Invariant neural representation of parts of speech in the human brain

Pranav Misra[1,5], Yen-Cheng Shih[2,3], Hsiang-Yu Yu[2,3], Daniel Weisholtz[4], Akshay Sharma[5], Demitre Serletis[5], Juan Bulacio[5], William Bingaman[5], Joseph R Madsen[6], Sceillig Stone[6], Gabriel Kreiman[6,7*]

[1]Harvard University, Cambridge, MA, USA
[2]Department of Neurology, Taipei Veterans General Hospital, Taipei, Taiwan
[3]School of Medicine, National Yang Ming Chiao Tung University College of Medicine, Taipei, Taiwan
[4]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[5]Cleveland Clinic, Cleveland, OH, USA
[6]Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[7]Center for Brains, Minds and Machines, Cambridge, MA, USA
*To whom correspondence should be addressed: Gabriel.kreiman@tch.harvard.edu

Number of figures: 5
Number of supplementary figures: 11
Number of supplementary tables: 12

Abstract

Elucidating the 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. Here we recorded invasive neurophysiological responses from 1,801 electrodes in 20 patients with epilepsy while they were presented with two-word phrases consisting of an adjective and a noun. We observed neural signals that distinguished between these two parts of speech. The selective signals were circumscribed within a small region in the left lateral orbitofrontal cortex. The representation of parts of speech showed invariance across visual and auditory presentation modalities, robustness to word properties like length, order, frequency, and semantics, and even generalized across different languages. Furthermore, we extended these ideas by evaluating how parts of speech are processed within full sentences. Recording activity from additional 1,593 electrodes in 17

participants, we found neural signals that separate nouns from verbs in sentences. This selective, invariant, and localized representation of parts of speech provides a foundation to understand how the brain orchestrates more complex aspects of language.

Introduction

Language plays a central role in almost all of our daily activities and is at the heart of how we interact with others[1-3]. Early neurological studies and subsequent work using electrical stimulation demonstrated that there exist specific brain regions that play essential roles in language understanding and production[4-9]. Despite the critical importance of language, progress towards elucidating the neural circuits underlying its representation has remained elusive, in part due to the difficulties in investigating animal models, and in part due to the challenges associated with examining the neurophysiological responses in the human brain.

Several neurophysiological experiments have begun to investigate neural signals associated with presentation of individual words or short phrases[10-20]. There has been work examining the orthographic features of real versus pseudowords[21-23], phonetic features of word comprehension[11,21,24,25] and production[26-30], retrieval of semantic information for audio-visual naming[11], and semantic encoding[31]. These studies have shed light on the early processes associated with detecting, comprehending and producing words. Beyond individual words, at the heart of linguistic structures is the notion that words serve specific functions within a sentence, including articles, nouns, adjectives, and verbs. These parts of speech (POS) are widely shared across languages, are combined according to defined grammatical rules, and play critical roles in natural language processing algorithms[1,12,18,19,32-38]. Furthermore, recent work has suggested that POS may be implicitly learned and represented in modern large language models[39,40]. Many studies in patients with brain lesions have shown deficits in the retrieval of individual nouns versus verbs[20,22,41-45]. However, previous studies could not resolve explicit neural circuits associated with POS processing due to insufficient spatial or temporal resolution[20].

What would a representation for parts of speech 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: (i) presentation modality (e.g., auditory versus visual), (ii) specific noun or adjective (e.g., green or red), (iii) position within a phrase (e.g., “green apple” versus “apple green”), (iv) specific language in bilingual speakers (e.g., “green apple” in English versus “manzana verde” in Spanish), (v) other word properties like their written length, number of syllables, and phoneme composition.

Here we set out to investigate the representation of parts of speech in the human brain by recording intracranial field potential responses with high spatiotemporal resolution and high signal-to-noise ratio from large dataset encompassing 3,394 electrodes implanted in 37 participants with pharmacologically-resistant epilepsy. We describe neural signals, especially in the left lateral orbitofrontal cortex, that selectively distinguish between nouns and adjectives. These part-of-speech 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. Furthermore, we extend the work to whole sentences where we show neural signals that distinguish between nouns and verbs.

Results

We conducted two experiments, one with minimal two-word phrases and a second one with full sentences composed of four words. We start by describing the experiment with two-word phrases and extend the work to full sentences in the last section. We recorded intracranial field potentials from 1,801 electrodes (840 in gray matter, 961 in white matter) implanted in 20 participants via stereoelectroencephalography. 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 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 1b-g** (see also **Tables S1-S2** and **Methods**). We use a bipolar reference, and we 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**).

