, pitch, or musicality, get correlated with linguistic or grammatical stimuli.

0.3 Why do we need intracranial evidence for language?

Compared to non-invasive methods like fMRI that offer a general activation map, intracranial experiments provide a much more precise picture. By placing electrodes directly within the brain, these experiments can pinpoint the specific areas involved in various aspects of language processing, such as word and grammar comprehension. Additionally, intracranial recordings directly measure the electrical activity of neurons, offering a deeper understanding of how language is encoded and processed within the brain.

The remarkable precision offered by intracranial studies opens doors to new possibilities in preventing and treating language impairments in:

1. Epilepsy: For epilepsy patients with language centers located near seizure zones, intracranial recordings can help identify these areas more precisely. This allows for targeted surgical intervention to minimize the risk of language impairment after surgery.

2. Stroke: Stroke can damage language areas (Caramazza and Hillis, 1991). Intracranial studies can help map the extent of damage and residual language function. This information is crucial for guiding rehabilitation efforts and predicting potential recovery.

Given the success of LLMs, which require more data and are less concerned with the 'cleanliness' of the data, neural recordings with movie-like stimuli can be very useful for creating foundational models for language and narrative perception. When I started my PhD project in 2019, ChatGPT had not yet been released, and there were no available foundational language models that could be repurposed for the small amounts of experimental neural data available per participant. I identified critical gaps that could be studied with invasive recordings, which led me in the direction of this thesis.

3. Neurodegenerative Disorders: Diseases like Alzheimer's and Pick's Disease can affect language abilities leading to Primary Progressive Aphasia (PPA) (Mesulam et al., 2022). By pinpointing the specific brain regions and linguistic processes affected, intracranial studies can aid in developing targeted therapies to slow or prevent language decline.

0.4 Summary of Thesis

This thesis is split into five chapters, and one appendix.

Chapter 1 goes directly into the results of the minimal phrase study.

Chapter 2 contains the discussion of the minimal phrase study.

Chapter 3 details the methods of analysis used in the minimal phrase study.

Chapter 4 summarizes my findings with the audiovisual sentence-meaning task. I started this work in June 2023, and am close to completing it with exciting results.

Chapter 5 contains the conclusions about my experience with human invasive neuroscience and the goals that I had set for myself to understand language from within the brain.

This thesis also contains one Appendix. Appendix A details the findings of the BrainTreeBank effort that was done in collaboration with Andrei Barbu, Adam Yaari and Christopher Wang at MIT.

"Words can travel where pictures cannot reach. While a picture may capture a fleeting moment, words can explore the endless nuances of that moment, giving it life, context, and meaning."

"Opposite of a picture is worth a thousand words",

GPT4

1 Results

Imagine yourself to be a caveperson right after the advent of language. You are telling your fellow caveperson whether the animal you found lurking by the watering hole was a "big animal" or a "small animal", or if the "red berry" or the "blue berry" was eatable. Would it not help to know all of this in just a few words? Imagine the multitude of simple adjective-noun word combinations that you could tell your friend that would help them avoid danger or find food. Such is the beauty of language: limitless possibilities with only a few words.

Here in my initial experiment, I investigated how POS are processed in the brain in minimal phrases, where an adjective is followed by a noun. What would a representation for POS, like nouns and adjectives, in the brain look like? Consider the adjective "green" and the noun "apple", combined to create the simple phrase "green apple." Fundamental constraints for such a representation should include the basic invariances underlying the cognitive understanding of this phrase. The basic desiderata for the representation of parts of speech in language includes invariance to:

1. Presentation modality (e.g., auditory versus visual),
2. Specific noun or adjective (e.g., green or red)
3. Position within a phrase (e.g., "green apple" versus "apple green"),
4. Specific language in bilingual speakers (e.g., "green apple" in English versus "manzana verde" in Spanish)
5. Superficial statistical word properties (e.g., written length, number of syllables, and phoneme composition)

With the guidance and collaboration of clinicians, I conducted experiments and recorded intracranial field potential responses with high spatiotemporal resolution and high signal-to-noise ratio from 1,801 electrodes implanted in 20 participants with pharmacologically-resistant epilepsy. I found neural signals, especially in the left lateral orbitofrontal cortex, that selectively distinguish between nouns and adjectives. These POS selective signals are robust when words are matched for orthography (e.g., word length), acoustic features (e.g., number of syllables), word sequence (e.g., noun or adjective at first or second position within a phrase), and frequency of occurrence. Interestingly, the representation of nouns versus adjectives generalizes across audio and visual modalities, across different semantic categories within each part of speech, and across different languages.

1.1 Experiment Design and Data for Parts of Speech (POS)

I recorded intracranial field potentials from 1,801 electrodes (840 in gray matter, 961 in white matter) implanted in 20 participants. Participants heard (auditory modality) or read (visual modality) two words that were sequentially presented and were asked to indicate whether the words were the same or not (**Figure 1.1a**, **Methods**). Participants performed the task correctly on 93.6±7.7% of the trials (here and throughout, mean±std, unless stated otherwise). All electrode locations are shown in **Figure 1.1b-g** (see also **Tables S1 S2** and **Methods**). I use a bipolar reference, and I focus on the intracranial field potential signals filtered in the high gamma frequency band, referred to as neural responses throughout and reported in the plots as gamma power (65-150 Hz, **Methods**).

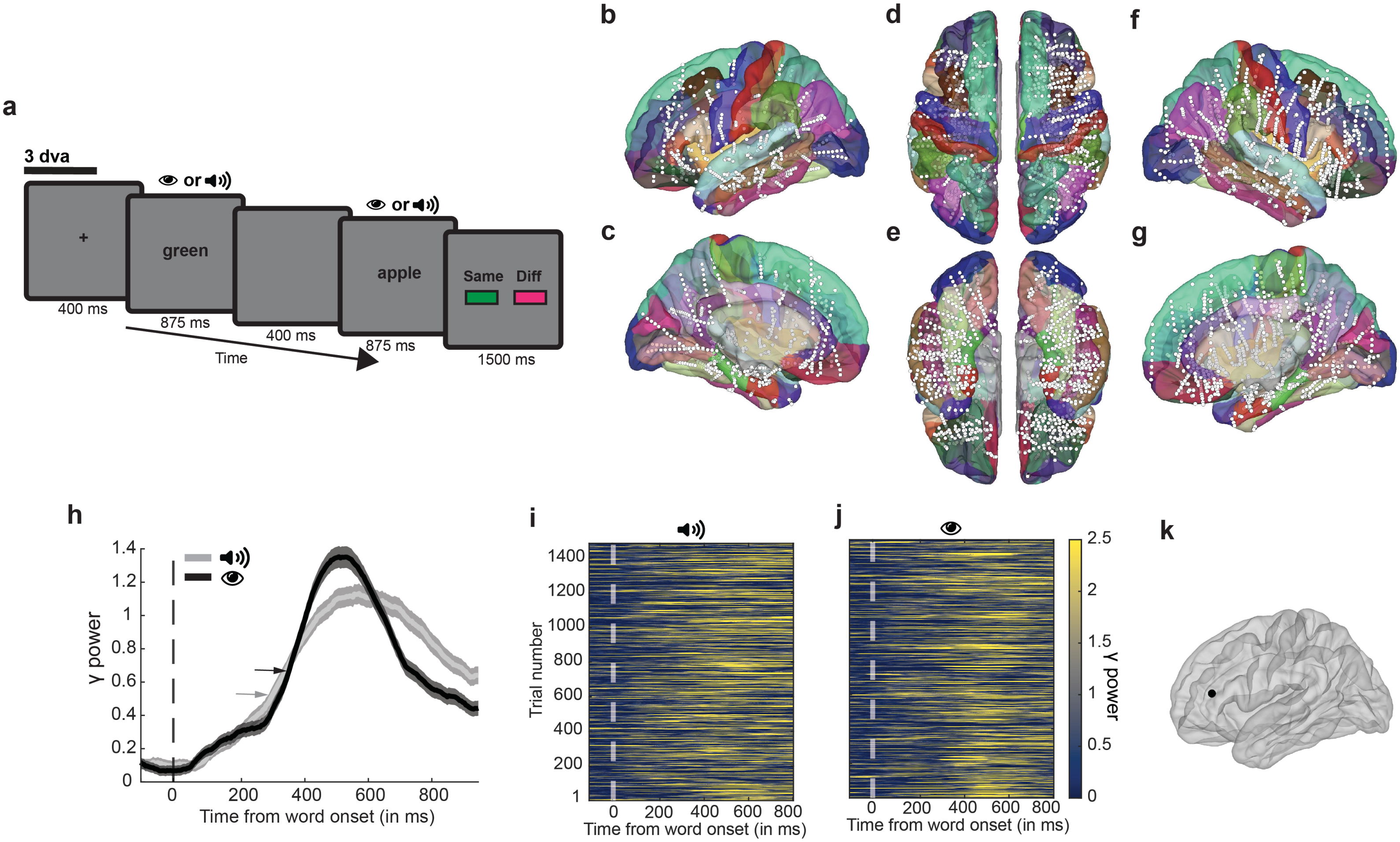

a
3 dva
or
+
400 ms
green
875 ms
400 ms
apple
875 ms
Same
Diff
1500 ms
Time
b
c
d
e
f
g
h
γ power
Time from word onset (in ms)
i
Trial number
Time from word onset (in ms)
j
Time from word onset (in ms)
γ power
k

Figure 1.1: Task design, electrode locations and multimodal responses.

a. Task schematic. Two words were sequentially presented either in visual modality or auditory modality. Participants indicated whether the two words were the same (e.g., “apple apple” or “green green”, 8% of trials of each type) or different (e.g., “green apple” or “apple green”: 42% of trials of each type, Methods). In the 84% of trials where the two-words were different, there was an adjective followed by a noun or a noun follower by an adjective.

b-g. Location of all electrodes overlayed on the Desikan-Killiany Atlas shown with different views. Each white circle shows one electrode. b. Left lateral view (n=693), c. Left medial view (n=693), d. Superior, whole brain view (n=1,801), e. Inferior, whole brain view (n=1,801), f. Right lateral view (n=1108) g. Right medial

h. Trial-averaged (± SEM) gamma power for responses to auditory (light grey) or visual (black) presentations for an example electrode in the left rostral middle frontal gyrus (electrode location shown in k). Responses are aligned to word onset (vertical dashed line). The arrows indicate the half-maximum time.

i,j. Raster plots showing each individual trial for the same electrode for each of the 1,496 words for auditor (i) and visual (j) presentations (see color scale on right).

1.1.1 Neural signals reflect visual, auditory and multimodal inputs

I observed 565 electrodes (31.4% of the total) that responded to auditory stimuli (**Figure 1.2 a-c, g-i**) and 532 electrodes (29.5% of the total) that responded to visual stimuli (**Figure 1.2 d-f, g-i**). The overall proportions and dynamics of visual and auditory responsive signals are consistent with previous work (Bansal et al., 2014, Forseth, 2018, Woolnough, 2021). Of these electrodes, there were 293 electrodes that responded to both auditory and visual stimuli (**Figure 1.2 g-i**). These 293 electrodes represent 16.3% of the total, 51.9% of the auditory responsive electrodes, and 55.0% of the visually responsive electrodes. This number of audiovisual electrodes is highly unlikely to arise by chance from the number of auditory and visual electrodes ($p<10^{-4}$, permutation test, $n=10^6$ iterations). Of these 293 electrodes, 147 (50.2%) were in the left hemisphere and 146 (49.8%) were in the right. Of the 41 the regions in the Desikan-Killiani Atlas where I had sampling (34 defined regions and 7 extra regions representing deep gray matter structures, **Methods**, **Figure 1.1**, **Tables S1 S2**), 13 regions had a significantly higher number of multimodal electrodes than from the number of audio or visual electrodes ($p<0.01$, permutation test, $n=10^6$ iterations). These regions are indicated in bold in **Table S2** . **Figure 1.1 h-j** shows the responses of an example audiovisual responsive electrode located in the left rostral middle-frontal gyrus (**Figure 1.1k**). This electrode showed strong evoked responses evident in the trial-average responses (**Figure 1.1h**), and even in individual trials for both auditory stimuli (**Figure 1.1i**) and visual stimuli (**Figure 1.1j**). To compare the response dynamics of auditory and visual responses, I calculated the time at which the neural signals reached half of the max amplitude (half-maximum time, arrows in Figure 1h, Methods) and the average area under the curve (AUC) for neural responses such as those in **Figure 1.1h**. **Figure 1.3a** shows the half-maximum time for auditory-only electrodes (left), visual-only electrodes (middle), and audiovisual electrodes on audio trials (right light-gray half) or visual trials (right black half). There was no significant difference between the half- maximum time for auditory-only electrodes (329±187 ms) and visual only electrodes (336±174 ms) ($p>0.05$, ranksum test). Similarly, there was no significant difference between the half- maximum time for the audio and visual responses of audiovisual electrodes (379±193 ms versus 341±174 ms, $p>0.05$, ranksum test). However, there was a small but significant difference between the half-maximum time for audio only electrodes and auditory responses of audiovisual electrodes ($p<0.01$, ranksum test). As expected, for the audio-only electrodes, the average response AUC to auditory stimuli (108±100 μV^2/Hz-ms) was larger than the average AUC response to visual stimuli (44±16 μV^2/Hz- ms) (p<10-4, ranksum test, **Figure 1.3b**). Similarly, for the visual-only electrodes, the average response AUC to auditory stimuli (40±23 μV^2/Hz-ms) was smaller than the average AUC response to visual stimuli (53±43 μV^2/Hz-ms) ($p<10^{-4}$, ranksum test, **Figure 1.3c**). For the audiovisual electrodes, the average response AUC to auditory stimuli (71±72 μV^2/Hz-ms) was slightly larger than the AUC of their responses to visual stimuli (54±39 μV^2/Hz-ms) ($p<0.01$, ranksum test, **Figure 1.3d**).

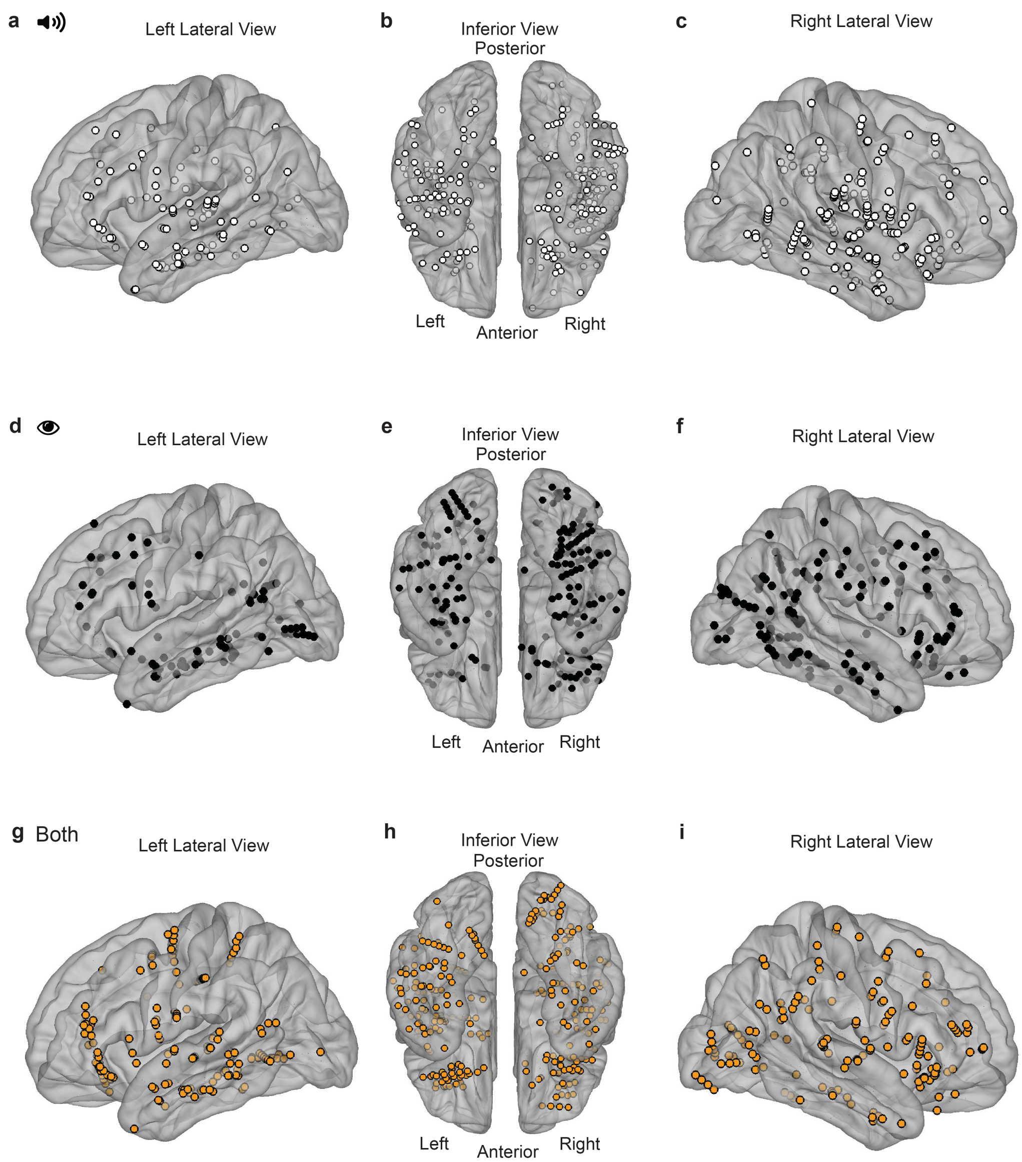


Figure 1.2: **a-c.** Only audio responsive electrodes (a: left hemisphere lateral view, n= 102; b: inferior view (n=272); c: right hemisphere lateral view, n= 170). **d-f.** Only visually responsive electrodes (d: n= 85; e: n=239; f n= 154).**g-i.** Audiovisual responsive electrodes (g: n= 147; h: n=293; i: n= 146). The same color scheme is followed throughout the paper to indicate vision-only, audio-only or audiovisual electrodes. iELVis pullout factor=20, opaqueness=0.6.

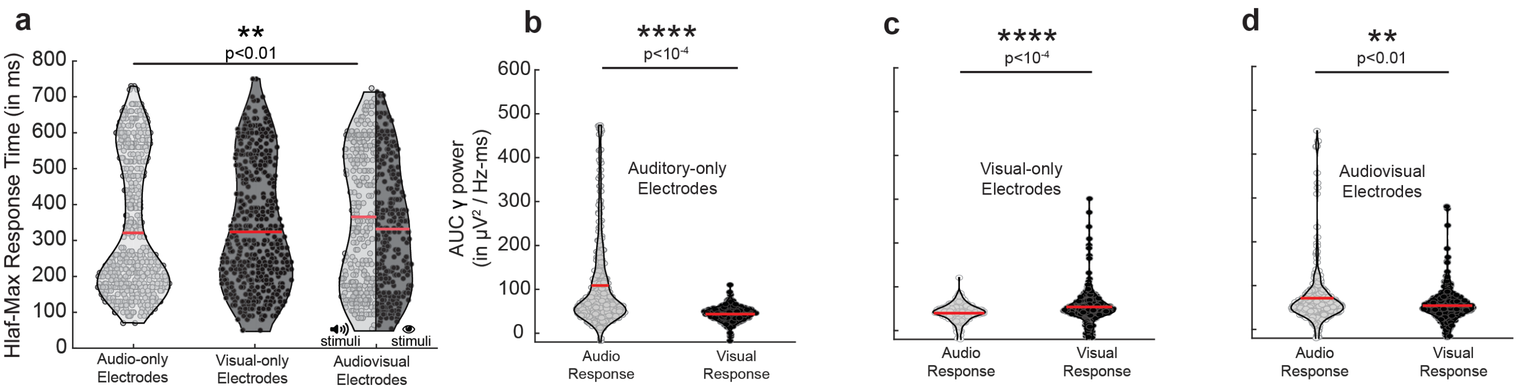


Figure 1.3: Half-maximum time and area under the curve for responsive electrodes.
a. Half-maximum time for audio-only electrodes (left, light-gray: 329±187 ms), visual-only electrodes (middle, black: 336±174 ms), and audiovisual electrodes (right; auditory stimuli in light-gray: 379±193 ms, visual stimuli in black: 341±174 ms). There was a small but significant difference between the half-maximum time for auditory-only electrodes and for auditory responses of audiovisual electrodes ($p<0.01$, ranksum test). Horizontal red bars indicate mean. Horizontal black bars indicate significant differences.
b-d. Area under the curve for the trial averaged response to auditory stimuli (light-gray violin plots) and visual stimuli (black violin plots) for audio-only electrodes (b, auditory stimuli: 108±100 μV2/Hz-ms, visual stimuli: 44±16 μV2/Hz-ms; p<10-4, ranksum test), visual-only electrodes (c, auditory stimuli: 40±23 μV2/Hz-ms, visual stimuli: 53±43 μV2/Hz-ms; p<10-4, ranksum test), and audiovisual electrodes (d, auditory stimuli: 71±72 μV2/Hz-ms, visual stimuli: 54±39 μV2/Hz-ms; p<0.01, ranksum test). Horizontal red bars indicate mean. Horizontal black bars indicate significant differences.

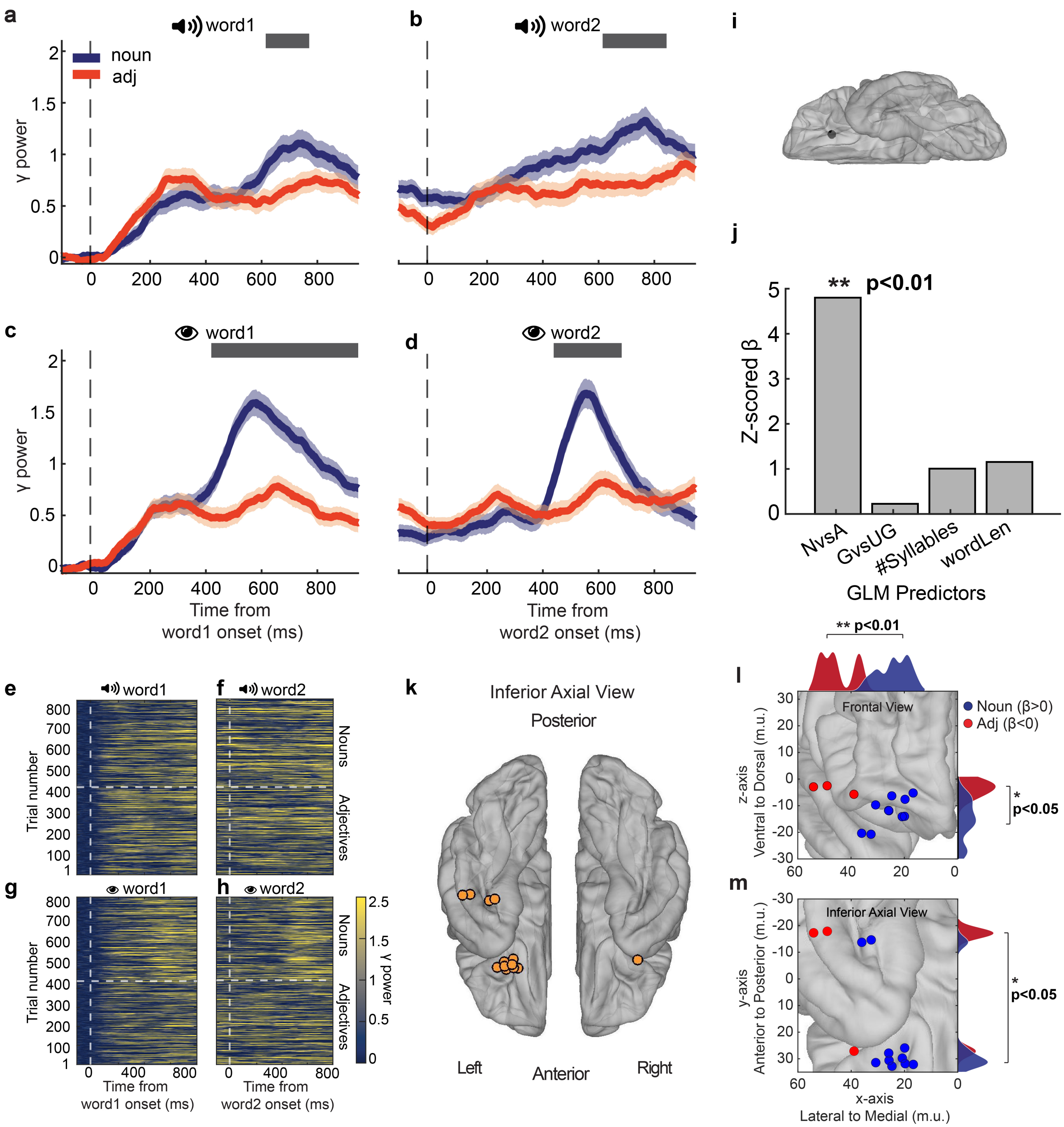
1.2 POS Encoding: Invariant and Localized

1.2.1 Multimodal neural signals distinguish different parts of speech

I evaluated whether the neural signals differentiated between nouns and adjectives. Nouns and adjectives were matched for their number of syllables, word length to control for potential confounds not specific to parts of speech (**Table S3**, **Methods**). **Figure 1.4** shows the responses of an example electrode located in the orbital H-shaped sulcus within the left lateral orbitofrontal cortex (**Figure 1.4i** depicts the electrode location). The orbital H-shaped sulcus lies above the bone of the eye socket where a butterfly-like gyrus can be seen, formed along H-shaped recessions of the sulcus. The neural responses are aligned to the word onset (vertical dashed line) for auditory presentation (**Figure 1.4 a, b**) or visual presentation (**Figure 1.4 c, d**), for the first (**Figure 1.4 a, c**), or second (**Figure 1.4 b, d**) word in each trial. This electrode showed multimodal responses triggered by both auditory and visual stimuli. The responses to nouns (blue) were stronger than adjectives (red) across all four conditions, including both word 1 and word 2, and both for visual and auditory stimuli. The differences between nouns and adjectives can be readily appreciated even in individual trials (**Figure 1.4 e-h**). These differences became significant at approximately 430 ms after word onset for visual presentation and about 610 ms for auditory presentation. In all, there were 89 electrodes, 97 electrodes, and 48 electrodes that showed a difference between nouns and adjectives for auditory stimuli only, visual stimuli only, or both modalities, respectively. The 48 electrodes cannot be ascribed to randomly sampling from the total of audio and visual electrodes ($p<10^{-4}$, permutation test, $n=10^6$ iterations).

Figure 1.4: Neural signals distinguish between different parts of speech.
a-d. Trial-averaged normalized gamma-band power of responses from an example electrode in the left lateral orbitofrontal cortex (see location in i) to nouns (blue) or adjectives (red) during presentation of auditory stimuli (a, b, n=435 grammatical and 432 ungrammatical trials) or visual stimuli (c, d, n=435 grammatical and 432 ungrammatical trials) aligned to the onset (vertical dashed line) of the first word (a, c) or second word (b, d). Shaded areas denote s.e.m. Horizontal gray lines denote windows of statistically significant differences between responses to nouns versus adjectives (t-test $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$).
(e-h) e-h. Raster plots showing the responses in each individual trial (see color scale on bottom right). The red and blue curves in a-d correspond to the averages of noun and adjective trials, respectively, in e- h. **(i)** Location of the example electrode in the left lateral orbitofrontal cortex.
j. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset, using four task predictors: Noun versus Adjectives, Grammatically correct versus incorrect, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients, corrected for multiple comparisons (**Methods**). **k.** Inferior axial view of both hemispheres showing electrodes that revealed statistically significant differences between nouns and adjectives for both audio and visual presentation (orange circles, n=13 electrodes). Electrodes whose responses were significantly explained only by the Nouns versus Adjective task predictor in the GLM are included in this plot.
l, m. All electrodes from k projected onto the left hemisphere are shown on the frontal plane (l) and the axial plane (m, same plane as k). All the electrodes that respond more strongly to nouns, i.e., Nouns versus Adjectives $\beta>0$ (n=10 electrodes), are shown in blue and electrodes that responded more strongly to adjectives ($\beta<0$, n=3 electrodes), are shown in red. All units are in MNI305 coordinates. Kernel density curves (bandwidth 2) outline the marginal distributions of noun-preferring (blue) and adjective-preferring (red) electrodes along the lateral-medial axis (l,m: x-axis, zero being more medial), ventral-dorsal axis (l: right z-axis) and anterior-posterior axis (m: right y-axis). P-values indicate significant differences between the coordinates for noun- and adjective-preferring electrodes (ranksum test).Raster plots showing each individual trial for the same electrode for each of the 1,496 words for auditor (i) and visual (j) presentations (see color scale on right).

Figure 1.4: (continued)



1.2.2 Neural selectivity for nouns versus adjectives was robust to word properties, phrase grammar, usage frequency, and word subcategory

Robustness to early sensory word presentation

Even though nouns and adjectives were matched in their average number of syllables and word length, I asked whether these variables could still contribute to the neural responses differentiating nouns and adjectives. Additionally, each trial could be grammatically correct (e.g., “green apple”), or incorrect (e.g., “apple green”) (Methods); therefore, I asked whether grammar could contribute to the neural differences between nouns and adjectives. To address these questions, I built a generalized linear model (GLM) for each electrode to predict its response AUC between 200 ms and 800 ms after word onset using four predictors: nouns versus adjectives, grammatically correct or not, and word length (vision) or number of syllables (audition) (Methods). The predictor coefficients in the GLM model for the example electrode in **Figure 1.4 a-d** show that only the nouns versus adjectives label significantly explained the neural responses for both auditory and visual presentation (**Figure 1.4 j**). A total of 14 electrodes showed nouns versus adjectives as the only statistically significant predictor in the GLM analysis; 13/14 (93%) of these electrodes distinguished nouns versus adjectives for both auditory and visual inputs, such as the example electrode in **Figure 1.4 a-j**.

The locations of these electrodes that robustly distinguished nouns and adjectives (orange in **Figure 1.4 k**) and reveal a cluster enriched in the left lateral orbitofrontal cortex (LOF). Within the left LOF, 8 out of the 8 (100%) electrodes were in the posterior part of the orbital H-shaped sulcus. I recorded from a total of 113 electrodes in the lateral orbitofrontal region, 38 electrodes in the left hemisphere and 75 electrodes in the right hemisphere (**Figure 1.4 b-g**, **Table S1** . Of the 38 left hemisphere electrodes, 21% distinguished nouns from adjectives during both audio and visual presentation. In stark contrast, only 1.3% of the 75 electrodes in the right hemisphere distinguished nouns from adjectives in both audio and vision (these hemispheric differences were statistically significant: $p<10^{-4}$, permutation test, $n=10^6$ iterations). **Table S4** shows the distribution of electrodes distinguishing part of speech between the left and right hemispheres for all brain regions and **Table S5** shows the distribution of electrodes separating nouns versus adjectives in different participants.

I had initially assumed that distinguishing parts of speech constitutes a core component of language and would therefore be reflected exclusively in both visual and auditory modalities. Indeed, 13/14 (93%) of electrodes differentiating nouns from adjectives in the GLM did so in both modalities. In addition to these 13 electrodes there was a small number of electrodes (2 auditory only and 1 visual only) that showed differences between nouns and adjectives in one modality but not the other. Unlike the electrodes in **Figure 1.4 k**, for the 2 auditory-only electrodes, the number of syllables also significantly contributed towards explaining the neural responses. **Figure 1.5** shows the responses of an example electrode located in the right insula that showed a difference between nouns and adjectives during auditory presentation but not during visual presentation. Conversely, **Figure 1.6** shows the responses of an example electrode located in the left lateral orbitofrontal cortex that showed a clear difference between nouns and adjectives during visual presentation but not during auditory presentation. **Figure 1.6 k,l** shows the locations of auditory only (white circles) and visual only electrodes (black circle) in the left and the right hemispheres, respectively.

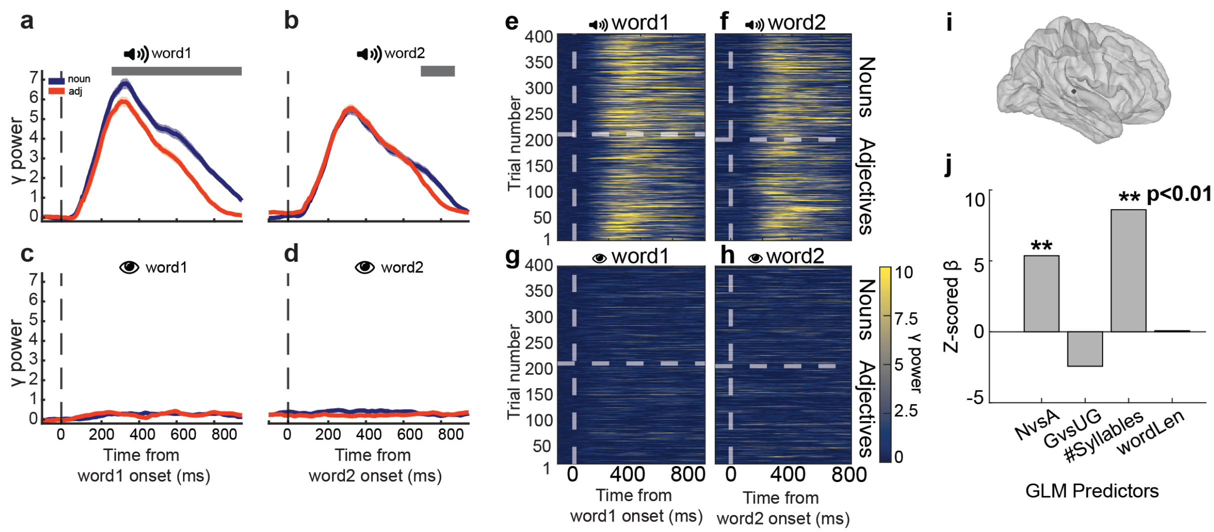


Figure 1.5: **Example electrode distinguishing parts of speech only for auditory stimuli.**

a-d. Trial averaged γ- power of neural responses to Taiwanese words, separated by nouns (blue) and adjectives (red). Neural responses are shown for auditory presentation (a, b), and visual presentation (c, d), aligned to word1 onset (a, c) or word2 onset (b, d). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between nouns and adjectives (t- test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There was a significant differences between noun and adjectives for auditory presentations shown with a gray horizontal line but no difference for visual presentations.

e-h. Raster plots showing the responses in individual trials (see color scale on bottom right). **(i)** Electrode location in the right insula.

j. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset using four task predictors: Noun versus Adjectives, Grammatically Correct versus Ungrammatical, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients.

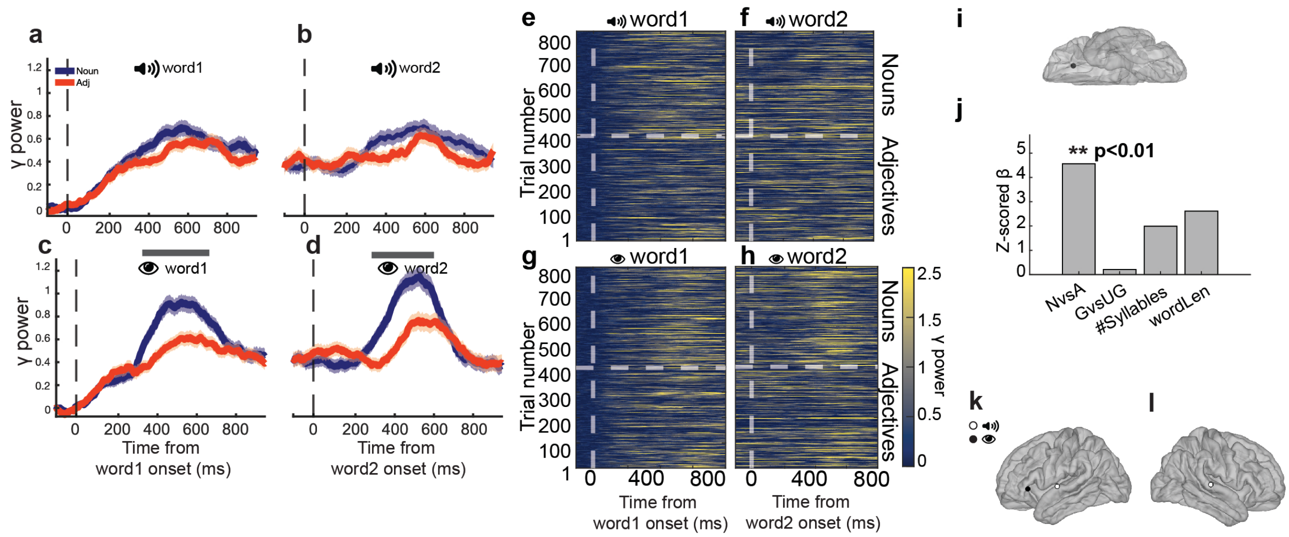


Figure 1.6: Example electrode distinguishing parts of speech only for auditory stimuli.

a-d. Trial averaged γ- power of neural responses to Taiwanese words, separated by nouns (blue) and adjectives (red). Neural responses are shown for auditory presentation (a, b), and visual presentation (c, d), aligned to word 1 onset (a, c) or word 2 onset (b, d). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between nouns and adjectives (t- test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There was a significant differences between noun and adjectives for visual presentations shown with a gray horizontal line but no difference for auditory presentations.

e-h. Raster plots showing the responses in individual trials (see color scale on bottom right). **(i)** Electrode location in the left lateral orbitofrontal.

j. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset using four task predictors: Noun versus Adjectives, Grammatically Correct versus Ungrammatical, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients.

k,l. Electrodes in the left (k) and right (l) hemispheres that showed significant differences between nouns and adjectives either only for auditory trials (white circles) or visual trials (black circles).

Robustness to frequency of occurence

Nouns and adjectives differ in their usage frequency. I asked whether the differences in the neural responses to nouns versus adjectives depended on usage frequency. To address this question, I randomly subsampled the trials to match the distribution of Google Ngram frequency (**Methods**). **Figure 1.7 a** shows matched noun and adjective distributions for the example electrode shown in **Figure 1.7 a-k**. This electrode showed differential responses between parts of speech for auditory (**Figure 1.7 b,c**) and visual (**Figure 1.7 d,e**) stimuli during word1 (**Figure 1.7 b,d**) and word2 (**Figure 1.7 c,e**), even after nouns and adjectives were matched for their frequency of occurrence. Of the 13 audiovisual electrodes where nouns versus adjectives was the only significant predictor in the GLM analysis, 6 electrodes (43%, 4 in the left-LOF, and 2 in left superior temporal gyrus) robustly distinguished nouns and adjectives matched for their frequency of occurrence, like the example electrode in **Figure 1.4** and **Figure 1.7** whereas the other electrodes maintained their selectivity in most but not all conditions.

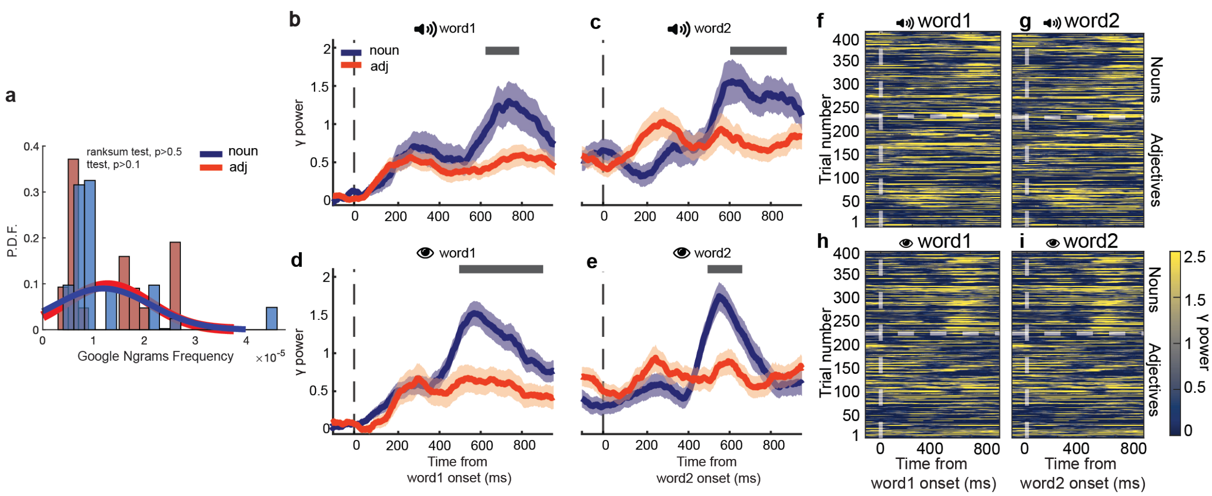


Figure 1.7:

Example electrode distinguishes parts-of-speech for nouns and adjectives matched for their frequency of occurrence.

a. Google Ngrams frequency distribution of nouns (blue) and adjectives (red) that were matched for their median ($p>0.05$, ranksum test) and mean ($p>0.05$, t-test).

b-e. Trial averaged γ-power of neural responses to word onsets, separated by nouns (blue) and adjectives (red). Neural responses are shown for auditory presentation (b, c), and visual presentation (d, e), aligned to word 1 onset (a, c) or word 2 onset (c, e). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between noun subcategories and adjective subcategories (t-test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$).

f-i. Raster plots showing the responses in individual trials (see color scale on bottom right).

Robustness to semantic subcategories

Within my stimulus set, there were two subcategories of nouns, animals and food, and there were two subcategories of adjectives, concrete and abstract (**Table S3**). I asked whether the electrodes that showed differential responses generalized across different word subcategories. The example electrode in **Figure 1.4 a-j** did not show differences between the two noun or adjective subcategories for either auditory stimuli (**Figure 1.8 a, b, f, g**), visual stimuli (**Figure 1.8 c,d,h,i**), word1 (**Figure 1.8 a,c,f,h**), or word2 (**Figure 1.8 b,d,g,i**). Of the13 audiovisual electrodes where nouns versus adjectives was the only significant predictor in the GLM analysis, 8 electrodes (62%) showed generalization across different noun or adjective subcategories. The remaining 6 electrodes (38%) showed a significant difference between the two noun subcategories or between the two adjective subcategories (**Table S5**). Figure S7 shows one of the exceptions, i.e., an electrode in the left LOF which showed a significant response only for food nouns. This selectivity was particularly pronounced for the visual stimuli (**Figure 1.9 c, d, h, i**), but was also apparent for auditory stimuli (**Figure 1.9 a, b, f, g**), and was evident both for word 1 and word 2. In sum, differences in selective responses to nouns versus adjectives were particularly prominent and clustered in the left lateral orbitofrontal cortex, persisted across different word lengths, whether the word was used in a grammatically correct phrase or not, after equalizing word occurrence frequency, and generalized across different noun or adjective subcategories.

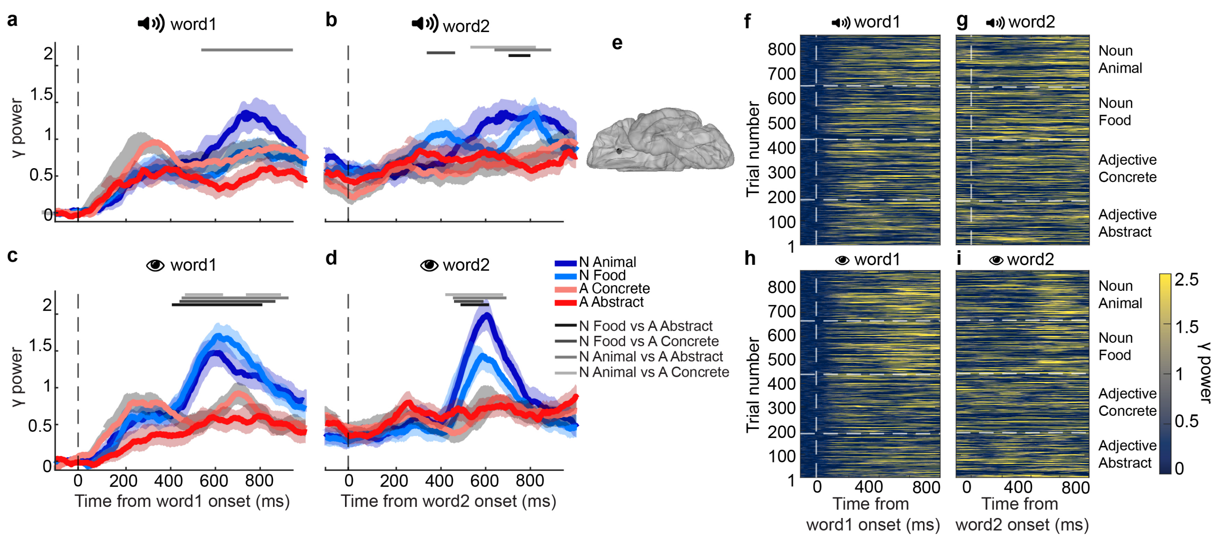


Figure 1.8:

Selective responses to nouns versus adjectives across different noun and adjective categories.

a-d. Trial averaged γ-power of neural responses to words, separated by animal nouns (dark blue), food nouns (light blue), concrete adjectives (light red), and abstract adjectives (dark red). Neural responses are shown for auditory presentation (a, b), and visual presentation (c, d), aligned to word 1 onset (a, c) or word 2 onset (b, d). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between noun subcategories and adjective subcategories (t-test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There were no significant differences between noun sub-categories or between adjective sub-categories.

e. Electrode location in left lateral orbitofrontal. **f-i.** Raster plots showing the responses in individual trials (see color scale on bottom right).

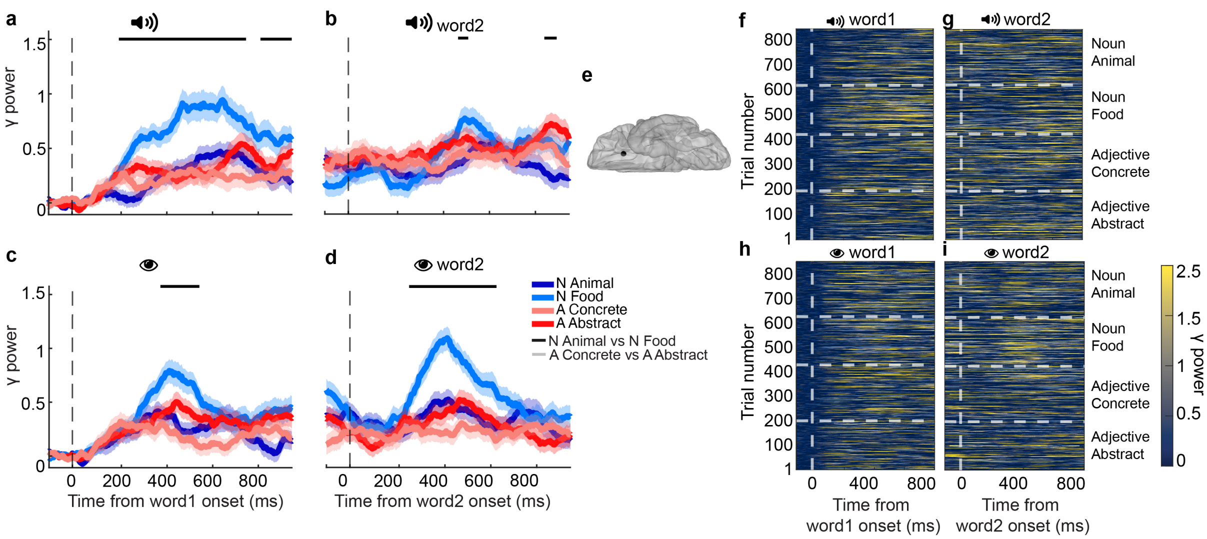


Figure 1.9:

Example electrode distinguishing different types of nouns.

a-d. Trial averaged γ-power of neural responses to words, separated by animal nouns (dark blue), food nouns (light blue), concrete adjectives (light red), and abstract adjectives (dark red). Neural responses are shown for auditory presentation (a, b), and visual presentation (c, d), aligned to word 1 onset (a, c) or word 2 onset (b, d). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between noun subcategories and adjective subcategories (t-test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There was a significant differences between noun sub-categories shown with a black horizontal line but no difference between adjective sub-categories.

e. Electrode location in left lateral orbitofrontal. **f-i.** Raster plots showing the responses in individual trials (see color scale on bottom right).