Neural signals reflect visual, auditory and multimodal inputs

We observed 565 electrodes (31.4% of the total) that responded to auditory stimuli (**Figure S1a-c, g-i**) and 532 electrodes (29.5% of the total) that responded to visual stimuli (**Figure S1d-f, g-i**). The overall proportions and dynamics of visual and auditory responsive signals are consistent with previous work[27,46]. Of these electrodes, there were 293 electrodes that responded to *both* auditory and visual stimuli (**Figure S1g-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 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 hemisphere. Of the 41 the regions in the Desikan-Killiani Atlas where we had sampling (34 defined regions and 7 extra regions representing deep gray matter structures, **Methods**, **Figure 1**, **Tables S1-S2**), 13 regions had a significantly higher number of multimodal electrodes than expected 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 1h-j** shows the responses of an example audiovisual responsive electrode located in the left rostral middle-frontal gyrus (**Figure 1k**). This electrode showed strong evoked responses evident in the trial-averaged responses (**Figure 1h**), and even in individual trials for both auditory stimuli (**Figure 1i**) and visual stimuli (**Figure 1j**).

To compare the response dynamics of auditory and visual responses, we 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 1h**. **Figure S2a** 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 response AUC to auditory stimuli (108±100 $\mu V^2/\text{Hz-ms}$) was larger than to visual stimuli (44±16 $\mu V^2/\text{Hz-ms}$) ($p<10^{-4}$, ranksum test, **Figure S2b**). Similarly, for the visual-only electrodes, the response AUC to auditory stimuli (40±23 $\mu V^2/\text{Hz-ms}$) was smaller than to visual stimuli (53±43 $\mu V^2/\text{Hz-ms}$) ($p<10^{-4}$, ranksum test, **Figure**

S2c). For the audiovisual electrodes, the response AUC to auditory stimuli (71±72 $\mu V^2/\mathrm{Hz\text{-}ms}$) was slightly larger than to visual stimuli (54±39 $\mu V^2/\mathrm{Hz\text{-}ms}$) (p<0.01, ranksum test, **Figure S2d**).

Multimodal neural signals distinguish different parts of speech

We evaluated whether the neural signals differentiated between nouns and adjectives. Nouns and adjectives were matched for their number of syllables and word length to control for potential confounds not specific to parts of speech (**Table S3**, **Methods**). **Figure 2** shows the responses of an example electrode located in the orbital H-shaped sulcus within the left lateral orbitofrontal cortex (**Figure 2i** 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 2a, b**) or visual presentation (**Figure 2c, d**), for the first (**Figure 2a, c**), or second (**Figure 2b, 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 (**Figures 2e-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).

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

Even though nouns and adjectives were matched in their average number of syllables and word length, we 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, we asked whether grammar could contribute to the neural differences between nouns and adjectives. To address these questions, we 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, word length (vision) or number of syllables (audition) (**Methods**). The predictor coefficients in the GLM model for the example electrode in **Figure 2a-d** show that only the nouns versus adjectives label significantly explained the neural responses for both auditory and visual presentation (**Figure 2j**). 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 2a-j**.

The locations of electrodes that robustly distinguished nouns and adjectives (orange in **Figure 2k**) revealed 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. We 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 1b-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.

We 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 2k**, for the 2 auditory-only electrodes, the number of syllables also significantly contributed towards explaining the neural responses. **Figure S3** 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 S4** 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 S4 k,l** shows the locations of auditory only (white circles) and visual only electrodes (black circle) in the left and the right hemispheres, respectively.

Nouns and adjectives differ in their usage frequency. We asked whether the differences in the neural responses to nouns versus adjectives depended on usage frequency. To address this question, we randomly subsampled the trials to match the distribution of Google Ngram frequency (**Methods**). **Figure S5a** shows matched noun and adjective distributions for the example electrode shown in **Figure 2a-k.** This electrode showed differential responses between parts of speech for auditory (**Figure S5b,c**) and visual (**Figure S5d,e**) stimuli during word1 (**Figure S5b,d**) and word2 (**Figure S5c,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 **Figures 2** and **S5** whereas the other electrodes maintained their selectivity in most but not all conditions.

Within our stimulus set, there were two subcategories of nouns, animals and food, and there were two subcategories of adjectives, concrete and abstract (**Table S3**). We asked whether the electrodes that showed differential responses generalized across different word subcategories. The example electrode in **Figure 2a-j** did not show differences between the two noun or adjective subcategories for either auditory stimuli (**Figure S6a, b, f, g)**, visual stimuli (**Figure S6c, d**, **h, i**), word 1 (**Figure S6a**, **c**, **f**, **h**), or word 2 (**Figure S6b**, **d**, **g**, **i**). Of the 13 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 S7c**, **d**, **h**, **i**), but was also apparent for auditory stimuli (**Figure S7a**, **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.

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 2a-j**. The remaining 23% showed responses that were stronger for adjectives compared to nouns ($\beta_{NvsA} < 0$) as illustrated by the example in **Figure S8 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 we displayed the electrode locations on the brain, we observed an anatomical separation between these two groups of responses (**Figure 2l,m,** x-axis: lateral to medial, y-axis: anterior to posterior, z-axis: ventral to dorsal). We compared noun- versus adjective- preferring electrodes along 3 axes of Montreal Neurological Institute 305 Coordinates (MNI305, units abbreviated as m.u.)[47]. Along the lateral to medial axis (x-axis in **Figure 2l,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 2l**), 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 2m**), 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% $\beta_{NvsA} > 0$, $p<10^{-4}$, permutation test, $n=10^6$ iterations, **Methods**).

A population of electrodes in the lateral orbitofrontal cortex can distinguish nouns from adjectives in individual trials and generalizes across words and modalities

To assess whether information about part of speech was available in individual trials, we used a machine learning pseudopopulation approach by combining electrodes within anatomically defined brain regions in the Desikan-Killiany Atlas[48]. We 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. We trained an SVM classifier with a linear kernel to distinguish between nouns and adjectives and tested the classifier on held-out data (**Methods**). **Figure 3** shows decoding accuracy for the left (**Figure 3a,d,g**) and the right (**Figure 3b,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 3a**). 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 1b-g**), decoding performance was much higher for the left LOF compared to the right LOF (compare **Figure 3a** versus **Figure 3b**). The differences between the left and right LOF persisted after randomly subsampling to equalize the number of electrodes across hemispheres for all regions (**Figure S9a,b**).

In **Figure 3a,b**, word 1 and word 2 are combined. Decoding performance in the left LOF was also high when separately considering word 1 (**Figure S10a-c**) and word 2 (**Figure S10d-f**). Furthermore, the machine learning classifier was able to generalize across words, as evidenced by the decoding performance when training on the responses to word 1 and testing on the responses to word 2 (**Figure 3d,e**), and vice versa (**Figure 3g,h**). Similarly, auditory and visual trials are combined in **Figure3a,b**. Decoding performance in the left LOF was also high when separately considering auditory trials (**Figure S10g-i**) and visual trials (**Figure S10j-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 S10m-o**) and vice versa (**Figure S10p-r**).

We extended the analyses in **Figure 3a,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 3c**). However, in contrast to the results for the left LOF, the decoding results for other regions were less robust

(**Figure S9c**) and did not generalize across words (**Figure 3f**,**i**) or across modalities (**Figure S10o,r**).

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 4a-h** shows the responses of an example electrode located in the left LOF (**Figure 4k**). This electrode showed a stronger response to nouns compared to adjectives for auditory stimuli (**Figure 4a**, **b**, **e**, **f**), for visual stimuli (**Figure 4c**, **d**, **g**, **h**), for Word 1 (**Figure 4a**, **c**, **e**, **g**), and for Word 2 (**Figure 4b**, **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 4e-h**). The GLM analysis showed that nouns versus adjectives was the only significant predictor in English trials (**Figure 4i**), and Spanish trials (**Figure 4j**). 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 4k,** 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 4k**, we show all electrodes from the left LOF that showed part-of-speech encoding from different participants (**Table S7**). We 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 4k** 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.

Multimodal neural signals distinguished nouns and verbs in full sentences

The experiment presented thus far concerned the responses to nouns and adjectives within minimal phrases. We 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. We 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 5a**, **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. We 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 S11 (**see also **Table S11, Methods).**

Following the procedures described in the analyses of neural responses to nouns versus adjectives, we evaluated whether neural signals differentiated between nouns and verbs. **Figure 5** shows the responses of an example electrode located in the pars triangularis (**Figure 5f** denotes the electrode location). The neural responses are aligned to word onset for auditory presentation (**Figure 5b**) or visual presentation (**Figure 5c**). 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 (**Figures 5d,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 we 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, e.g., “the girls ate cakes”) or not (NS, e.g., “the cakes ate girls). To evaluate whether word features and semantic features contributed to the neural signals underlying parts of speech, we 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 5b-e** show that only the nouns versus verbs label significantly explained the neural responses (**Figure 5g**). 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 5b-g**. The locations of these electrodes are shown in **Figure 5h,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 5b-e**. We observed an anatomical separation between these two groups of responses (**Figure 5j,k,** x-axis: lateral to medial, y-axis: anterior to posterior, z-axis: ventral to dorsal**)**. We compared noun- versus verb-preferring electrodes along 3 axes of Montreal Neurological Institute 305 Coordinates (MNI305, units abbreviated as m.u.)[47]. Along the anterior-posterior axis (y-axis in **Figure 5j,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 5j**), 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 5k,** 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).