1.2.3 Neural signals enhanced for nouns versus adjectives were anatomically segregated

Of those electrodes uniquely selective for part of speech, 77% showed responses that were significantly stronger for nouns compared to adjectives ($\beta NvsA > 0$) as illustrated by the example in **Figure 1.4a-j**. The remaining 23% showed responses that were stronger for adjectives compared to nouns ($\beta NvsA < 0$) as illustrated by the example in **Figure 1.10 a-i** (**Table S5**). For auditory stimuli, the difference in the onset time between nouns and adjectives was larger for noun-preferring electrodes (550 ± 107 ms) than adjective-preferring electrodes (312 ± 94 ms, ranksum test, $p<0.05$). For visual stimuli, the difference in the onset time between nouns and adjectives was not different between noun-preferring electrodes (425 ± 107 ms) and adjective- preferring electrodes (437 ± 134 ms, ranksum test, $p>0.05$). There was a significant correlation between auditory and visual difference onset times for noun-preferring electrodes (Pearson $R^2 = 0.80$, $p<0.01$) but not for adjective preferring electrodes (Pearson $R^2 = -0.70$, $p>0.05$).

When I displayed the electrode locations on the brain, I observed an anatomical separation between these two groups of responses (**Figure 1.4 l,m**). I compared noun- versus adjective- preferring electrodes along 3 axes of Montreal Neurological Institute 305 Coordinates (MNI305, units abbreviated as m.u.(Bansal et al., 2014)). Along the lateral to medial axis (x-axis in **Figure 1.4 l,m**, zero being more medial), noun-preferring electrodes had a mean of 25.3±6.2 m.u. and adjective- preferring electrodes had a mean of 47.3±7.7 m.u. ($p<0.01$, ranksum test). Along the ventral- dorsal axis (z-axis in **Figure 1.4 l**), noun electrodes had a mean of -12.17±5.3 m.u. and adjective electrodes had a mean of -3.7±1.7 m.u. ($p<0.05$, ranksum test). Along the posterior-anterior axis (y-axis in **Figure 1.4 m**), noun electrodes had a mean of 21.4±18.9 m.u. and adjective electrodes had a mean of -2.7±25.8 m.u. ($p<0.05$, ranksum test). Table S6 summarizes the locations of noun- vs adjective- preferring electrodes across brain regions. A permutation test combining all brain regions for these electrodes showed that that electrodes in the LOF tended to show stronger responses to nouns (90% βN-vsA > 0, $p<10^{-4}$, permutation test, $n=10^6$ iterations, Methods).

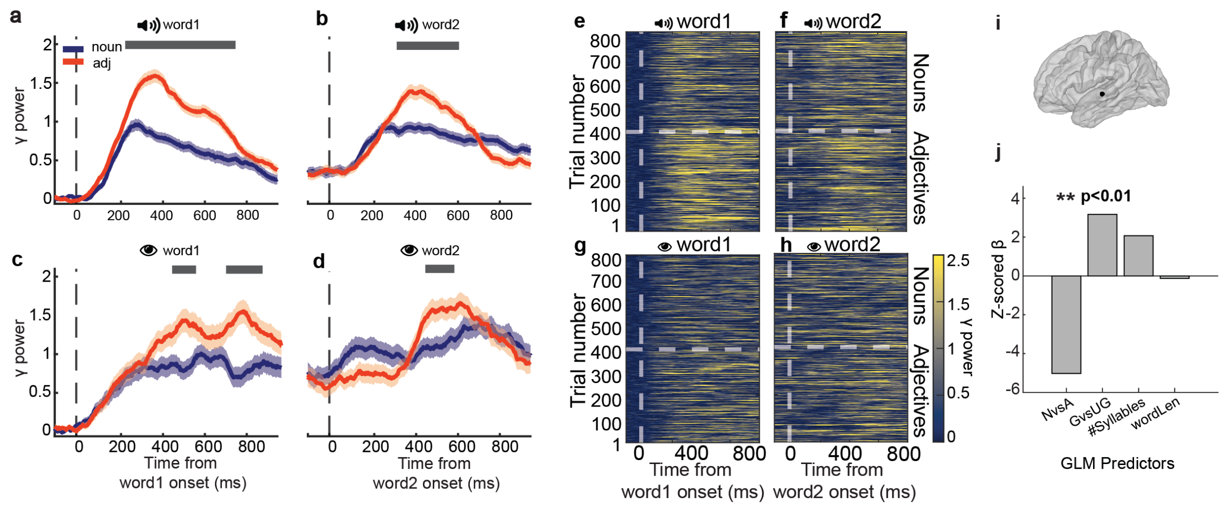


Figure 1.10:

Example electrode distinguishing nouns from adjectives with a preference for adjectives.

a-d.Trial averaged γ-power of neural responses to English words, separated by nouns (blue), and adjectives (red). Neural responses are shown for auditory presentation (a, b, , n=442 grammatical and 438 ungrammatical trials), and visual presentation (c, d, n=432 grammatical and 434 ungrammatical trials), aligned to word 1 onset (a, c) or word 2 onset (b, d). The vertical dashed lines show word onsets. Shaded areas represent s.e.m. Horizontal lines indicate time periods of statistically significant differences between nouns and adjectives (t-test, $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$).

(e-h) Raster plots showing the responses in individual trials (see color scale on bottom right).

i. Electrode location in the left superior temporal gyrus.

j-k. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word using four task predictors: Noun versus Adjectives, Grammatical versus Ungrammatical, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients. Only the Nouns vs Adjective task predictor was significant and showed a preference for adjectives ($p<0.01$, corrected for multiple comparisons and $\beta_{NvsA} < 0$)

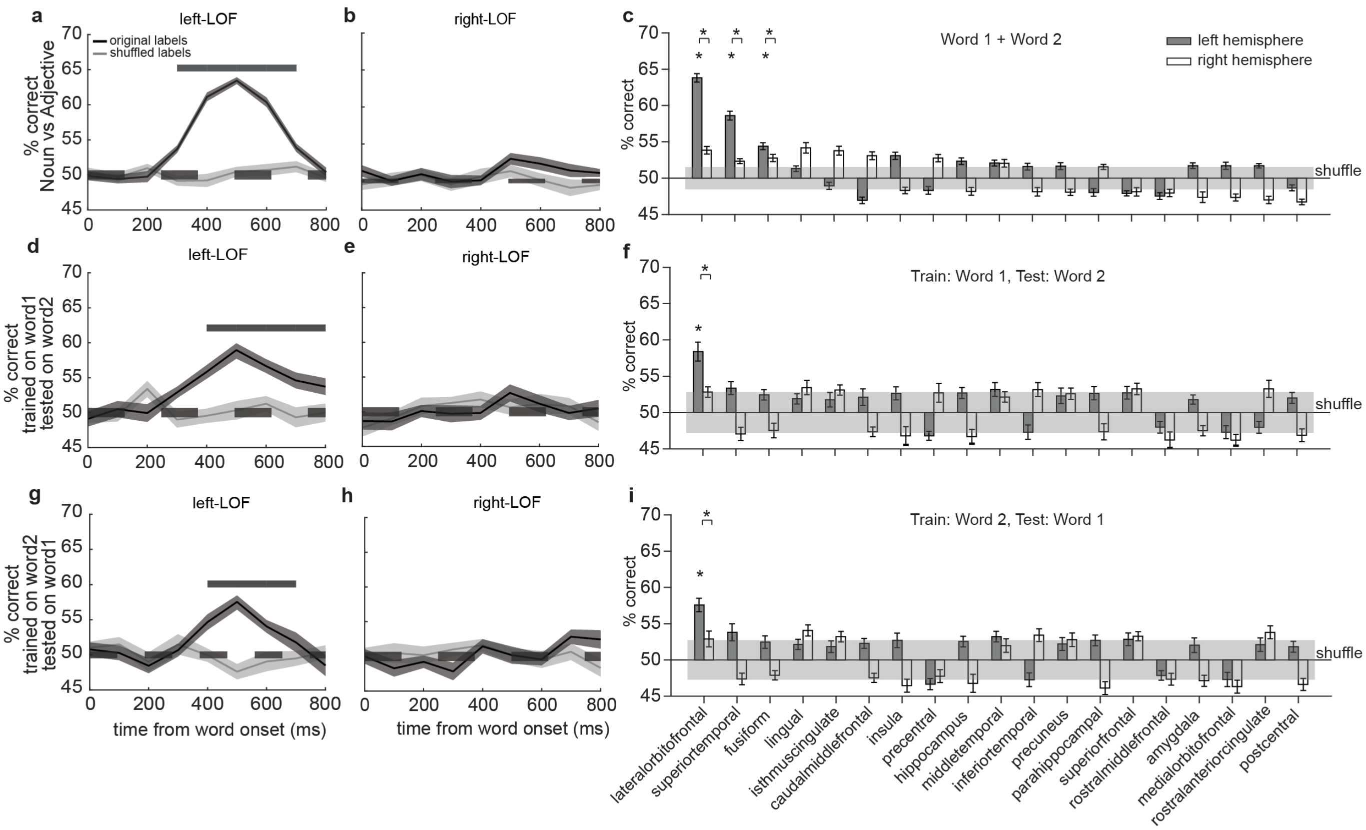
1.3 POS Decoding: Generalized and Robust

1.3.1 Electrodes in the left lateral orbitofrontal cortex can distinguish POS categories in individual trials and generalizes across words and modalities

To assess whether information about part of speech was available in individual trials, I used a machine learning pseudopopulation approach by combining electrodes within anatomically defined brain regions in the Desikan-Killiany Atlas. I binned the response in 100 ms time bins and used the top-N principal components that explained more than 70% of the variance in the training data for all the electrodes. I trained an SVM classifier with a linear kernel to distinguish between nouns and adjectives and tested the classifier on held-out data (**Methods**). **Figure 1.11** shows decoding accuracy for the left (**Figure 1.11 a,d,g**) and the right (**Figure 1.11 b,e,h**) LOF as a function of time from word onset. When trained using data from both word1 and word2 with combined auditory and visual features, there was a statistically significant decoding performance starting approximately at 300 ms after word onset and reaching a peak of 63.6±1.1% at 500 ms after word onset in the left LOF (**Figure 1.11 a**). Statistical significance was assessed by comparing with a control where noun and adjective labels were randomly shuffled (**Methods**). Even though there were almost twice as many electrodes in the right LOF compared to the left LOF (**Table S2**, **Figure 1.1 b-g**), decoding performance was much higher for the left LOF compared to the right LOF (compare **Figure 1.11 a** versus **Figure 1.11 b**). The differences between the left and right LOF persisted after randomly subsampling to equalize the number of electrodes across hemispheres for all regions (**Figure 1.12 a,b**).

In **Figure 1.11 a,b**, word 1 and word 2 are combined. Decoding performance in the left LOF was also high when separately considering word 1 (**Figure 1.13 a-c**) and word 2 (**Figure 1.13 d- f**). Furthermore, the machine learning classifier was able to generalize across words, as evidenced by the decoding performance when training on word 1 and testing on word 2 (**Figure 1.11 d,e**), and vice versa (**Figure 1.11 g,h**). Similarly, auditory and visual trials are combined in **Figure 1.11 a,b**. Decoding performance in the left LOF was also high when separately considering auditory trials (**Figure 1.13 g-i**) and visual trials (**Figure 1.13 j-l**). Furthermore, the machine learning classifier was able to generalize across modalities as evidenced by the decoding performance when training on auditory trials and testing on vision trials (**Figure 1.13 m-o**) and vice versa (**Figure 1.13 p-r**).

I extended the analyses in **Figure 1.11 a,b,d,e,g,h** to all other regions in the Desikan-Killiany atlas. In addition to the left LOF, the left superior temporal cortex and the left fusiform cortex also showed statistically significant decoding performance (**Figure 1.11 c**). However, in contrast to the results for the left LOF, the decoding results for other regions were less robust (**Figure 1.12 c**) and did not generalize across words (**Figure 1.11 f,i**) or across modalities (**Figure 1.13 o,r**).



a
left-LOF
original labels
shuffled labels
% correct
Noun vs Adjective
b
right-LOF
c
Word 1 + Word 2
left hemisphere
right hemisphere
% correct
shuffle
d
left-LOF
% correct
trained on word1
tested on word2
e
right-LOF
f
Train: Word 1, Test: Word 2
g
left-LOF
% correct
trained on word2
tested on word1
h
right-LOF
i
Train: Word 2, Test: Word 1
time from word onset (ms)
lateralorbitofrontal
superiortemporal
fusiform
lingual
isthmuscingulate
caudalmiddlefrontal
insula
precentral
hippocampus
middletemporal
inferiortemporal
precuneus
parahippocampal
superiorfrontal
rostralmiddlefrontal
amygdala
medialorbitofrontal
rostralanteriorcingulate
postcentral

Figure 1.11: Neural signals from left-LOF distinguish nouns and adjectives when number of electrodes were normalized across all regions and both hemispheres.
a, b, d, e, g, h. Average cross-validated performance of a support vector machine classifier (SVM, 80% training/20% test) decoding nouns versus adjectives for all electrodes in the left lateral orbitofrontal cortex (LOF) (a, d, g) or the right LOF (b, e, h). The dotted horizontal black line shows the chance level. Shaded areas denote s.e.m. Solid horizontal black bar shows time points where performance significantly differed from chance (100 random shuffles, ranksum test, $p<0.01$). The inputs to the SVM included the top-N principal components of the electrode response that explained >70% variance for the training data at each time bin (Methods). a, b: Features from auditory and visual responses were combined and used for training and testing on a dataset of both Word1 and Word2 trials. c, d: Generalization across word order was evaluated on a dataset where Word1 trials were used for training and word2 trials were used for testing. g, h: Training on Word2 and testing on Word1. Black: original labels; Gray: shuffled labels. (see Figure S9 for decoding performance when the number of electrodes was same across all regions and both hemispheres)
c, f, i. Summary of average of max-decoding performance for distinguishing nouns versus adjectives in each hemisphere (dark: left; white: right) for different brain regions. Bottom asterisks denote regions with significant decoding performance with respect to chance and performance from the real and null distribution do not overlap within 3 standard deviations of each other ($p<0.01$, ranksum test, corrected for multiple comparisons, Methods). Shaded box: maximum of the mean ± SD. for the null distribution across all regions. Top asterisks with a U-bracket denote significant differences between decoding accuracy of the left versus the right hemisphere ($p<0.01$, ranksum test, corrected for multiple comparisons). Regions are sorted in descending order of performance in panel c. c: Classifiers were trained and tested with features from both Word1 and Word2 trials. f: Classifiers were trained on Word1 trials and tested on Word2 trials. i: Classifiers were trained on Word2 trials and tested on Word1 trials. (see Figure S10 for controls on word1-only, word2-only, audio-only, visual-only, audio-to- vision and vision-to-audio performance.)c.

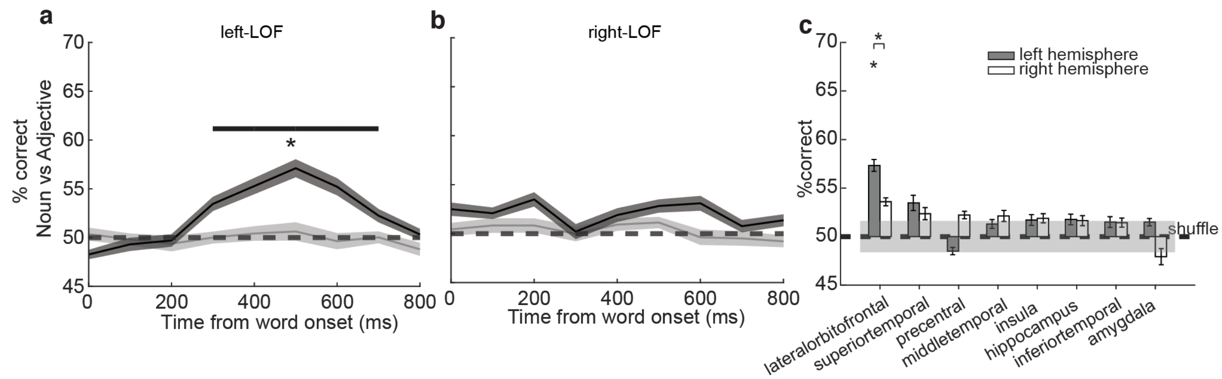
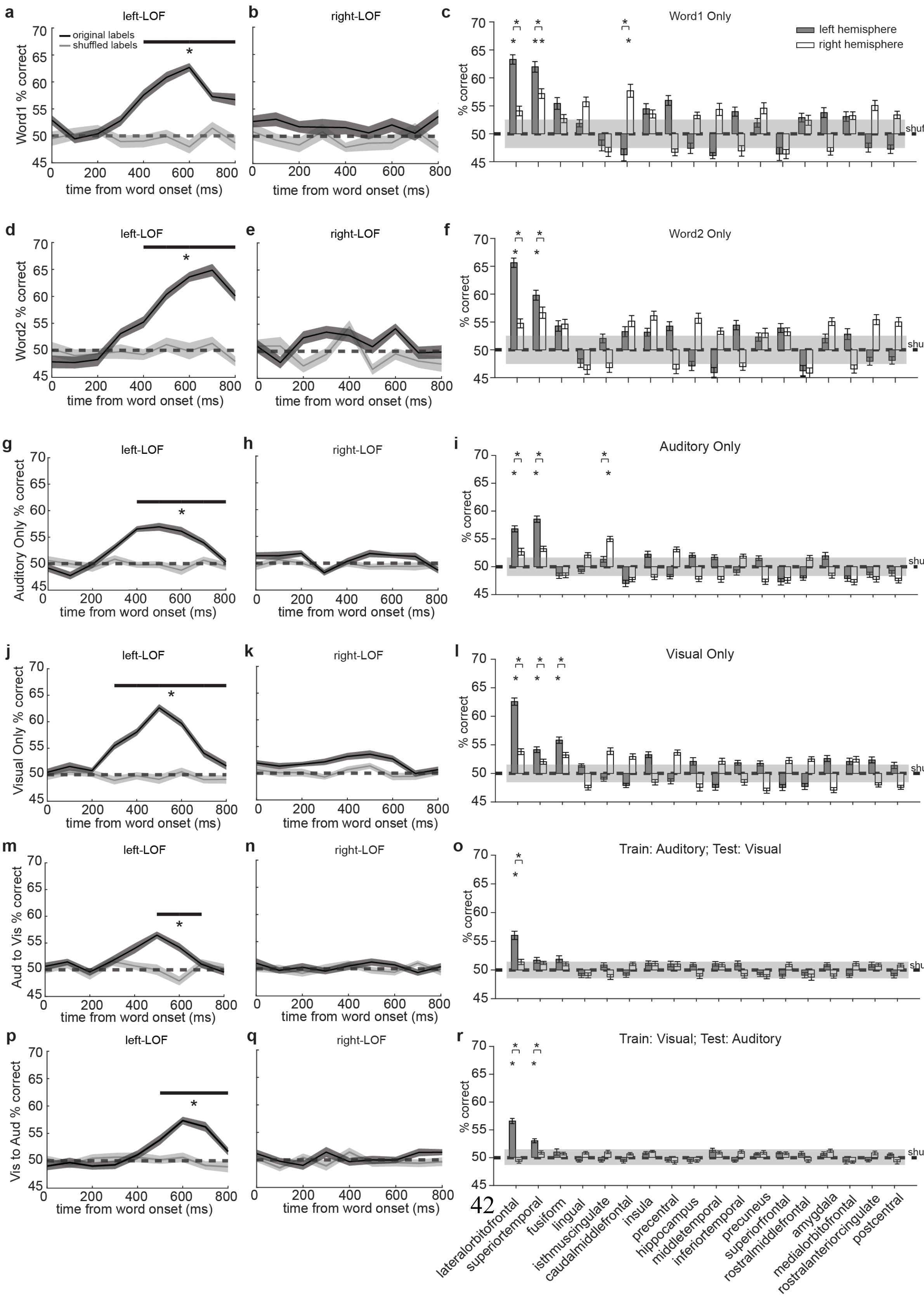


Figure 1.12: Neural signals from left-LOF distinguish nouns and adjectives when number of electrodes were normalized across all regions and both hemispheres.
a-b. Average cross-validated performance of a support vector machine classifier (SVM, 80% training/20% test) decoding nouns versus adjectives for 8 randomly subsampled electrodes in the left lateral orbitofrontal cortex (LOF) (a), and in the right LOF (b). Black: original labels; Gray: shuffled labels. The dotted horizontal black line shows the chance level. Solid horizontal gray bar shows time points where decoding from correct labels significantly differed from that of shuffled labels (100 random shuffles of the data, ranksum test, $p<0.01$). The inputs to the SVM were 100 ms time bins from word onset containing the top-N principal components of the electrode response at each bin that explained >70% variance for the training data (Methods).
(c). Summary of average of max-decoding performance for distinguishing nouns versus adjectives across both hemispheres (left hemisphere: dark gray bars; right hemisphere: white bars) for different brain regions when a total of 8 electrodes was taken from each hemisphere in each region for the decoding. Regions with less than 8 electrodes in either hemisphere were omitted. Asterisk: significant hemisphere within a Desikan-Killiani defined brain region ($p<0.01$, ranksum test, corrected for multiple comparisons, and performance from the real and null distribution do not overlap within 3 standard deviations of each other) (Methods). Gray box: maximum mean ± s.t.d. for the null distribution across all regions. Asterisk with a U-bracket: significant difference between decoding accuracy of the left versus the right hemisphere ($p<0.01$, ranksum test, corrected for multiple comparisons). Regions are sorted in descending order of performance in panel c.

Figure 1.13: Neural signals distinguish nouns and adjectives in single trials for word1-only, word2-only, audio-only features, vision-only features, and generalization from audio to vision or vice versa.

a, b, d, e, g, h, m, n, p, q. Average cross-validated performance of a support vector machine classifier (SVM, 80% training/20% test) decoding nouns versus adjectives for all electrodes in the left lateral orbitofrontal cortex (LOF) (a,d,g,j,m,p), and in the right LOF (b,e,h,k,n,q). The dotted horizontal black line shows the chance level. Shaded areas denote s.e.m. Solid horizontal black bar shows time points where performance significantly differed from chance (100 random shuffles, ranksum test, p<0.01). The inputs to the SVM included the top-N principal components of the electrode response that explained >70% variance for the training data at each time bin (Methods). Features from auditory and visual responses were combined and used for training and testing on datasets of word1 (a,b) and word2 trials (d,e). Using a combined dataset of word1 and word2 trials, the decoding performance was evaluated for audio-only (g,h) and vision-only features (j,k). The decoding performance generalized for audio to vision (m,n) and vice versa (p,q).
(c). Summary of average of max-decoding performance for distinguishing nouns versus adjectives across both hemispheres (left hemisphere: dark gray bars; right hemisphere: white bars) for different brain regions when a total of 8 electrodes was taken from each hemisphere in each region for the decoding. Regions with less than 8 electrodes in either hemisphere were omitted. Asterisk: significant hemisphere within a Desikan-Killiani defined brain region (p<0.01, ranksum test, corrected for multiple comparisons, and performance from the real and null distribution do not overlap within 3 standard deviations of each other) (Methods). Gray box: maximum mean ± s.t.d. for the null distribution across all regions. Asterisk with a U-bracket: significant difference between decoding accuracy of the left versus the right hemisphere (p<0.01, ranksum test, corrected for multiple comparisons). Regions are sorted in descending order of performance in panel c.



a
left-LOF
original labels
shuffled labels
Word1 % correct
time from word onset (ms)
b
right-LOF
c
Word1 Only
left hemisphere
right hemisphere
% correct
shuffle
d
Word2 % correct
e
f
Word2 Only
g
Auditory Only % correct
h
i
Auditory Only
j
Visual Only % correct
k
l
Visual Only
m
Aud to Vis % correct
n
o
Train: Auditory; Test: Visual
p
Vis to Aud % correct
q
r
Train: Visual; Test: Auditory
lateralorbitofrontal
superiortemporal
fusiform
lingual
isthmuscingulate
caudalmiddlefrontal
insula
precentral
hippocampus
middletemporal
inferiortemporal
precuneus
superiorfrontal
rostralmiddlefrontal
amygdala
medialorbitofrontal
rostralanteriorcingulate
postcentral

1.3.2 Multimodal neural signals distinguishing different parts of speech are conserved across languages

One of the participants was fluent in two languages, English and Spanish. Therefore, this patient provided an opportunity to ask whether the neural signals discriminating between different parts of speech were language-specific or showed invariance across languages. All the words were translated into Spanish by a native Spanish speaker and the task was repeated in both languages. **Figure 1.14 a-h** shows the responses of an example electrode located in the left LOF (**Figure 1.14 k**). This electrode showed a stronger response to nouns compared to adjectives for auditory stimuli (**Figure 1.14 a, b, e, f**), for visual stimuli (**Figure 1.14 c, d, g, h**), for Word 1 (**Figure 1.14 a, c, e, g**), and for Word 2 (**Figure 1.14 b, d, f, h**). Interestingly, the separation between nouns and adjectives was evident both when the words were presented in English (Figure 4a-d) and when the words were presented in Spanish (**Figure 1.14 e-h**). The GLM analysis showed that nouns versus adjectives was the only significant predictor in English trials (**Figure 1.14 i**), and Spanish trials (**Figure 1.14 j**). All in all, there were three electrodes in this participant that showed a multimodal response selective for part of speech. All three of these electrodes were in the left orbital H-shaped sulcus within the LOF (**Figure 1.14 k**, green).

In addition to this bilingual participant, the task was run in monolingual participants who spoke English (n=16 participants) and monolingual participants who spoke Taiwanese (n=3 participants, **Table S1**). In **Figure 1.14 k**, I show all electrodes from the left LOF that showed part-of-speech encoding from different participants (**Table**

S7). I also indicate the language in which this difference was observed whether it be English (pink), Taiwanese (brown) or bilingual English/Spanish (green). All participants in **Figure 1.14 k** were right-handed. Electrodes separating parts of speech from monolingual participants were also clustered in the same region. Thus, the left LOF distinguished between parts of speech for both auditory and visual presentations of stimuli across participants speaking different languages. The proportion of POS-selective electrodes in the left LOF in my study is consistent with previous studies analyzing selective electrodes for visual object processing in the human fusiform and inferotemporal cortex.

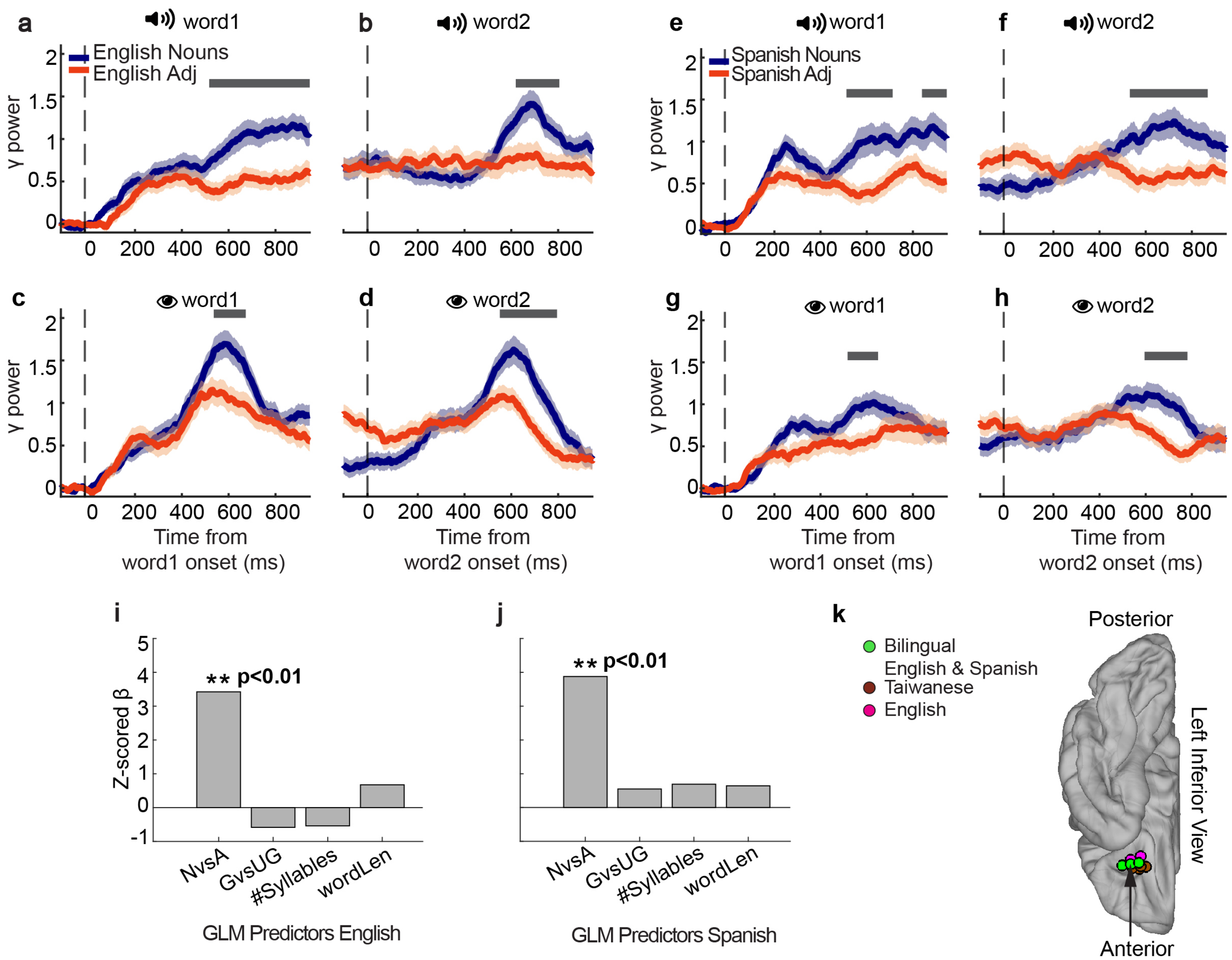

a
word1
English Nouns
English Adj
γ power
b
word2
c
word1
d
word2
Time from word1 onset (ms)
Time from word2 onset (ms)
e
word1
Spanish Nouns
Spanish Adj
f
word2
g
word1
h
word2
i
Z-scored β
** p<0.01
NvsA
GvsUG
#Syllables
wordLen
GLM Predictors English
j
** p<0.01
GLM Predictors Spanish
k
Bilingual English & Spanish
Taiwanese
English
Posterior
Anterior
Left Inferior View

Figure 1.14: Neural signals in left LOF generalize across languages in a bilingual subject and in monolingual subjects.

a-h. Trial averaged responses of an electrode in the left lateral orbitofrontal cortex from a bilingual patient. The format follows Fig. 2a-d. (a-d) English words (audio: n=190 grammatical and 185 ungrammatical trials; vision: n=189 grammatical and 191 ungrammatical trials). (e- h) Spanish words (audio: n=184 grammatical and ungrammatical trials; vision: 184 grammatical and 186 ungrammatical trials). Auditory responses (a, b, e, f) . Visual responses (c, d, g, h) . Word 1 (a, c, e, g) and Word 2 (b, d, f, h) .
i,j. Z-scored β coefficients for Generalized Linear Model to predict area under the curve (AUC) for the English experiment (i) and for the Spanish experiment (j). The AUC computed between 200 ms and 800 ms post word onset using four task predictors: Noun versus Adjectives, Grammatical versus Ungrammatical, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients corrected for multiple comparisons (Methods). The word order for grammatically correct trials in English is an adjective followed by a noun, such as “green apple”. This word order gets flipped in grammatically correct Spanish trials.
k. Inferior view of all the 9 out of 38 electrodes (8 audiovisual: Figure 2k, 1 visual-only: Figure S4, see Table S4 and S7) in the left lateral orbitofrontal cortex that showed noun versus adjective differences across different languages in which the experiment was conducted (significant Nouns versus Adjectives β, $p<0.01$ corrected for multiple comparisons). These electrodes come from 4 different subjects. Electrodes from the bilingual patient are in green with a black arrow indicating the example electrode. Electrodes from one monolingual English patient are in pink and those from 2 monolingual Taiwanese patients are in brown.

1.3.3 Electrodes distinguish parts-of-speech for nouns and adjectives matched for their frequency of occurrence in a bilingual participant.

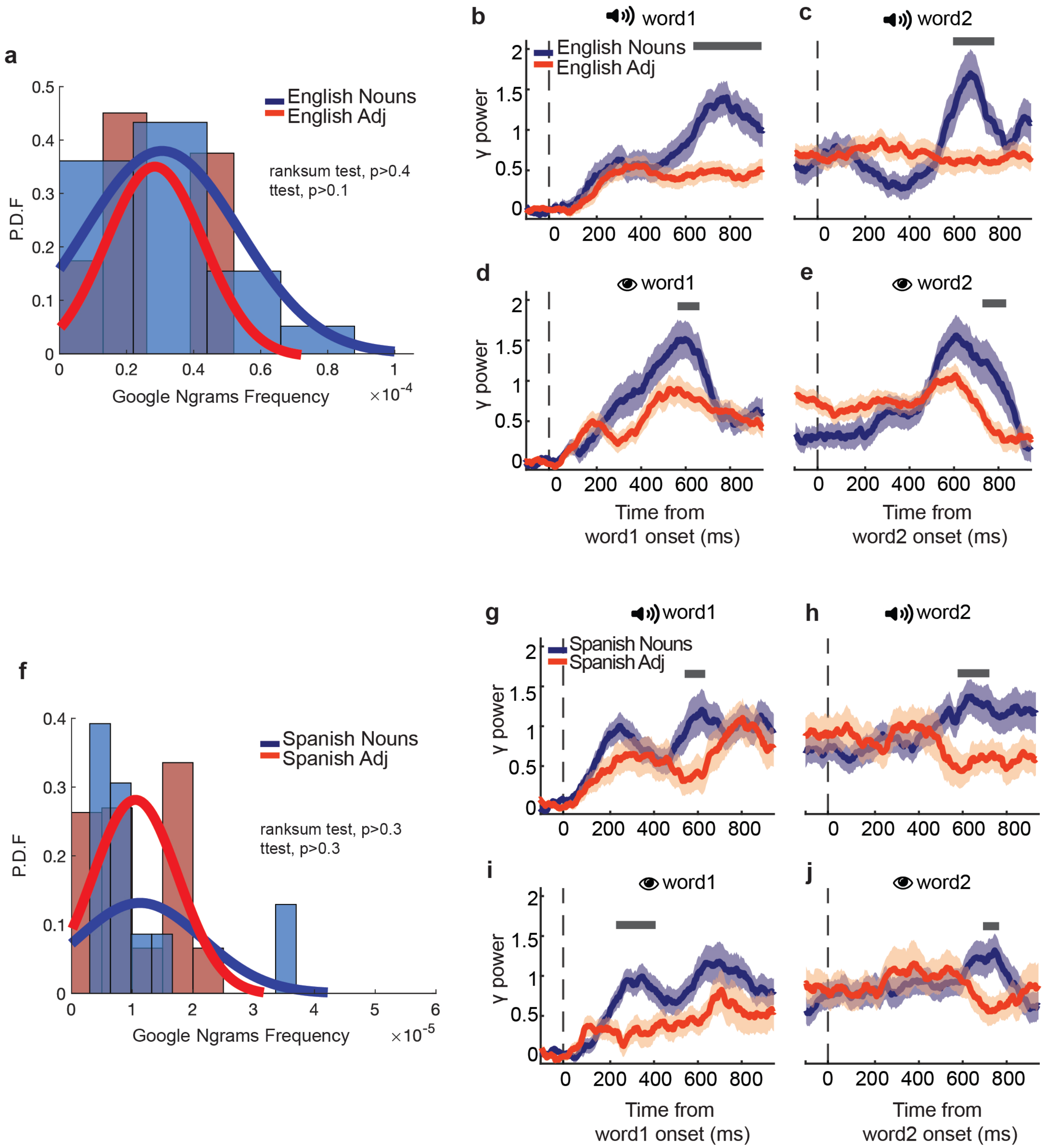

a
English Nouns
English Adj
ranksum test, p>0.4
ttest, p>0.1
P.D.F
Google Ngrams Frequency
×10-4
b
word1
English Nouns
English Adj
γ power
c
word2
d
word1
γ power
e
word2
Time from
word1 onset (ms)
Time from
word2 onset (ms)
f
Spanish Nouns
Spanish Adj
ranksum test, p>0.3
ttest, p>0.3
P.D.F
Google Ngrams Frequency
×10-5
g
word1
Spanish Nouns
Spanish Adj
γ power
h
word2
i
word1
γ power
j
word2
Time from
word1 onset (ms)
Time from
word2 onset (ms)

Figure 1.15: Example electrode distinguishes parts-of-speech for nouns and adjectives matched for their frequency of occurrence in a bilingual participant.
a-e Same format as **Figure 1.7 a-e** for English phrases.
a. Sub-sampled distribution of English with nouns and adjectives matched for their frequency of occurrence (both mean and median were matched with ttest and ranksum test, respectively).
b-e., Trial-averaged gamma power for auditory phrases (b,c) and visual phrases (d,e). Gray bars indicate time periods of significant differences between nouns and adjectives. (Methods)
f-j Same format as above for Spanish phrases.

1.4 POS in Full Sentences

1.4.1 Multimodal neural signals distinguishing nouns and verbs in sentences

The experiment presented thus far concerned the responses to nouns and adjectives within minimal phrases. I extended these results in two ways: (1) by evaluating whether there are multimodal signals that distinguish between nouns and verbs; (2) by evaluating the neural signals to words embedded within full sentences. I recorded intracranial field potentials from 1,563 electrodes (844 in gray matter, 719 in white matter) implanted in 17 patients via stereoelectroencephalography. Participants heard (auditory modality) or read (visual modality) four-word sentences that were sequentially presented (**Figure 1.17 a**, **Methods**). To assess comprehension, participants were asked to indicate whether the sentence adequately described an image that followed the last word after a 1,000 ms interval. Participants performed the task correctly on 85.7±14.3% of the trials. I considered two types of sentences, semantic (e.g., "the girls ate cakes") or non-semantic (e.g., "the cakes ate girls"). All electrode locations are shown in **Figure 1.16** (see also **Table S11**, **Methods**).

Following the procedures described in the analyses of neural responses to nouns versus adjectives, I evaluated whether neural signals differentiated between nouns and verbs. **Figure 1.17** shows the responses of an example electrode located in the pars triangularis (**1.17 f** denotes the electrode location). The neural responses are aligned to word onset for auditory presentation (**Figure 1.17 b**) or visual presentation (**Figure 1.17 c**). The responses to nouns (blue) were stronger than verbs (black) for auditory and visual stimuli. The differences between nouns and verbs can be readily appreciated even in individual trials (**Figure 1.17 d,e**). These differences became significant at approximately 140 ms after word onset for auditory presentation and about 320 ms for visual presentation. In all, there were 121 electrodes that showed selective responses distinguishing nouns from verbs both for auditory and visual presentation.

Even though I tested these electrodes for nouns versus verbs differences, it is possible that auditory features (like number of syllables) or orthographic features (like word length) could contribute to the neural responses. Further, each sentence could either be semantic (S: “the girls ate cakes”) or not (NS: “the cakes ate girls). To address evaluate whether word features and semantic features contributed to the neural signals underlying parts of speech, I built a GLM for each electrode to predicts its response AUC between 200 ms and 800 ms after word onset using four predictors: nouns versus verbs, semantic or not, number of syllables, and word length (**Methods**). The predictor coefficients in the GLM model for the example electrode in **Figure 1.17 b-e** show that only the nouns versus verbs label significantly explained the neural responses (**Figure 1.17 g**). A total of 41 audiovisual electrodes showed nouns versus verbs as the only statistically significant predictor in the GLM analysis, such as the example electrode **Figure 1.17 b-g**. The locations of these electrodes are shown in **Figure 1.17 h,i**. The electrode locations reveal two clusters enriched in the left pars triangularis and precentral regions. The difference in the number of significant electrodes between the right and left hemispheres was statistically significant: $p<10^{-4}$, permutation test, $n=10^6$ iterations, see **Table S12**).

Many electrodes (63%) showed responses that were significantly stronger for nouns compared to adjectives ($\beta NvsV > 0$), as illustrated by the example in **Figure 1.17 b-e**. I observed an anatomical separation between these two groups of responses (**Figure 1.17 j,k**). I compared noun- versus verb- preferring electrodes along 3 axes of Montreal Neurological Institute 305 Coordinates (MNI305, units abbreviated as m.u. (Quian Quiroga et al., 2005)). Along the anterior-posterior axis (y-axis in **Figure 1.17 j,k**), noun electrodes had a mean of -11.8±21.8 m.u. and verb electrodes had a mean of 10.1±26.4 m.u. ($p<0.01$, ranksum test). Along the ventral-dorsal axis (z-axis in **Figure 1.17 j**), noun electrodes had a mean of -5.1±29.5 m.u. and verb electrodes had a mean of 15.9±27.0 m.u. ($p<0.05$, ranksum test). Along the lateral to medial axis (x-axis in **Figure 1.17 k**, zero being more medial), noun-preferring electrodes had a mean of 38.3±16 m.u. and adjective-preferring electrodes had a mean of 40.3±11.6 m.u. (not significant, $p>0.05$, ranksum test).

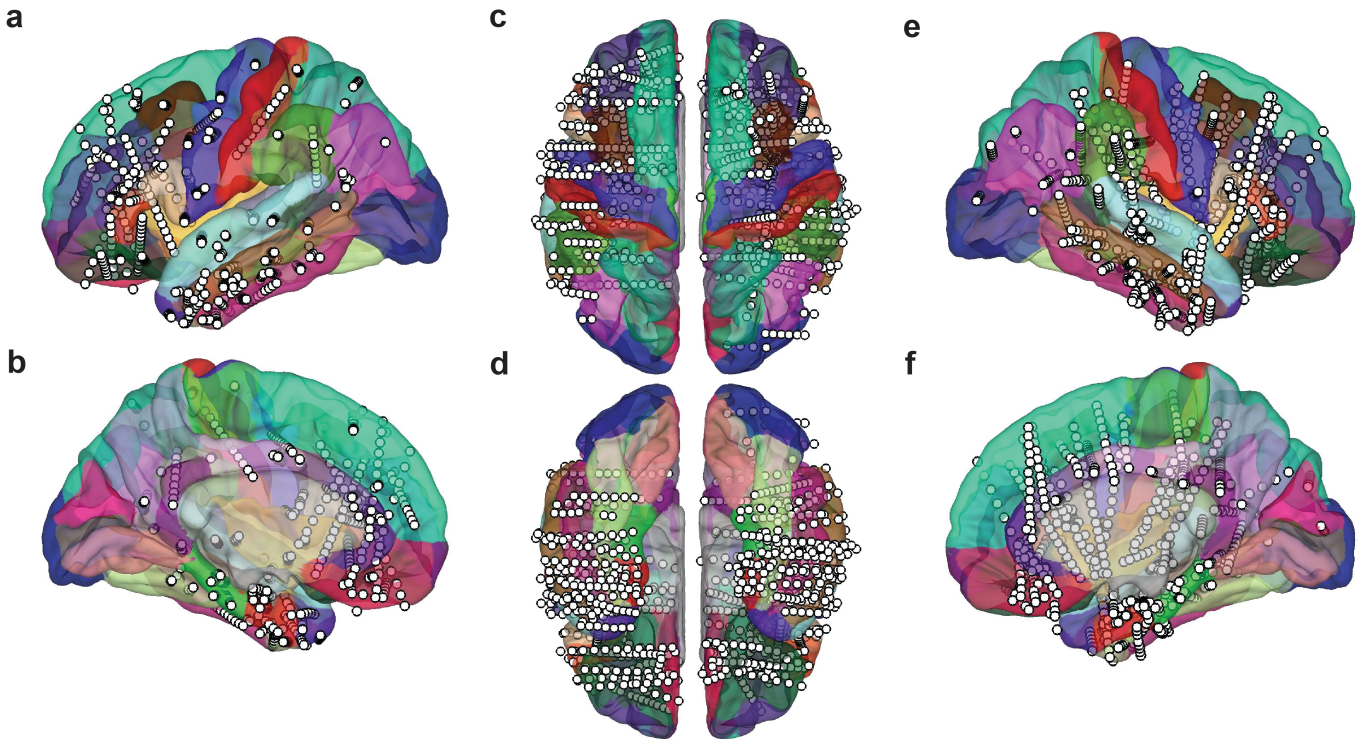
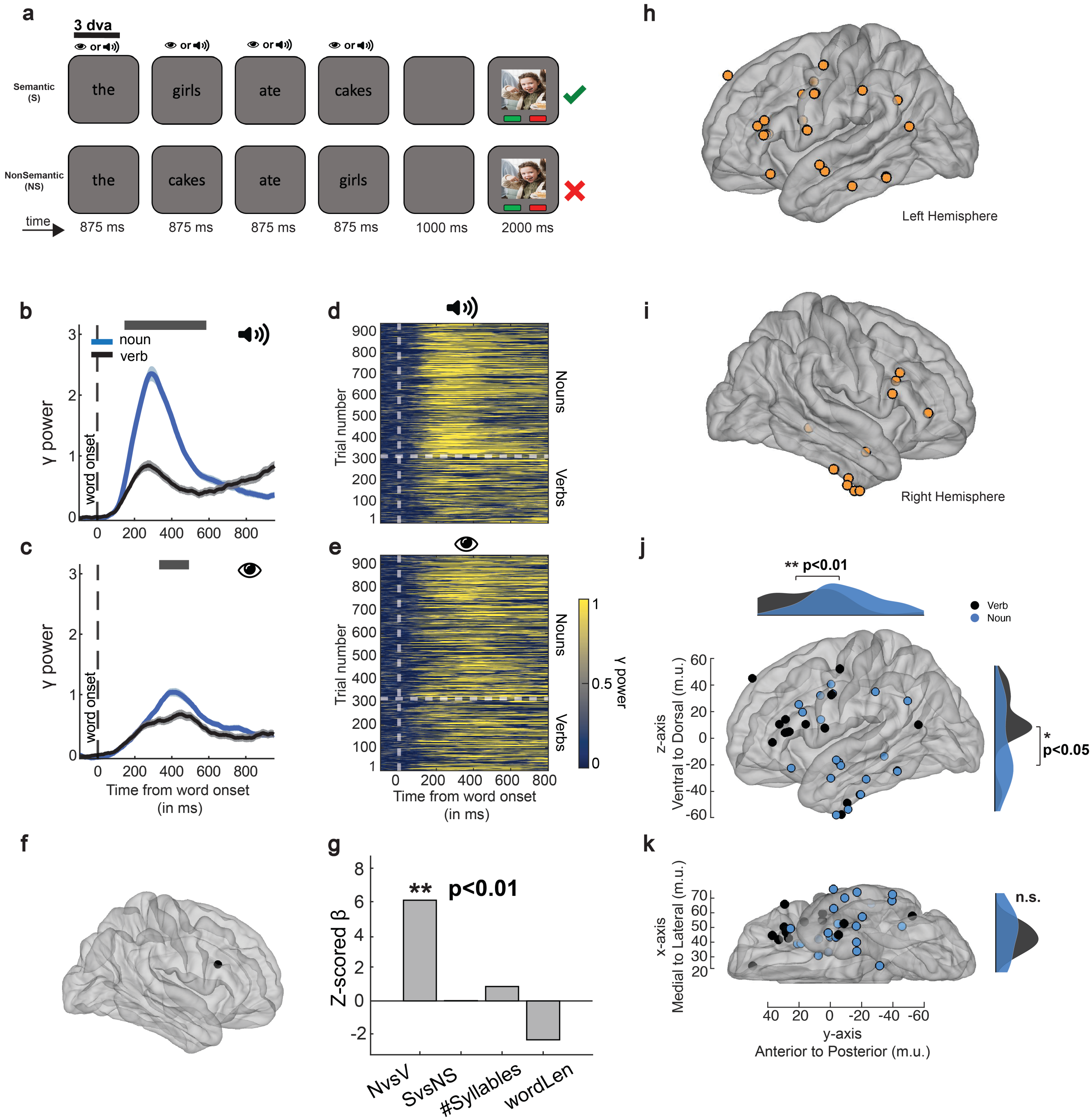


Figure 1.16: Electrode locations for sentence task.
a-f. Location of all electrodes overlayed on the Desikan-Killiany Atlas shown with different views. Each white circle shows one electrode. a. Left lateral view (n=760), b. Left medial view (n=760), c. Superior, whole brain view (n=1593), d. Inferior, whole brain view (n= 1593), e. Right lateral view (n=833) f. Right medial view (n=833).



a
3 dva
Semantic (S)
the
girls
ate
cakes
NonSemantic (NS)
the
cakes
ate
girls
time
875 ms
875 ms
875 ms
875 ms
1000 ms
2000 ms
h
Left Hemisphere
b
noun
verb
γ power
word onset
d
Trial number
Nouns
Verbs
i
Right Hemisphere
c
γ power
word onset
Time from word onset (in ms)
e
Trial number
Nouns
Verbs
γ power
Time from word onset (in ms)
j
** p<0.01
Verb
Noun
z-axis
Ventral to Dorsal (m.u.)
* p<0.05
f
g
** p<0.01
Z-scored β
NvsV
SvsNS
#Syllables
wordLen
k
x-axis
Medial to Lateral (m.u.)
n.s.
y-axis
Anterior to Posterior (m.u.)