Discussion

We described neurophysiological signals that selectively discriminate between two parts of speech, nouns and adjectives (**Figure 2**). This selectivity was robust to orthographic variables such as word length, phonetic features such as number of syllables, and word occurrence statistics (**Figure 2**). This selectivity for part of speech generalized across sensory modalities (**Figures 2**, **3**, **4**), word positions, grammatical correctness and motor outputs (**Figures 2**, **3**, **4**), and semantic groups of nouns and adjectives (**Figure S6**). These neurophysiological signals enable discrimination between parts of speech even in single trials (**Figures 2, 3**). Electrodes that uniquely distinguished nouns from adjectives were clustered within a small, circumscribed region of the lateral orbitofrontal cortex, lateralized to the left hemisphere (**Figure 2**, **4**). Neural discrimination of nouns from adjectives was apparent in the LOF in English-speaking and Taiwanese-speaking participants (**Figure 4**). Interestingly, in a bilingual participant, the same electrodes within the left LOF distinguished nouns and adjectives in both English and in Spanish (**Figure 4**). Extending the study of minimal phrases, we conducted an additional experiment where we showed neural signals that distinguished nouns from verbs within full sentences (**Figures 5**).

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 or verbs (**Table S8**). Thus, the words used in this study had a prototypical interpretation as either, noun, adjective, or verb. 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 results in **Figure 5** demonstrates a selective representation of parts of speech that extends to full sentences, invariant to changes in semantics.

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[49]). 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[49]. 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.

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.,[12-16,21,27,32,37,38]). While we 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 descriptions[50,51]. 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[52]. Indeed, there exist strong anatomical and functional connections between the medial temporal lobe and frontal regions that could link language and memory formation[53].

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[54-56]. Neurobiological tracings from rats, mice, and macaques have identified LOF as a nexus of many inputs[56] 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[53,55,57-61]. 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[55,56,62]. Interestingly, work on Primary Progressive Aphasia, and frontotemporal lesions implicate the orbitofrontal cortex in word and sentence comprehension deficits[8,62-64] (see also[31]). 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). 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.

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[65,66]. 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**, **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 we 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[67], 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.

Methods

Preregistration

This study was preregistered on the Open Science Framework (OSF) website. The preregistration DOI is: https://doi.org/10.17605/OSF.IO/8TU2G.

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.

Participants

We recorded data from 20 participants (9 male, 9-60 years old, 2 left-handed, 2 ambidextrous, **Table S1**) with pharmacologically resistant epilepsy for the minimal phrase task (**Fig. 1a**) and 17 participants (7 male, 13-50 years old, 3 left-handed, Table S9) for the full sentence task (**Fig. 5a**). All experiments were conducted while participants stayed at Children's Hospital Boston (CHB), Brigham and Women's Hospital (BWH), Taipei Veterans General Hospital (TVGH), or Cleveland Clinic (CC). All studies were approved by each hospital's institutional review boards and were carried out with the participants' informed consent.

Recordings and Electrode Locations

Participants were implanted with intracranial electrodes via stereo electroencephalography (sEEG) (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). The sampling rate was 2048 Hz at BCH and TVGH, 1024 Hz or 512 Hz at BWH, and 1000 Hz at CC. 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 for the minimal phrase task and 1,593 for the full sentence task[68].

Following implantation, electrodes were localized by co-registration of pre-operative T1 MRI and post-operative CT scans using the iELVis software[47]. We used FreeSurfer to segment MRI images, upon which post implant CT was rigidly registered[69]. Electrodes were marked in the CT aligned to pre-operative MRI using the Bioimage Suite[70]. The Desikan-Killiany (DK) atlas was used to assign the electrodes locations. **Figure 1b-g** and **Table S2** show the locations of all the electrodes.

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[71]. 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**. 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 participants) 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 5a**. Participants were presented with four-word sentences. There was a 600 ms fixation, followed by four words presented sequentially for 875 ms each. After the last word there was a 1,000 ms delay with a gray screen and then an image was presented.