Figure 1.17: Neural signals distinguish between different parts of speech in sentences.

a. Task Schematic. Sentences comprising four words sequentially presented either in visual or auditory modality were followed by an image. The sentences were either semantic (50% S sentences, e.g., “the girls ate cakes”) or non-semantic (50% NS sentences, e.g., “the cakes ate girls”). Participant were instructed to indicate via a button press whether the sentence described the image accurately or not (Methods).

b,c. Trial-averaged normalized gamma-band power of responses from an example electrode in the pars triangularis (see eletrode location in f) to nouns (blue) or verbs (black) during presentation of auditory stimuli (b, n=628 nouns and 314 verbs) or visual stimuli (c, n=628 nouns and 314 verbs) aligned to word onset (vertical dashed line). Shaded areas denote s.e.m. Horizontal gray lines denote windows of statistically significant differences between responses to nouns versus verb (t-test $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$).

d,e. Raster plots showing the responses in each individual trial (see color scale on bottom right). The blue and black curves in b,c correspond to the averages of noun and verb trials, respectively, in d,e.

f. Location of the example electrode in the pars triangularis.

g. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset, using four task predictors: Noun versus Verbs, Semantically correct versus incorrect, number of syllables (auditory presentation) and word length (visual presentation). Asterisks denote statistically significant coefficients, corrected for multiple comparisons (Methods).

h,i. Lateral view of left (h) and right (i) hemispheres showing electrodes that revealed statistically significant differences between nouns and verbs for both audio and visual presentation (orange circles, n=41 electrodes, 27 left). Electrodes whose responses were significantly explained only by the Nouns versus Verbs task predictor in the GLM are included in this plot.

j,k. All electrodes from h,i projected onto the left hemisphere are shown on the lateral plane (j) and the axial plane (k). All the electrodes that respond more strongly to nouns, i.e., Nouns versus Verbs $\beta>0$ (n=23 electrodes), are shown in blue and electrodes that responded more strongly to verbs ($\beta<0$, n=18 electrodes), are shown in black. All units are in MNI305 coordinates. Kernel density curves (bandwidth 2) outline the marginal distributions of noun-preferring (blue) and verb-preferring (black) electrodes along the anterior-posterior axis (j,k: y-axis), ventral-dorsal axis (j: left z-axis) and lateral- medial axis (k: left x-axis, zero being more medial). P-values indicate significant differences between the coordinates for noun- and verb-preferring electrodes (ranksum test).

1.5 Grammatical Versus Ungrammatical Phrases

1.5.1 Multimodal Neural Signals Distinguish Grammatical and Ungrammatical in Minimal Phrases

I evaluated whether the neural signals differentiated between grammatical and ungrammatical trials. **Figure 1.18** shows the responses of an example electrode located in the postcentral cortex (**Figure 1.18 i**). The responses of this electrode are aligned to the word onset (vertical dashed line) for auditory presentation (**Figure 1.18 a,b**), and visual presentation (**Figure 1.18 c,d**), for the first word in each trial (**Figure 1.18 a,c**), or the second word in each trial (**Figure 1.18 b,d**). This electrode showed multimodal responses triggered by both auditory and visual stimuli. There was a stronger response to grammatical trials (word1-red and word2-blue) compared to ungrammatical trials (word1-blue and word2-red) for all four conditions for both visual and auditory stimuli (horizontal gray bars denote periods with a statistically significant difference between nouns and adjectives). The differences between grammatical and ungrammatical trials for this electrode became significant at approximately 530 ms after word onset for visual presentation and at about 565 ms for auditory presentation. **Figure 1.18 e-h** shows the raster plot with individual trials for auditory presentation in the top row and for visual presentation in the bottom row.

Even though each this electrode could separate grammatically correct and incorrect trials, I asked whether trial-to-trial level responses could also be modulated by parts-of-speech (nouns vs adjectives), number of syllables in the word in auditory stimuli or word length for visual stimuli. To address this, I built a GLM similar to **Figure 1.4 j** (Methods). The predictor coefficients for the example electrode in **Figure 1.18 a-d** show that (**Figure 1.18 j**) only the grammatically correct versus ungrammatical label significantly explained the neural responses. A total of 16 electrodes showed grammatically correct versus ungrammatical as the only statistically significant predictor in the GLM analysis. 5 of these 16 electrodes distinguished grammatically correct versus ungrammatical trials for both visual and auditory inputs, such as the example electrode shown in **Figure 1.18 a-d** (10.42% of the multimodal electrodes). The locations of these electrodes are shown in orange in **Figure 1.18 k,l**.

As a null hypothesis, I had assumed that distinguishing grammatically correct phrases from ungrammatical phrases constitutes a core component of language and would therefore be reflected in both modalities. However, this was not the case. Differential responses between nouns and adjectives were observed in 8 electrodes only during visual presentation (50% of the grammar separating electrodes, their locations are shown in black in **Figure 1.18 k,l**) and 3 electrodes only during auditory presentation (18.75% of the grammar separating, their locations are shown in white in **Figure 1.18 k,l**).

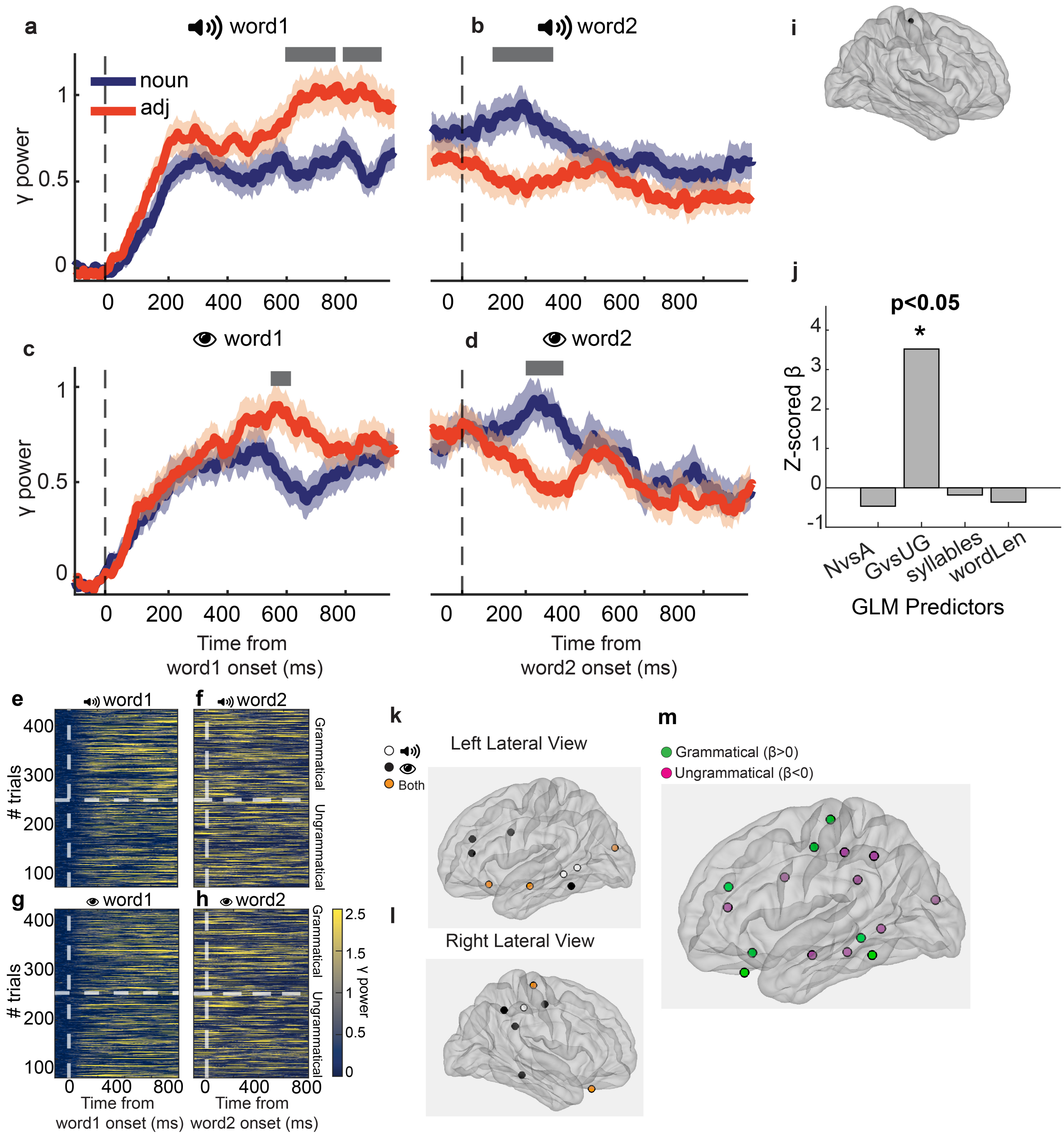

a
word1
noun
adj
γ power
b
word2
i
c
word1
d
word2
Time from
word1 onset (ms)
Time from
word2 onset (ms)
j
p<0.05
Z-scored β
NvsA
GvsUG
syllables
wordLen
GLM Predictors
e
word1
f
word2
trials
Grammatical
Ungrammatical
g
word1
h
word2
γ power
Time from
word1 onset (ms)
Time from
word2 onset (ms)
k
Both
Left Lateral View
l
Right Lateral View
m
Grammatical (β>0)
Ungrammatical (β<0)

Figure 1.18: Neural signals distinguish correct from incorrect syntax.
a-d. Normalized gamma-band power of responses from an example electrode in the right posterior parietal cortex (i). The format is the same as Fig. 2 a-d (audio: n=214 grammatical and 224 ungrammatical trials; vision: 219 grammatical and 221 ungrammatical trials).
e-h. Raster plots showing the responses in each individual trial (see color scale on bottom right).
i. Location of the example electrode in the right precentral gyrus.
j. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset. (Methods)
k,l. Lateral view of electrodes in the left [l, n=9 electrodes] or right [m, n=7 electrodes] hemisphere that showed significant activation for both words during audio presentation only (white circles , nLeft = 2, nRight = 1 electrodes). visual presentation only (black circles, nLeft = 4, nRight = 4 electrodes) or both audio and visual presentations (orange circles, nLeft = 3, nRight = 3 electrodes). Only electrodes whose responses were significantly explained by the Grammatical vs Ungrammatical label in the GLM are included in this plot. (see **Table S13** for distribution of electrodes across regions)
m. Distribution of electrodes with signals enhanced either by grammatically correct (green) or incorrect (magenta) phrases, projected on to the left lateral view.
(see **Table S13**)

1.6 Supplementary Tables

Subject	Age	Gender	Language	Handedness	#Trials	%Correct	#Electrodes
1	9	M	EN	R	1178	98.5	142
2	14	F	EN	R	1332	94.7	100
3	22	M	EN	R	1520	97.5	212
4	49	F	EN	L	760	83.3	44
5	18	F	EN	R	3573	99.6	100
6	20	M	EN	R	760	99.3	139
					760	67.5	
7	16	F	EN	L	760	97.9	88
8	12	F	EN	R	760	92.9	131
9	37	F	EN	R	1900	89.3	51
10	47	F	EN	R	1895	98.3	75
11	12	M	EN	R	950	95.5	135
12	13	F	EN	R	950	99.3	121
13	25	M	EN	L write, R throw	1520	99.3	84
14	26	F	EN	L write, R other	1900	97.4	32
15	32	F	EN	R	1521	98.4	29
16	22	M	EN & SP	R	EN:950 SP:950	97.3	77
17	42	M	TW	R	1068	89	57
18	36	M	TW	R	2429	87.1	59
19	53	F	EN	R	950	90.5	73
20	44	M	TW	R	1900	NA	52
						TOTAL	**1801**

Table S1 | Information about each participant including age, gender, language (ENglish, SPanish, TaiWanese), handedness, number of trials, behavioral performance and number of electrodes.

Region\nElecs	Total	GMLeft	WMLeft	Left	GMRight	WMRight	Right	rAud	rVis	rAV
Amygdala	52	21	0	21	31	0	31	9	7	1
Cerebellum-Cortex	4	3	0	3	1	0	1	2	1	1
Hippocampus	59	33	0	33	26	0	26	15	14	5
Inf-Lat-Vent	1	1	0	1	0	0	0	0	1	0
Lateral-Ventricle	4	0	0	0	4	0	4	0	1	0
Putamen	2	1	0	1	1	0	1	1	0	0
VentralDC	1	1	0	1	0	0	0	0	0	0
bankssts	24	2	4	6	9	9	18	20	11	9
caudalanteriorcingulate	17	3	1	4	3	10	13	0	0	0
caudalmiddlefrontal	52	2	12	14	21	17	38	11	14	7
cuneus	16	0	3	3	10	3	13	6	8	5
entorhinal	6	1	1	2	2	2	4	0	1	0
frontalpole	1	0	0	0	1	0	1	1	1	1
fusiform	108	19	32	51	27	30	57	40	41	21
inferiorparietal	73	3	17	20	31	22	53	14	35	11
inferiortemporal	119	14	38	52	31	36	67	22	26	15
insula	109	18	29	47	29	33	62	45	24	12
isthmuscingulate	32	7	4	11	11	10	21	6	4	1
lateraloccipital	41	2	8	10	13	18	31	16	26	9
lateralorbitofrontal	113	7	31	38	38	37	75	45	40	29
lingual	45	6	13	19	8	18	26	21	33	17
medialorbitofrontal	47	10	8	18	14	15	29	6	9	5
middletemporal	117	27	39	66	24	27	51	28	20	8
paracentral	4	0	0	0	2	2	4	0	0	0
parahippocampal	27	12	4	16	6	5	11	3	3	1
parsopercularis	18	5	4	9	4	5	9	10	7	5
parsorbitalis	20	2	5	7	7	6	13	4	7	4
parstriangularis	34	3	4	7	16	11	27	9	10	7
pericalcarine	13	3	2	5	3	5	8	8	8	6
postcentral	37	2	6	8	10	19	29	13	12	7
posteriorcingulate	16	4	3	7	1	8	9	6	4	4
precentral	87	10	10	20	33	34	67	47	40	29
precuneus	80	12	24	36	14	30	44	6	15	3
rostralanteriorcingulate	16	3	4	7	5	4	9	0	0	0
rostralmiddlefrontal	94	16	22	38	24	32	56	27	28	18
superiorfrontal	85	17	16	33	19	33	52	18	18	9
superiorparietal	49	4	6	10	15	24	39	13	16	7
superiortemporal	108	20	38	58	15	35	50	69	28	26
supramarginal	56	0	0	0	30	26	56	15	12	5
temporalpole	4	3	1	4	0	0	0	1	2	0
transversetemporal	10	2	5	7	2	1	3	8	5	5
TOTAL	**1801**	**299**	**394**	**693**	**541**	**567**	**1108**	**565**	**532**	**293**

Table S2 | Distribution of electrodes over the Desikan-Killiany Atlas

The number of electrodes for different brain regions of the DK atlas (rows) for different conditions (columns). From the left to right the columns represent the following: (1) Total electrodes, (2) Gray Matter Left, (3) White Matter Left, (4) Total Left, (5) Gray Matter Right, (6) White Matter Right, (7) Total Right, (8) Responsive Audio, (9) Responsive Visual, (10) Responsive Audiovisual. The regions that showed a significant percent of audiovisual electrodes that was statistically unlikely to get from a random intersection of audio or visual electrodes are highlighted in bold ($p<0.01$, permutation test, $n=10^6$ iterations, total electrodes >=20)

(a) ENGLISH

Noun Animal	Length	Syll.	Ngram Freq	Noun food	Length	Syll.	Ngram Freq
'fish'	4	1	6.42E-05	'water'	5	2	3.39E-04
'horse'	5	1	6.07E-05	'oil'	3	1	8.64E-05
'bear'	4	1	5.56E-05	'coffee'	6	2	3.94E-05
'dog'	3	1	5.33E-05	'salt'	4	1	3.70E-05
'bird'	4	1	3.20E-05	'fruit'	5	1	3.59E-05
'cat'	3	1	2.91E-05	'milk'	4	1	3.58E-05
'mouse'	5	1	2.31E-05	'sugar'	5	2	3.40E-05
'sheep'	5	1	1.95E-05	'tea'	3	1	3.33E-05
'turkey'	6	2	1.92E-05	'rice'	4	1	3.13E-05
'fox'	3	1	1.84E-05	'bread'	5	1	3.07E-05
'bull'	4	1	1.63E-05	'eggs'	4	1	2.13E-05
'rat'	3	1	1.49E-05	'corn'	4	1	2.00E-05
'wolf'	4	1	1.47E-05	'apple'	5	1	1.65E-05
'seal'	4	1	1.40E-05	'cheese'	6	1	1.53E-05
'lion'	4	2	1.36E-05	'butter'	6	2	1.51E-05
'deer'	4	1	1.24E-05	'pepper'	6	2	1.26E-05
'cow'	3	1	1.14E-05	'olive'	5	1	1.14E-05
'snake'	5	1	1.12E-05	'bean'	4	1	9.08E-06
'penguin'	7	2	1.05E-05	'garlic'	6	2	8.59E-06
'eagle'	5	2	9.73E-06	'salad'	5	2	8.55E-06
'dragon'	6	2	9.54E-06	'lemon'	5	2	8.49E-06
'pig'	3	1	9.35E-06	'onion'	5	2	6.42E-06
'bat'	3	1	9.28E-06	'berry'	5	1	6.39E-06
'tiger'	5	2	8.53E-06	'cherry'	6	1	6.27E-06
'rabbit'	6	2	8.41E-06	'pizza'	5	2	5.60E-06
'monkey'	6	2	6.86E-06	'nut'	3	1	5.20E-06
'duck'	4	1	6.70E-06	'pasta'	5	2	4.50E-06
'goat'	4	1	6.34E-06	'grape'	5	1	3.86E-06
'whale'	5	1	5.77E-06	'peas'	4	1	3.61E-06
'hawk'	4	1	5.56E-06	'peach'	5	1	3.22E-06
'spider'	6	2	5.46E-06	'plum'	4	1	2.78E-06
'ant'	3	1	5.33E-06	'lettuce'	7	2	2.68E-06

Adj concrete	Length	Syll.	Ngram Freq	Adj abstract	Length	Syll.	Ngram Freq
'long'	4	1	5.04E-04	'good'	4	1	5.76E-04
'small'	5	1	3.48E-04	'best'	4	1	2.69E-04
'large'	5	1	3.21E-04	'better'	6	2	2.63E-04
'low'	3	1	2.09E-04	'free'	4	1	2.17E-04
'short'	5	1	1.80E-04	'real'	4	1	2.10E-04
'clear'	5	1	1.74E-04	'poor'	4	1	1.42E-04
'hard'	4	1	1.59E-04	'bad'	3	1	1.09E-04
'strong'	6	1	1.47E-04	'serious'	7	2	8.12E-05
'big'	3	1	1.40E-04	'happy'	5	2	7.47E-05
'deep'	4	1	1.07E-04	'rich'	4	1	6.90E-05

'dark'	4	1	1.00E-04	'holy'	4	2	6.65E-05
'cold'	4	1	9.21E-05	'pretty'	6	2	5.84E-05
'round'	5	1	8.84E-05	'evil'	4	2	5.71E-05
'heavy'	5	2	6.79E-05	'wild'	4	1	5.37E-05
'hot'	3	1	6.75E-05	'pure'	4	1	4.79E-05
'fast'	4	1	6.44E-05	'sick'	4	1	3.45E-05
'dry'	3	1	5.33E-05	'busy'	4	2	2.96E-05
'soft'	4	1	5.22E-05	'sad'	3	1	2.54E-05
'slow'	4	1	4.84E-05	'proud'	5	1	2.53E-05
'solid'	5	2	4.74E-05	'calm'	4	1	2.47E-05
'huge'	4	1	4.71E-05	'gentle'	6	1	2.04E-05
'warm'	4	1	4.71E-05	'strict'	6	1	2.02E-05
'fat'	3	1	4.18E-05	'mad'	3	1	2.02E-05
'bright'	6	1	4.17E-05	'smart'	5	1	2.01E-05
'weak'	4	1	4.16E-05	'crazy'	5	2	1.70E-05
'thin'	4	1	4.10E-05	'brave'	5	1	1.55E-05
'sweet'	5	1	4.08E-05	'cheap'	5	1	1.49E-05
'silent'	6	1	3.65E-05	'ugly'	4	2	1.12E-05
'oval'	4	2	5.71E-06	'clever'	6	1	1.10E-05
'tiny'	4	2	2.93E-05	'jealous'	7	2	7.70E-06
'dirty'	5	2	1.58E-05	'shy'	3	1	7.70E-06
'massive'	7	2	2.52E-05	'lazy'	4	2	6.09E-06

(b) SPANISH

Noun Animal	Length	Syll.	Ngram Freq	Noun food	Length	Syll.	Ngram Freq
buho	4	2	6.00E-08	ajo	3	2	5.00E-06
burro	5	2	3.50E-06	arroz	5	2	1.30E-05
dragon	6	2	8.00E-08	café	4	2	5.00E-05
gallo	5	2	6.00E-06	coco	4	2	3.00E-06
gato	4	2	1.40E-05	frijol	6	2	4.50E-06
leon	4	1	2.20E-07	huevo	5	2	1.00E-05
lobo	4	2	7.80E-06	jamon	5	2	2.50E-08
mono	4	2	6.00E-06	jugo	4	2	5.90E-06
oso	3	2	4.80E-06	limon	5	2	6.40E-08
pato	4	2	2.50E-06	maiz	4	1	1.70E-06
pavo	4	2	3.00E-06	mango	5	2	4.50E-06
perro	5	2	3.70E-05	melon	5	2	3.00E-08
pez	3	1	8.00E-06	pan	3	1	3.70E-05
pulpo	5	2	1.00E-06	pastel	6	2	4.80E-06
raton	5	2	2.50E-08	postre	6	2	7.60E-06
tigre	5	2	5.00E-06	queso	5	2	1.10E-05
topo	4	2	1.60E-06	vino	4	2	8.00E-05
toro	4	2	9.30E-06	yogur	5	2	1.98E-06
Adj concrete	**Length**	**Syll.**	**Ngram Freq**	**Adj abstract**	**Length**	**Syll.**	**Ngram Freq**

alto	4	2	1.20E-04	bello	5	2	1.40E-05
ancho	5	2	2.50E-05	bueno	5	2	7.57E-05
bajo	4	2	2.88E-04	cruel	5	2	1.40E-05
claro	5	2	1.35E-04	feliz	5	2	4.84E-05
debil	5	2	1.20E-07	feo	3	2	5.56E-06
dulce	5	2	3.00E-05	guapo	5	2	7.44E-06
duro	4	2	3.30E-05	lindo	5	2	4.53E-06
fino	4	2	1.00E-05	listo	5	2	1.25E-05
frio	4	2	6.00E-07	loco	4	2	2.44E-05
fuerte	6	2	1.10E-05	malo	4	2	3.00E-05
grande	6	2	1.00E-04	pobre	5	2	5.00E-05
largo	5	2	2.22E-04	puro	4	2	2.50E-05
lento	5	2	1.70E-05	rico	4	2	2.32E-05
rojo	4	2	3.77E-05	sabio	5	2	1.38E-05
seco	4	2	1.80E-05	serio	5	2	3.96E-05
suave	5	1	3.00E-05	tonto	5	2	7.94E-06
sucio	5	2	6.78E-06	triste	5	2	3.36E-05
verde	5	2	3.78E-05	vago	4	2	5.00E-06

(c) TAIWANESE

Noun Animal	Length	Syll.	Ngram Freq	Noun food	Length	Syll.	Ngram Freq
乳牛-cow	2	2	1.00E-07	咖啡-coffee	2	2	4.60E-05
企鵝-penguir	2	2	1.40E-06	大蒜-garlic	2	2	8.00E-07
兔子-rabbit	2	2	3.30E-06	奶油-butter	2	2	7.60E-07
海豹-seal	2	2	5.10E-07	桃子-peach	2	2	4.60E-07
熊-bear	1	1	4.00E-06	水-water	1	1	2.00E-04
狐狸-fox	2	2	1.60E-06	沙拉-salad	2	2	3.80E-06
狗-dog	1	1	2.20E-05	洋蔥-onion	2	2	1.50E-06
狼-wolf	1	1	1.00E-05	牛奶-milk	1	1	7.80E-06
猴子-monkey	2	2	9.00E-06	玉米-corn	1	1	1.40E-05
獅子-lion	2	2	6.20E-06	米-rice	1	1	6.00E-05
綿羊-sheep	2	2	1.80E-06	糖-sugar	1	1	8.40E-06
老虎-tiger	2	2	7.00E-06	茶-tea	1	1	2.20E-05
老鼠-mouse	2	2	8.60E-06	葡萄-grape	2	2	3.70E-06
蜘蛛-spider	2	2	5.00E-06	蘋果-apple	2	2	2.30E-05
螞蟻-ant	2	2	4.60E-06	蛋-eggs	1	1	7.70E-06
貓-cat	1	1	9.10E-06	豆子-bean	2	2	4.70E-07
馬-horse	1	1	8.20E-05	辣椒-chili	2	2	9.00E-07
鯨魚-whale	2	2	1.80E-06	鳳梨-pineapp	2	2	4.50E-07
鳥-bird	1	1	1.30E-05	鹽-salt	1	1	1.20E-05
龍-dragon	1	1	2.00E-05	麵包-bread	2	2	2.50E-05
Adj concrete	**Length**	**Syll.**	**Ngram Freq**	**Adj abstract**	**Length**	**Syll.**	**Ngram Freq**
乾的-dry	2	2	1.00E-04	假的-fake	2	2	4.00E-05
低的-low	2	2	2.00E-04	傻的-silly	2	2	4.00E-06

冷的-cold	2	2	1.50E-05	壞的-bad	2	2	4.30E-05
大的-large	2	2	1.40E-03	好的-good	2	2	6.00E-04
小的-small	2	2	4.50E-04	帥的-handsor	2	2	2.50E-06
快的-fast	2	2	9.60E-05	忙的-busy	2	2	1.50E-05
慢的-slow	2	2	3.00E-05	怒的-angry	2	2	5.50E-06
濕的-wet	2	2	8.10E-06	懶的-lazy	2	2	3.80E-06
熱的-hot	2	2	5.50E-05	新的-new	2	2	1.40E-03
甜的-sweet	2	2	5.00E-06	病的-sick	2	2	4.30E-04
瘦的-thin	2	2	4.00E-06	瘋的-mad	2	2	4.50E-06
短的-short	2	2	5.00E-05	真的-real	2	2	1.40E-04
硬的-hard	2	2	2.00E-05	窮的-poor	2	2	1.80E-05
胖的-fat	2	2	6.30E-06	笨的-stupid	2	2	2.20E-06
軟的-soft	2	2	2.60E-05	累的-tired	2	2	1.70E-05
輕的-light	2	2	3.00E-05	美的-beautifu	2	2	2.00E-04
酸的-sour	2	2	3.80E-06	舊的-old	2	2	8.00E-05
重的-heavy	2	2	2.20E-04	貴的-expensiv	2	2	2.70E-05
長的-long	2	2	3.00E-04	醜的-ugly	2	2	6.30E-06
高的-tall	2	2	5.00E-04	難的-difficult	2	2	1.40E-04

Table S3 | List of all the words used in the experiment, their lengths, number of syllables, and occurrence frequency. (a) English. (b) Spanish. (c) Taiwanese

Region	Total	Left	Right
hippocampus	1	1	0
fusiform	1	1	0
lateralorbitofrontal	10	9	1
superiortemporal	2	2	0
TOTAL	**14**	**13**	**1**

Table S4 | Distribution of electrodes that showed modulation by part of speech across brain regions. Significant regions showing lateralization shown in bold. ($p<10^{-5}$, permutation test, $n=10^{6}$ iterations, regions with less than 4 electrodes were excluded).

subject#	#Total	NounEnhanced (β>0)	AdjEnhanced (β<0)	Generalize	SubCategory (=Total-Generalized)
5	2	2	0	0	2
14	3	2	1	1	2
16	3	3	0	3	0
18	2	2	0	2	0
20	4	2	2	2	2
TOTAL	**14**	**11**	**3**	**8**	**6**

Table S5 | Distribution of nouns- versus adjective-preferring electrodes and electrodes that generalize for parts-of-speech versus those that do not. Distribution across different subjects of electrodes that are more noun enhanced (column 3) versus more adjective enhanced (column 4), and that of electrodes that generalize to nouns and adjectives (column 5) versus those that showed differences between noun subcategories or adjective subcategories (column 6).

Region	RegionTotal	Noun	Adjective
hippocampus	1	1	0
fusiform	1	1	0
lateralorbitofrontal	10	9	1
superiortemporal	2	0	2
TOTAL	**14**	**11**	**3**

Table S6 | Distribution of nouns- versus adjective-preferring electrodes across brain regions. A permutation test combining all brain regions for these electrodes showed that that LOF was significantly noun preferring. ($p<10^{-5}$, permutation test, $n=10^6$ iterations, regions with less than 4 electrodes were excluded).

subject#	nLeftLOF	nPOS	%POS
6	6	0	0
9	3	0	0
13	6	0	0
14	4	3	**75**
15	4	0	0
16	5	3	**60**
18	5	2	**40**
20	5	2	**40**
Total	**38**	**10**	**Mean%All** = 27 ± 31
			Mean%POSsubjects = 54 ± 17

Table S7 | Distribution of part of speech encoding electrodes in the left lateralorbitofrontal cortex across subjects

Noun Animal	#N	#A	#V	(#N+1)/ (#A+1)	(#N+1)/ (#V+1)
'fish'	105	0	11	106.0	8.8
'horse'	126	0	0	127.0	127.0
'bear'	11	0	93	12.0	0.1
'dog'	124	0	0	125.0	125.0
'bird'	93	0	0	94.0	94.0
'cat'	55	0	0	56.0	56.0
'mouse'	28	0	0	29.0	29.0
'sheep'	30	0	0	31.0	31.0
'turkey'	0	0	0	-	-
'fox'	13	0	0	14.0	14.0
'bull'	12	0	0	13.0	13.0
'rat'	24	0	0	25.0	25.0
'wolf'	12	0	0	13.0	13.0
'seal'	15	0	15	16.0	1.0
'lion'	21	0	0	22.0	22.0
'deer'	0	0	0	-	-
'cow'	26	0	0	27.0	27.0
'snake'	12	0	0	13.0	13.0
'penguin'	0	0	0	-	-
'eagle'	18	0	0	19.0	19.0
'dragon'	13	0	0	14.0	14.0
'pig'	25	0	0	26.0	26.0
'bat'	13	0	0	14.0	14.0
'tiger'	13	0	0	14.0	14.0
'rabbit'	25	0	0	26.0	26.0
'monkey'	11	0	0	12.0	12.0
'duck'	19	0	0	20.0	20.0
'goat'	12	0	0	13.0	13.0
'whale'	13	0	0	14.0	14.0
'hawk'	0	0	0	-	-
'spider'	10	0	0	11.0	11.0
'ant'	10	0	0	11.0	11.0

Noun food	#N	#A	#V	(#N+1)/ (#A+1)	(#N+1) /(#V+1)
'water'	372	0	0	373	373
'oil'	110	0	0	111	111
'coffee'	68	0	0	69	69
'salt'	33	0	0	34	34
'fruit'	51	0	0	52	52
'milk'	48	0	0	49	49
'sugar'	38	0	0	39	39
'tea'	88	0	0	89	89
'rice'	16	0	0	17	17
'bread'	38	0	0	39	39
'eggs'	62	0	0	63	63
'corn'	12	0	0	13	13
'apple'	35	0	0	36	36
'cheese'	30	0	0	31	31
'butter'	21	0	0	22	22
'pepper'	11	0	0	12	12
'olive'	0	0	0	-	-
'bean'	18	0	0	19	19
'garlic'	0	0	0	1	1
'salad'	14	0	0	15	15
'lemon'	14	0	0	15	15
'onion'	12	0	0	13	13
'berry'	0	0	0	-	-
'cherry'	0	0	0	-	-
'pizza'	0	0	0	-	-
'nut'	15	0	0	16	16
'pasta'	0	0	0	-	-
'grape'	0	0	0	-	-
'peas'	33	0	0	34	34
'peach'	0	0	0	-	-
'plum'	0	0	0	-	-
'lettuce'	0	0	0	-	-

Adj concrete	#A	#N	#V	(#A+1)/ (#N+1)	(#A+1)/ (#V+1)
'long'	392	0	0	393.0	393.0
'small'	518	0	0	519.0	519.0
'large'	471	0	0	472.0	472.0
'low'	286	0	0	287.0	287.0
'short'	198	0	0	199.0	199.0

Adj abstract	#A	#N	#V	(#A+1)/ (#N+1)	(#A+1)/ (#V+1)
'good'	1276	25	0	49.1	1277.0
'best'	0	0	0	-	-
'better'	0	0	0	-	-
'free'	200	0	23	201.0	8.4
'real'	227	0	0	228.0	228.0

'clear'	239	0	55	240.0	4.3	'poor'	166	0	0	167.0	167.0
'hard'	176	0	0	177.0	177.0	'bad'	264	0	0	265.0	265.0
'strong'	196	0	0	197.0	197.0	'serious'	124	0	0	125.0	125.0
'big'	338	0	0	339.0	339.0	'happy'	129	0	0	130.0	130.0
'deep'	97	0	0	98.0	98.0	'rich'	79	0	0	80.0	80.0
'dark'	104	31	0	3.3	105.0	'holy'	30	0	0	31.0	31.0
'cold'	103	25	0	4.0	104.0	'pretty'	30	0	0	31.0	31.0
'round'	28	47	0	0.6	29.0	'evil'	15	16	0	0.9	16.0
'heavy'	105	0	0	106.0	106.0	'wild'	55	0	0	56.0	56.0
'hot'	94	0	0	95.0	95.0	'pure'	36	0	0	37.0	37.0
'fast'	50	0	0	51.0	51.0	'sick'	44	0	0	45.0	45.0
'dry'	56	0	28	57.0	2.0	'busy'	53	0	0	54.0	54.0
'soft'	66	0	0	67.0	67.0	'sad'	36	0	0	37.0	37.0
'slow'	56	0	23	57.0	2.4	'proud'	32	0	0	33.0	33.0
'solid'	35	0	0	36.0	36.0	'calm'	14	0	0	15.0	15.0
'huge'	79	0	0	80.0	80.0	'gentle'	29	0	0	30.0	30.0
'warm'	70	0	0	71.0	71.0	'strict'	24	0	0	25.0	25.0
'fat'	20	28	0	0.7	21.0	'mad'	32	0	0	33.0	33.0
'bright'	62	0	0	63.0	63.0	'smart'	16	0	0	17.0	17.0
'weak'	45	0	0	46.0	46.0	'crazy'	18	0	0	19.0	19.0
'thin'	56	0	0	57.0	57.0	'brave'	18	0	0	19.0	19.0
'sweet'	36	0	0	37.0	37.0	'cheap'	68	0	0	69.0	69.0
'silent'	38	0	0	39.0	39.0	'ugly'	14	0	0	15.0	15.0
'oval'	0	0	0	-	-	'clever'	25	0	0	26.0	26.0
'tiny'	56	0	0	57.0	57.0	'jealous'	0	0	0	-	-
'dirty'	27	0	0	28.0	28.0	'shy'	11	0	0	12.0	12.0
'massive'	44	0	0	45.0	45.0	'lazy'	0	0	0	-	-

Table S8 | List of all the words used in the experiment, the number of times they occurred in the British National Corpus as a noun (#N), as an adjective (#A), or a verb (#V), and the ratios of their frequency of occurence in their assigned part of speech versus their usage in other parts of speech. Dashes indicate words that were missing in the corpus.

Subject	Age	Gender	Language	Handedness	#Trials	%Correct	#Electrodes
1	13	M	HI	R	786	54%	134
2	15	F	TW	R	627	69%	35
3	19	M	EN	R	587	96%	174
4	37	F	EN	R	603	97%	78
5	40	M	EN	R	604	99%	73
6	21	M	TW	L	621	71%	64
7	30	M	TW	R	625	59%	78
8	42	F	EN	R	601	87%	154
9	27	F	EN	R	589	96%	62
10	32	F	EN	R	600	92%	141
11	20	F	EN	L	600	96%	72
12	25	M	EN	L	597	77%	92
13	33	F	EN	R	594	96%	65
14	50	F	EN	R	588	96%	117
15	20	M	EN	R	578	94%	65
16	20	F	EN	R	604	84%	81
17	20	F	EN	R	600	94%	78
						TOTAL	1563

Table S9 | Information about each participant for the sentence task including age, gender, language (ENglish, HIndi, TaiWanese), handedness, number of trials, behavioral performance and number of electrodes.

(a) English

Semantically Correct

W1	W2	W3	W4	W1	W2	W3	W4
the	boys	caugh	balls	the	girls	opene	cans
the	girls	ate	pies	the	wome	wrote	poems
the	kids	drank	milk	the	men	read	stories
the	men	cut	cakes	the	dogs	wagg	tails
the	girls	read	books	the	men	poste	letters
the	cows	ate	grass	the	kids	locke	doors
the	bees	flappe	wings	the	kids	playe	dice
the	horse:	ate	grass	the	boys	grew	plants
the	kids	broke	plates	the	guard	held	guns
the	boys	playec	soccer	the	dogs	shed	hair
the	girls	playec	hockey	the	banke	loane	money
the	men	ate	bread	the	kids	spillec	water
the	wome	plucke	fruits	the	wome	cleane	rooms
the	men	tried	shirts	the	wome	playe	songs
the	dogs	bit	shoes	the	boys	made	noises
the	birds	lay	eggs	the	friend	ran	races
the	snake	ate	eggs	the	ants	carrie	leaves
the	cows	gave	milk	the	horse:	made	sounds
the	chefs	made	pizza	the	peopl	loved	books
the	kids	playec	music	the	kids	closec	doors
the	girls	baked	cakes	the	sisters	board	planes
the	kids	broke	windows	the	teach	hande	papers
the	boys	poure	water	the	artists	painte	trees
the	girls	dug	holes	the	stude	raised	questions
the	girls	fried	eggs	the	grand	grew	chickens
the	men	kickec	balls	the	uncles	plante	seeds
the	boys	heard	music	the	father	paid	bills
the	men	rode	bikes	the	mothe	raised	babies
the	kids	rang	bells	the	fairies	gave	gifts
the	wome	stored	food	the	girls	playe	guitar
the	girls	carriec	books	the	grand	grew	lettuce
the	girls	stored	money	the	teach	answe	questions
the	boys	watch	movies	the	mothe	foldec	clothes
the	moms	baked	cakes	the	thieve	stole	food
the	birds	ate	seeds	the	men	move	chairs
the	cats	drank	milk	the	rocks	broke	windows
the	men	sippec	wine	the	monk	showe	teeth
the	kids	cleane	tables	the	chefs	cracke	eggs
the	kids	wore	hats	the	boys	pushe	rocks
the	boys	wore	pants	the	girls	finishe	homework
the	wome	boiled	eggs	the	stude	wrote	notes
the	men	clickec	pictures	the	friend	grasp	hands
the	men	chopp	trees	the	playe	heade	balls
the	boys	wrote	letters	the	teach	helpe	students
the	girls	drove	cars	the	kids	enjoy	parties
the	dogs	fetche	bones	the	salesr	knock	doors
the	bears	ate	honey	the	men	loade	trucks
the	dogs	drank	milk	the	wome	loved	shoes
the	cats	climbe	trees	the	wome	ordere	dinner
the	ducks	laid	eggs	the	men	placec	bets
the	monk	stole	mangoe	the	girls	ran	races
the	robbe	stole	money	the	paren	scolde	children
the	clouds	brougl	rain	the	girls	smelle	flowers
the	girls	cut	pizza	the	soldie	sprea	flags
the	girls	wrote	stories	the	wome	tried	shoes
the	boys	read	books	the	grand	taugh	lessons
the	men	build	walls	the	chefs	toaste	bread
the	girls	drove	trucks	the	monk	grabb	branches
the	girls	rode	bikes	the	touris	tossec	coins
the	wome	sippec	tea	the	parrot	used	tools
the	guard	kept	guns	the	girls	untiec	knots
the	kids	wore	masks	the	girls	wante	icecream
the	boys	watch	soccer	the	wome	warm	food
the	girls	watch	baseball	the	wome	washe	clothes
the	wome	watch	movies	the	chami	won	games
the	horse:	drank	milk	the	shopp	wrapr	presents
the	kids	wore	shirts	the	dogs	watch	trees
the	boys	carriec	rocks	the	cats	chase	lasers
the	boys	broke	eggs	the	teach	spelle	words
the	wome	bough	clothes	the	drago	spat	fire
the	girls	bough	fruits	the	girls	flickec	pages
the	boys	bough	food	the	turtles	ate	food
the	birds	made	nests	the	kids	carrie	backpacks
the	birds	hid	eggs	the	men	sang	songs
the	boys	threw	stones	the	boys	opene	doors
the	girls	playec	basketball				

Semantically Incorrect/Odd

W1	W2	W3	W4	W1	W2	W3	W4
the	balls	caught	boys	the	cans	opene	girls
the	pies	ate	girls	the	poem	wrote	women
the	milk	drank	kids	the	storie	read	men
the	cakes	cut	men	the	tails	wagg	dogs
the	books	read	girls	the	letters	poste	men
the	grass	ate	cows	the	doors	locke	kids
the	wings	flappe	bees	the	dice	playe	kids
the	grass	ate	horses	the	plants	grew	boys
the	plates	broke	kids	the	guns	held	guards
the	socce	playec	boys	the	hair	shed	dogs
the	hocke	playec	girls	the	mone	loane	bankers
the	bread	ate	men	the	water	spillec	kids
the	fruits	plucke	women	the	rooms	cleane	women
the	shirts	tried	men	the	songs	playe	women
the	shoes	bit	dogs	the	noises	made	boys
the	eggs	lay	birds	the	races	ran	friends
the	eggs	ate	snakes	the	leave	carrie	ants
the	milk	gave	cows	the	sound	made	horses
the	pizza	made	chefs	the	books	loved	people
the	music	playec	kids	the	doors	closec	kids
the	cakes	baked	girls	the	plane:	board	sisters
the	windo	broke	kids	the	paper	hande	teachers
the	water	poure	boys	the	trees	painte	artists
the	holes	dug	girls	the	questi	raised	students
the	eggs	fried	girls	the	chicke	grew	grandmas
the	balls	kicked	men	the	seeds	plante	uncles
the	music	heard	boys	the	bills	paid	fathers
the	bikes	rode	men	the	babie:	raised	mothers
the	bells	rang	kids	the	gifts	gave	fairies
the	food	stored	women	the	guitar	playe	girls
the	books	carriec	girls	the	lettuc	grew	grandmas
the	mone	stored	girls	the	questi	answe	teachers
the	movie	watch	boys	the	clothe	foldec	mothers
the	cakes	baked	moms	the	food	stole	thieves
the	seeds	ate	birds	the	chairs	move	men
the	milk	drank	cats	the	windo	broke	rocks
the	wine	sipped	men	the	teeth	showe	monkeys
the	tables	cleane	kids	the	eggs	cracke	chefs
the	hats	wore	kids	the	rocks	pushe	boys
the	pants	wore	boys	the	home	finishe	girls
the	eggs	boiled	women	the	notes	wrote	students
the	pictur	clickec	men	the	hands	grasp	friends
the	trees	chopp	men	the	balls	heade	players
the	letters	wrote	boys	the	stude	helpe	teachers
the	cars	drove	girls	the	partie	enjoy	kids
the	bones	fetche	dogs	the	doors	knock	salesmen
the	honey	ate	bears	the	trucks	loade	men
the	milk	drank	dogs	the	shoes	loved	women
the	trees	climbe	cats	the	dinne	ordere	women
the	eggs	laid	ducks	the	bets	placec	men
the	mang	stole	monkeys	the	races	ran	girls
the	mone	stole	robbers	the	childre	scolde	parents
the	rain	brougl	clouds	the	flowe	smelle	girls
the	pizza	cut	girls	the	flags	sprea	soldiers
the	storie	wrote	girls	the	shoes	tried	women
the	books	read	boys	the	lessor	taugh	grandpas
the	walls	build	men	the	bread	toaste	chefs
the	trucks	drove	girls	the	brancl	grabb	monkeys
the	bikes	rode	girls	the	coins	tossec	tourists
the	tea	sipped	women	the	tools	used	parrots
the	guns	kept	guards	the	knots	untiec	girls
the	mask:	wore	kids	the	icecre	wante	girls
the	socce	watch	boys	the	food	warm	women
the	baseb	watch	girls	the	clothe	washe	women
the	movie	watch	women	the	game	won	champion
the	milk	drank	horses	the	prese	wrapr	shoppers
the	shirts	wore	kids	the	trees	watch	dogs
the	rocks	carriec	boys	the	lasers	chase	cats
the	eggs	broke	boys	the	words	spelle	teachers
the	clothe	bough	women	the	fire	spat	dragons
the	fruits	bough	girls	the	pages	flickec	girls
the	food	bough	boys	the	food	ate	turtles
the	nests	made	birds	the	backp	carrie	kids
the	eggs	hid	birds	the	songs	sang	men
the	stone:	threw	boys	the	doors	opene	boys
the	baske	playec	girls				

(b) Taiwanese

Semantically Correct

W1	W2	W3	W4	Translation
男孩	正在	接	球	the.boys.are.catching.balls
女孩	正在	吃	派	the.girls.are.eating.pies

Semantically Incorrect/Odd

W1	W2	W3	W4
球	正在	接	男孩
派	正在	吃	女孩

小孩	正在	喝	牛奶	the.kids.are.drinking.milk
男人	正在	切	蛋糕	the.men.are.cutting.cakes
女孩	正在	看	書	the.girls.are.reading.books
牛	正在	吃	草	the.cows.are.eating.grass
蜜蜂	正在	拍	翅膀	the.bees.are.flapping.wings
馬	正在	吃	草	the.horses.are.eating.grass
小孩	正在	打破	盤子	the.kids.are.breaking.plates
男孩	正在	玩	足球	the.boys.are.playing.soccer
女孩	正在	玩	棒球	the.girls.are.playing.baseball
男人	正在	吃	麵包	the.men.are.eating.bread
女人	正在	採集	水果	the.women.are.plucking.fruits
男人	正在	試穿	襯衫	the.man.is.trying.shirts
狗	正在	咬	鞋子	the.dogs.are.biting.shoes
鳥	正在	下	蛋	the.birds.are.laying.eggs
蛇	正在	吃	蛋	the.snakes.are.eating.eggs
乳牛	正在	產	乳	the.cows.are.giving.milk
廚師	正在	做	披薩	the.chefs.are.making.pizza
小孩	正在	播	音樂	the.kids.are.playing.music
女孩	正在	烘培	蛋糕	the.girls.are.baking.cakes
小孩	正在	打破	窗戶	the.kids.are.breaking.windows
男孩	正在	倒	水	the.boys.are.pouring.water
女孩	正在	挖	洞	the.girls.are.dugging.holes
女孩	正在	煎	蛋	the.girls.are.frying.eggs
男人	正在	踢	球	the.men.are.kicking.balls
男孩	正在	聽	音樂	the.boys.are.hearing.music
男人	正在	騎	單車	the.men.are.riding.bikes
小孩	正在	搖	鈴	the.kids.are.ringing.bells
女人	正在	擺放	食物	the.women.are.storing.food
女孩	正在	搬	書	the.girls.are.carrying.books
女孩	正在	存	錢	the.girls.are.storing.money
男孩	正在	看	電影	the.boys.are.watching.movies
媽媽	正在	烘培	蛋糕	the.mom.is.baking.cakes
鳥	正在	吃	種子	the.bird.is.eating.seeds
貓	正在	喝	牛奶	the.cat.is.drinking.milk
男人	正在	喝	酒	the.man.is.sipping.wine
小孩	正在	清理	桌子	kid.is.cleaning.tables
小孩	正在	戴	帽子	the.kid.is.wearing.a.hat
男人	正在	搬	磚頭	the.man.is.moving.bricks
男孩	正在	穿	長褲	the.boy.is.wearing.pants
女人	正在	水煮	雞蛋	the.woman.is.boiling.eggs
男人	正在	拍	照片	the.man.is.clicking.pictures
男人	正在	砍	樹	the.man.is.chopping.trees
男孩	正在	寫	字	the.boy.is.writing.words
女孩	正在	開	車	the.girl.is.driving.a.car
狗	正在	接	骨頭	the.dog.is.fetching.bones
熊	正在	吃	蜂蜜	the.bear.is.eating.honey
狗	正在	喝	牛奶	the.dog.is.drinking.milk
貓	正在	爬	樹	the.cat.is.climbing.a.tree
鴨子	正在	下	蛋	the.duck.is.laying.eggs
猴子	正在	偷	芒果	the.monkey.is.stealing.mangoes
強盜	正在	偷	錢	the.robber.is.stealing.money
雲	正在	帶來	雨水	the.cloud.is.bringing.rain
女孩	正在	切	披薩	the.girl.is.cutting.pizza
女孩	正在	寫	故事	the.girl.is.writing.a.story
男孩	正在	看	書	the.boy.is.reading.books
男人	正在	築	牆	the.man.is.building.walls
女孩	正在	開	卡車	the.girl.is.driving.a.truck
女孩	正在	騎	單車	the.girls.are.riding.bikes
女人	正在	喝	茶	the.women.are.sipping.tea
警衛	正在	持	槍	the.guards.are.keeping.guns
小孩	正在	戴	口罩	the.kids.are.wearing.masks
男孩	正在	看	足球	the.boy.is.watching.soccer
女孩	正在	看	棒球	the.girl.is.watching.baseball
女人	正在	看	電影	the.woman.is.watching.movies
馬	正在	喝	奶	the.horse.is.drinking.milk
小孩	正在	穿	襯衫	the.kid.is.wearing.shirts
男孩	正在	搬	石頭	the.boy.is.moving.rocks
男孩	正在	打破	雞蛋	the.boy.is.breaking.eggs
女人	正在	買	衣服	the.woman.is.buying.clothes
女孩	正在	買	水果	the.girl.is.buying.fruits
男孩	正在	買	食物	the.boy.is.buying.food
鳥	正在	築	巢	the.bird.is.making.nests
鳥	正在	藏	蛋	the.bird.is.hiding.eggs
男孩	正在	丟	石頭	the.boy.is.throwing.stones
女孩	正在	打	籃球	the.girl.is.playing.basketball
男人	正在	寄	信	the.man.is.posting.letter
小孩	正在	鎖	門	the.kid.is.locking.a.door
男孩	正在	種	樹	the.boy.is.growing.a.tree
衛兵	正在	持	槍	the.guard.is.holding.a.gun
狗	正在	脫	毛	the.dog.is.sheding.hairs
行員	正在	借出	錢	the-bankers-loaned-money
小孩	正在	打翻	水	the.kid.is.spilling.the.water
女人	正在	打掃	房間	the.woman.is.cleaning.rooms
女人	正在	播放	歌曲	the.woman.is.playing.songs
男孩	正在	製造	噪音	the.boy.is.making.noise

牛奶	正在	喝	小孩
蛋糕	正在	切	男人
書	正在	看	女孩
草	正在	吃	牛
翅膀	正在	拍	蜜蜂
草	正在	吃	馬
盤子	正在	打破	小孩
足球	正在	玩	男孩
棒球	正在	玩	女孩
麵包	正在	吃	男人
水果	正在	採集	女人
襯衫	正在	試穿	男人
鞋子	正在	咬	狗
蛋	正在	下	鳥
蛋	正在	吃	蛇
乳	正在	產	乳牛
披薩	正在	做	廚師
音樂	正在	播	小孩
蛋糕	正在	烘培	女孩
窗戶	正在	打破	小孩
水	正在	倒	男孩
洞	正在	挖	女孩
蛋	正在	煎	女孩
球	正在	踢	男人
音樂	正在	聽	男孩
單車	正在	騎	男人
鈴	正在	搖	小孩
食物	正在	擺放	女人
書	正在	搬	女孩
錢	正在	存	女孩
電影	正在	看	男孩
蛋糕	正在	烘培	媽媽
種子	正在	吃	鳥
牛奶	正在	喝	貓
酒	正在	喝	男人
桌子	正在	清理	小孩
帽子	正在	戴	小孩
磚頭	正在	搬	男人
長褲	正在	穿	男孩
雞蛋	正在	水煮	女人
照片	正在	拍	男人
樹	正在	砍	男人
字	正在	寫	男孩
車	正在	開	女孩
骨頭	正在	接	狗
蜂蜜	正在	吃	熊
牛奶	正在	喝	狗
樹	正在	爬	貓
蛋	正在	下	鴨子
芒果	正在	偷	猴子
錢	正在	偷	強盜
雨水	正在	帶來	雲
披薩	正在	切	女孩
故事	正在	寫	女孩
書	正在	看	男孩
牆	正在	築	男人
卡車	正在	開	女孩
單車	正在	騎	女孩
茶	正在	喝	女人
槍	正在	持	警衛
口罩	正在	戴	小孩
足球	正在	看	男孩
棒球	正在	看	女孩
電影	正在	看	女人
奶	正在	喝	馬
襯衫	正在	穿	小孩
石頭	正在	搬	男孩
雞蛋	正在	打破	男孩
衣服	正在	買	女人
水果	正在	買	女孩
食物	正在	買	男孩
巢	正在	築	鳥
蛋	正在	藏	鳥
石頭	正在	丟	男孩
籃球	正在	打	女孩
信	正在	寄	男人
門	正在	鎖	小孩
樹	正在	種	男孩
槍	正在	持	衛兵
毛	正在	脫	狗
錢	正在	借出	行員
水	正在	打翻	小孩
房間	正在	打掃	女人
歌曲	正在	播放	女人
噪音	正在	製造	男孩

朋友	正在	比賽	跑步	friends.are.running.races	跑步	正在	比賽	朋友
螞蟻	正在	搬	樹葉	the-ants-moved-leaves	樹葉	正在	搬	螞蟻
馬	正在	製造	聲音	the.horse.is.making.sounds	聲音	正在	製造	馬
人們	正在	買	書	people.are.buying.books	書	正在	買	人們
小孩	正在	關	門	the.kid.is.closing.a.door	門	正在	關	小孩
老虎	正在	走	路	the.tiger.is.taking.a.walk	路	正在	走	老虎
姐妹	正在	搭	飛機	the.sisters.are.boarding.a.plane	飛機	正在	搭	姐妹
老師	正在	發	紙	the.teacher.is.handing.papers	紙	正在	發	老師
畫家	正在	畫	樹	the.artists.painted.trees	樹	正在	畫	畫家
學生	正在	提出	問題	the.student.is.raising.questions	問題	正在	提出	學生
奶奶	正在	養	雞	the.grandma.is.growing.chickens	雞	正在	養	奶奶
叔叔	正在	種	菜	the.uncle.is.growing.vegetables	菜	正在	種	叔叔
爸爸	正在	付	帳	the.father.is.paying.bills	帳	正在	付	爸爸
媽媽	正在	抱	小孩	the.mother.is.holding.babies	小孩	正在	抱	媽媽
仙女	正在	送	禮物	fairies.are.giving.gifts	禮物	正在	送	仙女
女孩	正在	彈	吉他	the.girl.is.playing.guitar	吉他	正在	彈	女孩
奶奶	正在	種	萵苣	grandma.is.growing.lettuce	萵苣	正在	種	奶奶
老師	正在	回答	問題	the.teacher.is.ansering.quesetions	問題	正在	回答	老師
小偷	正在	偷	食物	the.thieve.is.stealing.food	食物	正在	偷	小偷
男人	正在	搬	椅子	the.man.is.moving.chairs	椅子	正在	搬	男人
石頭	正在	撞擊	窗戶	the.rock.is.hitting.the.window	窗戶	正在	撞擊	石頭
廚師	正在	打	蛋	the.chef.is.cracking.eggs	蛋	正在	打	廚師
男孩	正在	推	石頭	the.boy.is.pushing.rocks	石頭	正在	推	男孩
女孩	正在	完成	作業	the.girl.is.finishing.homework	作業	正在	完成	女孩
學生	正在	寫	筆記	the.student.is.writing.notes	筆記	正在	寫	學生
朋友	正在	牽	手	friends.are.grasping.hands	手	正在	牽	朋友
球員	正在	踢	球	the.players.are.kicking.balls	球	正在	踢	球員
老師	正在	幫助	學生	the.teacher.is.helping.students	學生	正在	幫助	老師
小孩	正在	參加	派對	the.kid.is.joining.the.party	派對	正在	參加	小孩
郵差	正在	敲	門	the.mailman.is.knocking.the.door	門	正在	敲	郵差
男人	正在	開	卡車	the.man.is.driving.trucks	卡車	正在	開	男人
女人	正在	試穿	鞋子	the.woman.is.trying.shoes	鞋子	正在	試穿	女人
女人	正在	訂	晚餐	the.woman.is.ordering.dinner	晚餐	正在	訂	女人
男人	正在	下	注	the.man.is.placing.bets	注	正在	下	男人
女孩	正在	比賽	跑步	the.girl.is.running.race	跑步	正在	比賽	女孩
家長	正在	罵	小孩	the.parent.is.scolding.the.child	小孩	正在	罵	家長
女孩	正在	聞	花香	the.girl.is.smelling.flowers	花香	正在	聞	女孩
士兵	正在	拿	國旗	the.soldiers.are.holding.a.flag	國旗	正在	拿	士兵
女人	正在	丟	鞋子	the.woman.is.throwing.shoes	鞋子	正在	丟	女人
朋友	正在	交換	禮物	friends.are.exchanging.gifts	禮物	正在	交換	朋友
爺爺	正在	教	課	grandpa.is.teaching.lessons	課	正在	教	爺爺
廚師	正在	烤	麵包	the.chef.is.toasting.bread	麵包	正在	烤	廚師
猴子	正在	抓	樹枝	the.monkey.is.grabbing.branches	樹枝	正在	抓	猴子
遊客	正在	擲	銅板	tourists.are.tossing.coins	銅板	正在	擲	遊客
鸚鵡	正在	使用	工具	the.parrot.is.using.tools	工具	正在	使用	鸚鵡
女孩	正在	解開	繩結	the.girl.is.untiing.knots	繩結	正在	解開	女孩
女孩	正在	吃	冰棒	the.girl.is.eating.an.ice.bar	冰棒	正在	吃	女孩
女人	正在	加熱	食物	the.woman.is.heating.food	食物	正在	加熱	女人
女人	正在	洗	衣服	the.woman.is.washing.clothes	衣服	正在	洗	女人
選手	正在	爭奪	冠軍	the.player.is.competing.for.the.first	冠軍	正在	爭奪	選手
店員	正在	包裝	禮物	the.shoopper.is.wrapping.the.prese	禮物	正在	包裝	店員
狗	正在	爬	樹	the.dog.is.climbing.the.tree	樹	正在	爬	狗
貓	正在	追	球	the.cat.is.chasing.the.ball	球	正在	追	貓
老師	正在	教	單字	the.teacher.is.teaching.words	單字	正在	教	老師
龍	正在	噴	火	dragon.is.spatting.fire	火	正在	噴	龍
女孩	正在	翻	書	the.girl.is.flicking.pages	書	正在	翻	女孩
烏龜	正在	吃	食物	the.turtle.is.eating.food	食物	正在	吃	烏龜
小孩	正在	背	背包	the.kid.is.carrying.a.backpack	背包	正在	背	小孩
奶奶	正在	烘培	餅乾	grandma.is.baking.cookies	餅乾	正在	烘培	奶奶
護士	正在	照顧	病人	the.nurse.is.taking.care.of.patients	病人	正在	照顧	護士
技師	正在	修理	煞車	the.mechanics.is.fixsing.brakes	煞車	正在	修理	技師
蛇	正在	下	蛋	the.snake.is.laying.eggs	蛋	正在	下	蛇
女孩	正在	彈	鋼琴	the.girl.is.playing.piano	鋼琴	正在	彈	女孩
男孩	正在	吹	長笛	the.boy.is.playing.flute	長笛	正在	吹	男孩
男人	正在	唱	歌	the.man.is.singing.a.song	歌	正在	唱	男人
男孩	正在	開	門	the.boy.is.opening.a.door	門	正在	開	男孩
女孩	正在	開	罐頭	the.girl.is.opening.a.can	罐頭	正在	開	女孩
女人	正在	寫	詩	the.woman.is.writing.a.poem	詩	正在	寫	女人
男人	正在	讀	詩	the.man.is.reading.a.poem	詩	正在	讀	男人
狗	正在	搖	尾巴	the.dog.is.wagging.its.tail	尾巴	正在	搖	狗

(c) Hindi

Semantically Correct					Semantically Incorrect/Odd			
W1	**W2**	**W3**	**W4**	**Translation**	**W1**	**W2**	**W3**	**W4**
लडकी	ने	गेंद	पकडी	the girl caught the ball	गेंद	ने	लडकी	पकडी
लडकी	ने	किताब	पढी	the girl read the book	किताब	ने	लडकी	पढी
गाय	ने	घास	खाई	the cow ate the grass	घास	ने	गाय	खाई
बच्ची	ने	प्लेट	तोडी	the youngGirl broke the plate	प्लेट	ने	बच्ची	तोडी
आदमी	ने	फल	बेचा	the man sold the fruit	फल	ने	आदमी	बेचा
गाय	ने	रोटी	खाई	the cow ate the bread	रोटी	ने	गाय	खाई
नारी	ने	रोटी	बेली	the woman rolled the bread	रोटी	ने	नारी	बेली
बकरी	ने	रस्सी	काटी	the goat cut the rope (with its teeth	रस्सी	ने	बकरी	काटी
टीचर	ने	किताब	लिखी	the teacher wrote the book	किताब	ने	टीचर	लिखी
घोडी	ने	लात	मारी	the horse hit the kick	लात	ने	घोडी	मारी
लडकी	ने	लस्सी	पी	the girl drank the lassi	लस्सी	ने	लडकी	पी

बचचा	न	फालम	दखा	the youngGirl watched the film	फालम	न	बचचा	दखा
बालल	न	डरस	फाडा	the cat tore the dress	डरस	न	बालला	फाडा
टाचर	न	कवात	लाखा	the teacher wrote the poem	कवात	न	टाचर	लाखा
दादा	न	साडा	पहना	sister wore the saree	साडा	न	दादा	पहना
दादा	न	खार	बनाय	grandma made the kheer	खार	न	दादा	बनाया
रानी	न	चाय	पा	the queen drank tea	चाय	न	रानी	पा
लडका	न	शरट	साला	the girl stitched the shirt	शरट	न	लडका	साला
भाल	न	शहद	खाया	the bear ate the honey	शहद	न	भाल	खाया
आदम	न	आल	बचा	the man sold the potato	आल	न	आदमा	बचा
भाड	न	कदद	काटा	brother cut the pumpkin	कदद	न	भाड	काटा
बदर	न	शार	काया	the monkey made noise	शार	न	बदर	काया
बदर	न	फल	ताडा	the monkey plucked the flower	फल	न	बदर	ताडा
जानव	न	दध	पाया	the animal drank milk	दध	न	जानव	पाया
गारड	न	गट	खाला	the guard opened the gate	गट	न	गारड	खाला
लडका	न	फालम	बनाय	the girl made the film	फालम	न	लडका	बनाया
बचचा	न	गद	फका	the youngGirl threw the ball	गद	न	बचचा	फका
लडका	न	गाडा	चलाय	the girl drove the car	गाडा	न	लडका	चलाया
हाथा	न	पड	ताडा	the elephant broke the tree	पड	न	हाथा	ताडा
माला	न	पड	ताडा	the gardner cut the tree	पड	न	माला	ताडा
माला	न	पाना	डाला	the gardner gave water (to the gar	पाना	न	माला	डाला
आदम	न	गाना	गया	the man sang the song	गाना	न	आदमा	गया
दादा	न	गाना	बजाय	grandpa played the song	गाना	न	दादा	बजाया
आदम	न	सातार	बजाय	the man played sitar	सातार	न	आदमा	बजाया
बालल	न	आवाज	लगाय	the cat made the call (meow)	आवाज	न	बालला	लगाया
दादा	न	दवा	दा	grandma gave medicine	दवा	न	दादा	दा
डाकट	न	डलाज	करा	the doctor did the treatment	डलाज	न	डाकटर	करा
नारा	न	पढाड	करा	the woman did the study	पढाड	न	नारा	करा
बालल	न	घटा	बजाय	the cat rang the bell	घटा	न	बालला	बजाया
दादा	न	खाडक	खाला	grandma opened the window	खाडक	न	दादा	खाला
भाल	न	गडढा	खादा	the bear dug the hole	गडढा	न	भाल	खादा
लडका	न	टरक	चलाय	the girl drove the truck	टरक	न	लडका	चलाया
शर	न	आराम	करा	the lion did rest (noun)	आराम	न	शर	करा
बचचा	न	हाका	खला	the youngGirl played hockey	हाका	न	बचचा	खला
आदम	न	जता	पहना	the man wore the shoe	जता	न	आदमा	पहना
बाबा	न	सट	पहना	old-man wore a suit	सट	न	बाबा	पहना
लडका	न	चटना	खाड	the girl ate the chutney	चटना	न	लडका	खाड
लडका	न	फाटा	खाचा	the girl clicked the photograph	फाटा	न	लडका	खाचा
दादा	न	घडा	पहना	grandma wore the watch	घडा	न	दादा	पहना
शरना	न	चाय	पा	the lioness drank tea	चाय	न	शरना	पा
बकरा	न	शराब	पा	the goat drank wine	शराब	न	बकरा	पा
लडका	न	चाय	पा	the girl drank tea	चाय	न	लडका	पा
लडका	न	खार	खाड	the girl ate the pudding	खार	न	लडका	खाड
बकरा	न	राटा	खाड	the goat ate bread	राटा	न	बकरा	खाड
घाडा	न	आवाज	करा	the horse made sound (neigh)	आवाज	न	घाडा	करा
भाल	न	फल	खाया	the bear ate the fruit	फल	न	भाल	खाया
बदर	न	आम	चसा	the monkey sucked a mango	आम	न	बदर	चसा
बकरा	न	घास	चबाड	the goat chewed grass	घास	न	बकरा	चबाड
बकरा	न	जाभ	दाखार	the goat showed the tougue	जाभ	न	बकरा	दाखाया
घाडा	न	जाभ	दाखार	the horse showed the tongue	जाभ	न	घाडा	दाखाया
बचचा	न	जाभ	दाखार	the youngGirl showed the tongue	जाभ	न	बचचा	दाखाया
बदर	न	आम	ताडा	the monkey plucked the mango	आम	न	बदर	ताडा
बचचा	का	चाट	लगा	the youngGirl got hurt	चाट	का	बचचा	लगा
बालल	का	ठड	लगा	the cat felt the cold	ठड	का	बालला	लगा
नारा	का	गरमा	लगा	the woman felt the heat	गरमा	का	नारा	लगा
बचचा	न	नाव	चलाय	the youngGirl rowed the boat	नाव	न	बचचा	चलाया
लडका	न	बदक	चलाय	the girl operated the gun	बदक	न	लडका	चलाया
आदम	न	पाना	डाला	the man put water (on the plants)	पाना	न	आदमा	डाला
आदम	न	कक	बनाय	the man made the cake	कक	न	आदमा	बनाया
आदम	न	मासक	पहना	the man wore a mask	मासक	न	आदमा	पहना
आदम	न	फान	काया	the man dialed the phone	फान	न	आदमा	काया
आदम	न	अखबा	पढा	the man read the newspaper	अखबा	न	आदमा	पढा
आदम	न	खत	पढा	the man read the letter	खत	न	आदमा	पढा
आदम	न	खत	लाखा	the man wrote the letter	खत	न	आदमा	लाखा
नरस	न	जाभ	दखा	the nurse checked the tongue	जाभ	न	नरस	दखा
नरस	न	दवा	दा	the nurse gave the medicine	दवा	न	नरस	दा
लडका	न	दवा	खाड	the girl ate the medicine	दवा	न	लडका	खाड
बाबा	न	जस	पाया	the oldMan drank the juice	जस	न	बाबा	पाया
बदर	न	जस	पाया	the monkey drank the juice	जस	न	बदर	पाया
भाल	न	जस	पाया	the bear drank the juice	जस	न	भाल	पाया
बदर	न	दध	पाया	the monkey drank the milk	दध	न	बदर	पाया
नारा	न	खार	बनाय	the woman made the pudding	खार	न	नारा	बनाया
नारा	न	सबजा	बनाय	the woman made the vegetableCu	सबजा	न	नारा	बनाया
नारा	न	आग	लगाय	the woman put the fire	आग	न	नारा	लगाया
नारा	न	आग	बझाय	the woman extinguished the fire	आग	न	नारा	बझाया
नारा	न	राटा	खाड	the woman ate bread	राटा	न	नारा	खाड
नारा	न	डरस	पहना	the woman wore the dress	डरस	न	नारा	पहना
नरस	न	जान	बचाय	the nurse saved life	जान	न	नरस	बचाया
लडका	न	पढाड	करा	the girl did the study	पढाड	न	लडका	करा
आदम	न	पहाड	चढा	the man climbed the hill	पहाड	न	आदमा	चढा
लामडा	न	गद	फका	the fox threw the ball	गद	न	लामडा	फका
लामडा	न	बरड	खाड	the fox ate the bread	बरड	न	लामडा	खाड
आदम	न	शाशा	ताडा	the man broke the windowGlass	शाशा	न	आदमा	ताडा
सचान	का	डनाम	दाया	the player got the prize	डनाम	का	सचान	दाया
बतत	न	डबका	मारा	the duck took the dip	डबका	न	बततख	मारा
लडका	न	डबका	मारा	the girl took the dip	डबका	न	लडका	मारा

लड़की	ने	आँख	मारी	the girl winked the eye	आँख	ने	लड़की	मारी
बिल्ली	ने	आँख	मारी	the cat winked the eye	आँख	ने	बिल्ली	मारी
नारी	ने	गुस्सा	किया	the woman made anger	गुस्सा	ने	नारी	किया
आदमी	ने	गुस्सा	किया	the man made anger	गुस्सा	ने	आदमी	किया
आदमी	ने	सेब	खाया	the man ate the apple	सेब	ने	आदमी	खाया
टीचर	ने	सवाल	पूछा	the teacher asked the question	सवाल	ने	टीचर	पूछा
स्टूडेंट	ने	जवाब	दिया	the student answered the question	जवाब	ने	स्टूडेंट	दिया
नर्स	ने	दवा	लगाया	the nurse applied the medicine	दवा	ने	नर्स	लगाया
नर्स	ने	इलाज	करा	the nurse did the treatment	इलाज	ने	नर्स	करा
नर्स	ने	सुई	लगाया	the nurse gave the injection	सुई	ने	नर्स	लगाया
लड़की	ने	फुटबा	खेली	the girl played soccer	फुटबा	ने	लड़की	खेली
आदमी	ने	क्रिकेट	खेला	the man played cricket	क्रिकेट	ने	आदमी	खेला
बंदर	ने	क्रिकेट	खेला	the monkey played cricket	क्रिकेट	ने	बंदर	खेला
बंदर	ने	इनाम	जीता	the monkey won the prize	इनाम	ने	बंदर	जीता
शेर	ने	पहाड़	चढ़ा	the lion climbed the hill	पहाड़	ने	शेर	चढ़ा
बंदर	ने	पेड़	चढ़ा	the monkey climbed the tree	पेड़	ने	बंदर	चढ़ा
बंदर	ने	मास्क	पहना	the monkey wore the mask	मास्क	ने	बंदर	पहना
बंदर	ने	फोन	किया	the monkey called the phone	फोन	ने	बंदर	किया
लड़की	ने	प्लेट	तोड़ी	the girl broke the plate	प्लेट	ने	लड़की	तोड़ी
लड़की	ने	खिड़की	तोड़ी	the girl broke the window	खिड़की	ने	लड़की	तोड़ी
लड़की	ने	गलती	करी	the girl made the mistake	गलती	ने	लड़की	करी
टीचर	ने	डाँट	लगाया	the teacher gave the scolding	डाँट	ने	टीचर	लगाया
बच्ची	ने	खीर	खाई	the youngGirl ate the pudding	खीर	ने	बच्ची	खाई
बंदर	ने	सूट	पहना	the monkey wore the suit	सूट	ने	बंदर	पहना
बंदर	ने	गिलास	पकड़ा	the monkey held the glass	गिलास	ने	बंदर	पकड़ा
भालू	ने	गिलास	पकड़ा	the bear held the glass	गिलास	ने	भालू	पकड़ा
भालू	ने	दूध	पीया	the bear drank the milk	दूध	ने	भालू	पीया
बकरी	ने	टोपी	पहनी	the goat wore the hat	टोपी	ने	बकरी	पहनी
गाय	ने	शर्ट	पहनी	the cow wore the shirt	शर्ट	ने	गाय	पहनी
गाय	ने	घास	खाई	the cow ate grass	घास	ने	गाय	खाई
गाय	ने	पत्ती	खाई	the cow ate the leaf	पत्ती	ने	गाय	खाई
बकरी	ने	पत्ती	खाई	the goat ate the leaf	पत्ती	ने	बकरी	खाई
गाय	ने	साड़ी	पहनी	the cow wore the saree	साड़ी	ने	गाय	पहनी
लड़की	ने	साड़ी	पहनी	the girl wore the saree	साड़ी	ने	लड़की	पहनी
माली	ने	पेड़	लगाया	the gardner planted the tree	पेड़	ने	माली	लगाया

Table S10 | List of all sentences used in the experiment. (a) English, (b) Taiwanese, and (c) Hindi. The semantically incorrect/odd sentences were formed by swapping the nouns of the correct sentences, without changing the grammatical correctness of the sentence.

Region\nElecs	Total	GMLeft	WMLeft	Left	GMRight	WMRight	Right
'Amygdala'	44	22	0	22	22	0	22
'Hippocampus'	65	29	0	29	36	0	36
'Inf-Lat-Vent'	7	3	0	3	4	0	4
'Lateral-Ventricle'	3	2	0	2	1	0	1
'Putamen'	6	3	0	3	3	0	3
'VentralDC'	1	0	0	0	1	0	1
'bankssts'	17	4	0	4	8	5	13
udalanteriorcingula	27	8	4	12	7	8	15
caudalmiddlefrontal	54	7	23	30	3	21	24
'entorhinal'	14	8	1	9	4	1	5
'fusiform'	44	7	9	16	14	14	28
'inferiorparietal'	57	10	7	17	24	16	40
'inferiortemporal'	75	16	10	26	23	26	49
'insula'	82	14	7	21	29	32	61
'isthmuscingulate'	14	2	3	5	1	8	9
'lateraloccipital'	13	2	0	2	6	5	11
'lateralorbitofrontal'	79	20	21	41	12	26	38
'lingual'	3	0	0	0	3	0	3
medialorbitofrontal	57	15	22	37	9	11	20
'middletemporal'	175	54	30	84	61	30	91
'paracentral'	4	0	0	0	1	3	4
'parahippocampal'	9	2	1	3	5	1	6
'parsopercularis'	39	8	7	15	13	11	24
'parsorbitalis'	23	12	4	16	1	6	7
'parstriangularis'	66	21	19	40	8	18	26
'pericalcarine'	4	0	0	0	1	3	4
'postcentral'	24	9	3	12	3	9	12
'posteriorcingulate'	27	4	10	14	2	11	13
'precentral'	59	23	21	44	3	12	15
'precuneus'	19	1	8	9	4	6	10
stralanteriorcingula	29	4	16	20	4	5	9
rostralmiddlefrontal	87	24	33	57	14	16	30
'superiorfrontal'	74	20	20	40	8	26	34
'superiorparietal'	15	1	1	2	4	9	13
'superiortemporal'	128	43	15	58	37	33	70
'supramarginal'	70	24	9	33	7	30	37
'temporalpole'	38	21	5	26	11	1	12
transversetemporal	11	3	5	8	1	2	3
TOTAL	1563	446	314	760	398	405	803

Table S11 | Distribution of electrodes for sentence task over the Desikan-Killiany Atlas
The number of electrodes for different brain regions of the DK atlas (rows) for different

conditions (columns). From the left to right the columns represent the following: (1) Total electrodes, (2) Gray Matter Left, (3) White Matter Left, (4) Total Left, (5) Gray Matter Right, (6) White Matter Right, (7) Total Right.

	Total	Left	Right	Noun	Verb
'caudalmiddlefrontal'	2	2	0	2	0
'fusiform'	2	0	2	1	1
'inferiorparietal'	1	1	0	1	0
'inferiortemporal'	2	2	0	2	0
'insula'	2	1	1	1	1
'middletemporal'	4	1	3	**4**	0
'parahippocampal'	1	0	1	1	0
'parsopercularis'	3	2	1	1	2
'parsorbitalis'	1	1	0	1	0
'parstriangularis'	6	**4**	2	**5**	1
'postcentral'	1	1	0	1	0
'precentral'	7	**7**	0	1	**6**
'rostralmiddlefrontal'	2	1	1	1	1
'superiorfrontal'	1	1	0	0	1
'superiortemporal'	3	2	1	2	1
'supramarginal'	1	1	0	0	1
'Hippocampus'	2	0	2	2	0
TOTAL	**41**	**27**	**14**	**26**	**15**

Table S12 | Electrode locations of all the electrodes on the Desikan-Kiliany Atlas, that had a significant contribution of the Nouns versus Verbs predictor in the GLM. From the left to right the columns represent the following: (1) Total electrodes, (2) left or, (3) right hemisphere, (4) noun- or, (5) verb-preferring.Significant regions showing lateralization shown in bold. ($p<10^{-5}$, permutation test, $n=10^{6}$ iterations, regions with less than 4 electrodes were excluded).

Region	Total	Left	Right	G>	UG>
'caudalmiddlefrontal'	1	1	0	0	1
'cuneus'	1	1	0	0	1
'inferiorparietal'	1	0	1	0	1
'inferiortemporal'	2	2	0	2	0
'lingual'	1	1	0	0	1
'medialorbitofrontal'	2	1	1	2	0
'parahippocampal'	1	0	1	0	1
'precentral'	2	0	2	2	0
precuneus'	1	0	1	0	1
'rostralmiddlefrontal'	2	2	0	1	1
'superiortemporal'	1	1	0	0	1
'supramarginal'	1	0	1	0	1
TOTAL	**16**	**9**	**7**	**7**	**9**

Table S13 | Distribution of grammatical versus ungrammatical phrase enhanced electrodes across the Desikan-Killiany Atlas. Bolded regions had audiovisual electrodes.

"What we observe is not nature itself, but nature exposed to our method of questioning."

Werner Heisenberg

2 Discussion

I described neurophysiological signals that selectively discriminate between two parts of speech, nouns and adjectives (**Figure 1.4**). This selectivity was robust to orthographic variables such as word length, phonetic features such as number of syllables, and word occurrence statistics (**Figure 1.4**). This selectivity for part of speech generalized across sensory modalities (**Figure 1.4, 1.11, 1.14**), word positions, grammatical correctness and motor outputs (**Figure 1.4, 1.11, 1.14**), and semantic groups of nouns and adjectives (Figure S6). These neurophysiological signals enable discrimination between parts of speech even in single trials (**Figure 1.4, 1.11**). Electrodes that uniquely distinguished nouns from adjectives were particularly clustered within a small, circumscribed region of the lateral orbitofrontal cortex, lateralized to the left hemisphere (**Figure 1.4, 1.14**). Neural discrimination of nouns from adjectives was apparent in the LOF in English-speaking and Taiwanese-speaking participants (**Figure 1.4**). Interestingly, in a bilingual participant, the same electrodes within the left LOF distinguished nouns and adjectives in both English and in Spanish (**Figure 1.4**). Extending the study of minimal phrases to the full sentences data, I conducted an additional experiment where I showed neural signals that distinguished nouns from verbs within full sentences (**Figure 1.17** and **1.16**).

2.1 Effect of Task Design on Findings

In English and other languages, some words can be used both as a noun or as an adjective (e.g., long race versus race horse). In most instances, one usage is more frequent than the other. In particular, the nouns and adjectives in this study are highly overrepresented in their labeled part of speech (**Table S8**). Similarly, some words can be used both as a noun or as a verb (e.g., long race versus race you to the top); all the nouns in this study are highly overrepresented in their usage as nouns (**Table S8**). Thus, the words used in this study had a prototypical interpretation as either a noun or an adjective. The distinction between POS includes their grammatical roles but also their associated semantic connotations (e.g., nouns typically refer to things and adjectives to the attributes of those things).

In languages like English, nouns and adjectives follow a specific grammatical order (i.e., adjectives precede nouns). Other languages reverse this order. In Spanish, adjectives typically follow nouns, though the English order can also be used. It is thus interesting to observe that many electrodes demonstrated strong selectivity for nouns versus adjectives, irrespective of their position within the two-word phrases. Furthermore, in the bilingual participant, the neural responses separated nouns and adjectives in both languages despite the fact that the grammatical order is typically reversed between English and Spanish. It is conceivable that the strong part-of-speech selectivity independent of grammar shown here could be linked to the two-word phrase structures. Another possibility is that the representation of nouns versus adjectives is invariant to grammatical usage rules. The experiment in **Figure 1.17** demonstrates a selective representation of parts of speech that extends to full sentences, invariant to changes in semantics.

2.2 Relation with previous MEG and EEG studies

Non-invasive scalp electroencephalography and magnetoencephalography signals have revealed correlates of language processing with a wide range of onset times from approximately 100 ms all the way to well over 600 ms (for a review, see (Quian Quiroga et al., 2005)). The earliest onset signals commencing between 100 and 300 ms after stimulus onset, sometimes referred to as early left anterior negativity, have been associated with grammatical violations, but previous studies have not documented any invariance in the representation of parts of speech and there is disagreement about whether these early signals are even associated with language (Quian Quiroga et al., 2005). Our work reports an invariant distinction between nouns and adjectives in the LOF commencing at approximately 400 ms after stimulus onset, which is consistent with part-of-speech being represented well after the onset of modality-specific purely visual and auditory signals.

2.3 Unimodal versus Multimodal Approach

A remarkable hallmark of language is its universality. We can interpret the word cat when uttering the word, writing it, listening to it, reading it, and even when examining a photograph of a cat. It is therefore tempting to speculate that there may be an invariant representation of language concepts in the brain. Several studies have examined putative correlates of language processing using only unimodal signals (e.g.,Keshishian (2023), Sinai (2005), Ding et al. (2016), Cometa (2023), Artoni (2020), Gwilliams et al. (2022), OpenAI (2023), Calinescu et al. (2023), Woolnough (2021), Aflalo et al. (2020)). While I observed electrodes that distinguished between parts of speech only in the auditory stimuli or only in the visual stimuli, the responses of those electrodes could be partly explained by other variables including number of syllables, word frequency, or grammar. Using strict criteria and after controlling for confounding variables, most electrodes that distinguished nouns from adjectives showed selectivity during both auditory and visual presentation. Future work should evaluate whether the same electrodes also distinguish parts of speech when participants utter words, write them, or when examining images. An intriguing study described neurons in the human medial temporal lobe that respond selectively to images and their corresponding text and sound descriptions47,48. However, these medial temporal lobe neurons do not seem to distinguish between different parts of speech and their responses seem to be connected with the formation of memories rather than the internal representation of language (Geva-Sagiv et al., 2023). Indeed, there exist strong anatomical and functional connections between the medial temporal lobe and frontal regions that could link language and memory formation (Xiao et al., 2023).

2.4 The Lateral Orbitofrontal Cortex

The lateral orbitofrontal (LOF) cortex constitutes a large expanse of neocortex within the frontal lobe, spanning Brodmann areas (BA) 10, 11, 12 (called BA47 in humans due to cytoarchitectural differences from monkeys) and 13 (Kringelbach, 2005, Ongur and Price, 2000, Wojtasik et al., 2020). Neurobiological tracings from rats, mice, and macaques have identified LOF as a nexus of many inputs53 conveying olfactory, gustatory, visual, auditory, somatosensory, and visceral-sensory information. The LOF has been associated with many cognitive functions, including multisensory integration, working memory, long-term memory consolidation, reward processing, social interactions, memory, decision making, and emotion processing (Ongur and Price, 2000, Xiao et al., 2023, Hunt et al., 2018, Noonan et al., 2010, de Araujo et al., 2003, Dronkers et al., 2004, Mesulam et al., 2014). This heterogeneity might be partly ascribed to investigations probing different cognitive tasks, as in the case of the proverbial blind men sampling different parts of an elephant. Given the prominent role of language in cognition, it is conceivable that previous studies that describe other roles of the LOF did not probe its possible associations with language. However, it is even more likely that descriptors like LOF that refer to such large brain areas would inevitably fail to uncover specific functionality. The current results point to a rather well circumscribed location within LOF, the posterior part of the H-shaped sulcus in the left hemisphere. In humans, this location overlaps with BA13-lateral and BA 47-medial and has been shown to have a strong convergence of auditory and visual inputs (Hunt et al., 2018, Xiao et al., 2023, Mesulam et al., 2022). Interestingly, work on Primary Progressive Aphasia, and frontotemporal lesions implicate the orbitofrontal cortex in word and sentence comprehension deficits (Warrington and Shallice, 1984, Mesulam et al., 2014, Fried et al., 2014, Mukamel and Fried, 2012b). In these studies, the orbitofrontal cortex, dorsal premotor cortex, temporoparietal junction (canonical Wernicke's area), and pars opercularis were associated with sentence comprehension and grammatical production aphasias (evaluated with complex grammatical output requiring planning and motor production). Word comprehension and naming deficits were assessed using binary perceptual choice tasks, implicating the orbitofrontal cortex and the anterior temporal lobe (ATL). However, no single region was found to necessary or sufficient for either grammatical, word or sentence comprehension deficits suggesting a distributed network for linguistic representations. In addition, single neuron studies within the dorsomedial prefrontal cortex (dmPFC) reveal signals for semantic clusters of words within semantically related and grammatically correct sentences. I found significant and invariant electrodes in the ATL and the LOF for the minimal phrase task, but not in other regions, likely due to the task's simplicity, which does not require grammatical production or because the task lacked the necessary context required to activate the representations in the dorsal frontal cortex. Overlapping with the PPA studies, the ATL and pars opercularis also showed selective and consistent responses for the sentence task, but there weren't enough electrodes to form a significant cluster for either task. However, I found significant clusters in the pars triangularis and precentral gyrus activations for the sentence task that were associated with grammatical and sentence production deficits in previous studies. Consistent with extensive work documenting the lateralization of language functions, the results presented here also show a strong predominance of the left hemisphere in the representation of part of speech, despite the fact that there were more electrodes sampling signals from the right hemisphere.

2.5 Limitations of sEEG Recordings

Several limitations in the current work are worth noting. First, all the results reported here are derived from patients with epilepsy. The invasive study of epilepsy patients constitutes the predominant way to access neurophysiological signals from the human brain (Dale et al., 1999, Joshi et al., 2011). Neurophysiological studies in other patient populations (e.g., paraplegic patients, Parkinson's patients, brain tumor patients), typically target specific regions that are not known to be associated with language processing. Caution should be exercised in the interpretation of results from patient populations. To the best of our knowledge, all patients used language fluently and had no language impediments, but one should be aware of the possibility that epilepsy could potentially impact the representation of language. Second, the electrode locations are strictly dictated by clinical criteria. Our sampling of brain activity is extensive but not exhaustive (**Figure 1.1**, **Tables S1-S2**). It is quite possible that other areas not examined here may also reveal neural correlates of parts of speech and that the regions I found interact with other relevant brain areas. A critical goal of cortical resections in epilepsy patients is to cure seizures without interfering with cognitive function. As such, given the strong lateralization and ubiquitous role for language in cognition, it is extremely important to precisely understand the neural structures that support language in these patients and the current results could help guide surgical approaches for epilepsy. Third, the current work focuses on three parts of speech. Nouns, adjectives, and verbs do not constitute an exhaustive list of parts of speech and future work should examine the representation of pronouns, adverbs, prepositions, and conjunctions. Finally, our work provides a correlate of the representation of POS, but future work should evaluate whether any such signals are causally required for online language interpretation.

These results provide initial glimpses into highly localized structures that represent a fundamental component of language that has been extensively studied by linguists for decades, the functional role of different words within a sentence. The representation of nouns versus adjectives in the human brain is invariant to the presentation modality, word properties, grammar, and semantics. Furthermore, the representation even generalizes across different languages. These observations open the doors to begin to elucidate the neural representation of more complex language concepts and to bridge the extensive work in language and linguistics to their underlying neural representations.

"You make experiments and I make theories. Do you know the difference? A theory is something nobody believes, except the person who made it. An experiment is something everybody believes, except the person who made it."

Albert Einstein

3 Methods

3.1 Preregistration

When one preregisters their research, they are simply specifying their research plan in advance of their study and submitting it to a registry. Preregistration separates **hypothesis-generating (exploratory)** from **hypothesis-testing (confirmatory)** research. Both are important. But the same data cannot be used to generate and test a hypothesis, which can happen unintentionally and reduce the credibility of your results. Addressing this problem through planning improves the quality and transparency of your research. This helps you clearly report your study and helps others who may wish to build on it.

Pre-registered Hypotheses:

1. There exists a common neural representation of words presented via visual cues or auditory cues

2. Neural responses to nouns are different from neural responses to adjectives (irrespective of presentation modality)

3. Presentation order matters. The neural representation of a noun followed by an adjective is different from the neural representation of an adjective followed by a noun (irrespective of presentation modality)

4. The neural representation of nouns is category-specific

5. The neural representation of adjectives is category-specific

This study was preregistered on the Open Science Framework (OSF) website. The preregistration DOI is: https://doi.org/10.17605/OSF.IO/8TU2G. To the best of my knowledge, this is the first pre-registered study in human neurophysiology. I found neural correlates for all five hypotheses in the preregistration. The pre-registration is included in the thesis **as-is**, so the findings can be compared directly with the goals outlined in the pre-registration.

Design Plan

Hypotheses

1. There exists a common neural representation of words presented via visual cues or auditory cues
2. Neural responses to nouns are different from neural responses to adjectives (irrespective of presentation modality)
3. Presentation order matters. The neural representation of a noun followed by an adjective is different from the neural representation of an adjective followed by a noun (irrespective of presentation modality)
4. The neural representation of nouns is category specific
5. The neural representation of adjectives is category specific

Design Plan

We study patients with pharmacologically intractable epilepsy (Liu et al Neuron 2009), implanted with invasive surface and/or depth electrodes for the purposes of locating the focus of seizures.
Subjects with intracranial electrodes will be sequentially shown two words as shown in Figure 1. Each trial involves fixation (400 ms), presentation of Word 1 (875 ms), delay (400 ms),

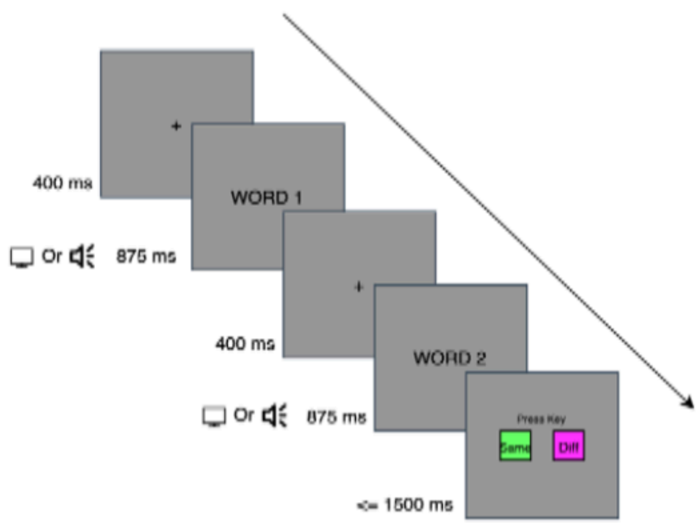


Figure 1: Task Design

presentation of Word 2 (875 ms), followed by a two-alternative forced choice question where subjects have to determine whether the two words are the same or different. The words are presented visually (50% of the trials) or via sound (50% of the trials).
The trials are presented in blocks. Within each block, we present 5 visual stimuli and five auditory stimuli, selected randomly from the stimuli assigned for each block. In a small fraction of trials (16%), Word1=Word2.

The subject will be required to press a key indicating whether the two words presented were same or different. This is done to ensure that the subject is paying attention to the words presented. The subject indicates this by pressing either one of two buttons using a response box device. The two buttons have a green sticker (for choice Same) and a pink sticker (for choice Different). The subjects use only one hand for the task, and have 1,500 ms to respond. If they do not respond within 1500 ms, the experiment moves to the next trial.

The font size is 50. Words are displayed in the center of the screen in lower case letters and each word subtends approximately 5 degrees of visual angle. The background is gray (50% white 50% black) and is displayed at 80% brightness on a MacBook Pro 15-inch laptop, placed approximately 50 cm away from the subject.

The sound stimuli are delivered using the speakers of the laptop. The interval between onsets for Word1 and Word2 for both auditory and visual trials is the same.

The task is coded in MATLAB using the psychophysics toolbox.

The words pairs belong to any of the three categories:

A. Noun-Adjective – Grammatically correct [42% of trials]
B. Adjective-Noun – Grammatically incorrect [42% of trials]
C. Repeated-Pair: either Noun-Noun [8% of trials] or Adjective-Adjective [8% of trials]

The order of A, B, and C trial types is randomized.

The nouns are uniformly distributed within 3 categories: animal, food, and object (Table 1). The adjectives are uniformly distributed within 3 categories: physical, quantitative, and subjective (Table 1). Any noun can be paired with any adjective. If a particular pairing is included in a given block (e.g., proud dog), then the reverse pair is also included in the same block (dog proud). The trial order is randomized.

For each block we use a single seed to create a random set of word pairs using a 'v5uniform' random number generator in MATLAB. Within each block the seed for the random number generator remains fixed. For each consecutive block a new seed corresponding to the block number is picked from a set of seeds stored in MATLAB. This is done so that different subjects see the same stimuli with different word pairs in exactly the same order. [*MATLAB command:* `s = rng(curr_state.active_block_number, 'v5uniform');`].

Each word pair (e.g., proud dog) is presented 5 times using visual stimuli and 5 times using auditory stimuli. The different word pairs are created using pseudorandom seeds, and then shuffled within each block of presentation. Each block of the experiment has 8 Noun-Adj pairs (which automatically implies 8 reversed Adj-Noun pairs using the same words), and 3 repeated pairs. Therefore, the total number of trials per block of data is: (8+8+3) x (5+5) = 190 trials.

Within each block, after every 10 trials, or 20 trials for some patients, we provide a brief visual feedback. We also show them a brief gif with dogs to keep them engaged; this gif is purely for entertainment and is uncorrelated to the hypotheses and task. The maximum time per trial is (0.4+0.875) x 2 + 1.5 = 4.05 s. Time per block = 4.05*190 = 769.5 s = 12.825 minutes.

Noun-Animal	Noun-Food	Noun-Object	Adjective-Physical	Adjective-Quantitative	Adjective-Subjective
'fish'	'water'	'table'	'long'	'one'	'good'
'horse'	'oil'	'book'	'small'	'all'	'best'
'bear'	'coffee'	'door'	'large'	'some'	'better'
'dog'	'salt'	'key'	'low'	'two'	'free'
'bird'	'fruit'	'car'	'short'	'first'	'real'
'cat'	'milk'	'bed'	'clear'	'any'	'poor'
'mouse'	'sugar'	'window'	'hard'	'most'	'bad'
'sheep'	'tea'	'phone'	'strong'	'many'	'serious'
'turkey'	'rice'	'glass'	'big'	'each'	'happy'
'fox'	'bread'	'chair'	'deep'	'three'	'rich'
'bull'	'eggs'	'plate'	'dark'	'little'	'holy'
'rat'	'corn'	'cup'	'cold'	'second'	'pretty'
'wolf'	'apple'	'bag'	'round'	'few'	'evil'
'seal'	'cheese'	'desk'	'heavy'	'four'	'wild'
'lion'	'butter'	'mirror'	'hot'	'whole'	'pure'
'deer'	'pepper'	'wire'	'fast'	'enough'	'sick'
'cow'	'olive'	'shirt'	'dry'	'five'	'busy'
'snake'	'bean'	'knife'	'soft'	'full'	'sad'
'penguin'	'garlic'	'bowl'	'slow'	'half'	'proud'
'eagle'	'salad'	'pen'	'solid'	'third'	'calm'
'dragon'	'lemon'	'lock'	'huge'	'six'	'gentle'
'pig'	'onion'	'belt'	'warm'	'ten'	'strict'
'bat'	'berry'	'brush'	'fat'	'million'	'mad'
'tiger'	'cherry'	'jacket'	'bright'	'seven'	'smart'
'rabbit'	'pizza'	'couch'	'weak'	'none'	'crazy'
'monkey'	'nut'	'pencil'	'thin'	'eight'	'brave'
'duck'	'pasta'	'oven'	'sweet'	'nine'	'cheap'
'goat'	'grape'	'curtain'	'silent'	'empty'	'ugly'
'whale'	'peas'	'shoe'	'oval'	'fifty'	'clever'
'hawk'	'peach'	'hammer'	'tiny'	'billion'	'jealous'
'spider'	'plum'	'fork'	'dirty'	'plenty'	'shy'
'ant'	'lettuce'	'pillow'	'massive'	'ample'	'lazy'

Table 1. List of nouns (columns 1-3) and adjectives (columns 4-6) used in this study

All the words in this table are “common” in the sense of having a Google-N-Gram frequency $>10^{-6}$. The number of words in each category is the same. The words are balanced in terms of the number of syllables across categories (paired ttest, p-value > 0.05).

Randomization

The words used for different subjects are the same. For each subject, we generate a seed. That seed dictates the order of presentation within a block and the order of the blocks. Both subjects and the investigator are unaware of the order of presentation.

III. Sampling Plan

Data collection procedures*

Inclusion criteria: We will work with subjects with pharmacologically intractable epilepsy who are implanted with invasive surface and/or depth electrodes for the purpose of localizing their seizures (Liu et al 2009). Subjects must be between 6 and 80 years of age. Subjects must be able to understand and perform the task. If subjects don't speak English fluently, we will do every effort to translate the words and task instructions to the subject's native language.

Exclusion criteria: Subjects who fail to understand the task or who cannot perform the task due to cognitive deficits will be excluded.

Population: Patients with Phase II invasive approach to epilepsy. Both males and females. There are no exclusions based on race or gender. Subject ages range from 6 to 80. Electrode locations are strictly dictated by clinical needs and does not depend on the scientific questions studied here (Liu et al 2009, Fried et al 2014, MIT Press)

Recruitment efforts: There are no recruitment efforts. Patients arrive to our study via previous meetings with the clinical team including neurologists and neurosurgeons. All the clinical work is independent of the scientific study. Subjects provide consent or assent to participate. All the protocols are approved by the IRB.

Payment for participation: Subjects are presented with an age-appropriate token of appreciation for participation in the studies.

Subject selection: There is no subject selection within the pool of patients undergoing invasive implantation of electrodes to treat epilepsy.

Study timeline: We collect data during the stay of the patients at the hospital. The duration of the patient's stay in the hospital is strictly dictated by clinical needs. Typically, patients stay in the hospital for about one week. Subjects can stop the tests at any time if they feel uncomfortable or tired.

Sample size*

We will collect data from 20 subjects

Sample size rationale

It is very hard to work with a large number of subjects in this type of study. We see approximately one patient per month. Based on our previous work (e.g. Liu et al 2009, Madhavan et al 2019, among many others), and the work of many others in the field, and based on power analyses for similar experiments, we expect that 20 subjects will yield enough brain coverage to answer the hypotheses posed in this study.

Variables

Manipulated Variables

Constants: Duration between word onset times (see Figure 1 in Study Design).

Variables:
Level 1: Sound vs. Visual stimuli
Level 2: Word order: noun-adjective (e.g., “red door”), or adjective-noun (reversed pair, e.g., “door red”), repeated pair (e.g. “door door”).
Level 3: Choice of Noun from fixed list (see table below)
Level 4: Choice of adjective from fixed list (see table below)

Noun-Animal	Noun-Animal	Noun-Object	Adjective-Physical	Adjective-Quantitative	Adjective-Subjective
'fish'	'water'	'table'	'long'	'one'	'good'
'horse'	'oil'	'book'	'small'	'all'	'best'
'bear'	'coffee'	'door'	'large'	'some'	'better'
'dog'	'salt'	'key'	'low'	'two'	'free'
'bird'	'fruit'	'car'	'short'	'first'	'real'
'cat'	'milk'	'bed'	'clear'	'any'	'poor'
'mouse'	'sugar'	'window'	'hard'	'most'	'bad'
'sheep'	'tea'	'phone'	'strong'	'many'	'serious'
'turkey'	'rice'	'glass'	'big'	'each'	'happy'
'fox'	'bread'	'chair'	'deep'	'three'	'rich'
'bull'	'eggs'	'plate'	'dark'	'little'	'holy'
'rat'	'corn'	'cup'	'cold'	'second'	'pretty'
'wolf'	'apple'	'bag'	'round'	'few'	'evil'
'seal'	'cheese'	'desk'	'heavy'	'four'	'wild'
'lion'	'butter'	'mirror'	'hot'	'whole'	'pure'
'deer'	'pepper'	'wire'	'fast'	'enough'	'sick'
'cow'	'olive'	'shirt'	'dry'	'five'	'busy'
'snake'	'bean'	'knife'	'soft'	'full'	'sad'
'penguin'	'garlic'	'bowl'	'slow'	'half'	'proud'
'eagle'	'salad'	'pen'	'solid'	'third'	'calm'
'dragon'	'lemon'	'lock'	'huge'	'six'	'gentle'
'pig'	'onion'	'belt'	'warm'	'ten'	'strict'
'bat'	'berry'	'brush'	'fat'	'million'	'mad'
'tiger'	'cherry'	'jacket'	'bright'	'seven'	'smart'
'rabbit'	'pizza'	'couch'	'weak'	'none'	'crazy'
'monkey'	'nut'	'pencil'	'thin'	'eight'	'brave'
'duck'	'pasta'	'oven'	'sweet'	'nine'	'cheap'
'goat'	'grape'	'curtain'	'silent'	'empty'	'ugly'
'whale'	'peas'	'shoe'	'oval'	'fifty'	'clever'
'hawk'	'peach'	'hammer'	'tiny'	'billion'	'jealous'
'spider'	'plum'	'fork'	'dirty'	'plenty'	'shy'
'ant'	'lettuce'	'pillow'	'massive'	'ample'	'lazy'

Measured Variables

(1) Behavior
    a. Reaction times
    b. Correct / incorrect answers
(2) Electrode locations
    a. Pre-operative MR images
    b. Post-operative CT images
(3) Neural data
    Voltage as a function of time for each electrode implanted in the patient's brain

Indices

All data processing steps are described in the Analysis Plan

Analysis Plan

Transformations

Data pre-processing

Intracranial field potentials (IFP) are pre-processed according to the following steps (Liu et al 2009, Madhavan et al 2019):

1. We use a notch filter at 60 Hz and harmonics in each channel to remove line noise with a 2nd order Butterworth filter.
2. We use a 5th order Butterworth filter to lowpass the data at 200 Hz with steepness 0.995. [edit here: not sure if it is butterworth filter →using MATLAB 2019a's lowpass function]
3. We use a global reference or bipolar montage reference
4. For each electrode and each trial *i*, we compute the amplitude of the IFP in the window [50,500] ms with respect to word onset: $A_i = \max(V_i(t)) - \min(V_i(t))$ with $50 \leq t \leq 500$ ms. We compute the distribution of amplitudes A_i across all the trials for each electrode (irrespective of the condition of each trial). We exclude from analyses trials where: $A_i > \mu - 4\sigma$ where μ and σ are the mean and standard deviation of A_i, respectively. This procedure typically amounts to removing 1% of the trials.

Neural features

We consider two types of neural features
(1) raw broadband IFP signals
(2) signals in the gamma frequency band (Madhavan et al 2019). The gamma frequency band is defined by band-pass filtering the data with a 5th-order Butterworth filter between 30 and 100 Hz.

Time windows

We evaluate the responses in the window [50,800] ms with respect to stimulus onset.
We define the baseline as the interval [-350,0] ms with respect to first stimulus onset.

Statistical models*

We perform analyses at the level of individual electrodes and at the level of electrode populations. These two types of analyses are separately discussed next.

Single electrode analyses

In all cases, we use a two-tailed non-parametric permutation test to assess statistical significance at the $p<0.01$ level. For example, consider the definition of visually responsive electrode below. We consider the neural responses during visual stimulus presentation $s_1, s_2, \ldots, s_n$ and the neural responses during baseline $b_1, b_2, \ldots, b_n$. We define the visual responsivity index:

$v = (\mu_s - \mu_b) / (\mathrm{sqrt}(\sigma_s^2 + \sigma_b^2))$.

We next compute the distribution of the visual responsivity index v' under the null hypothesis obtained by shuffling the labels s, b. This procedure is repeated 10,000 times. The probability that the actual index v comes from the shuffled distribution v' is given by:

$p = \{\#\ \text{iterations} \mid |v'|>v\}$

A similar procedure is followed for all the other statistical tests.

<u>Visually responsive electrodes</u>

An electrode is *visually responsive* if the neural response during visual stimulus presentation is different from baseline (e.g., Liu et al 2009).

<u>Auditory responsive electrodes</u>

An electrode is *auditory responsive* if the neural response during auditory stimulus presentation is different from baseline.

<u>Audiovisual responsive electrodes</u>

An *audiovisual responsive electrode* is a visually responsive *and* an auditory responsive electrode.

Hypothesis 1. *There exists a common neural representation of words presented via visual cues or auditory cues.*

A common representation is defined by:

1. Audiovisual responsive electrodes
2. A statistically significant correlation between the responses to the same stimuli when presented in an audio format versus visual format. (how to do this? What about time offsets?)

Hypothesis 2. *Neural responses to nouns are different from neural responses to adjectives (irrespective of presentation modality)*

An electrode is selective for part-of-speech (noun / adjective) if the responses to nouns and adjectives are different.

An electrode could be selective for part-of-speech during visual presentation, audio presentation, or both.

An electrode that is selective for part-of-speech and which is involved in language representation should show selectivity during both visual and auditory presentation.

Hypothesis 3. *Presentation order matters. The neural representation of noun followed by adjective is different from the neural representation of adjective followed by noun (irrespective of presentation modality)*

An electrode is modulated by the presentation order if the neural response to noun-adjective is different from the neural response to adjective-noun.
An electrode can be modulated by the presentation order during visual presentation, auditory presentation, or both.
An electrode that is modulated by the presentation order and which is involved in language representation should show modulation by presentation order during both visual and auditory presentation.

Hypothesis 4: *The neural representation of nouns is category specific.*

An electrode is category-selective for nouns if the neural responses differ between the 3 possible noun categories. Note that this question requires a non-parametric analysis of variance (instead of all the previous comparisons that depend on comparing only 2 groups).
An electrode can show category specificity for nouns during visual presentation, during auditory presentation or during both.
An electrode that shows category-specificity for nouns and which is involved in language representation should show category-specificity for nouns during both visual and auditory presentation.
Furthermore, the selectivity for nouns during visual presentation should correlate with that during auditory presentation.

Hypothesis 5: *The neural representation of adjectives is category specific.*

An electrode is category-selective for adjectives if the neural responses differ between the 3 possible adjective categories. Note that this question requires a non-parametric analysis of variance (instead of all the previous comparisons that depend on comparing only 2 groups).
An electrode can show category specificity for adjectives during visual presentation, during auditory presentation or during both.
An electrode shows category-specificity for adjectives and which is involved in language representation should show category-specificity for adjectives during both visual and auditory presentation.
Furthermore, the selectivity for adjectives during visual presentation should correlate with that during auditory presentation.

Electrode population analyses

The analyses presented so far examine each electrode separately. In addition, we build "pseudo-populations" of electrodes for different electrode locations. We use the term "pseudo-population" to refer to merging electrodes from different subjects and to distinguish this type of analyses from a population where all the signals are simultaneously recorded (discussed in Hung et al 2005, Liu et al 2009).

Populations are built based on the electrode locations to assess the spatial specificity of the findings. We only consider regions where we have at least 10 electrodes.
The population analyses are based on using a machine learning classifier. Activity from all the electrodes in the population are concatenated into a vector for each trial. We use a support vector machine classifier with a linear kernel. In all cases, we use cross-validation by randomly selecting 70% of the trials for training the classifier and reporting test performance on the remaining 30% of the trials. More details about the use of machine learning classifiers can be found in Hung et al 2005, Liu et al 2009, Kriegeskorte and Kreiman 2011. We illustrate the process by considering hypothesis 2, but the same procedure is used for the other hypotheses.

Consider hypothesis 2: *Neural responses to nouns are different from neural responses to adjectives (irrespective of presentation modality).*
We build a machine learning classifier to evaluate whether we can distinguish whether a given presentation is a noun or an adjective in single trials.
Condition 1: train classifier on visual responses, test on visual responses
Condition 2: train classifier on auditory responses, test on auditory responses
Condition 3: same classifier as condition 1, test on auditory responses
Condition 4: same classifier as condition 2, test on visual responses
Null hypothesis: shuffle the trial labels (noun or adjective), repeat 10,000 iterations. Show distribution of classification performance (which we expect to be centered around 50%).

A population (brain area) shows selectivity for nouns versus adjectives if the classifier is above chance in condition 1 or condition 2.
A population (brain area) shows selectivity for nouns versus adjectives and invariance to the presentation modality if the classifier is above chance in all four conditions.

Similar steps are followed for all the other hypothesis.
We note that in hypothesis 4 and 5, there are 3 possible labels, and hence we expect chance to be centered around 33.3%.

Inference criteria

Inference will be drawn based on $p<0.01$.
All tests are two-tailed.
In the population-level analyses the comparisons are based on the performance of the classifier under the null hypothesis. This is converted into a p value by assessing the proportion of iterations where the null distribution reaches a classification performance above the actual population.
All tests are non-parametric.
Control for multiple comparisons is performed by ensuring an experiment wide false discovery rate < 0.01.

Data exclusion

The criteria for detecting potential artifacts and excluding trials for analyses was described above (see Transformations)

The behavioral task is very easy and is mostly used to ensure that the subject pays attention to the stimuli. If a subject's performance is below 20%, we will exclude that subject from analyses.

Whenever possible, we will collect eye movements during the task. Trial in which subjects deviate by more than 2 degrees from the fixation center will be excluded from analyses.

Interictal discharges and seizure events will also be excluded from analyses. The epileptogenic focus is defined by clinical criteria by a teach of experts. Electrodes from the epileptogenic focus are not considered for analyses. Data within a window of 30 minutes before the onset of a seizure to 30 minutes after the offset of a seizure will not be considered for analyses.

Missing data

We do not expect to have any missing data.
Patients may not finish the entire experiment in one session. In these cases, we will run additional sessions as needed. To a reasonable first approximation, invasive neurophysiological recordings are stable over periods of days (Bansal et al 2012). Therefore, we will merge data from multiple sessions for a given patient unless we detect any non-stationarities in the recordings. Occasionally, it is possible that some subjects may not finish all the trials. As long as we have a minimum of 200 trials, we will include those incomplete datasets in the analyses.
Whenever the number of trials is a potential confound in the analyses, we will randomly subsample the data to assess the impact of number of trials on the results.

Additional analyses

Electrode localization

Electrode locations are computed by co-registering the preoperative magnetic resonance imaging with the post-operative computed tomography scans. For each subject, the 3D brain surface is reconstructed and then an automatic parcellation is performed using Freesurfer (Destrieux et al., 2010). The electrode positions ae mapped onto 74 brain areas (Destrieux et al., 2010). For examples of the electrode localization maps, see Madhavan et al 2019.

Response latencies

In addition to the single electrode and population level responses discussed above, we are also especially interested in the dynamics of the neural responses. We will use methods developed in our previous work (Tang et al 2014) to evaluate the response latencies. We will compare the response latencies for:

(1) Visual versus auditory presentation
(2) Nouns versus adjectives
(3) Noun-adjective versus adjective-noun

Brain interactions

In addition to evaluating single electrode responses, and population-level responses assuming independent electrodes, we will also quantify pairwise brain interactions as described in the study of Madhavan et al 2019. The analyses will follow the ones outlined above for the single electrode case, except that here we will use neural interactions between areas as the neural response metric.

Exploratory analysis

Additional analyses will be reported as exploratory results

3.2 Data and Recordings

3.2.1 Data availability

All data and code will be made publicly available through the following link: https://klab.tch.harvard.edu/resources/Misraetal_POS.html The pseudocode can be found within the Readme.docx file in the above link. There is a script for each figure and sub-figure. The README.docx file details how to run those scripts to plot each of them and also redo the analysis for the entire study.

3.2.2 Participants

Minimal Phrase Task: We recorded data from 20 participants (9 male, 9-60 years old, 2 left-handed, 2 ambidextrous, **Table S1**) with pharmacologically resistant epilepsy. All experiments were conducted while participants stayed at Children's Hospital Boston (CHB), Brigham and Women's Hospital (BWH), or Taipei Veterans General Hospital (TVGH). All studies were approved by each hospital's institutional review boards and were carried out with the participants' informed consent.
Sentence Task: We recorded data from 17 participants (7 male, 13-50 years old, 3 left-handed, **Table S9**) with pharmacologically resistant epilepsy. All experiments were conducted while participants stayed at Children's Hospital Boston (CHB), Cleveland Clinic, Ohio, or Taipei Veterans General Hospital (TVGH). All studies were approved by each hospital's institutional review boards and were carried out with the participants' informed consent.

3.2.3 Recordings and Electrode Locations

Participants were implanted with intracranial depth electrodes (Ad-Tech, Racine, WI, USA). Neurophysiological data were recorded using XLTEK (Oakville, ON, Canada), Bio-Logic (Knoxville, TN, USA), Nihon Kohden (Tokyo, Japan), and Natus (Pleasanton, CA, USA). The sampling rate was 2048 Hz at BCH and TVGH, and 1024 Hz or 512 Hz at BWH. All data were referenced in a bipolar montage. There were no seizure events in any of the sessions. Electrode locations were decided based on clinical criteria for each participant. Electrodes in the epileptogenic foci, as well as pathological areas, were removed from analyses. The total number of electrodes after bipolar referencing and removing electrodes with no signal, line noise or recording artifacts was 1,801.

Following implantation, electrodes were localized by co-registration of pre-operative T1 MRI and post-operative CT scans using the iELVis software (Groppe et al., 2017). We used FreeSurfer to segment MRI images, upon which post implant CT was rigidly registered (Desikan et al., 2006). Electrodes were marked in the CT aligned to pre-operative MRI using the Bioimage Suite (Joshi et al., 2011, Dale et al., 1999). The Desikan-Killiany (DK) atlas was used to assign the electrodes locations. **Figure 1.1 b-g** and **Table S2** show the locations of all the electrodes.

3.3 Experiment Design

All visual stimuli were displayed on a 15.4 inch 2,880 × 1,800 pixel LCD screen using the Psychtoolbox in MATLAB (Natick, MA) and a MacBook Pro laptop (Cupertino, CA). The stimuli were positioned at eye level at about 80 cm from the participant and each word subtended approximately 3 degrees of visual angle. Sounds were played from the speakers of a MacBook Pro 15.4 at 80% loudness using the Psychtoolbox in MATLAB (Brainard, 1997). We used the USB-1208FS-Plus device from Measurement Computing Corporation (Norton, Massachusetts) to send trigger pulses that enabled us to align stimuli onsets and behavioral responses to neural recordings.

Minimal Phrase Task: A schematic of the task is shown in **Figure 1.1**. Participants were presented two words, 875 ms presentation time, with a 400 ms blank screen between them. At the end of each trial, participants were asked to indicate via a button press whether the two words were same or different. Word presentation was either visual or auditory. On average, we presented 1500 ± 710 trials (**Table S1** shows the number of trials per participant).

There were three types of trials: Noun followed by Adjective (42% of trials, e.g., "apple green"), Adjective followed by Noun (42% of trials, e.g., "green apple"), Repeated Noun (8% of trials, e.g., "apple apple"), and Repeated Adjective (8% of trials, e.g., "green green"). The order of trials (stimulus presentation modality and noun/adjective structure) was randomly interleaved. Each word combination was presented in a randomized manner 5 times in the audio modality and 5 times in the visual modality. The nouns belonged to two categories, animals (e.g., "cat") and food (e.g., "apple"). The adjectives belonged to two categories, concrete adjectives (e.g., "big") and abstract adjectives (e.g., "good"). A list of all the nouns and adjectives is included in **Table S3**. We selected only high frequency English words that were more frequent than 10^{-6} in Google Ngram and were shorter than 7 letters and had no more than 1 or

2 syllables. We used the max frequency of a word between 2006 and 2019. Finally, we created a balanced selection of nouns and adjectives such that noun and adjectives were indistinguishable from each other using word length or number of syllables ($p>0.05$ ranksum test). We conducted the experiment in 3 languages, English (16 monolingual and 1 bilingual participants), Spanish (1 bilingual participant) and Taiwanese (3 monolingual participants). Two bilingual international scholars whose native language was Spanish (MAG) and Taiwanese (YLK) translated the words in the task. For non-English languages, we also kept nouns and adjectives indistinguishable based on word-length and number of syllables.

Participants had to indicate whether the two words in a trial were the same or not. The motor responses were the same for nouns or adjectives. The motor responses were also the same for noun followed by adjective or adjective followed by noun trials. Thus, the motor responses were orthogonal to parts of speech and grammar and differences between nouns and adjectives cannot be attributed to motor signals.

Sentence Task: A schematic of the task is shown in **Figure 1.17 a**. Participants were presented four-word sentences. There was a 600 ms fixation, followed by four words with 875 ms presentation time. After the last word there was a 1s delay with gray screen and then an image was presented.

There were two types of trials: semantically correct (50% of trials, e.g., “the girls ate cakes”), and semantically incorrect (50% of trials, e.g., “the cakes ate girls”). Thus, the semantically incorrect/odd sentences were formed by swapping the nouns of the correct sentences, without changing the grammatical correctness of the sentence. The order of trials (stimulus presentation modality and semantically correct/incorrect structure) was randomly interleaved. The subjects were instructed to indicate via a button-press whether the sentence described the image (green button) or not (red button), ignoring notions of singular or plural. Note, an accurately described image was only possible for semantically correct sentences when an image corresponding to the sentence was presented on the screen (60% trials of the semantically correct sentences with related images that were described by the sentence, 40% unrelated images which were not described by the sentence). For the semantically incorrect/odd sentences, there was no image that the sentence described accurately (i.e., the response was always the red button). However, for incorrect sentences, 60% sentences were followed by images that were related to the sentence but not accurately described by it (e.g., the sentence “the cakes ate girls” followed by an image of a girl eating a cake). The remaining 40% sentences were followed by unrelated images (e.g., the sentence “the cakes ate girls” followed by an image of a dog chasing a ball). We conducted the experiment in 3 languages, English (14 monolingual), Spanish (3 bilingual participants) and Hindi (1 monolingual participant). See **Table S10** for example sentences.

3.4 Data Analyses

Preprocessing

A total of 2,428 electrode contacts were implanted, 627 of which were excluded from analysis due to bipolar referencing, presence of line noise or recording artifacts (Wang et al., 2021). We removed 60 Hz line noise and its harmonics using a fifth-order Butterworth filter. We focus on the high-gamma band of the intracranial field potential signals obtained by bandpass filtering raw data of each electrode in the 65–150

Hz range (fifth-order Butterworth filter). The high gamma band (65-150 Hz) power was computed using the Chronux toolbox (Mitra and Bokil, 2008). We used a time-bandwidth product of 3 and 4 leading tapers, a moving window size of 200 ms, and a step size of 5 ms. For every trial, we computed the normalized high gamma activity by subtracting the mean activity from -150 to 50 ms from the onset of the first fixation and then dividing by the standard deviation. This normalized response is reported as "gamma power" on the y-axis when showing electrode responses.

Responsive Electrodes

We evaluated whether an electrode was responsive to visual or auditory stimuli by comparing the 100 to 400 ms post stimulus onset to the -400 to -100 ms before stimulus onset (e.g., **Figure 1.2**). The responsiveness threshold was set using Cohen's d prime coefficient and based on the number of trials for a statistical power of 80% and $p<0.01$ (one-tailed z-test). We also computed the time at which the neural signals reach half of the maximum amplitude.

Part-of-speech selectivity

We compared the neural responses to nouns versus adjectives. Periods of significant selective activation were tested using a one-tailed t-test with $p<0.05$ at each time point to differentiate between nouns and adjectives and were corrected for multiple comparisons with a Benjamini-Hochberg false detection rate (FDR) corrected threshold of $q<0.05$, separately for auditory and visual trials. After fixing the FDR with $q<0.05$, an electrode was considered to be selective for part of speech if there was a significant difference between nouns and adjectives for a minimum contiguous window of 65 ms.

General Linear Model (GLM)

Minimal Phrase Task: We created a GLM to tease out the experiment variables that significantly contribute for explaining the responses of a given channel. The equation for a GLM is as follows:

$$\begin{aligned} AUC = \beta_0 + \beta_{NvsA} NvsA + \beta_{GvsUG} GvsUG + \beta_{NumberOfSyllables} NumberOfSyllables \\ + \beta_{WordLength} WordLength \end{aligned} \tag{3.1}$$

where AUC is the area under the response curve (e.g., **Figure 1.4 a**) from 200 ms to 800 ms after the onset of word1 and word2, β_0 is a constant additive term, NvsA is 1 for Nouns and -1 for Adjectives, GvsUG is 1 for Grammatical trials and -1 for Ungrammatical trials, NumberOfSyllables is 1 or 2 (and 0 for visual trials), or WordLength goes from 3 to 7 (and 0 for auditory trials) as the task predictors. We fit this GLM model for each electrode separately using the MATLAB function fitglm and report the corresponding β coefficients (e.g., **Figure 1.4 j**). We assessed whether each coefficient was significantly different from zero when compared to β coefficients generated from shuffled labels (p<0.01, corrected for multiple comparisons).

Sentence Task: The equation for a GLM is as follows:

$$\begin{aligned} AUC = \beta_0 + \beta_{NvsV} NvsV + \beta_{SvsNS} SvsNS + \beta_{NumberOfSyllables} NumberOfSyllables \\ + \beta_{WordLength} WordLength \end{aligned} \tag{3.2}$$

where AUC is the area under the response curve (e.g., **Figure 1.17 b**) from 200 ms to 800 ms after the onset of word1 and word2, β_0 is a constant additive term, NvsV is 1 for Nouns and -1 for Verbs, SvsNS is 1 for Semantically correct trials and -1 for incorrect trials, NumberOfSyllables is 1, 2 or 3 (and 0 for visual trials), or WordLength goes from 3 to 10 (and 0 for auditory trials) as the task predictors.

Anatomical comparisons

To assess the degree of anatomical specificity in the neural responses, we compared the percentage of significant electrodes in each brain region to the null distribution expected given the number of electrodes in each area using a permutation test ($p<0.01$, 10^6 iterations). A similar approach was followed to compare the same region between the left and right hemispheres.

Decoding Analysis

We performed a machine learning decoding analysis (Bansal et al., 2014) to decode parts of speech in individual words combining all the electrodes in each brain region as defined by the Desikan-Killiany atlas (Desikan et al., 2006) (**Figure 1.12**). The top-N principal components of all electrodes that explained more than 70% of the variance in the training data for the area under curve of non-overlapping 100 ms time-windows of the signal following word onset were used for decoding. The signal for decoding comprised of features from different frequency bands (beta: 12-30 Hz, low gamma: 30-65 Hz, and high gamma power: 65-150 Hz). The analysis was repeated for 30 random splits of the data with 80% of the data used for training a Support Vector Machine with a linear kernel. Significant decoding performance was found by comparing performance from the original data at each time-window with a null distribution obtained by shuffling labels ($p<0.01$, ranksum test). Regions with statistically significant decoding performance were found by comparing the average of the maximum decoding performance across time for 30 random iterations of the original data with that of the null distribution, separately for both hemispheres ($p<0.01$, ranksum test corrected for multiple comparisons) (**Figure 1.11 c,f,i**, **Figure 1.12 c**, **Figure 1.13 c,f,i,l,o,r**). We also applied a threshold such that for a given region R

$$[\mu_R - 3\sigma_R]_{OriginalData} > [\mu_R - 3\sigma_R]_{NullData} \tag{3.3}$$

where μ and σ represent the average and standard deviation in region R. For the significant regions, the average max-performance between the left and right hemispheres was compared to find if decoding performance was lateralized ($p<0.01$, ranksum test, corrected for multiple comparisons).

"A preposition is a word you mustn't end a sentence with!"

Berton Braley

4

Audiovisual Sentences

A remarkable hallmark of language is its universality. This universality goes beyond the comprehension of individual words and facilitates a robust interpretation of sentence meaning and grammar. We can comprehend the sentence *the boys play soccer* when uttering it, writing it, listening to it, reading it and even when examining a picture. In previous experiments, I demonstrated multimodal and invariant representations for parts of speech (POS) – the building blocks of grammar (Misra et al., 2024). There have been invasive neurophysiological studies examining other aspects of multimodal and invariant representations for memories (Quian Quiroga et al., 2005) and semantic word retrieval (Forseth, 2018). Recent, neurophysiological have examined grammatical processing (Ding et al., 2016), and single neuron responses reflecting the word meaning based on their specific sentence context and independent of their phonetic form. (Jamali et al., 2024). However, multimodal representations for grammar and semantics processing of sentences remain unknown.

4.1 Results

To address these gaps, I designed a task where I presented sentences to patient participants in both auditory and visual modalities. These sentences were composed of four words and belonged to three categories:

1. Grammatically and Semantically Correct (GS): *the girls ate cakes*,
2. Semantically Incorrect (NS): *the cakes ate girls*,
3. Grammatically Incorrect (NG): *the ate girls cakes*

A schematic of the task is shown in **Figure 4.1**. Participants were presented four-word sentences. There was a 600 ms fixation, followed by four words with 875 ms presentation time for each word. After the last word there was a 1s delay with gray screen and then an image was presented. The participants were asked to indicate via a button press whether the sentence accurately described the image or not.

I recorded intracranial field potentials from 1,563 electrodes (844 in gray matter, 719 in white matter) implanted in 17 patients via stereoelectroencephalography. Participants heard (auditory modality) or read (visual modality) four-word sentences that were sequentially presented (**Figure 4.1**, **Methods**). To assess comprehension, participants were asked to indicate whether the sentence adequately described an image that followed the last word after a 1,000 ms interval. Participants performed the task correctly on 86±13% of the trials. I considered three types of sentences, semantic (e.g., “the girls ate cakes”, called GS sentences), non-semantic (e.g., “the cakes ate girls”, called NS sentences) and ungrammatical (e.g. “the are girls cake”, called NG sentences). All electrode locations are shown in **Figure 1.16** (see also **Table S11**, **Methods**).

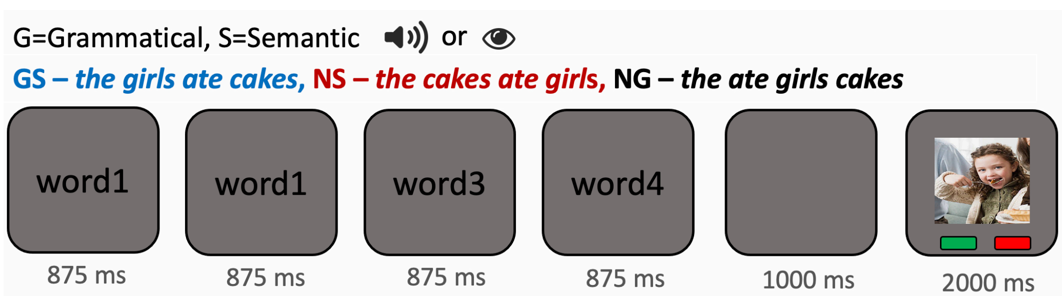


Figure 4.1: **Sentence Task Design.** Sentences comprising four words sequentially presented either in visual or auditory modality were followed by an image. The sentences were semantically correct (1/3rd GS sentences, e.g., “the girls ate cakes”), non-semantic (1/3rd NS sentences, e.g., “the cakes ate girls”) or non-grammatical (1/3rd NG sentences, e.g., “the ate girls cakes”). Participant were instructed to indicate via a button press whether the sentence described the image accurately or not (**Methods 4.4.1**).

4.1.1 Multimodal Neural Signals Distinguish Semantic and Non-semantic sentences

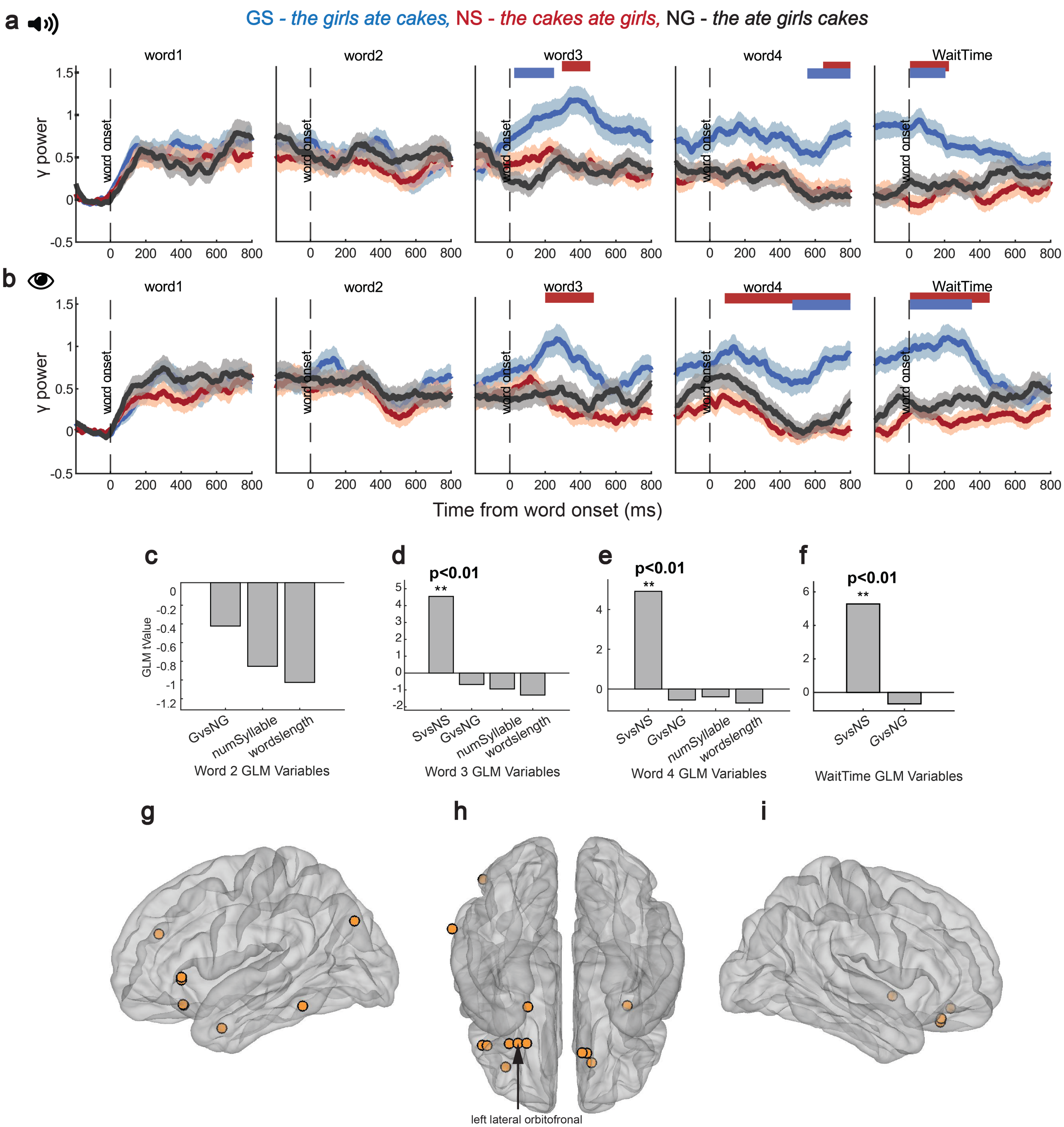
I evaluated whether neural signals differentiated between semantically correct and incorrect sentences. **Figure 4.2** shows the responses of an example electrode located in the left lateral orbitofrontal cortex (**Figure 4.2 h** depicts the electrode location). The neural responses are aligned to the word onset (vertical dashed line) for auditory presentation (**Figure 4.2 a**) or visual presentation (**Figure 4.2 b**) for each trial. This electrode showed multimodal responses triggered by both auditory and visual stimuli. The responses to semantically correct sentences (blue) were stronger than incorrect (red) and ungrammatical (black) sentences for visual and auditory stimuli. A total of 125 electrodes showed a difference across any of the 8 conditions word2 onset onwards. Out of them 37 were multimodal, 46 were audio only and 42 were visual only. The 37 electrodes cannot be ascribed to randomly sampling from the total of audio and visual electrodes ($p<10^{-4}$, permutation test, $n=10^6$ iterations).

4.1.2 Neural selectivity for semantically distinguishable sentences was robust to word properties and phrase grammar

I asked whether grammar, auditory properties (like number of phonemes) and orthographic properties (like word length) could contribute to the neural differences between semantically correct and incorrect sentences. To address these questions, I built a generalized linear model (GLM) for each electrode to predict its response AUC between 200 ms and 800 ms after each word onset using the following predictors: semantically correct or not, grammatically correct or not, and word length (vision) or number of syllables (audition) (**Methods 4.4.1**). I did not include the semantically correct or incorrect label as a predictor during word2 and included it after the onset of word3 because the semantically correct versus incorrect meaning of NS sentences becomes obvious only after the onset of word3 and cannot be distinguished at word2.

The predictor coefficients in the GLM model for the example electrode in **Figure 4.2 a,b** show that only the semantically correct versus incorrect label significantly explained the neural responses for both auditory and visual presentation (**Figure 4.2 c-f**). A total of 13 electrodes showed semantically correct versus incorrect as the only statistically significant predictor in the GLM analysis (**Methods** 4.4.1); 13/13 (100%) of these electrodes distinguished semantically correct versus incorrect for both auditory and visual inputs, such as the example electrode in **Figure 4.2 a,b**.

Figure 4.2 g-i shows the locations of these 13 electrodes (see also **Table S15 S16**).



a
GS - the girls ate cakes, NS - the cakes ate girls, NG - the ate girls cakes
word1
word2
word3
word4
WaitTime
γ power
word onset
b
Time from word onset (ms)
c
GLM tValue
GvsNG
numSyllable
wordslength
Word 2 GLM Variables
d
p<0.01
SvsNS
Word 3 GLM Variables
e
Word 4 GLM Variables
f
WaitTime GLM Variables
g
h
i
left lateral orbitofronal

Figure 4.2: **a,b.** Trial-averaged normalized gamma-band power of responses from an example electrode in the left lateral orbitofrontal cortex (see location in **h**) to semantically correct (GS: blue; one example sentence of each kind is shown in **a**), incorrect (NS: red) or ungrammatical (NG: black) sentences during presentation of auditory stimuli (**a**, n=440 trials) or visual stimuli (**b**, n=446 trials) aligned to the onset (vertical dashed line) of each word and the wait period before image onset. Shaded areas denote s.e.m. Horizontal colored lines denote windows of statistically significant differences between responses to nouns versus adjectives (t-test $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There is a significant difference between the semantically correct (GS) and incorrect (NS) conditions (red horizontal bars), and between semantically correct (GS) and ungrammatical (NG) sentences (blue horizonal bars) following word3, word4, and wait-time onset in auditory (**a**) and visual (**b**) presentations of stimuli.
c-f. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset for word2 (**c**), word3 (**d**), word4 (**e**), or wait-time (**f**). I used the following labels for prediction: Grammatically correct versus incorrect (GvsNG), Semantically correct versus incorrect (SvsNS), number of syllables and word length. Note that the semantically incorrect meaning of NS sentences becomes obvious after the onset of word3. Thus, I included SvsNS as a predictor word3 onwards. This electrode showed SvsNS as the only significant predictor (**d-f**) (**Methods 4.4.1**).
g-i. All electrodes that showed audiovisual differences between GS and NS, or GS and NG and had SvsNS as the only significant predictor during word2, word3, word4 or wait time (Methods, $p<0.01$, Bonferroni corrected). The electrodes are shown on the lateral view (**g,i**) for the left (**g**, n=7) and right hemispheres (**i**, n=6). The example electrode is shown on the inferior view for both hemispheres with an arrow (**h**, n=13, see also **Table S15 16**). This electrode was in the left lateral orbitofrontal cortex.

4.1.3 Multimodal Neural Signals Distinguish Grammatically Correct and Incorrect sentences

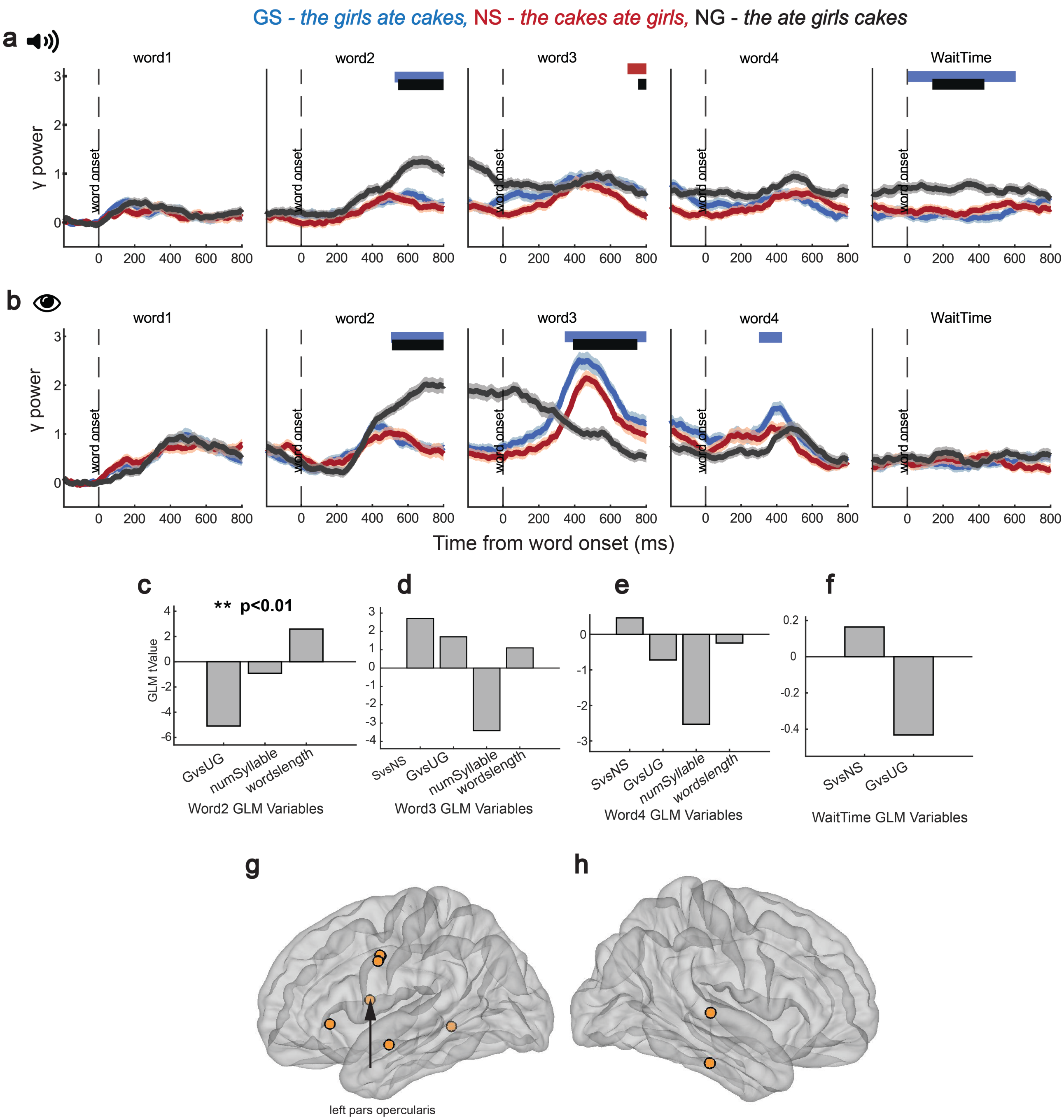
I evaluated whether neural signals differentiated between grammatically correct and incorrect sentences. **Figure 4.3** shows the responses of an example electrode located in the left pars opercularis (**Figure 4.3 h** depicts the electrode location). The neural responses are aligned to the word onset (vertical dashed line) for auditory presentation (**Figure 4.3 a**) or visual presentation (**Figure 4.3 b**) for each trial. This electrode showed multimodal responses triggered by both auditory and visual stimuli. The responses to grammatically incorrect sentences (black) were stronger than semantically correct (blue) and semantically incorrect but grammatically correct (red) sentences for visual and auditory stimuli.

4.1.4 Neural selectivity for grammatically distinguishable sentences was robust to word properties and semantics

I asked whether semantic coherence, auditory properties (like number of phonemes) and orthographic properties (like word length) could contribute to the neural differences between grammatically correct and incorrect sentences. To address these questions, I built a generalized linear model (GLM) for each electrode to predict its response AUC between 200 ms and 800 ms after each word onset using the following predictors: semantically correct or not, grammatically correct or not, and word length (vision) or number of syllables (audition) (**Methods 4.4.1**). I included semantically correct or not as a predictor following the onset of word3 because the semantically correct versus incorrect meaning of NS sentences becomes obvious after the onset of word3.

The predictor coefficients in the GLM model for the example electrode in **Figure 4.3 a,b** show that only the grammatically correct versus incorrect label significantly explained the neural responses for both auditory and visual presentation (**Figure 4.3 c**). A total of 10 electrodes showed grammatically correct versus incorrect as the only statistically significant predictor in the GLM analysis (**Methods** 4.4.1); 10/10 (100%) of these electrodes distinguished grammatically correct versus incorrect for both auditory and visual inputs, such as the example electrode in **Figure 4.3 a,b**.

Figure 4.3 g,h shows the locations of these 10 electrodes (see also **Table S17 S18**).



GS - the girls ate cakes, NS - the cakes ate girls, NG - the ate girls cakes
a
word1
word2
word3
word4
WaitTime
γ power
word onset
0 200 400 600 800
b
Time from word onset (ms)
c
** p<0.01
GLM tValue
GvsUG
numSyllable
wordslength
Word2 GLM Variables
d
SvsNS
Word3 GLM Variables
e
Word4 GLM Variables
f
WaitTime GLM Variables
g
h
left pars opercularis

Figure 4.3:
a,b. Trial-averaged normalized gamma-band power of responses from an example electrode in the left pars opercularis (see location in **g**) to semantically correct (GS: blue; one example sentence of each kind is shown in **a**), incorrect (NS: red) or ungrammatical (NG: black) sentences during presentation of auditory stimuli (**a**, n=438 trials) or visual stimuli (**b**, n=432 trials) aligned to the onset (vertical dashed line) of each word and the wait period before image onset. Shaded areas denote s.e.m. Horizontal colored lines denote windows of statistically significant differences between responses to nouns versus adjectives (t-test $p<0.05$, Benjamini-Hochberg false detection rate, $q<0.05$). There is a significant difference between the semantically correct (GS) and ungrammatical (NG) sentences (blue horizontal bars), and between semantically incorrect (NS) and ungrammatical (NG) sentences (black horizonal bars) at word2 in auditory (**a**) and visual (**b**) presentations of stimuli. There is a significant difference between the semantically correct (GS) and incorrect (NS) conditions (red horizontal bars) following word3 for auditory stimuli only (**a**).
c-f. Z-scored β coefficients for Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset for word2 (**c**), word3 (**d**), word4 (**e**), or wait-time (**f**). I used the following labels for prediction: Grammatically correct versus incorrect (GvsNG), Semantically correct versus incorrect (SvsNS), number of syllables and word length. Note that the semantically incorrect meaning of NS sentences becomes obvious after the onset of word3. Thus, I included SvsNS as a predictor word3 onwards. This electrode showed GvsNG as the only significant predictor (**c**) (**Methods**).
g-h. All electrodes that showed audiovisual differences between GS and NS, or GS and NG and had GvsNG as the only significant predictor during word2, word3, word4 or wait time (Methods, $p<0.01$, Bonferroni corrected). The electrodes are shown on the lateral view for the left (**g**, n=8) and right hemispheres (**h**, n=2) (see also **Table S17 S18**). This electrode was in the left pars opercularis.

4.2 Supplementary Tables

Subject	Age	Gender	Language	Handedness	#Trials	%Correct	#Electrodes
1	13	M	HI	R	457	50%	134
2	15	F	TW	R	942	79%	35
3	19	M	EN	R	901	97%	174
4	37	F	EN	R	913	97%	78
5	40	M	EN	R	906	99%	73
6	21	M	TW	L	904	80%	64
7	30	M	TW	R	952	70%	78
8	42	F	EN	R	909	78%	154
9	27	F	EN	R	882	94%	62
10	32	F	EN	R	900	84%	141
11	30	F	EN	L	906	97%	72
12	25	M	EN	L	900	84%	92
13	33	F	EN	R	900	97%	65
14	50	F	EN	R	972	96%	117
15	20	M	EN	R	862	86%	65
16	20	F	EN	R	909	78%	81
17	20	F	EN	R	907	95%	78
						TOTAL	1563

Table S14 | Information about each participant for the sentence task including age, gender, language (ENglish, HIndi, TaiWanese), handedness, number of trials, behavioral performance and number of electrodes.

subject#	#SvsNS
3	2
4	2
5	2
7	1
14	6
TOTAL	**13**

Table S15 | Distribution of multimodal electrodes that showed the SvsNS label as the only significant predictor following onset of word2, word3, word4 or wait time, across participants.

Region	Total	Left	Right
'parstriangularis'	2	2	0
'medialorbitofrontal'	3	0	3
'insula'	1	0	1
'entorhinal'	1	1	0
'inferiortemporal'	1	1	0
'rostralmiddlefrontal'	1	1	0
'lateralorbitofrontal'	3	1	2
'inferiorparietal'	1	1	0
TOTAL	**13**	**7**	**6**

Table S16 | Distribution of multimodal electrodes on the Desikan-Killiany Atlas that showed the SvsNS label as the only significant predictor following onset of word2, word3, word4 or wait time.

subject#	#GvsNG
3	1
4	1
6	1
12	3
14	4
TOTAL	**10**

Table S17 | Distribution of multimodal electrodes that showed the GvsNG label as the only significant predictor following onset of word2, word3, word4 or wait time, across participants.

Region	Total	Left	Right
'parstriangularis'	1	1	0
'middletemporal'	2	1	1
'superiortemporal'	2	1	1
'precentral'	4	4	0
'parsopercularis'	1	1	0
TOTAL	**10**	**8**	**2**

Table S18 | Distribution of multimodal electrodes on the Desikan-Killiany Atlas that showed the GvsNG label as the only significant predictor following onset of word2, word3, word4 or wait time.

4.3 Discussion

To the best of my knowledge, these are the first invasive results showing multimodal signals reflecting semantics (**Figure 4.2**) and grammar processing (**Figure 4.3**) for sentences.

In addition, this is the first study to demonstrate a role for the frontal orbital cortex in tracking the semantic content of sentences. Combining the medial and lateral orbitofrontal cortices, 46% of the electrodes distinguishing semantically correct versus incorrect sentences were in the orbital regions (see **Table S16**). Previous works on PPA (Mesulam et al., 2014, 2022, 2015) have shown the orbital regions (especially the left LOF and frontal pole) and the left LATL to be linked with word and sentence comprehension deficits. However, I tread carefully in interpreting the results for this brain region because the orbital region captures a vast real estate within the brain. It has been shown to be involved in a variety of functions such as working memory (Kringelbach, 2005), cognitive control (Xiao et al., 2023), memory consolidation (Geva-Sagiv et al., 2023), and reward expectation encoding (Kringelbach, 2005). Given the design of my task, it is possible that the subjects' neural responses reflect the future uncertainty in classifying the impeding image for semantically correct sentences (60% accurately described images and 40% unrelated images, **Methods 4.4.1**). This uncertainty is not present for semantically incorrect or ungrammatical sentences. Despite 60% related images and 40% unrelated images for these sentences, the answer is always the same, i.e. “not accurately described”, indicated via a button press (**Methods 4.4.1**). However, if such is this future uncertainty was indeed involved then one would expect to see differences between grammatically correct and incorrect stimuli at Word2 for the SvsNS encoding electrodes in **Figure 4.2 g-i** (for this to be true the blue and red curves should be elevated compared to the black curve at Word2 reflective of future uncertainty, which is not observed). This is because for both grammatically correct sentence types future uncertainty in equally likely when compared to ungrammatical sentences. But, I did not observe any such differences and the electrode responeses resembled the example electrode in **Figure 4.2**.

On the contrary, the ambiguity I face in interpreting my findings for semantically correct versus incorrect are absent for grammatically correct versus ungrammatical comparisons. Thus making the findings more compelling. The motor and anticipatory responses between NS (semantically incorrect) and NG (grammatically incorrect) sentences are the same. Still, I observed elevated and distinguishable responses only for the grammatically incorrect sentences (see **Figure 4.3**).

In summary, these findings are the first demonstration of multimodal and robust representations for multimodal semantics and grammar. The results I present are very new, and they require further ironing out before making further claims.

4.4 Methods

4.4.1 Design

There were three types of trials: semantically correct (GS, 1/3rd of trials, e.g., “the girls ate cakes”), semantically incorrect (NS, 1/3rd of trials, e.g., “the cakes ate girls”) or ungrammatical (NG, 1/3rd of trials, e.g., “the ate girls cakes”). Thus, the semantically incorrect/odd sentences were formed by swapping the nouns of the correct sentences, without changing the grammatical correctness of the sentence. The order of trials (stimulus presentation modality, and the semantically and grammatically correct/incorrect structure) was randomly interleaved. The subjects were instructed to indicate via a button-press whether the sentence described the image (green button) or not (red button), ignoring notions of singular or plural. Note, an accurately described image was only possible for semantically correct sentences when an image corresponding to the sentence was presented on the screen (60% trials of the semantically correct sentences with related images that were described by the sentence, 40% unrelated images which were not described by the sentence). For the semantically incorrect/odd sentences, there was no image that the sentence described accurately (i.e., the response was always the red button). However, for incorrect sentences, 60% sentences were followed by images that were related to the sentence but not accurately described by it (e.g., the sentence “the cakes ate girls” followed by an image of a girl eating a cake). The remaining 40% sentences were followed by unrelated images (e.g., the sentence “the cakes ate girls” followed by an image of a dog chasing a ball). I conducted the experiment in 3 languages, English (14 monolingual), Spanish

(3 bilingual participants) and Hindi (1 monolingual participant). See **Table S10** for example sentences.

4.4.2 Participants

I recorded data from 17 participants (7 male, 13-50 years old, 3 left-handed, **Table S9**) with pharmacologically resistant epilepsy. All experiments were conducted while participants stayed at Children's Hospital Boston (CHB), Cleveland Clinic, Ohio, or Taipei Veterans General Hospital (TVGH). All studies were approved by each hospital's institutional review boards and were carried out with the participants' informed consent.

4.4.3 Recordings and Electrode Locations

Participants were implanted with intracranial depth electrodes (Ad-Tech, Racine, WI, USA). Neurophysiological data were recorded using XLTEK (Oakville, ON, Canada), Bio-Logic (Knoxville, TN, USA), Nihon Kohden (Tokyo, Japan), and Natus (Pleasanton, CA, USA). The sampling rate was 2048 Hz at BCH and TVGH, and 1024 Hz or 512 Hz at BWH. All data were referenced in a bipolar montage. There were no seizure events in any of the sessions. Electrode locations were decided based on clinical criteria for each participant. Electrodes in the epileptogenic foci, as well as pathological areas, were removed from analyses. The total number of electrodes after bipolar referencing and removing electrodes with no signal, line noise or recording artifacts was 1,801.

Following implantation, electrodes were localized by co-registration of pre-operative

T1 MRI and post-operative CT scans using the iELVis software (Groppe et al., 2017). I used FreeSurfer to segment MRI images, upon which post implant CT was rigidly registered (Desikan et al., 2006). Electrodes were marked in the CT aligned to pre-operative MRI using the Bioimage Suite (Joshi et al., 2011, Dale et al., 1999). The Desikan-Killiany (DK) atlas was used to assign the electrodes locations. **Figure 1.16 a-f** and **Table S11** show the locations of all the electrodes.

4.4.4 Experiment Presentation

All visual stimuli were displayed on a 15.4 inch 2,880 × 1,800 pixel LCD screen using the Psychtoolbox in MATLAB (Natick, MA) and a MacBook Pro laptop (Cupertino, CA). The stimuli were positioned at eye level at about 80 cm from the participant and each word subtended approximately 3 degrees of visual angle. Sounds were played from the speakers of a MacBook Pro 15.4 at 80% loudness using the Psychtoolbox in MATLAB (Brainard, 1997). I used the USB-1208FS-Plus device from Measurement Computing Corporation (Norton, Massachusetts) to send trigger pulses that enabled us to align stimuli onsets and behavioral responses to neural recordings.

4.4.5 Data Analysis

Grammar & Semantic Selectivity

I compared the neural responses to the three sentence categories. Periods of significant selective activation were tested at each word using a one-tailed t-test with $p<0.05$ at each time point to differentiate between GS, NS and NG were corrected for multiple comparisons with a Benjamini-Hochberg false detection rate (FDR) corrected threshold of q<0.05, separately for auditory and visual trials. After fixing the FDR with q<0.05, an electrode was considered to be selective for sentence category if there was a significant difference between GS, NS and NG for a minimum contiguous window of 100 ms.

General Linear Model (GLM)

The equation for a GLM was defined for each word as follows. At Word2 of English sentences it was :

$$\begin{aligned} AUC = \beta_0 + &\beta_{GvsNG} GvsNG + \beta_{NumberOfSyllables} NumberOfSyllables \\ &+ \beta_{WordLength} WordLength \end{aligned} \tag{4.1}$$

where AUC is the area under the response curve (e.g., **Figure 4.2 c**) from 200 ms to 800 ms after the onset of word2, β_0 is a constant additive term, GvsNG is 1 for grammatically correct trials and -1 for non-grammatical trials, NumberOfSyllables is 1, 2 or 3 (and 0 for visual trials), or WordLength goes from 3 to 10 (and 0 for auditory trials) as the task predictors. I include SvsNS as a predictor word3 onwards because the semantically incorrect meaning of NS sentences becomes obvious after the onset of word3.

For Word3 and Word4 in English sentences, the GLM was defined as :

$$\begin{aligned} AUC = &\beta_0 + \beta_{SvsNS} SvsNS + \beta_{GvsNG} GvsNG \\ &+ \beta_{NumberOfSyllables} NumberOfSyllables \\ &+ \beta_{WordLength} WordLength \end{aligned} \tag{4.2}$$

where AUC is the area under the response curve (e.g., **Figure 4.2 d,e**) from 200 ms to 800 ms after the onset of word3 and word4, β_0 is a constant additive term, SvsNS is 1 for semantically correct trials and -1 for semantically incorrect trials and 0 for ungrammatical trials, GvsNG is 1 for grammatically correct trials and -1 for non-grammatical trials, NumberOfSyllables is 1, 2 or 3 (and 0 for visual trials), or WordLength goes from 3 to 10 (and 0 for auditory trials) as the task predictors.

For wait time period before picture onset, the GLM was defined as :

$$AUC = \beta_0 + \beta_{SvsNS} SvsNS + \beta_{GvsNG} GvsNG \tag{4.3}$$

where AUC is the area under the response curve (e.g., **Figure 4.2 f**) from 200 ms to 800 ms after the onset of wait time, β_0 is a constant additive term, SvsNS is 1 for semantically correct trials and -1 for semantically incorrect trials and 0 for ungrammatical trials, GvsNG is 1 for grammatically correct trials and -1 for non-grammatical trials, as the task predictors.

Semantic versus non-semantic encoding electrodes

An electrode was defined as encoding for semantic versus non-semantic information if it satisfied:

1. It showed a selectivity between GS versus NS, or GS versus NG sentences.

2. In addition, it must show a significant contribution of only the SvsNS predictor at word3, word4, or wait time, but not for other predictors (p<0.01, corrected, Bonferroni corrected for the number of selective electrodes and across the three

words).

Grammatically correct versus incorrect encoding electrodes

An electrode was defined as encoding for grammatically correct versus incorrect information if it satisfied:

1. It showed a selectivity between GS versus NS, or GS versus NG sentences.
2. In addition, it must show a significant contribution of only the GvsNG predictor at word3, word4, or wait time, but not for other predictors ($p<0.01$, corrected, Bonferroni corrected for the number of selective electrodes and across the three words).

5
Conclusion

My ultimate aim was to understand how the very organ that created language actually accomplishes this feat. I wanted to piece together the internal code with which the brain communicates with itself—through neural firing and voltage—to gain a clearer picture of language representations within the brain.

5.1 Audiovisual Language Processes with Intracranial recordings

During my rotations, I realized that questions about language processes in the brain required controlled experiments that had the necessary statistical power to contrast key theories about linguistic processes across a variety of controls (Chomsky et al., 2019, Misra et al., 2024). Prof. Kreiman's lab had expertise in analyzing and interpreting the neural code with which the brain talks to itself. I wanted to utilize this opportunity to unravel the mysteries of language processing in the brain.

Thus, I created an audiovisual experiment that had the necessary controls to study minimal phrase processing across a variety of critical controls. This work led to the evidence for an invariant POS representation in the human brain (Misra et al., 2024). The also results of my initial work became the bulk of my thesis. I also extended this work my by designing another experiment to study multimodal grammar and meaning processing in sentences.

5.2 Technical Summary of Thesis

5.2.1 Chapter 1

Here, I describe the results of my work on minimal phrases. I found evidence for a multimodal, robust and invariant representation for parts-of-speech in the left lateralorbitofronal cortex. For millennia (Panini 500 BC, Sakaṭayana 814–760 BCE, *Nirukta Texts*, 2nd millenium B.C. (Mondal, 2020)) *, linguists have decomposed language into words with defined functions called parts-of-speech, like nouns, verbs, and

*Panini refers to other grammarians older by 5 centuries. Only glimpses of their works are known and the bodies remain lost.

adjectives. Several theories have proposed that these concepts are represented in the brain (Chomsky, 1995). Our work provides rigorous neurobiological evidence from invasive neural recordings in support of Parts-of-Speech in the brain.

5.2.2 Chapter 2

I discuss how to interpret the results for POS and how my findings complement previous studies of language. In addition, I provide lots of details about the LOF and describe how LOF is related to a variety of multimodal processes and diseases that can result in language aphasias. I also discuss the shortcomings of my experiment design, and electrode coverage.

5.2.3 Chapter 3

I provide the pre-registration document and detail the methods used in this thesis.

5.2.4 Chapter 4

This part of the thesis describes the recent work to study multimodal grammar and sentence processing in sentences. I collected data from 1,563 electrode contacts from 17 participants in a record time of 6 months. I learned how to forge and lead collaborations with neurosurgeons to advance human cognitive neuroscience. Finally, I showed neurophysiological evidence for multimodal semantic and grammar processing in sentences.

5.2.5 Appendix A

This section describes the findings from the BrainTreeBank dataset.

I completed a rotation with Prof. Boris Katz and Prof. Kreiman, in collaboration with Dr. Andrei Barbu and Adam Yaari, at MIT, which led to the creation of the BrainTreeBank (see Appendix A). The primary objective was to acquire neural data while patients watched movies to study naturalistic language stimuli. The BrainTreeBank is a large-scale dataset of electrophysiological neural responses recorded from intracranial probes while 10 subjects watched one or more Hollywood movies. On average, subjects watched 2.6 Hollywood movies, with a total viewing time of 4.3 hours and an aggregate of 43 hours. The audio track for each movie was transcribed with manual corrections. Word onsets were manually annotated on spectrograms of the audio track for each movie. Each transcript was automatically parsed and manually corrected into the Universal Dependencies (UD) formalism, assigning a part of speech to every word and a dependency parse to every sentence. In total, subjects heard 36,000 sentences (205,000 words), while they had an average of 1,670 electrodes implanted. I collected data from 6 out of 10 participants for this study, performed the preliminary preprocessing, and created an outline for the analysis in collaboration with Andrei and Adam. Eventually, Chris Wang took over the project and brought it to completion.

5.3 Reflection: Methods Advancement for Human Neuroscience

Across the three projects that I worked on, involving intracranial recordings (1. Minimal Phrases, 2. Audiovisual Sentence Task, and 3. BrainTreeBank), I interacted extensively with the clinical environment. I recognized different players that make electrophysiological studies of human cognition possible: patient participants, doctors, nurses, research collaborators (such as the Kreiman Lab), regulatory staff, research programs, and funding bodies. This extended interaction brought me into contact with the goals of the different people involved, especially the patients and doctors.

The goals of researchers and clinicians are complementary. The study of human cognition has the potential to inform diagnoses and treatment. A recent review article (Lee et al., 2024) discusses the potential use of ex-vivo brain tissue from epilepsy and stroke surgeries to link function and cell types, especially for single neuron studies. I think a better solution would be to design electrodes with microtubules embedded within them [†]. These microtubules can be used to extract brain tissue in the form of extracellular or intracellular fluid via capillary action, circumventing the need for surgical resections to be the "rate-determining step" for downstream sequencing studies. Another advantage of using microtubules with electrode contacts to sequence brain tissue is that it will preserve the neural composition within the vicinity of the electrode contacts, which can get mixed with other brain tissues during removal [‡] The neural fluid extracted from the microtubules can then be used with modern sequencing technologies to create cell type identities that contain not only the anatomical and sequencing coordinates but also the functional coordinates, as identified by cognitive tasks. This can pave the way towards identifying proteins and genes that are involved in the cellular or tissue identities associated with particular brain functions.

[†]These novel ideas were discussed with my colleagues in the Kreiman Lab, especially with Elisa Pavarino. I am grateful for the inspiring discussions.

[‡]The number of sEEG surgeries is greater than the number of single neuron recordings for patients, and one could always extend the microtubules to single cells.

Such information can be used to design neuro-pharmacological or neurogenetic therapies for a variety of neurodegenerative disorders. These functional coordinates can be combined with cell atlases to identify upstream and downstream regions connected with the functional cell and anatomical identities.

A

Brain TreeBank

This work done in collaboration with Andrei Barbu, Chris Wang and Adam Yaari at MIT over a rotation project.

Brain Treebank: Large-scale intracranial recordings from naturalistic language stimuli

Christopher Wang*[1,2], Adam Yaari*[1,2], Aaditya K Singh[1,5], Vighnesh Subramaniam[1,2], Dana Rosenfarb[1,2], Jan DeWitt[1], Pranav Misra[2,3,4], Joseph R Madsen[4], Scellig Stone[4], Gabriel Kreiman[2,3,4], Boris Katz[1,2], Ignacio Cases[1,2], and Andrei Barbu[1,2]

[1]Computer Science and Artificial Intelligence Laboratory, MIT
[2]Center for Brains, Minds and Machines, MIT
[3]Center for Brain Science, Harvard University
[4]Boston Children's Hospital, Harvard Medical School
[5]Gatsby Computational Neuroscience Unit, University College London

Abstract

We present the Brain Treebank, a large-scale dataset of electrophysiological neural responses, recorded from intracranial probes while 10 subjects watched one or more Hollywood movies. Subjects watched on average 2.6 Hollywood movies, for an average viewing time of 4.3 hours, and a total of 43 hours. The audio track for each movie was transcribed with manual corrections. Word onsets were manually annotated on spectrograms of the audio track for each movie. Each transcript was automatically parsed and manually corrected into the universal dependencies (UD) formalism, assigning a part of speech to every word and a dependency parse to every sentence. In total, subjects heard 36,000 sentences (205,000 words), while they had on average 167 electrodes implanted. This is the largest dataset of intracranial recordings featuring grounded naturalistic language, one of the largest English UD treebanks in general, and one of only a few UD treebanks aligned to multimodal features. We hope that this dataset serves as a bridge between linguistic concepts, perception, and their neural representations. To that end, we present an analysis of which electrodes are sensitive to language features while also mapping out a rough time course of language processing across these electrodes. The Brain Treebank is available at `https://BrainTreebank.dev/`

1 Introduction

A single theory of language understanding that encompasses how our brains process language, how linguists understand language, and how machines process language is still beyond our reach. Despite numerous attempts to understand how the brain processes language through investigations of compositionality [1–4], semantic categories [5–7], and surprisal [8–12], a mechanistic understanding of language processing the brain is also lacking. Our hypothesis is that this is in part because studies of language processing in the brain often focus on small data regimes, since gathering large-scale neural recordings can be extremely laborious. Yet NLP and ML research in general has shown that scale matters. In particular, even probing experiments on artificial networks require fairly large scale to produce reliable results, certainly larger than a few hundred sentences [13, 14]. NLP would

Submitted to the 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks. Do not distribute.

Data	Quantity	Data	Quantity
Total subjects	10	Total sentences	36,433
Total hours	43.5	Unique sentences	27,988
Total electrodes	1,688	Avg. words per sentence	6.5
Avg. electrodes per subject	167	Total words	205,670
Total movies	21	Unique words	12,069
Unique movies	26	Unique speakers	937
Number of scenes	46,935	Unique part of speech labels	17

Table 1: Quantitative overview of Brain Treebank

not have progressed without large-scale resources, so to enable the same kind of progress and new discoveries we collect a new large-scale neuroscience dataset, which has naturalistic stimuli, is multimodal, and uses intracranial recordings — a high-spatial and high-temporal resolution recording method.

The Brain Treebank is foremost a treebank, like the Penn Treebank, annotated in the universal dependencies (UD) format. What distinguishes it, is that it is accompanied by both multimodal annotations and by neural recordings collected from 10 subjects who heard 205,670 annotated words while they watched Hollywood films. Subjects watched a total of 26 films (55 hours) as data was recorded from a total of 1,688 electrodes. To this, we add manual and automated annotations.

Scene labels: every scene in the movie was labeled according to the Places365 schema, [15], resulting in 46,935 scenes total. **Word onsets and offset**: while automatic speech recognition performs acceptably, errors are common which were manually corrected. In addition, automated systems are simply not trained to offer extreme accuracy, at the millisecond level, when determining the start and end of words. Word onsets had to be manually annotated on spectrograms for every word to ensure alignment with the neural recordings. **Part of speech tags and parses**: Sentences were automatically parsed into the Universal Dependencies framework and then each part of speech tag and dependency relationship was manually corrected. While POS tagging is fairly accurate, numerous parser errors existed. **Speaker identity**: A unique identifier, which can be traced back to a given character, was given to every speaker in every movie. This was done manually as no automated system exists to do so with any reasonable accuracy. Finally, we also curated a list of 16 automated video, audio, and language features that we provide to save computing time (see table 4). We release all our data with a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Large scale stimuli for the neuroscience of language and multimodal understanding can enable natural experiments: the kind of post-hoc analysis of large-scale datasets that has propelled NLP and machine learning in general forward. In the long term we hope that treebanks such as ours coupled with neural recordings will help the creation of theories of language understanding that span linguistics, neuroscience, and NLP. To demonstrate the utility of the dataset, in addition to providing the raw data, we also take new steps toward understanding language in the brain; our contributions are:

1. A dataset of intracranial recordings across 26 different movie viewings (43 hours total).
2. Localization of electrode positions and alignment with common brain atlases.
3. Multiple layers of manual annotations to enable numerous experiments: scene labels, word onsets and offsets, part of speech tags, parses in universal dependencies format, and speaker identity.
4. Multiple automated annotations for 16 other language, audio, and visual features.
5. Quantitative results that show neural responsiveness to word onset and differential activation based on the position of a word within a sentence.

2 Related work

Previous works have studied language processing in the context of Magnetoencephalography (MEG) [17, 18], Electroencephalography (EEG) [19, 20], and functional magnetic resonance imaging (fMRI) [21–26]. In this work, we present data with both high temporal resolution and naturalistic stimuli.

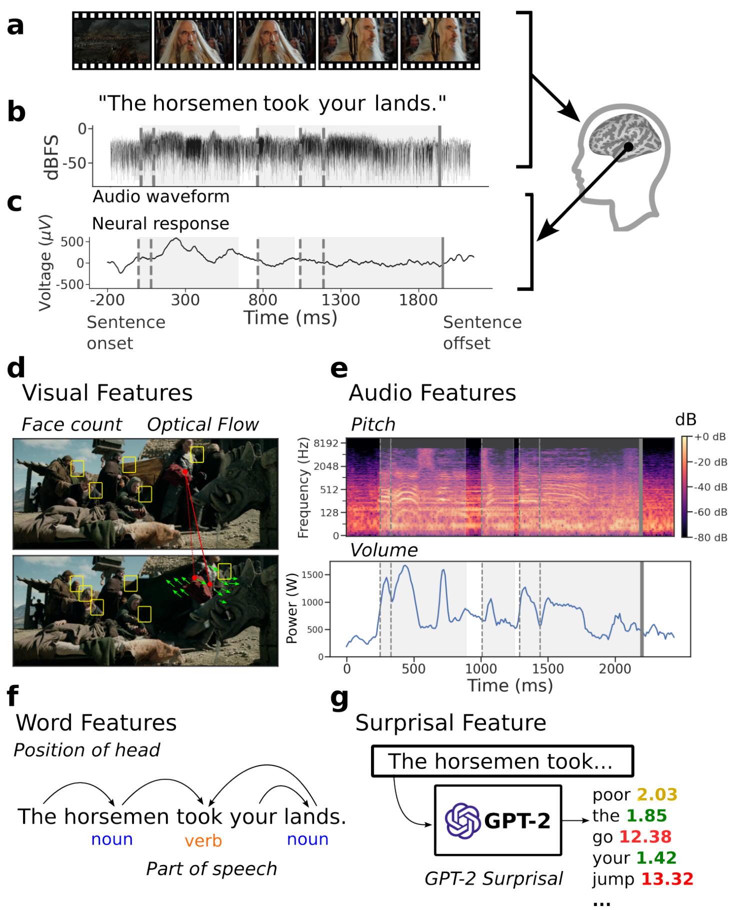


Figure 1: **Schematic of the approach. Top**: A film (**a-b**) was presented as visual and audio stimulus to the subject. Invasive neural recordings were performed while subjects watched the movie. A transcript of speech in the film is aligned to both the audio (**b**) and neural (**c**) signals. Shown here is a short signal segment from an exemplar electrode in the left superior temporal gyrus aligned to sentence onset at $t = 0$ ms. Word locations are shown as shaded regions between dashed lines. **Bottom**: Schematic overview of selected visual (**d**), audio (**e**), and language (**f-g**) features used for the General Linear Model (GLM) for each word. See Table 4 for a full list and description of features. Visual features **(d)** include the number of faces (yellow boxes), and the magnitude and angle of optical flow (green arrows). Audio features **(e)** include the average pitch (top) and volume (bottom) during each word (shaded gray regions between dashed lines). Word features **(f)** include part-of-speech and the position of each word's dependency head. A surprisal feature, **(g)**, computed using GPT-2 [16], a large language model, is the negative log probability of the word given its sentence context.

Recording the brain's response to naturalistic stimuli is critical to neuroscientific progress [27]. There exist fMRI datasets for naturalistic speech [28–30], vision [31, 32], and movies [33, 34]. And similar data has been collected for the EEG modality: speech [35], vision [36], and movies [37]. There are also movie datasets that cover both modalities [38]. However, when it comes to intracranial recordings, which provide better temporal resolution, but require invasive surgery to implant probes, data is much more sparse. There exist intracranial datasets for pose [39], speech production [40], and parts of speech [41], but none of these involve the complex natural language and concomitant visual inputs available from movie stimuli. The most similar work to our dataset, Berezutskaya et al. [42], presents participants with vastly less stimuli: a 6.5 minute short movie, compared to our average of 5.5 hours of movie per patient.

Already, the brain-recordings themselves, without annotation have proven useful for representation learning, as we show in Wang et al. [43]. And combined with the transcribed audio tracks, these have allowed for successful study of multimodal integration in the brain, as we show in Subramaniam et al. [44]. Now, for the first time, we release the complete annotated recordings for all subjects, as well as the accompanying Universal Dependency parse trees.

3 Data

Dataset construction Stereoelectroencephalography (SEEG) neural recordings were collected from 10 subjects (5 male, 5 female), aged 4-19 years ($\mu \approx 11.9$, $\sigma \approx 4.6$), under treatment for epilepsy at Boston Children's Hospital (BCH); see supplementary table 2 for per-subject statistics. All subjects were implanted with intracranial electrodes to localize seizure foci for potential surgical resection. All experiments were approved by BCH/Harvard IRB and were carried out with the subjects' informed consent. IRB documents are available upon request, but are otherwise sensitive. Electrode types, number, and position were driven solely by clinical considerations. Recorded data was anonymized, and identifying patient information was redacted.

Task and stimuli Stimuli consisted of 21 recent animated/action Hollywood movies; see supplementary table 3 for per-movie statistics. On average, movies were 2.07 hours long ($\sigma \approx$0.68) and contained 1,322 sentences ($\sigma \approx$303), 8,927 total words ($\sigma \approx$2104), 1,769 unique words ($\sigma \approx$324), 1,358 unique lemmas ($\sigma \approx$259), 1,219 nouns ($\sigma \approx$282), 615 unique nouns ($\sigma \approx$133), 1,334 verbs ($\sigma \approx$299), and 504 unique verbs ($\sigma \approx$100). Each subject was given a choice of which movies to watch, viewing an average of 2.6 movies ($\sigma \approx$1.7) corresponding to 4.3 hours ($\sigma \approx$3.6). For further details, see appendix A.3.

Data acquisition and signal processing Clinicians implanted subjects with intracranial stereo-electroencephalographic (sEEG) depth probes containing 6-16 0.8 mm diameter 2 mm long contact electrodes recording Intracranial Field Potentials (IFPs). Each subject had multiple (12 to 18) such probes implanted in locations determined by clinical concerns entirely unrelated to the experiment, informed by a functional analysis [45]. The number of electrodes per subject ranged between 106 and 246 ($\mu \approx$167, $\sigma \approx$40) for a total of 1,688 total electrodes; see Extended Figures table 2 for a per-subject breakdown. Data collected during periods of seizures or immediately following a seizure was discarded. For each electrode, a notch filter was applied at 60 Hz and harmonics. No other processing (downsampling, filtering specific frequency bands, etc.) was performed on the neural recordings. For further details, see appendix A.4. Finally, the location of all electrodes was identified and mapped to the common brain atlases Desikan et al. [46] and Destrieux et al. [47]. See appendix A.5 for further details.

Audio transcription and alignment For each movie, the timestamps for all words in the audio were transcribed and timestamps for each word were found programmatically and then manually corrected by trained annotators (see appendix A.1 for further details). The pipeline developed for this audio transcription and alignment effort is an independently useful source of annotated stimuli, which can now be used for further experiments. We described this pipeline more completely in a separate technical paper: Yaari et al. [48]. Part of speech tags and dependency parses were manually corrected and speaker identity and scene labels were manually annotated from scratch by an in-house expert hired at MIT.

Feature annotation To model the neural responses during the complex movies, we considered a series of 16 features (table 4). These features include 6 visual attributes (pixel brightness, global optical flow magnitude, global optical flow angle, optical flow magnitude, optical flow angle, and number of faces, fig. 1d), 4 auditory attributes (volume, mean pitch, delta volume, and delta pitch, fig. 1e), and 6 language attributes (GPT-2 surprisal, word time length, word time difference, index in sentence, word head, and part of speech tag, fig. 1f-g). All of these features were aligned to and computed for each word. Table 4 provides a brief description of each feature, and their calculation is described in appendix A.2. Additionally, scenes and speakers were labeled for each movie. Scenes were extracted based on camera cuts using PySceneDetect [49]. Each scene was labeled based on the corresponding image environment and labels were extracted from the Places365 dataset [15].

Finally, for each sentence in the audio transcript, the speaker identity was manually annotated (see appendix A.2).

4 Quantitative analyses of language function with the dataset

Word onsets triggered strong neural responses After aligning the neurophysiological data to the occurrence of words (fig. 1a-c), we assessed whether the neural responses were modulated by word onset by comparing the mean activity in 5 consecutive windows of 100ms duration before (-500 ms to 0 ms) versus after (500 to 1000ms) word onset. We defined an electrode to be *word-responsive* if it yielded a statistically significant difference in at least one of the 5 windows (paired t-test, p<0.05, Bonferroni corrected, see appendix A.6). Figure 2a shows the neural responses of an example electrode located in the left superior temporal sulcus (fig. 2a inset). The raster plot (top) and average activity (bottom) show strong activation triggered by the onset of each word. This activity can be readily appreciated for almost every word in the more than 6,000 words (raster plot) of a single movie. Interestingly, the activity of this electrode begins to show a slight deviation from baseline *before* the onset of words at time 0.

The complex nature of natural stimuli like the movie implies that multiple variables could in principle drive the neural responses. Indeed, the responses to individual words in fig. 2a show a strong degree of heterogeneity. To gain insight into what could drive these diverse responses, we considered a set of 16 visual, auditory, and language features (table 4, fig. 1d-g). We built a Generalized Linear Model (GLM) that included all 16 features. The coefficients for each feature indicated how much each annotation contributes to explaining the neural responses (fig. 2b). For this example electrode, visual, auditory, and language features all showed a statistically significant contribution to explaining the neural response. The strongest contributors were the four language features shown in fig. 2d-g. The average of all coefficients across the 339 electrodes in the temporal lobe is shown in fig. 2c, for which we note that the features with the highest averaged coefficients were the index in sentence, part of speech, and delta volume (further regions shown in fig. 8).

To better understand the contribution of the four language features with the largest coefficients for the example electrodes, we plotted the neural responses for words that had different values for those coefficients. In fig. 2d, we separated the words of a movie into those that appeared early in a sentence (quartile with lowest index in sentence, light gray) and those that appeared late in a sentence (quartile with highest index in sentence, dark gray). The average neural responses for this example electrode revealed notable differences between these two groups of words. Common to both groups, there was a deflection from baseline well before t=0. Words with high indices led to reduced voltages and words with low indices led to high voltages after t=0. In a similar fashion, we observed responses separated by nouns versus verbs (fig. 2e), high and low word length (fig. 2f), and high and low GPT-2 surprisal (fig. 2g). In all of these cases, words elicited neural responses across different features even as the neural responses were modulated by those features. Similar conclusions for this electrode can be drawn when considering auditory features (supplementary fig. 4a) or visual features (supplementary fig. 4b).

Next, we asked whether the neural responses are due purely to language, or whether an audio and/or visual explanation can be ruled out. Across all electrodes, we found that there exist 251 ($\approx 16\%$) electrodes for which there was a significant ($p < 0.05$, Bonferroni corrected) word response, after controlling for all audio and visual features. The fraction of such electrodes per region is shown in fig. 2h and the locations of these electrodes are shown in fig. 2i.

Sentence position modulates neural activity The results for the example electrode in fig. 2d suggest that the position of a word within a sentence can have a strong impact on the neural responses. To systematically evaluate whether neural signals are dependent on word position, we first categorized words according to their linear position (fig. 3a), separating them into sentence *onsets*, sentence *offsets*, and sentence *midsets*, which are the words that occur in between. Figure 3a shows the neural responses from an example electrode located in left superior temporal gyrus. This electrode showed stronger responses to sentence onsets (left) compared to midsets (middle) and offsets (right). These

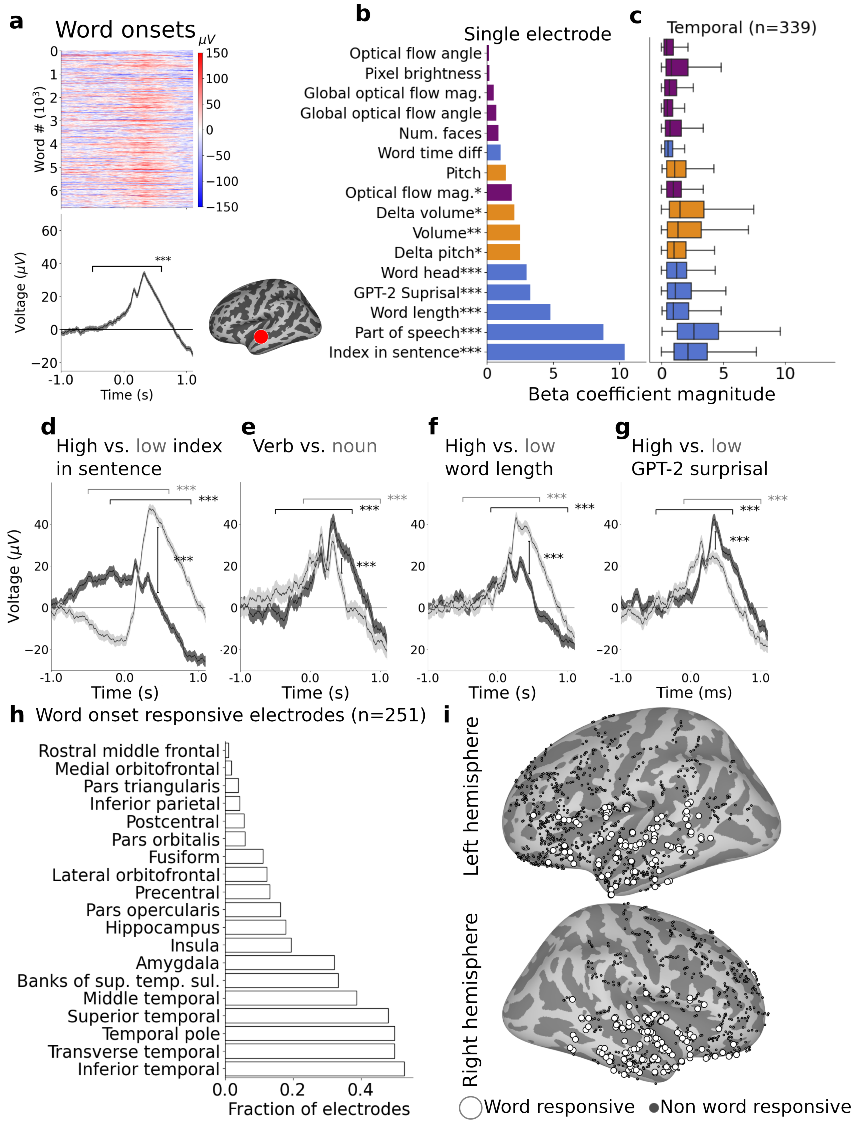


Figure 2: **Alignment to word onsets reveals strong neural responses. a.** Raster (top) and mean (bottom) plots of neural activity aligned to word onsets ($t = 0$ ms) for an exemplar electrode (inset; shown in red) in the left superior temporal sulcus. Each line in the raster is a separate word ($> 6,000$ words) in the movie. Shading in the mean plot indicates standard error. Asterisks indicate the significance (double-tailed paired t-test) of the response, measured by comparing mean activity in pre- and post- word-onset intervals (see section 4). A GLM was fitted to predict the average response in the 500ms window after word onset (section 4). The magnitude of the beta coefficients for all features is shown for the same example electrode (**b**) and averaged across all electrodes in the temporal lobe (**c**). Features are shown colored by category (blue: language, orange: audio, purple: visual). Asterisks indicate statistical significance of the beta coefficient for the example electrode (see section 4). Neural responses are shown from the same example electrode separated by **(d)** index in sentence, **(e)** part of speech, **(f)** word length, **(g)** GPT-2 surprisal. Asterisks on horizontal brackets indicate the significance of the neural response, i.e., the difference between pre- and post- word-onset activity, as in (**a**). Vertical brackets show the differences in mean sub-sampled activity (see section 4). In **h**, the fraction of electrodes per regions for which a significant word-onset response can be observed even after sub-sampling for visual and audio features is shown. The precise location of these electrodes is shown in **i**.

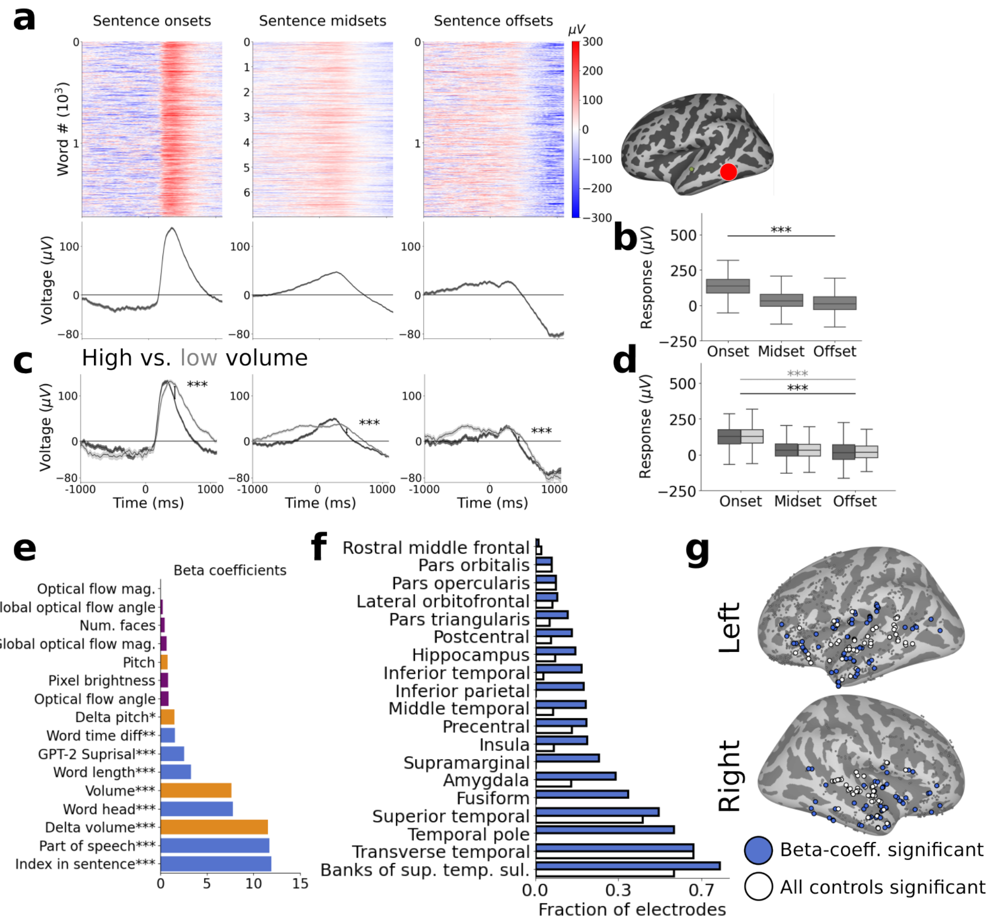


Figure 3: **Neural signals distinguish between different positions within the sentence. a.** Raster (top) and mean (bottom) neural responses for an example electrode in the left superior temporal gyrus (see electrode location on right) for words occurring at sentence onset (left), offset (right), or in between (midset, middle). The format and conventions follow fig. 2a. The box-plots (**b**) show the mean activity in a 100ms window. Asterisks show the significance of the difference between activities (f-test, Bonferroni corrected). **c.** Neural responses from the same electrode separated by trials with high volume (dark grey) or low volume (light grey). Vertical brackets and asterisks show the difference between the two conditions (two-tailed t-test). In both cases, the difference due to sentence position persists (shown by horizontal brackets and asterisks in **d**). **e.** Beta coefficients from a fitted GLM for all features, colored by category (format as in fig. 2b). Coefficients shown here are for the same electrode as in **(a)**. **f.** Per region, the fraction of electrodes (shown as blue bars) in each region for which there is a significant ($p < 0.05$, Bonferroni corrected) beta coefficient for position in sentence and the fraction of electrodes (white bars) which exhibit a significant ($p < 0.05$, f-test, Bonferroni corrected) difference in activity due to sentence position after controlling for all confounds. **g.** The exact location of the electrodes from f, shown as blue and white points respectively, projected onto the surface of the brain.

differences were evident even in single words (raster plots, top), as well as in the average responses (bottom) and are summarized in fig. 3b, which shows the mean neural activity for onsets, midsets, and offsets in a 100ms window. Similar to our analysis in the previous section, we evaluated mean neural activity at five evenly spaced 100ms windows, starting from the word onset. The activity shown in fig. 3b was taken from the window with the most significant difference between onset, midset, and offset activity (f-test, $p < 0.05$, Bonferroni corrected). Asterisks in fig. 3b denote the significance of this difference.

It might be the case that sentence onsets could be associated with a confounding feature, such as increased volume. We therefore separately plotted the responses to words in different sentence positions for cases with high and low volume. The strong modulation by part of sentence persisted across different volume levels (fig. 3c-d). Next, in addition to volume, we considered all the 16 features that we annotated in the movie, using a GLM model as illustrated in the previous section. The feature with the highest coefficient in the GLM model was the index in sentence (fig. 3e). Running the GLM analysis for all electrodes revealed 242 electrodes (15.3% of total electrodes) for which the sentence position feature has a significant ($p < 0.05$, Bonferroni corrected) beta coefficient in the fitted GLMs (fig. 3f, fig. 3g blue dots).

Among these electrodes which we identified to be modulated by position in sentence, we also used a different, more stringent test to determine the influence that the position in sentence has on mean activity. For each of these electrodes, the analysis that was discussed previously with respect to fig. 3c-d was repeated for all features. Across these electrodes, controlling for all co-occurring features, revealed 117 electrodes (6.5% of total electrodes) that showed a significant modulation by sentence position (f-test, $p < 0.05$, Bonferroni corrected). These electrodes were predominantly located in the transverse temporal cortex and the banks of the superior temporal sulcus (fig. 3f, fig. 3g white dots).

The temporal-course of speech decodability reveals the dynamics of language processing We also used a linear decoder to answer questions about when and where certain language induced activity is available in the neural signal. To that end, we trained a decoder for every 250ms interval in a [-1000ms,1000ms] window. As discussed in the previous sections, we had observed language responses to be stronger at sentence onsets, so we first considered the case of trying to decode whether or not a sentence onset was occurring. However, the case for generic word onsets was also considered (see supplementary fig. 2), and is discussed below as well.

For each electrode, we created a training dataset of neural activity (see appendix A.8). Positive examples consisted of sentence onsets and negative examples were taken from portions of the movie where no dialogue is occurring. We train our model per electrode, and evaluate using 10-fold cross validation. Decoding performance for a given electrode is then the average ROC-AUC, where the mean has been taken across cross validation folds. Figure 4a. shows the *peak* decoding performance per electrode. Here, the peak performance is the maximum performance achieved over the course of the entire considered time interval. Figure 4c shows the decoding per time interval in the temporal and frontal lobe, averaged across the 10 electrodes with the highest peak decoding performance on the train set. In the frontal region, decoding peaked later than in the temporal region (300ms vs 100ms). We performed the same decoding for generic word onsets (see fig. 2). Here we found a similar pattern as in figure fig. 4. Decoding in the temporal lobe reached a peak at 100ms, compared to 400ms in the frontal lobe.

5 Conclusion

The Brain Treebank has a unique combination of large scale, high temporal resolution, high spatial resolution, naturalistic stimuli, and may layers of manual annotation. Because naturalistic stimuli contain many uncontrolled co-occurring features, scale is critical in order to find natural experiments with controls post-hoc. We demonstrate two such an experiments: first, how response to words and sentences can be identified, even after controlling for co-occurring features, and second, how linear decoding reveal the time course of word and sentence processing. This only begins to explore what

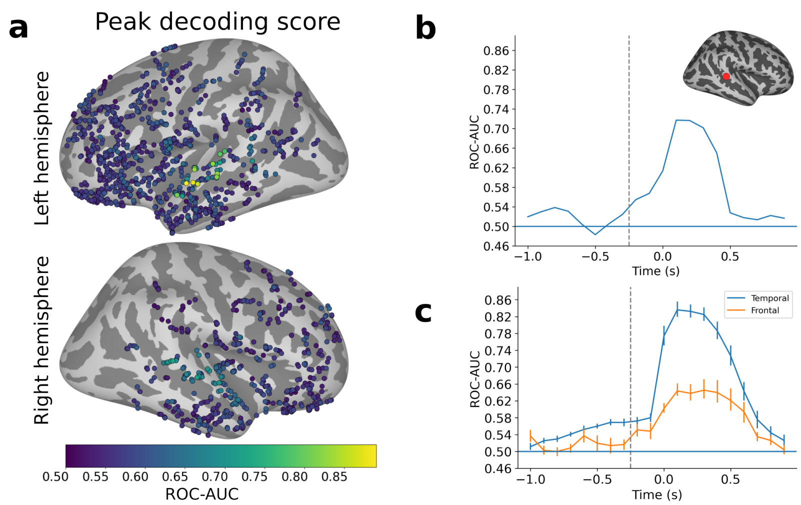


Figure 4: **Sentence onsets are linearly decodable.** A linear decoder is trained to classify portions of the movies according to whether or not a sentence onset is occurring, based on the corresponding neural activity. This decoding is done for activity in 0.25s windows, shifted in 0.1s increments, from -1s before the sentence onset to 1s after the sentence onset. The *peak* decoding performance for an electrode is the max ROC-AUC achieved across all increments. **a.** The spatial distribution of peak decoding scores. **b.** Decodability, as a function of time for an electrode in the banks of the superior temporal sulcus on the right hemisphere. **c.** The time course of decodability on the test set, for the top 10 electrodes that had the highest peak ROC-AUC score on the train set, in the temporal lobe and the frontal lobe. The test set is balanced between positive and negative examples so that chance performance is 0.5. Together, these curves reveal that for sentence onsets, information is processed before word onset enters the decoding window (dashed grey line). Error bars show a 95% confidence interval over performance per electrode. Comparing the curves reveals the mirrored time course of language processing in the frontal and temporal lobes. See supplementary fig. 2 for the same analysis, performed for word-onsets.

can be done with these data and annotations, and it remains to be seen what is detectable if more powerful decoding tools are applied.

Limitations Subjects only watched each movie once, thus one cannot simply average over repetitions of exactly the same stimulus. Although, each movie does repeat the same words and often shows the same characters, naturalistic stimuli are harder to work with than controlled experiments. Subjects all saw different movies making the cross-subject analysis more difficult. At the same time, this means that there are more opportunities to find interesting phenomena because of the diversity of the movies that subjects saw. As with all studies that involve naturalistic stimuli, controlling for confounds can be difficult. Intracranial recordings are only possible because subjects require neurosurgery for some condition, in this case epilepsy; it is possible that this could result in some sampling bias. Additionally, the corpus includes only movies in English, although we are adding Spanish movies and subjects shortly. In this vein, we are actively working on collecting more data and hope that others who intend to collect data can collect it for the movies we have annotated here. Tools and techniques to run experiments on naturalistic data are much newer and more limited at the moment.

We have not begun to scratch the surface of the kinds of analyses possible with the Brain Treebank, for example, we have never used the speaker identities and hardly exploited multimodality, nor have we made use of the parses aside from the POS tags. We hope that the Brain Treebank will enable the development of new tools and new kinds of neuroscientific experiments at scale with natural stimuli. As well as bring the neuroscience, NLP, and linguistics communities closer together with a shared resource that has components from each.

References

[1] Douglas K Bemis and Liina Pylkkänen. Flexible composition: Meg evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PloS one*, 8(9): e73949, 2013.

[2] Esti Blanco-Elorrieta, Itamar Kastner, Karen Emmorey, and Liina Pylkkänen. Shared neural correlates for building phrases in signed and spoken language. *Scientific reports*, 8(1):1–10, 2018.

[3] Marianne Schell, Emiliano Zaccarella, and Angela D Friederici. Differential cortical contribution of syntax and semantics: An fmri study on two-word phrasal processing. *Cortex*, 96: 105–120, 2017.

[4] Masha Westerlund, Itamar Kastner, Meera Al Kaabi, and Liina Pylkkänen. The latl as locus of composition: Meg evidence from english and arabic. *Brain and Language*, 141:124–134, 2015.

[5] Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195, 2008. doi: 10.1126/science.1152876.

[6] Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532 (7600):453–458, 2016.

[7] Yizhen Zhang, Kuan Han, Robert Worth, and Zhongming Liu. Connecting concepts in the brain by mapping cortical representations of semantic relations. *Nature communications*, 11(1):1–13, 2020.

[8] Hale JT Brennan JR. Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE*, 2019. doi: 10.1371/journal.pone.0207741.

[9] Jonathan R. Brennan, Edward P. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John T. Hale. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157-158:81–94, 2016. ISSN 0093-934X. doi: https://doi.org/10.1016/j.bandl.2016.04.008. URL `https://www.sciencedirect.com/science/article/pii/S0093934X15300687`.

[10] Ariel Goldstein, Zaid Kokaja Zada, Eliav Buchnik, Mariano Schain, Amy Rose Price, Bobbi Aubrey, Samuel A. Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, Aren Jansen, Harshvardhan Gazula, Gina Choe, Aditi Rao, Catherine Kim, Colton Casto, Fanda Lora, Adeen Flinker, Sasha Devore, Werner K. Doyle, Daniel Friedman, Patricia Dugan, Avinatan Hassidim, Michael P. Brenner, Y. Matias, Kenneth A. Norman, Orrin Devinsky, and Uri Hasson. Thinking ahead: spontaneous prediction in context as a keystone of language in humans and machines. *bioRxiv*, 2020.

[11] Harm Brouwer, Francesca Delogu, Noortje J. Venhuizen, and Matthew W. Crocker. Neurobehavioral correlates of surprisal in language comprehension: A neurocomputational model. *Frontiers in Psychology*, 12, 2021. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.615538. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2021.615538`.

[12] Shohini Bhattasali and Philip Resnik. Using surprisal and fMRI to map the neural bases of broad and local contextual prediction during natural language comprehension. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3786–3798, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-acl.332. URL `https://aclanthology.org/2021.findings-acl.332`.

[13] Ian Tenney, Dipanjan Das, and Ellie Pavlick. Bert rediscovers the classical nlp pipeline. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4593–4601, 2019.

[14] Yena Han, Tomaso A Poggio, and Brian Cheung. System identification of neural systems: If we got it right, would we know? In *International Conference on Machine Learning*, pages 12430–12444. PMLR, 2023.

[15] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017.

[16] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *Manuscript distributed by OpenAI*, 2019.

[17] Riitta Salmelin. Clinical neurophysiology of language: the meg approach. *Clinical Neurophysiology*, 118(2):237–254, 2007.

[18] Marinella Cappelletti, Felipe Fregni, Kevin Shapiro, Alvaro Pascual-Leone, and Alfonso Caramazza. Processing nouns and verbs in the left frontal cortex: A transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 20(4):707–720, 2008.

[19] Anna M Beres. Time is of the essence: A review of electroencephalography (eeg) and event-related brain potentials (erps) in language research. *Applied psychophysiology and biofeedback*, 42:247–255, 2017.

[20] John Hale, Chris Dyer, Adhiguna Kuncoro, and Jonathan Brennan. Finding syntax in human encephalography with beam search. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2727–2736, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1254. URL `https://aclanthology.org/P18-1254`.

[21] Ned T Sahin, Steven Pinker, and Eric Halgren. Abstract grammatical processing of nouns and verbs in broca's area: evidence from fmri. *Cortex*, 42(4):540–562, 2006.

[22] Rachel L Moseley and Friedemann Pulvermüller. Nouns, verbs, objects, actions, and abstractions: Local fmri activity indexes semantics, not lexical categories. *Brain and language*, 132: 28–42, 2014.

[23] Anna Mestres-Missé, Antoni Rodriguez-Fornells, and Thomas F Münte. Neural differences in the mapping of verb and noun concepts onto novel words. *NeuroImage*, 49(3):2826–2835, 2010.

[24] Manuela Berlingeri, Davide Crepaldi, Rossella Roberti, Giuseppe Scialfa, Claudio Luzzatti, and Eraldo Paulesu. Nouns and verbs in the brain: Grammatical class and task specific effects as revealed by fmri. *Cognitive neuropsychology*, 25(4):528–558, 2008.

[25] Evelina Fedorenko, Alfonso Nieto-Castanon, and Nancy Kanwisher. Lexical and syntactic representations in the brain: an fmri investigation with multi-voxel pattern analyses. *Neuropsychologia*, 50(4):499–513, 2012.

[26] Yasmeen Faroqi-Shah, Rajani Sebastian, and Ashlyn Vander Woude. Neural representation of word categories is distinct in the temporal lobe: An activation likelihood analysis. *Human brain mapping*, 39(12):4925–4938, 2018.

[27] Saurabh Sonkusare, Michael Breakspear, and Christine Guo. Naturalistic stimuli in neuroscience: critically acclaimed. *Trends in cognitive sciences*, 23(8):699–714, 2019.

[28] Amanda LeBel, Lauren Wagner, Shailee Jain, Aneesh Adhikari-Desai, Bhavin Gupta, Allyson Morgenthal, Jerry Tang, Lixiang Xu, and Alexander G Huth. A natural language fmri dataset for voxelwise encoding models. *Scientific Data*, 10(1):555, 2023.

[29] Samuel A Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J Honey, Yaara Yeshurun, Mor Regev, et al. The “narratives” fmri dataset for evaluating models of naturalistic language comprehension. *Scientific data*, 8(1):250, 2021.

[30] Jixing Li, Shohini Bhattasali, Shulin Zhang, Berta Franzluebbers, Wen-Ming Luh, R Nathan Spreng, Jonathan R Brennan, Yiming Yang, Christophe Pallier, and John Hale. Le petit prince multilingual naturalistic fmri corpus. *Scientific data*, 9(1):530, 2022.

[31] Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1):116–126, 2022.

[32] Zhengxin Gong, Ming Zhou, Yuxuan Dai, Yushan Wen, Youyi Liu, and Zonglei Zhen. A large-scale fmri dataset for the visual processing of naturalistic scenes. *Scientific Data*, 10(1): 559, 2023.

[33] Matteo Visconti di Oleggio Castello, Vassiki Chauhan, Guo Jiahui, and M Ida Gobbini. An fmri dataset in response to “the grand budapest hotel”, a socially-rich, naturalistic movie. *Scientific Data*, 7(1):383, 2020.

[34] S Aliko, J Huang, F Gheorghiu, S Meliss, and JI Skipper. A naturalistic neuroimaging database for understanding the brain using ecological stimuli. scientific data. 7, 347, 2020.

[35] Shohini Bhattasali, Jonathan Brennan, Wen-Ming Luh, Berta Franzluebbers, and John Hale. The alice datasets: fmri & eeg observations of natural language comprehension. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 120–125, 2020.

[36] Tijl Grootswagers, Ivy Zhou, Amanda K Robinson, Martin N Hebart, and Thomas A Carlson. Human eeg recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, 9(1):3, 2022.

[37] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015. doi: 10.1109/TAMD.2015.2431497.

[38] Qawi K Telesford, Eduardo Gonzalez-Moreira, Ting Xu, Yiwen Tian, Stanley J Colcombe, Jessica Cloud, Brian E Russ, Arnaud Falchier, Maximilian Nentwich, Jens Madsen, et al. An open-access dataset of naturalistic viewing using simultaneous eeg-fmri. *Scientific Data*, 10(1): 554, 2023.

[39] Nancy Wang, Ali Farhadi, Rajesh Rao, and Bingni Brunton. Ajile movement prediction: Multimodal deep learning for natural human neural recordings and video. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[40] Maxime Verwoert, Maarten C Ottenhoff, Sophocles Goulis, Albert J Colon, Louis Wagner, Simon Tousseyn, Johannes P Van Dijk, Pieter L Kubben, and Christian Herff. Dataset of speech production in intracranial electroencephalography. *Scientific data*, 9(1):434, 2022.

[41] Pranav Misra, Yen-Cheng Shih, Hsiang-Yu Yu, Daniel Weisholtz, Joseph R Madsen, Stone Sceillig, and Gabriel Kreiman. Invariant neural representation of parts of speech in the human brain. *bioRxiv*, pages 2024–01, 2024.

[42] Julia Berezutskaya, Mariska J Vansteensel, Erik J Aarnoutse, Zachary V Freudenburg, Giovanni Piantoni, Mariana P Branco, and Nick F Ramsey. Open multimodal ieeg-fmri dataset from naturalistic stimulation with a short audiovisual film. *Scientific Data*, 9(1):91, 2022.

[43] Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. Brainbert: Self-supervised representation learning for intracranial recordings. In *The Eleventh International Conference on Learning Representations*, 2022.

[44] Vighnesh Subramaniam, Colin Conwell, Christopher Wang, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. Revealing vision-language integration in the brain using multimodal networks. In *International conference on machine learning*. PMLR, 2024.

[45] Sallie Baxendale. The wada test. *Current opinion in neurology*, 22(2):185–189, 2009.

[46] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

[47] Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1): 1–15, 2010. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2010.06.010. URL `https://www.sciencedirect.com/science/article/pii/S1053811910008542`.

[48] Adam Yaari, Jan DeWitt, Henry Hu, Bennett Stankovits, Sue Felshin, Yevgeni Berzak, Helena Aparicio, Boris Katz, Ignacio Cases, and Andrei Barbu. The aligned multimodal movie treebank: An audio, video, dependency-parse treebank. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9531–9539, 2022.

[49] Brandon Castellano. PySceneDetect v0.6.1, 2022. URL `https://github.com/Breakthrough/PySceneDetect`.

[50] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[51] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, 2015.

[52] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*, 2020.

[53] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/2020.emnlp-demos.6`.

[54] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968 – 980, 2006. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2006.01.021. URL `http://www.sciencedirect.com/science/article/B6WNP-4JFHF4P-1/2/0ec667d4c17eafb0a7c52fa3fd5aef1c`.

[55] Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H. Salat, Evelina Busa, Larry J. Seidman, Jill Goldstein, David Kennedy, Verne Caviness,

Nikos Makris, Bruce Rosen, and Anders M. Dale. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22, 2004. doi: 10.1093/cercor/bhg087. URL `http://cercor.oxfordjournals.org/content/14/1/11.abstract`.

[56] Bruce Fischl, David H. Salat, André J.W. van der Kouwe, Nikos Makris, Florent Ségonne, Brian T. Quinn, and Anders M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23(Supplement 1):S69 – S84, 2004. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2004.07.016. URL `http://www.sciencedirect.com/science/article/B6WNP-4DCMGVT-2/2/7eee26326dc63f931b826eac33becc8b`. Mathematics in Brain Imaging.

[57] Bruce Fischl and Anders M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20):11050–11055, 2000.

[58] B. Fischl, A. Liu, and A. M. Dale. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Medical Imaging*, 20(1):70–80, Jan 2001.

[59] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33:341–355, 2002.

[60] Bruce Fischl, Martin I. Sereno, Roger B.H. Tootell, and Anders M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284, 1999. ISSN 1097-0193. doi: 10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4. URL `http://dx.doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4`.

[61] Jorge Jovicich, Silvester Czanner, Douglas Greve, Elizabeth Haley, Andre van der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, Bruce Fischl, and Anders Dale. Reliability in multi-site structural mri studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, 30(2):436 – 443, 2006. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2005.09.046. URL `http://www.sciencedirect.com/science/article/B6WNP-4HM7S0B-2/2/4fa5ff26cad90ba3c9ed12b7e12ce3b6`.

[62] G. R. Kuperberg, M. Broome, P. K. McGuire, A. S. David, M. Eddy, F. Ozawa, D. Goff, W. C. West, S.C.R. Williams, Andre van der Kouwe, David Salat, Anders Dale, and Bruce Fischl. Regionally localized thinning of the cerebral cortex in Schizophrenia. *Archives of General Psychiatry*, 60:878–888, 2003.

[63] H. D. Rosas, A. K. Liu, S. Hersch, M. Glessner, R. J. Ferrante, D. H. Salat, A. van der Kouwe, B. G. Jenkins, A. M. Dale, and B. Fischl. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology*, 58(5):695–701, 2002. URL `http://www.neurology.org/content/58/5/695.abstract`.

[64] David Salat, R.L. Buckner, A.Z. Snyder, Douglas N. Greve, R.S. Desikan, Evelina Busa, J.C. Morris, Anders Dale, and Bruce Fischl. Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14:721–730, 2004.

[65] F. Segonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in mri. *NeuroImage*, 22(3):1060 – 1075, 2004. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2004.03.032. URL `http://www.sciencedirect.com/science/article/B6WNP-4CF5CNY-1/2/33cc73136f06f019b2c11023e7a95341`.

[66] Anders Dale, Bruce Fischl, and Martin I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179 – 194, 1999.

[67] Bruce Fischl, Martin I. Sereno, and Anders Dale. Cortical surface-based analysis: Ii: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195 – 207, 1999.

[68] Xiao Han, Jorge Jovicich, David Salat, Andre van der Kouwe, Brian Quinn, Silvester Czanner, Evelina Busa, Jenni Pacheco, Marilyn Albert, Ronald Killiany, Paul Maguire, Diana Rosas, Nikos Makris, Anders Dale, Bradford Dickerson, and Bruce Fischl. Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32(1):180–194, 2006.

[69] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans Med Imaging*, 17:87–97, 1998.

[70] F. Segonne, J. Pacheco, and B. Fischl. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans Med Imaging*, 26:518–529, 2007.

[71] Martin Reuter, Herminia Diana Rosas, and Bruce Fischl. Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, 53(4):1181–1196, 2010. doi: 10.1016/j.neuroimage.2010.07.020. URL `http://dx.doi.org/10.1016/j.neuroimage.2010.07.020`.

[72] Martin Reuter and Bruce Fischl. Avoiding asymmetry-induced bias in longitudinal image processing. *NeuroImage*, 57(1):19–21, 2011. doi: 10.1016/j.neuroimage.2011.02.076. URL `http://dx.doi.org/10.1016/j.neuroimage.2011.02.076`.

[73] Martin Reuter, Nicholas J. Schmansky, Herminia Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61(4): 1402–1418, 2012. doi: 10.1016/j.neuroimage.2012.02.084. URL `http://dx.doi.org/10.1016/j.neuroimage.2012.02.084`.

[74] David M. Groppe, Stephan Bickel, Andrew R. Dykstra, Xiuyuan Wang, Pierre Mégevand, Manuel R. Mercier, Fred A. Lado, Ashesh D. Mehta, and Christopher J. Honey. ielvis: An open source matlab toolbox for localizing and visualizing human intracranial electrode data. *Journal of Neuroscience Methods*, 281:40–48, 2017. ISSN 0165-0270. doi: https://doi.org/10.1016/j.jneumeth.2017.01.022. URL `https://www.sciencedirect.com/science/article/pii/S0165027017300365`.

[75] A. Joshi, D. Scheinost, H. Okuda, D. Belhachemi, I. Murphy, L. H. Staib, and X. Papademetris. Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics*, 9(1):69–84, Mar 2011.

[76] A. I. Yang, X. Wang, W. K. Doyle, E. Halgren, C. Carlson, T. L. Belcher, S. S. Cash, O. Devinsky, and T. Thesen. Localization of dense intracranial electrode arrays using magnetic resonance imaging. *Neuroimage*, 63(1):157–165, Oct 2012.

[77] Noam Chomsky. *Syntactic structures*. De Gruyter Mouton, 2009.

[78] Liina Pylkkänen. The neural basis of combinatory syntax and semantics. *Science*, 366(6461): 62–66, 2019. doi: 10.1126/science.aax0050.

[79] Evelina Fedorenko, Terri L Scott, Peter Brunner, William G Coon, Brianna Pritchett, Gerwin Schalk, and Nancy Kanwisher. Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, 113(41):E6256–E6262, 2016.

6 Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] We release our full dataset of intracranial recordings.
   (b) Did you describe the limitations of your work? [Yes] See conclusion.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] This work focuses on specification of the released data.
   (b) Did you include complete proofs of all theoretical results? [N/A] See above.
3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] A quickstart guide is available at the URL
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See appendix A.8.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Statistical significance is reported with respect to our GLM and error is reported to reflect variance between individual words (see figs. 2 to 4).
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See appendix A.8
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A] This is a novel dataset collected by the authors.
   (b) Did you mention the license of the assets? [Yes] See section 1.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] The URL is given in the abstract.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See Section 3.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See section 3.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [Yes] See section 3.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] Per the IRB, participation in research was voluntary, and compensation was not permitted.

A Appendix

A.1 Audio transcription and alignment

The audio track of each movie was first annotated by commercial services (`Rev.com` and `HappyScribe.com` depending on the movie) and manually corrected by trained annotators. A custom tool was developed to refine the alignment via an auditory spectrogram of 4 seconds at a time and slowed-down audio track. Annotators were instructed to adjust the onset and offset of every word to align with the spectrogram and their perception of when the word started and ended. The audio annotation tool automatically played the audio segment corresponding to each word to allow annotators to verify their work. As the audio was played a line marked the location of the audio sample in the spectrogram in real time.

Since speech recognizers often misused or missed critical punctuation marks, these were inserted by annotators manually. Sentences were then manually segmented. Annotators were instructed not to use abbreviations, even if they are common. Annotators marked audio segments that consisted of overlapping speech or signing. These were removed from the dataset. All foreign language was marked and removed from the dataset. Annotators were instructed to transcribe literally, i.e, contractions were used in the transcript only when spoken as such. Similarly, foreshortened words, e.g., goin' vs going, were transcribed as such when used by speakers. Cardinal numbers were spelled out. Longer numbers were spelled out as spoken, including conjunctions such as "and". All overheard words were transcribed, even when they could not easily be localized on the spectrogram, for example, short words such as "to" can sometimes be heard but no specific segment of the spectrogram seems to correspond uniquely to such words. In this case annotators were asked to mark their onset and offset as they heard the words. Transcripts are as spoken, without correction, even when the speaker erred omitting a word or using a word inappropriately.

A.2 Feature annotation

We extract 16 features that were included in the analyses (see Extended Figures table 4).

Visual features The visual scene scalar features were extracted from the middle frame presented during a word utterance via OpenCV 4.4.0 [50]. Brightness was quantified as the average pixel HSV value channel. Flow vectors were computed as dense optical flow over grey-scale frames via the OpenCV `calcOpticalFlowFarneback` function (pyramid scale 0.5, 5 levels, window size 11, 5 iterations, pixel neighborhood of 5, and smoothing of 1.1). Number of faces per-frame was estimated via the OpenCV `CascadeClassifier` function with the Haar cascade frontal face default classifiers over gray-scale frames (scale factor: 1.1, minimum neighbours: 4). The first frame of every word utterance was mean-normalized and than passed through a pretrained ResNet-50 object detector (Torchvision 0.6.1) to compute a visual vector image embedding (size 2,048) as the last feature layer of the model.

Auditory features The auditory scalar features were collected with the Python Librosa package (0.7.2) [51], an open source audio analysis library. Sound intensity and mean frequency of the audio track during word utterance were estimated, as well as their change relatively to the preceding 500*ms* window. The average intensity of the audio segment was computed as the root-mean-square (RMS) (`rms` function, frame and hop lengths 2048 and 512 respectively) of that segment. Pitch was extracted using Librosa's `piptrack` function over a Mel-spectrogram (sampling rate 48,000 Hz, FFT window length of 2048, hop length of 512, and 128 mel filters). Auditory vector embeddings were computed as the flattened log-Mel-spectrogram of the 500*ms* word utterance window (size $128 \times 47 = 6016$).

Language features We used a state-of-the-art syntactic parser, Stanford NLP Group's Stanza [52], to parse every sentence. POS tags were recorded for every word. Surprisal was quantified as the negative-log word probability. Word probabilities were estimated by a transformer model. GPT-2 probabilities were computed via GPT-2 large using the Hugging Face Transformers 3.0.0 library [53]. Word particle surprisal were combined by summation.

All Universal Dependency features were inferred using the standard English model of the Stanza Natural Language Processing toolkit [52] and then manually corrected via a single trained annotator over the course of a year.

Speaker annotation Annotators doing speaker identification were instructed to use the characters' full names, insofar as they are known. If a character is unnamed, the annotator may identify them with a brief description of their role.

Occasionally, a character had another identity that they went by. In Spider-Man: Homecoming, the AI in Peter's suit is known for more than half the movie as "suit lady," until Peter finally decides to give her the name "Karen." In such situations, the annotator marked both identities, with whichever identity they decide is primary listed first, and the secondary identity in parentheses. So, in the above example, Peter's AI is annotated as "Karen (suit lady)"

Because of our data set, we deal with quite a lot of super heroes with secret identities. If a super hero was in costume, annotators identified them by their super hero name. Out of costume, they were identified by their birth name. When they are partially in costume (say, they're in costume, but they've taken off their mask), annotators identified them by their super hero name, followed by their birth name, separated by a forward slash: e.g. Spider-Man / Peter Parker

In situations where one character is pretending to be another, the guidelines bear some resemblance to the guidelines for heroes that are partially in costume. Annotators identified them by the person being imitated, followed by the true identity of the character, separated by a percent symbol. So, for a good part of the movie Megamind, the titular character is pretending to be a museum curator named Bernard. Dialog spoken by him during these moments should be annotated as "Bernard % Megamind."

Lines that had problems and therefore that need special attention can be identified using an asterisk. Two of the most common situations where this cropped up were when multiple characters were speaking in unison, or when a "sentence" actually contains utterances from multiple characters. In the former situation, these were identified with the line with `* multiple speakers`. In the latter situation, both speakers were annotated, with an asterisk between them e.g. "Peter Parker * Tony Stark," and an asterisk was added to the line of dialog at the point where one of them stops speaking and the other begins.

A.3 Task and stimuli

Movies were extracted from DVDs and are unchanged other than being re-encoded to a fixed frame rate (23.976 fps). Transcripts, and all annotations described in this work will be made publicly available. Due to copyrights prohibiting the release of the raw stimuli (movies) source material, multiple audio-visual sample clips and tools allowing users to verify alignment of their own movie copies will be publicly provided.

Movies were shown in full to each subject. Movies were displayed via a custom video player created in Matlab 2018b. The player ensured that the presentation was at a fixed frame rate to keep the audio and video synchronized. The presentation of movies was accompanied by regular electrical triggers sent to the neural recording system to enable accurate temporal alignment between the movie and the neural data. A 15.4 inch (resolution 2880×1800) Apple MacBook Pro Retina was placed 60-100cm in front of the subject. Subjects adjusted the volume and paused/resumed the movie as needed. The movie was paused by the experimenter any time someone entered the room or when subjects were distracted and was resumed when subjects could direct their full attention back to the movie. Subjects could freely change position, but were instructed by the experimenter, who watched the movies with the subjects, to remain focused on the stimulus or pause the movie. Subjects did not speak during the presentation of the movie nor did they overhear any other speech other than that found in the movie.

A.4 Data acquisition and signal processing

Clinicians implanted subjects with intracranial stereo-electroencephalographic (SEEG) depth probes containing 6-16 0.8 mm diameter 2 mm long contact electrodes (Ad-Tech, Racine, WI, USA) recording Intracranial Field Potentials (IFPs) with 1.5 mm separation. Each subject had multiple (12 to 18) such probes implanted in locations determined by clinical concerns entirely unrelated to the experiment. Data was recorded using XLTEK (Oakville, ON, Canada) and BioLogic (Knoxville, TN, USA) hardware with a sampling rate of 2048 Hz.

During movie presentation, triggers were sent to a separate channel on the neural recording device via a USB connection to a dedicated trigger box (Measurement Computing USB-1208FS) using the Psychtoolbox 3 Matlab package. Each pulse was logged with both its wall-lock timestamp and its movie timestamp. Individual triggers were sent every $100ms$. Specific events (movie start, pause, resume, and end) were marked by bursts of triggers (10, 8, 9, and 11 respectively) separated by 15ms. All triggers consisted of a 15ms electrical burst at a magnitude of 80mV. An automated tool found triggers and aligned the movie and neural data.

A.5 Cortical surface extraction and electrode visualization

For each subject, pre-operative T1 MRI scans without contrast were processed with FreeSurfer's `recon-all` function with `-localGI`, which performed skull stripping, white matter segmentation, surface generation, and cortical parcellation [54–73]. iELVis [74] was used to co-register a post-operative fluoroscopy scan to the preoperative MRI. Electrodes were manually identified using BioImageSuite [75], and then assigned to one of 68 regions (according to the Desikan-Killiany atlas [46]) using FreeSurfer's automatic parcellation. The alignment to the atlas was manually verified for each subject. One subject had a large frontal lesion in the right hemisphere that prevented alignment to an atlas. Electrodes from this subject were included in all analyses except for region analyses and they were not plotted on the brain.

Corrupted signal electrodes ($n = 114$) with extensive durations of static signal recordings were manually removed from consideration prior to any downstream analysis. For depth electrodes in the white matter, if they were within 1.5 mm of the gray-white matter boundary, they were projected to the nearest point on that boundary, and were labeled as coming from that region (for the purposes of region significance analyses). Of the 1,688 total electrodes, 1,414 of the electrodes were able to placed in this way into a particular region. The relevant region analyses are shown in fig. 2h-i, fig. 3f-h, fig. 12e-f, fig. 4b, fig. 2b, fig. 3b, fig. 5e.

This procedure is very similar to the post brain-shift correction methods used for electrocorticography electrodes [76]. For solely visualization purposes, all electrodes identified to lie in the gray matter or on the gray-white matter boundary were first projected to the pial surface (using nearest neighbors), and then mapped to an average brain (using Freesurfer's fsaverage atlas) for the visualizations shown in the main text.

A.6 Word responsiveness

To determine the word responsiveness of an electrode, we compared pre-onset windows to post-onset windows (fig. 10). Precisely, we compared the mean activity in a 100ms window before word onset to the activity in a 100ms window after word onset with a two-tailed paired t-test. The windows were separated by an interval of 1s. This test was performed for absolute offsets of $[-0.5s, -0.4s, -.3s, -.2s, -.1s]$ (fig. 10). This is done to account for the fact that any one offset may "miss" the neural response by chance. An electrode is *word responsive* if at least one of the tests shows a significant (after correction for multiple comparisons) difference between pre- and post-onset activity. In such cases, we report the significance of the t-test with the lowest p-value.

A.7 Testing difference between conditions

When determining the significance of the difference between two conditions (fig. 2c, fig. 3b, fig. 12a,c,d), we used a two-tailed t-test to compare the mean activity in a 100ms window for the two conditions. Five t-tests are performed, at absolute offsets of $[0s, 0.1s, 0.2s, 0.3s, 0.4s]$ and we say that the two conditions result in different neural responses if there exists a test for which there is a significant difference, after correction for multiple comparisons. In such cases, we report the significance of the tests with the lowest p-value. As in the above section, this is done to account for the fact that any one of the tests may miss the difference between the two conditions by chance.

A.8 Linear decoding

Model The model is a logistic regression.

Data pre-processing Neural is decimated by a factor of 10. Data is normalized to 0 mean and unit standard deviation. Normalization is done such that no data-leakage occurs (see below).

Dataset The sentence-onset decoding task requires the model to distinguish between neural activity from an interval in the movie during which a sentence is beginning versus an interval during which no speech is occurring. To obtain positive examples, for every sentence onset, we extract 2s of neural activity, centered on the sentence onset. To obtain negative examples, we divide the movies into 3s segments, and filter for segments that do not overlap with any speech time-stamps. The size of 3s guarantees that there is at least a 500ms buffer between every positive example and every negative example (see below). The dataset is balanced so that an equal number of negative and positive examples occur. Data is drawn from all recorded movies per subject.

Training We are interested in answering the question, how does decodability vary across time? To this end, we divide each example into 250ms intervals. Per each time interval, per electrode, we train our model. Training was done on a single NVIDIA Titan RTXs (24GB GPU Ram) with 80 CPU cores.

Evaluation Per electrode, we create an 80/20 train/test split. The model performance is reported on the test set. Train/test splits are shared between electrodes in the same subject. In fig. 4b,d, and e, we select the top 10 electrodes with the highest score on the train-set (5-fold cross-validation) per region, and report the performance of these electrodes on the test set. The same is done in fig. 2b,d-e and fig. 3b,d-e.

A.9 Part of speech modulates activity

Parts of speech are of particular importance for their fundamental role in linguistics and natural language processing (NLP). Indeed, the two word classes, nouns and verbs, are widely recognized to be among the few linguistic universals [77, 78]. Part-of-speech was a significant predictor in the example electrodes shown in fig. 2 and fig. 3. Given their importance in language, we directly compared the responses to nouns versus verbs (fig. 12). fig. 12a shows the responses of an example electrode located in the left superior temporal gyrus (inset) which showed stronger responses to verbs compared to nouns.

The GLM analysis showed that there were no electrodes which exhibited activity exclusively modulated by part-of-speech. Instead, the neural activity was captured by multiple features as shown in the previous examples. fig. 12b shows that the main feature for this electrode is the index in sentence, followed by the part-of-speech and volume. Indeed, after separating the responses according to the position in the sentence, there was a small but significant difference between nouns and verbs for sentence midsets and offsets but not for sentence offsets (fig. 12c). The differences between nouns and verbs persisted across high and low volumes (fig. 12d). There were no electrodes for which a difference in part-of-speech was observed across all sub-samplings for all features. But there were 83 electrodes for which part of speech has a significant ($p < 0.05$, Bonferroni corrected) beta coefficient in the GLM analysis. fig. 12e shows the exact location of these electrodes and fig. 12f shows the fraction, per region, of the part of speech significant electrodes. We also found that the noun-verb distinction is linearly decodable fig. 3, with significant decoding performance distributed across the brain fig. 3a, and with the highest decoding performance observed in the frontal lobe and cingulate (fig. 3b-e).

Finally, we observed a difference in the magnitude and timing of the peak neural response between nouns and verbs (fig. 13). For each electrode, we computed the mean of the neural response, averaged across all words. Restricting our attention to those electrodes which show at least a moderate neural response (Cohen's $d > 0.1$, see section 3), we can compute the peak of that mean response (fig. 13b) and observe that it is lower in the case of verbs at sentence onsets. ($\mu \approx 32.6, \sigma = 27.7\ \mu V$ for verbs, $\mu \approx 35.5, \sigma = 29.7\ \mu V$ for nouns), but higher in the case of verb midsets ($\mu = 34.1, \sigma = 25.9\ \mu V$ for verbs, $\mu = 30.5, \sigma = 26.4\ \mu V$ for nouns). We also find the timing (fig. 13c) of the sentence midset peaks and observe that it is earlier in the case of verbs ($\mu \approx 293, \sigma = 255\ ms$ for verbs, $\mu \approx 426, \sigma = 315\ ms$ for nouns).

Subj.	Age	Sex	Movies	Time (h)	# Sentences	# Words	# Lemmas	# Electrodes	# Probes
1	19	M	7, 18, 19	6.14	4054	29468	5908	154	13
2	12	M	2, 3, 4, 8, 9, 17, 21	15.49	9092	60958	12243	162	47
3	18	F	5, 11, 12	9.50	4845	32959	6156	134	12
4	12	F	10, 13, 15	5.06	3758	25394	5300	188	15
5	6	M	7	1.45	1162	8457	1892	156	12
6	9	F	6, 13, 20	8.02	3524	21455	4544	164	12
7	11	F	5, 13	3.36	3152	20237	3808	246	18
8	4	M	14	0.96	718	4218	804	162	13
9	16	F	1	1.95	1412	9846	1956	106	12
10	12	M	5, 16	3.93	3506	23408	4048	216	17

Table 2: All subjects language, electrodes and personal statistics. Columns from left to right are the subject's ID and information (age and gender), the the IDs of the movies they watched (corresponding to Extended Figures table 3), the cumulative movie time (hours), number of sentences, number of words (tokens) and number of unique lemmas (canonical word forms), as well as the number of probes the subject had and their corresponding number of electrodes..

756 # B Supplementary figures

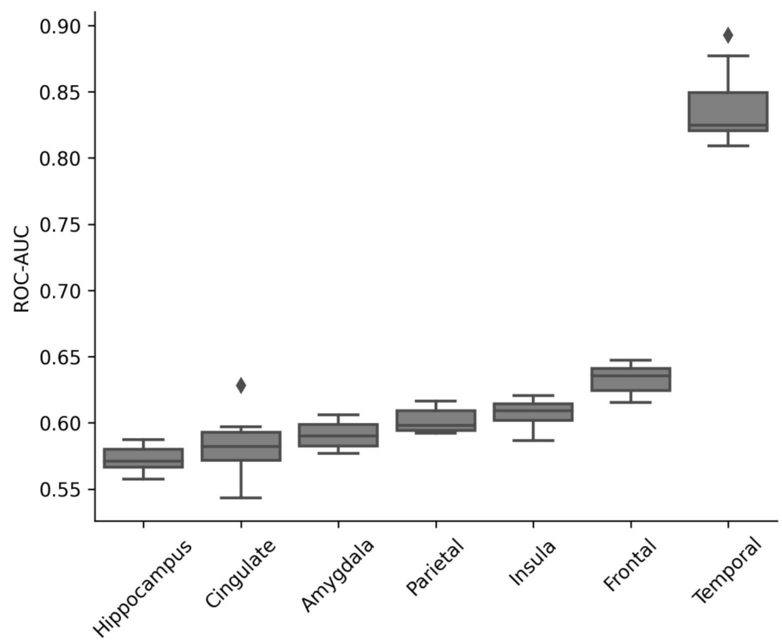


Figure 1: **Decodability of sentence onsets per region.** After decoding sentence onsets per electrodes (see fig. 4), we find distribution of the peak *test* ROC-AUC scores in each region, for the 10 electrodes in each region with the highest cross-validation ($k_{\text{folds}} = 5$) ROC-AUC on the *train* set. Boxes show quartiles and whiskers show 1.5$\times$ the interquartile range. Outliers shown as points beyond the whiskers.

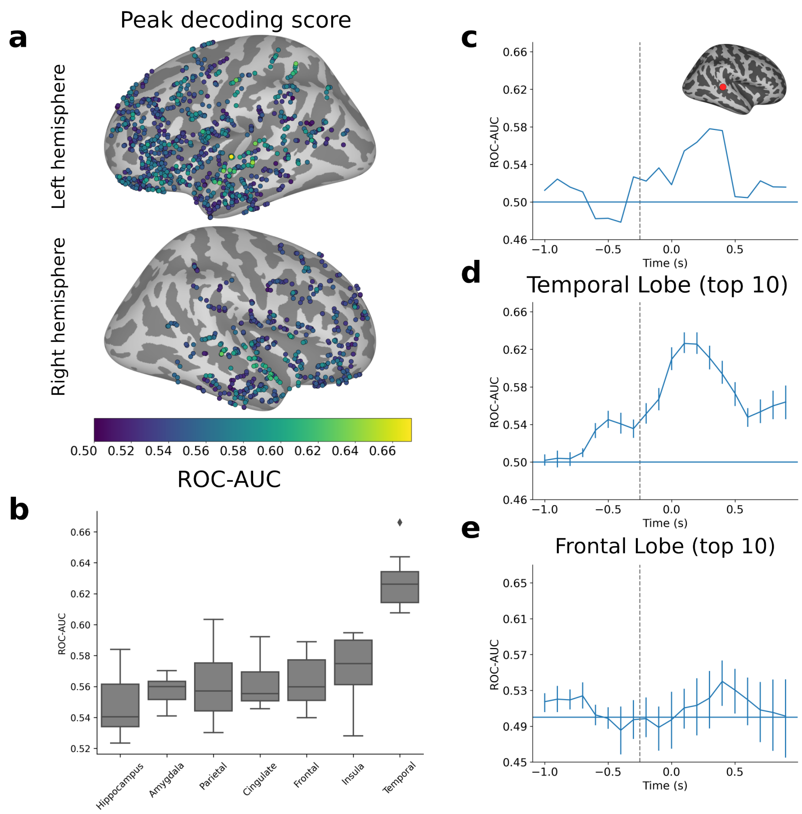


Figure 2: **Word onsets are linearly decodable and reveal the time course of language processes in the brain.** We perform the same analysis as shown in fig. 4, but for word-onsets, instead of sentence-onsets only. A linear decoder is trained to classify portions of the movies according to whether or not speech is occurring, based on the corresponding neural activity. This decoding is done for activity in a 0.25s window, which shifts in 0.1s increments from -1s before word-onset to 1s after word-onset. The spatial distribution of decoding scores, shown in (**a**) and (**b**), after a max has been taken over all windows, shows that word onsets are most decodable in the temporal and frontal lobes. Decodability, as a function of time, shown in (**c**), (**d**), and (**e**), reveal that some word onset information is processed before word onset enters the decoding window (dashed grey line). Averaging over time across the top 10 electrodes per region, as in (**d**) and (**e**), reveals the mirrored time course of language processing.

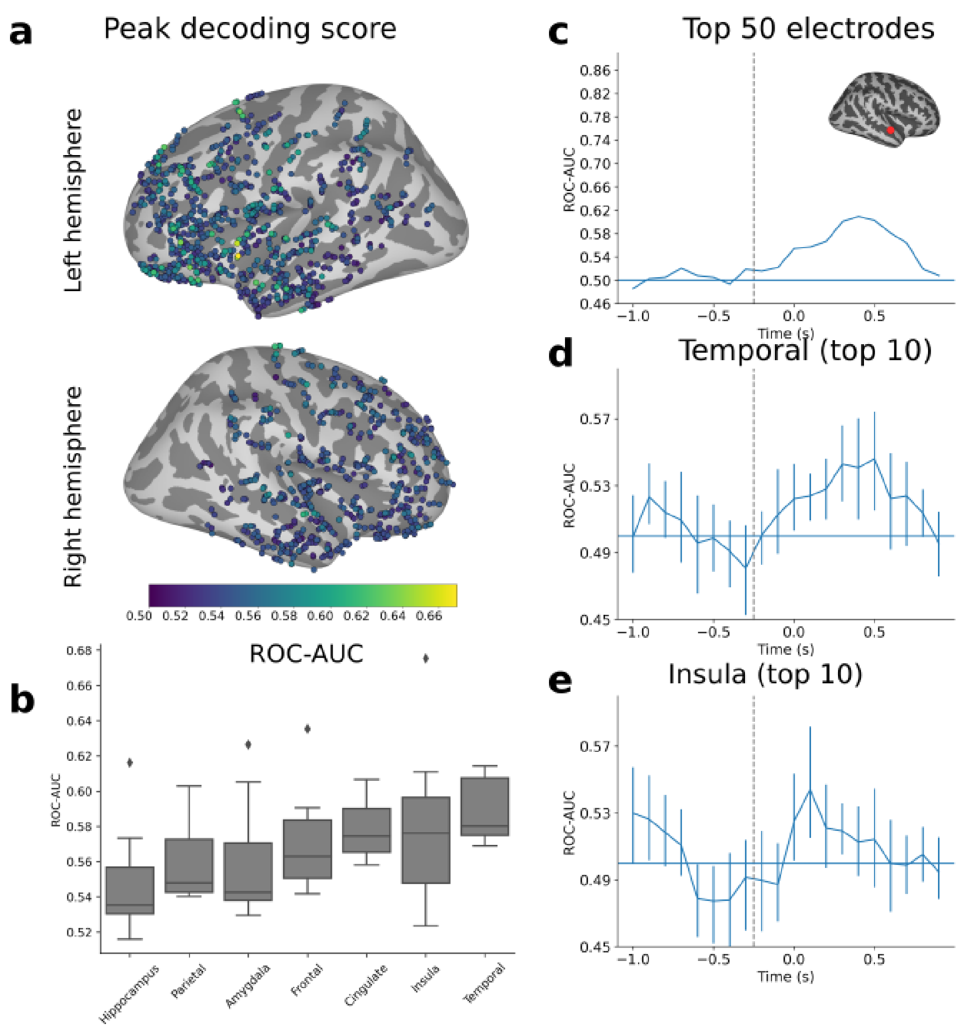


Figure 3: **Part of speech information is linearly decodable.** We perform the same analysis as shown in fig. 4 for nouns and verbs. A linear decoder is trained to classify words as either nouns or verbs, based on the corresponding neural activity. This decoding is done for activity in a 0.25s window in 0.1s increments. The spatial distribution of decoding scores, shown in (a) and (b), after a max has been taken over all windows, shows that part of speech is most decodable in the frontal, cingulate, insula, and temporal regions. Decodability, as a function of time, shown in (c, for an electrode in the superior temporal lobe), (d), and (e), reveal that some part of speech information is processed before word onset enters the decoding window (dashed grey line). Averaging over time across the top 10 electrodes per region, as in (c) and (d), reveals the time course of processing.

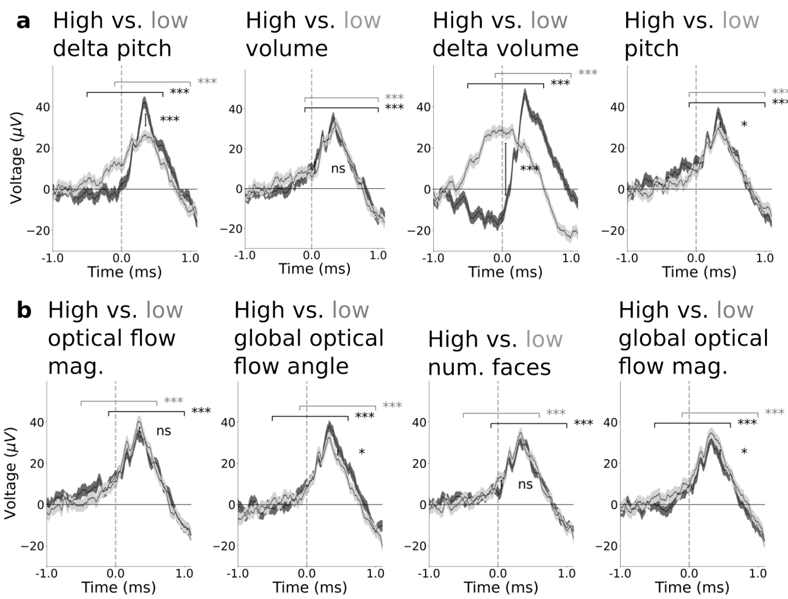


Figure 4: **Neural responses to word onsets are observable, even after controlling for visual and audio features. a**. Mean response to word onsets, after controlling for audio features for the same example electrode as shown in fig. 2. The same conventions as fig. 2c are followed. Vertical brackets and corresponding asterisks show the difference between conditions. Horizontal brackets and asterisks show the significance of the word onset response. **b.** Mean response to word onsets, after controlling for visual features. In both (a) and (b), significant response to word onset can be observed, even after controlling for audio and visual features respectively.

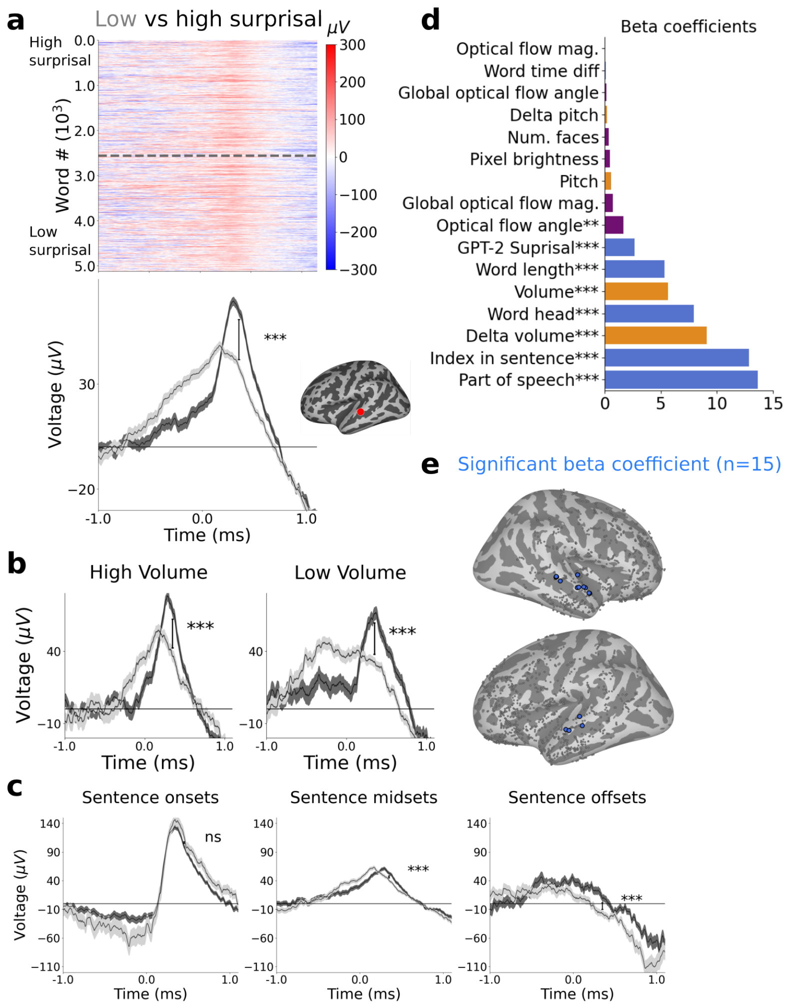


Figure 5: **Neural responses distinguish high and low surprisal. a.** Raster and mean plots aligned to word onsets for an example electrode in the right superior temporal gyrus (see inset in **d**; this is the same electrode as shown in fig. 12) separated by high and low surprisal. The difference between high and low surprisal words remains even after controlling for other features, such as volume (b) and position in sentence (c). GLM analysis reveals that activity in this electrode is modulated in part by surprisal, as well as by other features (d). There are 10 electrodes where part of speech has a significant beta-coefficient; these are all located in the superior temporal lobe (e).

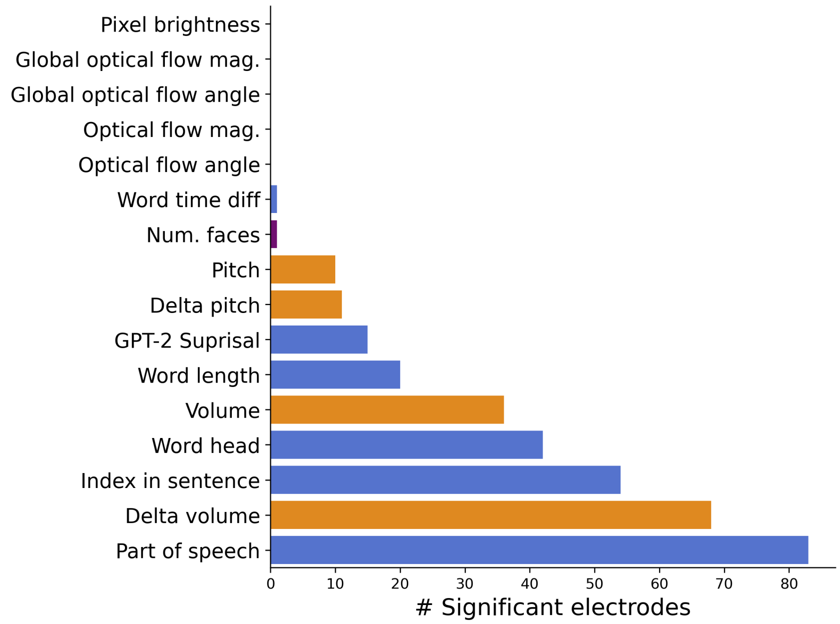


Figure 6: **The other factors which influence activity in part-of-speech-sensitive electrodes**. An electrode is said to be sensitive to part-of-speech, if a GLM fitted to mean neural activity has a significant beta coefficient ($p < 0.05$, after corrections for multiple comparisons) for the part-of-speech feature. Among all such part-of-speech sensitive electrodes ($n = 83$) , the number of electrodes that have other significant beta coefficients is shown.

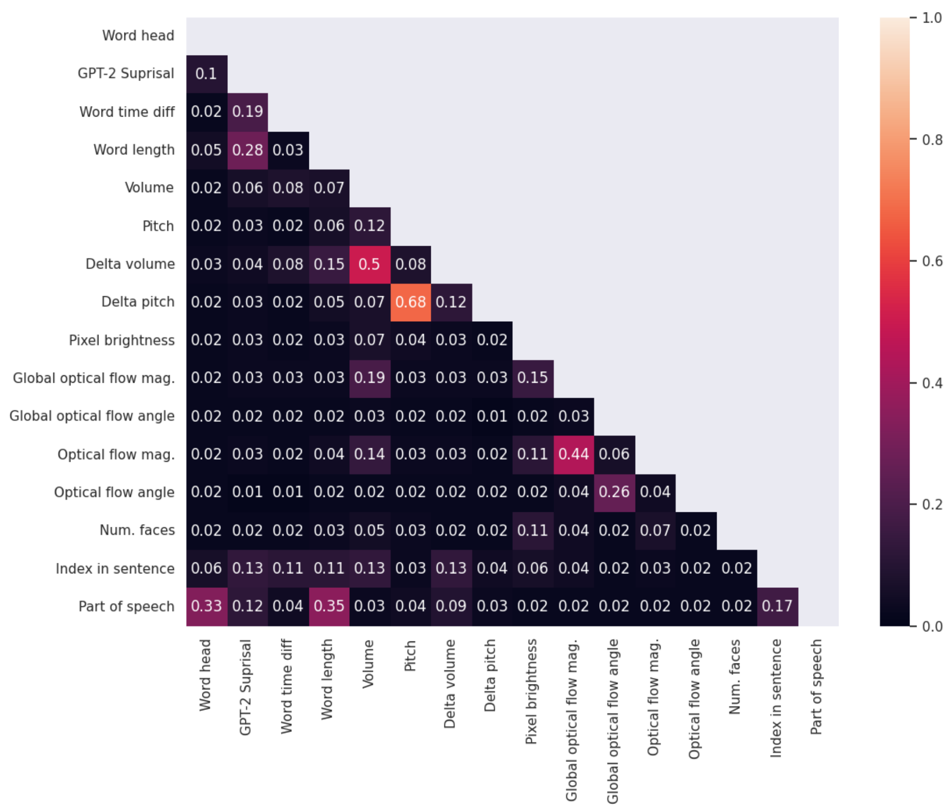


Figure 7: The (absolute value) of Pearson's r between input features, averaged across movies.

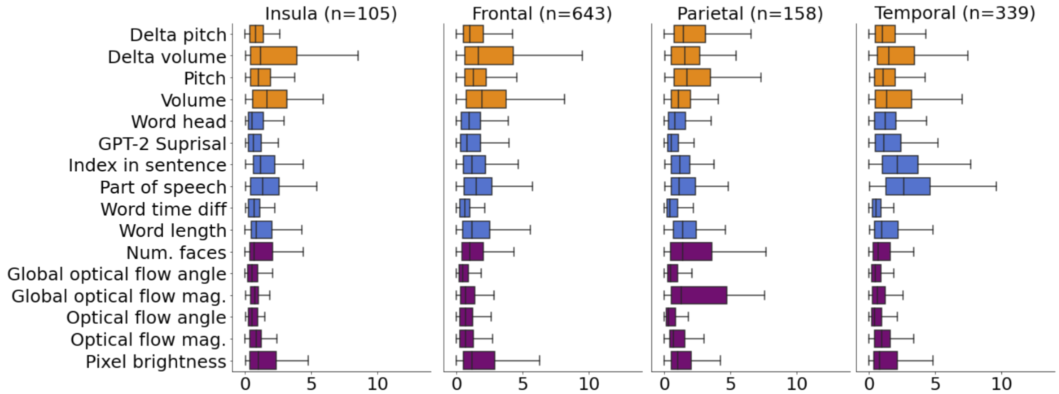


Figure 8: Magnitude of beta coefficients, averaged per region.

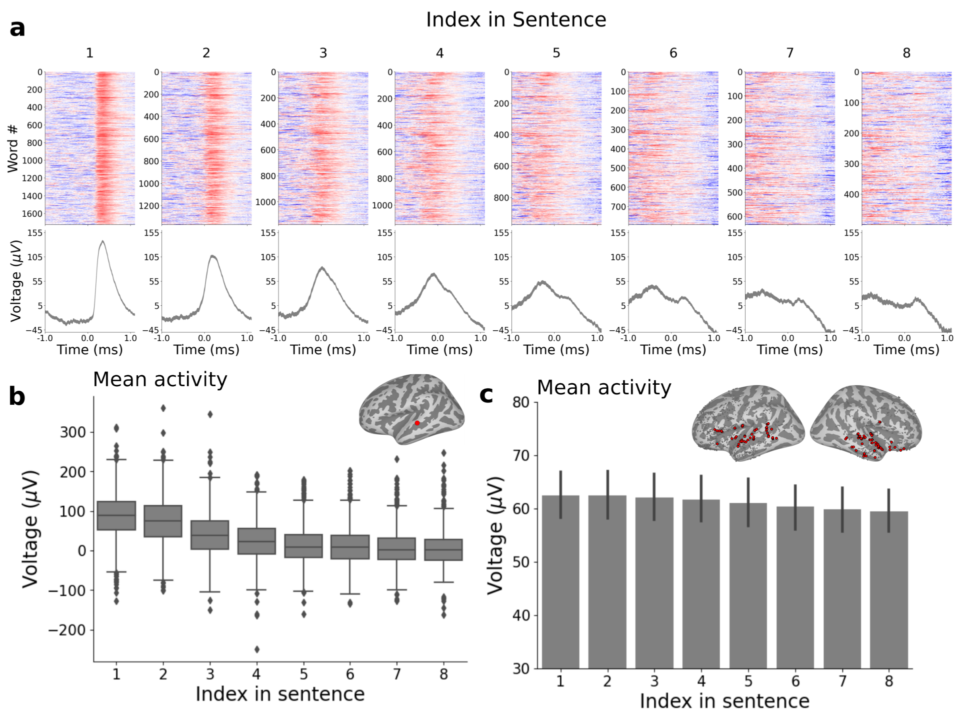


Figure 9: **Neural response decreases as a function of position in the sentence**. Making a more fine-grained examination of sentence position, we observed a trend in which mean activity decreased monotonically with the index in the sentence. (**a**) The neural response per index in sentence is shown for the first eight sentence positions for an electrode in the left temporal lobe (same electrode as shown in fig. 12). (**b**) The mean activity for this same electrode (location shown in inset) is taken for a [0ms,500ms] window after word onset. The box shows the quartiles, while the whiskers show 1.5 $\times$ the interquartile range, over all words at a given position. (**c**) Taking the mean of the magnitude over this same window for all word responsive electrodes shows the same trend. Error bars show a 95% confidence interval over electrodes. A word-responsive electrode is defined, as in fig. 2, as an electrode that shows a significant difference between pre- and post- onset activity. Here we restrict our attention to those electrodes ($n = 111$, locations shown in inset) for which this difference has at least a moderate effect size (Cohen's $d > 0.1$). Note that we do not believe this result stands in opposition to previous findings, such as in [79], foremost because we consider a much different distribution of sentences in our work. The sentences shown to subjects in this work cover a wide variety of forms, and importantly, are usually part of a longer dialogue. To make a direct comparison with previous studies of sentence processing, a more fine-grained inventory of sentence types should be made over the movie transcripts.

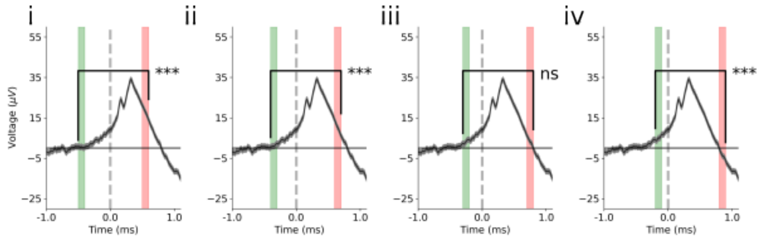


Figure 10: **Schematic of word-responsiveness testing procedure**. We test for word responsiveness at five different points (i-v). The grey line shows mean neural response, averaged across a movie. Shading shows standard error. At each point, a two-tailed paired t-test is performed between the mean activity in a pre-onset (green) and a post-onset (red) window of 100ms. We use multiple tests to account for the fact that sometimes the difference in activity may be 0 simply due to the absolute offset of the windows (this is the case for iii). We say that an electrode is word-responsive, if there is at least one test for which there is a significant difference between pre- and post- onset activity, after correcting for multiple comparisons.

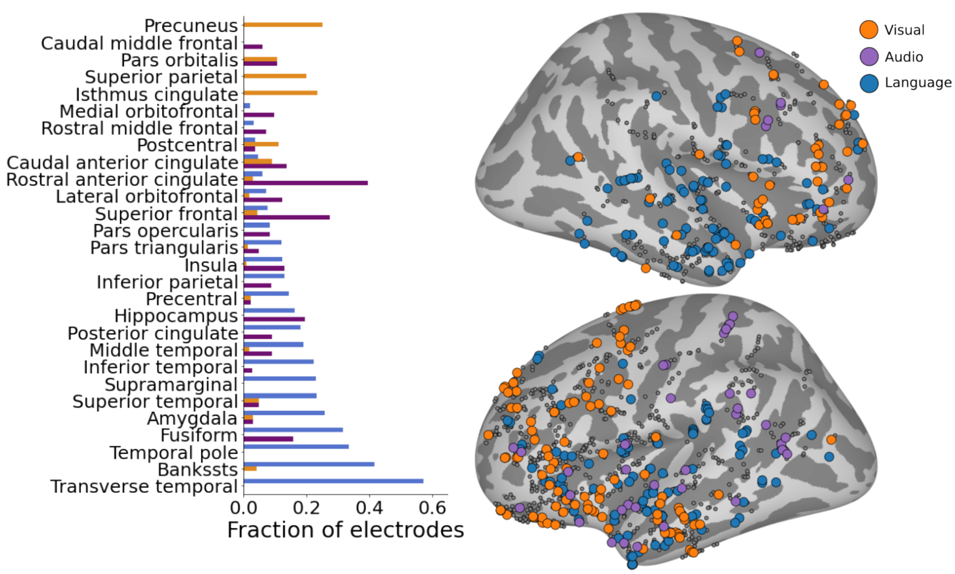


Figure 11: **Unimodal responsive electrodes**. We categorize features as either *visual*, *audio*, or *language*. For each electrode, we use the GLM analysis to determine whether a given electrode's activity has a significant (after Bonferroni correction) response for features from a single category, to the exclusion of the other categories.

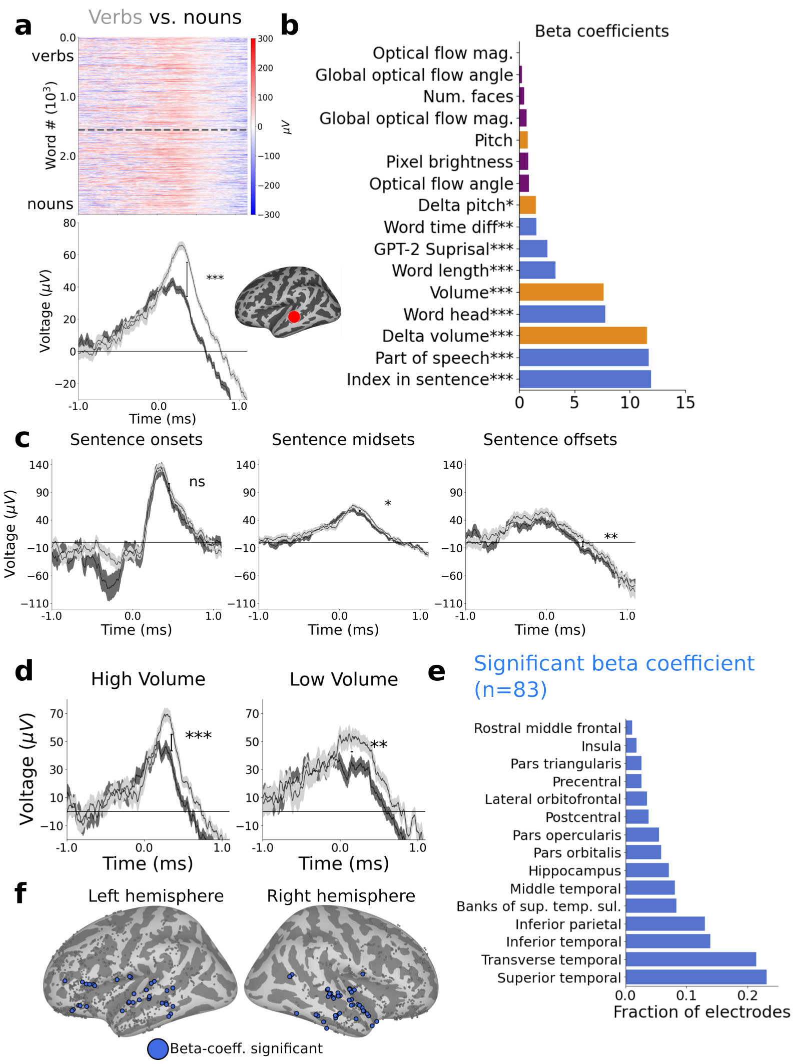


Figure 12: **Neural responses distinguish nouns and verbs. a.** Raster and mean plots aligned to word onsets for an example electrode in the left superior temporal gyrus (see inset) separated by nouns (bottom in raster plot, light grey in mean plot) and verbs (top in raster plot, dark grey in mean plot). **b.** GLM analysis reveals that activity in this electrode is modulated by part of speech, as well as by other features. **c.** For this electrode, a significant difference between nouns and verbs does not remain for the sentence onsets condition, after sub-sampling over sentence position. **d.** But, a difference does remain for all sub-sampled conditions, when controlling for other features, such as volume. Using the GLM analysis, allows us to judge the influence of part-of-speech on a per-word basis. **e.** The fraction of electrodes, per region, of electrodes where part of speech has a significant beta-coefficient (total $n = 83$); these are mainly located in the temporal and frontal lobes. **f.** The exact location of these electrodes (blue) projected onto the surface of the brain.

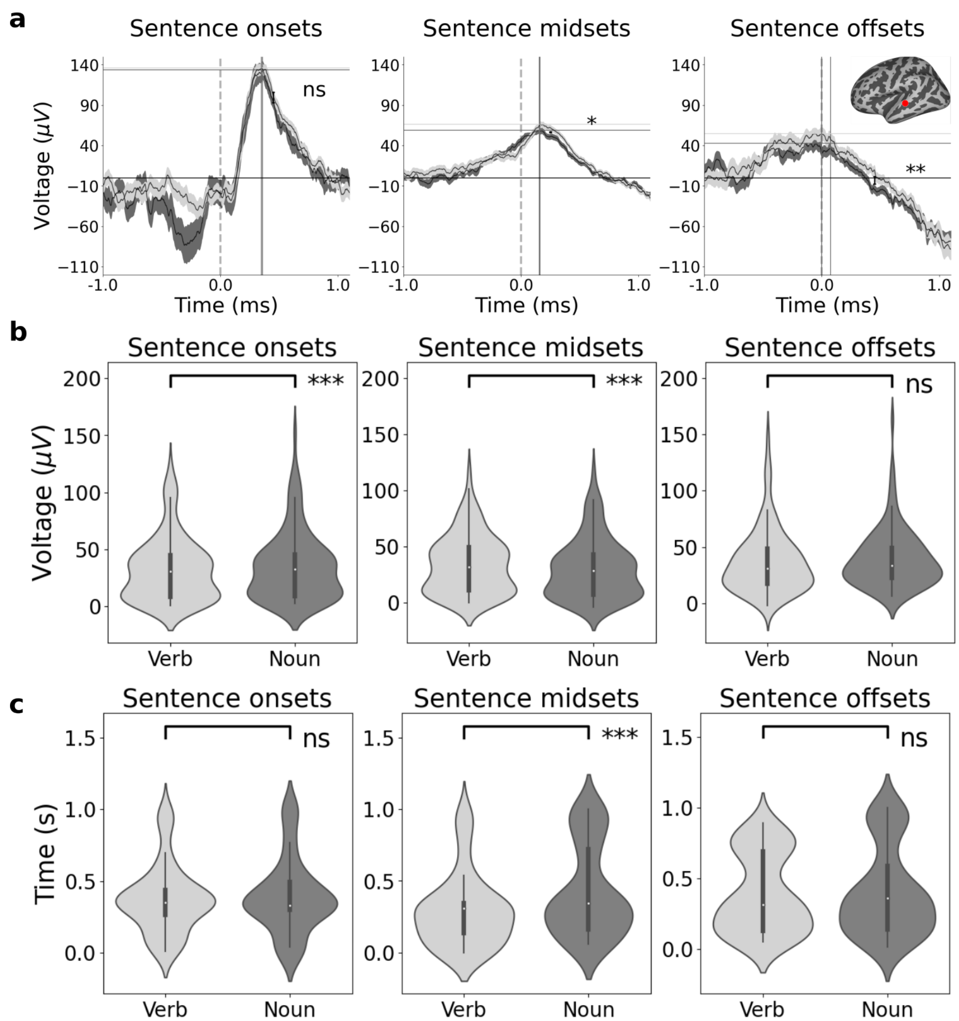


Figure 13: **Noun vs. verb peak amplitude and timing.**. For each electrode, we consider the mean signal. See, for example, (a) which shows the mean activity for an electrode in the STG (the same electrode shown in fig. 12). For an electrode, we find the amplitude (horizontal lines) of the peak mean activity and the timing of the peak (vertical lines). Across many electrodes, we observe a difference in the peak amplitudes such that nouns induce a higher response than verbs for sentence onsets, while verbs induce a higher response for offsets and midsets. The electrodes in (b) and (c) are those electrodes which respond to language (see fig. 2d), with the additional condition that the language response have at least moderate effect size (Cohen's d > 0.1).

#	Movie	Year	Time (s)	# Sentences	# Words	Unique words	Nouns	Unique nouns	Verbs	Unique verbs
1	Antman	2015	7027	1412	9846	1956	1370	712	1538	581
2	Aquaman	2018	8601	1003	7218	1563	1066	517	1094	508
3	Avengers: Infinity War	2018	8961	1372	8479	1780	1081	608	1294	485
4	Black Panther	2018	8073	1139	7571	1628	1084	544	1199	506
5	Cars 2	2011	6377	1801	11404	2060	1576	737	1649	563
6	Coraline	2009	6036	933	5428	1251	759	407	817	353
7	Fantastic Mr. Fox	2009	5205	1162	8457	1892	1240	690	1240	490
8	Guardians of the Galaxy 1	2014	7251	1104	8241	1799	1101	615	1235	521
9	Guardians of the Galaxy 2	2017	8146	1180	9332	1839	1210	623	1368	533
10	Incredibles	2003	6926	1408	9369	1966	1234	659	1545	582
11	Lord of the Rings 1	2001	13699	1424	10538	2011	1470	681	1480	595
12	Lord of the Rings 2	2002	14131	1620	11017	2085	1593	760	1587	631
13	Megamind	2010	5735	1351	8833	1748	1183	610	1340	496
14	Sesame Street Ep. 3990	2016	3440	718	4218	804	716	233	674	211
15	Shrek the Third	2007	5568	999	7192	1586	989	568	1072	418
16	Spiderman: Far From Home	2019	7764	1705	12004	1988	1442	660	1755	555
17	Spiderman: Homecoming	2017	8008	1993	12258	2107	1591	795	1794	569
18	The Martian	2015	9081	1421	11360	2210	1781	826	1686	630
19	Thor: Ragnarok	2017	7831	1471	9651	1806	1183	604	1440	546
20	Toy Story 1	1995	4863	1240	7194	1545	1039	561	1015	388
21	Venom	2018	6727	1301	7859	1527	892	509	1200	427

Table 3: Language statistics for all movies. Columns from left to right are the movie's ID, name, year of production, length (seconds), number of sentences, number of words (tokens), number of unique words (types), number of nouns, number of unique nouns, number of verbs and number of unique verbs.

#	Feature	Category	Description
1	Pixel brightness	Visual	The mean brightness computed as the average HSV value over all pixels
2	Global optical flow magnitude	Visual	A camera motion proxy. The maximal average dense optical flow vector magnitude
4	Optical flow magnitude	Visual	A large displacement proxy. The maximal optical flow vector magnitude
5	Optical flow angle	Visual	The orientation (degrees) of the above flow vector
6	Number of faces	Visual	The maximal number of faces per frame
7	Volume	Auditory	Average root mean squared watts of the audio
8	Mean pitch	Auditory	Average pitch of the audio
9	Delta volume	Auditory	The difference in average RMS of the 500ms windows pre and post word onset
10	Delta pitch	Auditory	The difference in average pitch of the 500ms windows pre and post word onset
11	GPT-2 surprisal	Language	Negative-log transformed GPT-2 word probability (given sentence preceding context)
12	Word time length	Language	Word length (ms)
13	Word time difference	Language	Difference between previous word offset and current word onset (ms)
14	Index in sentence	Language	The word index in its context sentence
15	Word head	Language	The relative position (left/right) of the word's dependency tree head
16	Part of speech tag	Language	The word Universal Part-of-Speech (UPOS) tag

Table 4: **Extracted visual, auditory, and language features used to model the neural responses**. All scalar type features were used as regressors in the GLM analysis and all scalar and vector features were used as test set balancing features in the multi-confounds CNN analysis. The difference between 2 and 4 is that 2 is the magnitude of the averaged optical flow vector, with the average being taken over all optical flow vectors on the screen, whereas 4 is the magnitude of the largest individual optical flow vector on the screen.

C Data documentation

The brain recordings and annotations and annotations are released at the subject level, and can be thought of as the raw source, from which derivative machine learning datasets may be created, and for this reason we do not include any croissant meta-data. An example of a dataset derivation could be: segmenting the audio track by word boundaries and then training a decoding model to map for neural recordings to word identity. Another example could involve segmenting the recording into uniform intervals and then training a decoding model to predict average color on screen. We release the recordings in their entirety to allow for this flexibility.

The website contains the following assets:

1. quickstart.ipynb A quickstart IPython notebook
2. localization.zip Spatial position of electrodes
3. subject_timings.zip Wall clock time of triggers used for synchronization with movie
4. subject_metadata.zip Movie metadata
5. electrode_labels.zip Semantic ID for electrodes
6. speaker_annotations.zip Speaker IDs for movie audio
7. scene_annotations.zip Scene cut annotations for movies
8. transcripts.zip Pre-computed features for movies
9. trees.zip Universal Dependency parse trees for movie dialogue
10. sub_<sub_id>_trial<trial_id>.h5.zip Neural recordings in HDF5 format

D Responsibility, License, Hosting Plan

Authors bear all responsibility in case of privacy violations. Authors release the data under a CC BY 4.0 license.

Data will be hosted on MIT CSAIL servers and will be accessible at the url https://braintreebank.dev/. Backups will be kept across multiple machines. Hardware will be maintained by the MIT CSAIL Infrastructure Group: https://tig.csail.mit.edu/.

References

K. Abe and D. Watanabe. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, 14(8):1067–1074, 2011. doi: 10.1038/nn.2869.

T. Aflalo et al. A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. *Sci Adv*, 6, 2020. doi: 10.1126/sciadv.abb3984.

F. e. a. Artoni. High gamma response tracks different syntactic structures in homophonous phrases. *Sci Rep*, 10:7537, 2020. doi: 10.1038/s41598-020-64375-9.

A. Bansal et al. Neural dynamics underlying target detection in the human brain. *Journal of Neuroscience*, 34:3042–3055, 2014. doi: 10.1523/JNEUROSCI.3781-13.2014.

I. Bhaya-Grossman and E. F. Chang. Speech computations of the human superior temporal gyrus. *Annual review of psychology*, 73:79–102, 2022. doi: 10.1146/annurev-psych-022321-035256.

D. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997.

J. Cai, A. Hadjinicolau, A. Paulik, Z. Williams, and S. Cash. Natural language processing models reveal neural dynamics of human communication. *Biorxiv*, 2023. doi: 10.1101/2023.03.10.531095.

L. Calinescu, G. Ramchand, and G. Baggio. How (not) to look for meaning composition in the brain: A reassessment of current experimental paradigms. *Frontiers in Language Sciences*, 2:1096110, 2023.

A. Caramazza and A. E. Hillis. Lexical organization of nouns and verbs in the brain. *Nature*, 349:788–790, 1991. doi: 10.1038/349788a0.

G. A. Castellucci, C. K. Kovach, M. A. I. Howard, J. D. W. Greenlee, and M. A. Long. A speech planning network for interactive language use. *Nature*, 602:117–122, 2022. doi: 10.1038/s41586-021-04270-z.

N. Chomsky. *The minimalist program*. MIT Press, 1995.

N. Chomsky, A. Gallego, and D. Ott. Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan Journal of Linguistics*, pages 226–261, 2019.

A. e. a. Cometa. Event-related causality in stereo-eeg discriminates syntactic processing of noun phrases and verb phrases. *J Neural Eng*, 20, 2023. doi: 10.1088/1741-2552/accaa8.

D. Crepaldi, M. Berlingeri, E. Paulesu, and C. Luzzatti. A place for nouns and a place for verbs? a critical review of neurocognitive data on grammatical-class effects. *Brain Lang*, 116:33–49, 2011. doi: 10.1016/j.bandl.2010.09.005.

A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage*, 9:179–194, 1999.

A. R. Damasio and D. Tranel. Nouns and verbs are retrieved with differently distributed neural systems. *Proc Natl Acad Sci U S A*, 90:4957–4960, 1993. doi: 10.1073/pnas.90.11.4957.

I. E. de Araujo, E. T. Rolls, M. L. Kringelbach, F. McGlone, and N. Phillips. Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *Eur J Neurosci*, 18:2059–2068, 2003. doi: 10.1046/j.1460-9568.2003.02915.x.

R. S. Desikan et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31:968–980, 2006.

N. Ding, L. Melloni, H. Zhang, X. Tian, and D. Poeppel. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci*, 19:158–164, 2016. doi: 10.1038/nn.4186.

N. F. Dronkers, D. P. Wilkins, J. Van Valin, R. D., B. B. Redfern, and J. J. Jaeger. Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92:145–177, 2004. doi: 10.1016/j.cognition.2003.11.002.

A. Dubey and S. Ray. Cortical electrocorticogram (ecog) is a local signal. *Journal of Neuroscience*, 39:4299–4311, 2019.

Y. Elazar, S. Ravfogel, A. Jacovi, and Y. Goldberg. Amnesic probing: behavioral explanation with amnesic counterfactuals. *arXiv*, (2006.00995), 2021.

W. T. Fitch and M. D. Hauser. Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656):377–380, 2004. doi: 10.1126/science.1089401.

K. Forseth, X. Pitkow, S. Fischer-Baum, and N. Tandon. What the brain does as we speak. *bioRxiv*, 2021. doi: 10.1101/2021.02.05.429841.

K. J. Forseth, G. Hickok, P. S. Rollo, and N. Tandon. Language prediction mechanisms in human auditory cortex. *Nat Commun*, 11:5240, 2020. doi: 10.1038/s41467-020-19010-6.

K. J. e. a. Forseth. A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain*, 141:2112–2126, 2018. doi: 10.1093/brain/awy120.

I. Fried, M. Cerf, U. Rutishauser, and G. Kreiman. *Single neuron studies of the human brain: Probing cognition*. MIT Press, 2014.

T. Q. Gentner, K. M. Fenn, D. Margoliash, and H. C. Nusbaum. Recursive syntactic pattern learning by songbirds. *Nature*, 440:1204–1207, 2006. doi: 10.1038/nature04675.

M. Geva-Sagiv et al. Augmenting hippocampal-prefrontal neuronal synchrony during sleep enhances memory consolidation in humans. *Nat Neurosci*, 26:1100–1110, 2023. doi: 10.1038/s41593-023-01324-5.

A. Goldstein et al. Shared computational principles for language processing in humans and deep language models. *Nat Neurosci*, 25:369–380, 2022. doi: 10.1038/s41593-022-01026-4.

D. M. Groppe et al. ielvis: An open source matlab toolbox for localizing and visualizing human intracranial electrode data. *J Neurosci Methods*, 281:40–48, 2017. doi: 10.1016/j.jneumeth.2017.01.022.

L. Gwilliams, J. R. King, A. Marantz, and D. Poeppel. Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nat Commun*, 13: 6606, 2022. doi: 10.1038/s41467-022-34326-1.

P. Hagoort and P. Indefrey. The neurobiology of language beyond single words. *Annu Rev Neurosci*, 37:347–362, 2014. doi: 10.1146/annurev-neuro-071013-013847.

L. S. Hamilton, Y. Oganian, J. Hall, and E. F. Chang. Parallel and distributed encoding of speech across human auditory cortex. *Cell*, 184:4626–4639 e4613, 2021. doi: 10.1016/j.cell.2021.07.019.

L. A. Heimbauer, C. M. Conway, M. H. Christiansen, M. J. Beran, and M. J. Owren. Visual artificial grammar learning by rhesus macaques (macaca mulatta): exploring the role of grammar complexity and sequence length. *Animal Cognition*, 21(3): 447–457, 2018. doi: 10.1007/s10071-018-1164-4. URL https://doi.org/10.1007/s10071-018-1164-4.

L. T. Hunt et al. Triple dissociation of attention and decision computations across prefrontal cortex. *Nat Neurosci*, 21:1471–1481, 2018. doi: 10.1038/s41593-018-0239-5.

M. Jamali, B. Grannan, J. Cai, et al. Semantic encoding during language comprehension at single-cell resolution. *Nature*, 2024. doi: 10.1038/s41586-024-07643-2. URL https://doi.org/10.1038/s41586-024-07643-2.

X. Jiang, T. Long, W. Cao, J. Li, S. Dehaene, and L. Wang. Production of supra-regular spatial sequences by macaque monkeys. *Current Biology*, 28(12):1851–1859.e4, 2018. doi: 10.1016/j.cub.2018.04.047. Epub 2018 Jun 7.

E. L. Johnson and R. T. Knight. Intracranial recordings and human memory. *Current Opinion in Neurobiology*, 31:18–25, 2015.

A. Joshi et al. Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics*, 9:69–84, 2011. doi: 10.1007/s12021-010-9092-8.

M. e. a. Keshishian. Joint, distributed and hierarchically organized encoding of linguistic features in the human auditory cortex. *Nat Hum Behav*, 7:740–753, 2023. doi: 10.1038/s41562-023-01520-0.

A. R. Khanna, W. Muñoz, Y. J. Kim, et al. Single-neuronal elements of speech production in humans. *Nature*, 626:603–610, 2024. doi: 10.1038/s41586-023-06982-w. URL https://doi.org/10.1038/s41586-023-06982-w.

M. L. Kringelbach. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat Rev Neurosci*, 6:691–702, 2005. doi: 10.1038/nrn1747.

A. T. Lee, E. F. Chang, M. F. Paredes, et al. Large-scale neurophysiology and single-cell profiling in human neuroscience. *Nature*, 630:587–595, 2024. doi: 10.1038/s41586-024-07405-0. URL https://doi.org/10.1038/s41586-024-07405-0.

R. Malassis, S. Dehaene, and J. Fagot. Baboons (papio papio) process a context-free but not a context-sensitive grammar. *Scientific Reports*, 10:7381, 2020. doi: 10.1038/s41598-020-64244-5. Published online 2020 Apr 30.

M. M. Mesulam, C. K. Thompson, S. Weintraub, and E. J. Rogalski. The wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*, 138:2423–2437, 2015. doi: 10.1093/brain/awv154.

M. M. Mesulam et al. Primary progressive aphasia and the evolving neurology of the language network. *Nat Rev Neurol*, 10:554–569, 2014. doi: 10.1038/nrneurol.2014.159.

M. M. Mesulam et al. Neuropathological fingerprints of survival, atrophy and language in primary progressive aphasia. *Brain*, 145:2133–2148, 2022. doi: 10.1093/brain/awab410.

P. Misra, Y.-C. Shih, H.-Y. Yu, D. Weisholtz, J. R. Madsen, S. Sceillig, and G. Kreiman. Invariant neural representation of parts of speech in the human brain. *arXiv preprint arXiv:2024.01.15.575788*, 2024. doi: https://doi.org/10.1101/2024.01.15.575788.

P. Mitra and H. Bokil. *Observed brain dynamics*. Oxford University Press, 2008.

A. Mondal. The development of sanskrit grammar on pre-pāṇinian stage. *International Journal of Sanskrit Research*, 6(6):154–158, 2020.

R. Mukamel and I. Fried. Human intracranial recordings and cognitive neuroscience. *Annual Review of Psychology*, 63:511–537, 2012a.

R. Mukamel and I. Fried. Human intracranial recordings and cognitive neuroscience. *Annual review of psychology*, 63:511–537, 2012b. doi: 10.1146/annurev-psych-120709-145401.

E. e. a. Murphy. Minimal phrase composition revealed by intracranial recordings. *J Neurosci*, 42:3216–3227, 2022. doi: 10.1523/JNEUROSCI.1575-21.2022.

M. P. Noonan et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci U S A*, 107:20547–20552, 2010. doi: 10.1073/pnas.1012246107.

K. V. e. a. Nourski. Gamma activation and alpha suppression within human auditory cortex during a speech classification task. *J Neurosci*, 42:5034–5046, 2022. doi: 10.1523/JNEUROSCI.2187-21.2022.

D. Ongur and J. L. Price. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex*, 10:206–219, 2000. doi: 10.1093/cercor/10.3.206.

OpenAI. GPT-4 Technical Report. *arXiv*, (2303.08774), 2023.

R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435:1102–1107, 2005. doi: 10.1038/nature03687.

B. Rapp and A. Caramazza. Selective difficulties with spoken nouns and written verbs: A single case study. *Journal of neurolinguistics*, 15:373–402, 2002.

A. S. Reber. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6):855–863, 1967. doi: 10.1016/S0022-5371(67)80149-X. URL https://doi.org/10.1016/S0022-5371(67)80149-X.

J. Saffran, M. Hauser, R. Seibel, J. Kapfhamer, F. Tsao, and F. Cushman. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, 107(2):479–500, 2008. doi: 10.1016/j.cognition.2007.10.010. Author manuscript; available in PMC 2009 May 1. Published in final edited form as: Cognition. 2008 May; 107(2): 479–500. Published online 2007 Dec 20.

A. Salehi, P. H. Yang, and M. D. Smyth. Single-center cost comparison analysis of stereoelectroencephalography with subdural grid and strip implantation. *Journal of Neurosurgery: Pediatrics*, 29:568–574, 2022.

A. e. a. Sinai. Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain*, 128:1556–1570, 2005. doi: 10.1093/brain/awh491.

N. Stobbe, G. Westphal-Fitch, U. Aust, and W. T. Fitch. Visual artificial grammar learning: comparative research on humans, kea (nestor notabilis) and pigeons (columba livia). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367:1995–2006, 2012. doi: 10.1098/rstb.2012.0096. URL http://doi.org/10.1098/rstb.2012.0096.

I. Tenney, D. Dipanjan, and E. Pavlick. Bert rediscovers the classical nlp pipeline. *arXiv*, (1905.05950), 2019.

C. A. A. van Heijningen, J. de Visser, W. Zuidema, and C. ten Cate. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48):20538–20543, 2009. doi: 10.1073/pnas.0908113106. URL https://doi.org/10.1073/pnas.0908113106. Epub 2009 Nov 16.

G. Vigliocco, D. P. Vinson, J. Druks, H. Barber, and S. F. Cappa. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci Biobehav Rev*, 35:407–426, 2011. doi: 10.1016/j.neubiorev.2010.04.007.

S. Waldert. Invasive vs. non-invasive neuronal signals for brain-machine interfaces: Will one prevail? *Frontiers in Neuroscience*, 10:295, 2016. doi: 10.3389/fnins.2016.00295. URL https://doi.org/10.3389/fnins.2016.00295.

C. Wang, A. Yaari, A. K. Singh, V. Subramaniam, D. Rosenfarb, J. DeWitt, P. Misra, J. R. Madsen, S. Stone, G. Kreiman, B. Katz, I. Cases, and A. Barbu. Brain treebank: Large-scale intracranial recordings from naturalistic language stimuli. *Unpublished*, 2024. □ denotes equal contribution.

J. Wang, A. Tao, W. S. Anderson, J. R. Madsen, and G. Kreiman. Mesoscopic physiological interactions in the human brain reveal small-world properties. *Cell Rep*, 36: 109585, 2021. doi: 10.1016/j.celrep.2021.109585.

E. Warrington and T. Shallice. Category specific semantic impairments. *Brain*, 107: 829–854, 1984.

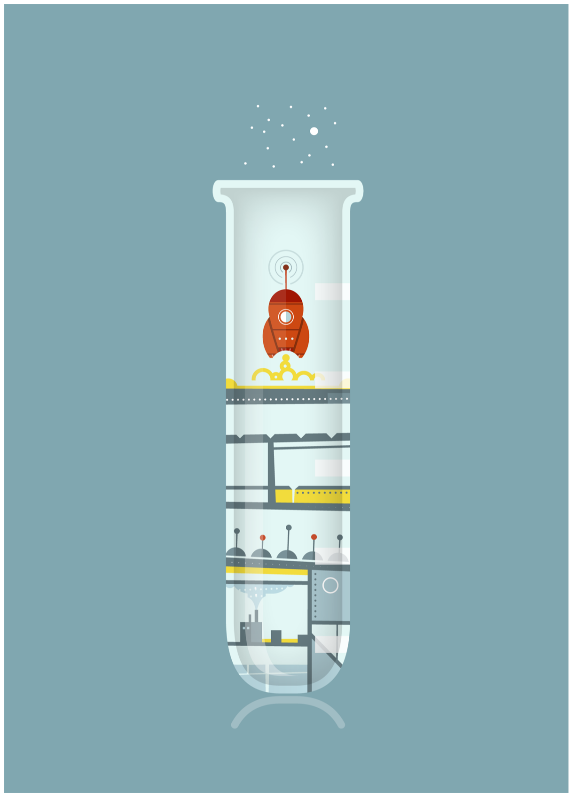
B. Wilson, H. Slater, Y. Kikuchi, A. E. Milne, W. D. Marslen-Wilson, K. Smith, and C. I. Petkov. Auditory artificial grammar learning in macaque and marmoset monkeys. *Journal of Neuroscience*, 40(34):6510–6520, 2020. doi: 10.1523/JNEUROSCI.0152-20.2020.

M. Wojtasik et al. Cytoarchitectonic characterization and functional decoding of four new areas in the human lateral orbitofrontal cortex. *Front Neuroanat*, 14:2, 2020. doi: 10.3389/fnana.2020.00002.

O. e. a. Woolnough. Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nat Hum Behav*, 5:389–398, 2021. doi: 10.1038/s41562-020-00982-w.

Y. Xiao et al. Integration of recognition, episodic, and associative memories during complex human behavior. *Biorxiv*, 2023. doi: 10.1101/2023.03.27.534384.

H. G. Yi et al. Learning nonnative speech sounds changes local encoding in the adult human cortex. *Proc Natl Acad Sci U S A*, 118, 2021. doi: 10.1073/pnas.2101777118.



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under cc by-nc-nd 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive mit (x11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.