There were two types of trials: semantical (50% of trials, e.g., “the girls ate cakes”), and non-semantic (50% of trials, e.g., “the cakes ate girls”). The non-semantic sentences were formed by swapping the nouns of the correct sentences, without changing the grammatical correctness of the sentence (see **Table S10,** for example sentences). The order of trials (stimulus presentation modality and semantical/non-semantic structure) was randomly interleaved. Participants 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. An accurately described image was *only* possible for semantical sentences We conducted the experiment in 3 languages, English (13 monolingual), Taiwanese (3 monolingual) and Hindi (1 monolingual participant).

Data Analyses

Preprocessing

For the minimal phrase task, 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[68]. Similarly, for the sentence task, a total of 2,507 electrode contacts were implanted, 914 of which were excluded from the analysis. 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[72]. 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 S1**). 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 to explaining the responses of a given electrode.

$$AUC = \beta_0 + \beta_{NvsA} NvsA + \beta_{GvsUG} GvsUG + \beta_{NSyllables} NSyllables + \beta_{WordLength} WordLength \quad (1)$$

where AUC is the area under the response curve (e.g., **Figure 2a**) from 200 ms to 800 ms after word onset, $\beta_0$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 2j**). 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 GLM followed the one in the minimal phrase task. The GLM in this case was:

$$AUC = \beta_0 + \beta_{NvsV} NvsV + \beta_{SvsNS} SvsNS + \beta_{NSyllables} NSyllables + \beta_{WordLength} WordLength \quad (2)$$

where AUC is the area under the response curve (e.g., **Figure 5b**) from 200 ms to 800 ms after the word onset, $\beta_0$is a constant additive term, NvsV is 1 for Nouns and -1 for Verbs, SvsNS is 1 for semantic trials and -1 for non-semantic trials, NumberOfSyllables is 1, 2 or 3 (and 0 for visual trials), and WordLength goes from 3 to 10 (and 0 for auditory trials).

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[73] to decode parts of speech in individual words combining all the electrodes in each brain region as defined by the Desikan-Killiany atlas[48] (**Figure S9**). 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: 65-150Hz). 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 3c,f,i, Figure S9c**, **Figure S10c,f,i,l,o,r**). We also applied a threshold such that for a given region R

$$[\mu_R - 3 * \sigma_R]_{original\ data} > [\mu_R + 3 * \sigma_R]_{null\ data}$$

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).

Acknowledgments

We thank Marcelo Armendariz and Yen-Ling Kuo for helping translate the task to Spanish and Taiwanese, respectively. We thank Professors Alfonso Caramazza and Boriz Katz for suggestions and insightful comments throughout this work.

References

1 Chomsky, N. *The minimalist program*. (MIT Press, 1995).

2 Scott, S. K. From speech and talkers to the social world: The neural processing of human spoken language. *Science* **366**, 58-62, doi:10.1126/science.aax0288 (2019).

3 Pylkkanen, L. The neural basis of combinatory syntax and semantics. *Science* **366**, 62-66, doi:10.1126/science.aax0050 (2019).

4 Heilman, K. M. & Valenstein, E. *Clinical Neuropsychology*. (Oxford University Press, 1993).

5 Ojemann, G., Ojemann, J., Lettich, E. & Berger, M. Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *J Neurosurg* **71**, 316-326, doi:10.3171/jns.1989.71.3.0316 (1989).

6 Petrides, M. *Neuroanatomy of language regions in the human brain*. (Elsevier, 2014).

7 Mahon, B. Z. & Caramazza, A. Concepts and categories: a cognitive neuropsychological perspective. *Annual review of psychology* **60**, 27-51, doi:10.1146/annurev.psych.60.110707.163532 (2009).

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

9 Warrington, E. & Shallice, T. Category specific semantic impairments. *Brain* **107**, 829-854 (1984).

10 Nourski, K. V. *et al.* Gamma Activation and Alpha Suppression within Human Auditory Cortex during a Speech Classification Task. *J Neurosci* **42**, 5034-5046, doi:10.1523/JNEUROSCI.2187-21.2022 (2022).

11 Forseth, K. J. *et al.* A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* **141**, 2112-2126, doi:10.1093/brain/awy120 (2018).

12 Murphy, E. *et al.* Minimal Phrase Composition Revealed by Intracranial Recordings. *J Neurosci* **42**, 3216-3227, doi:10.1523/JNEUROSCI.1575-21.2022 (2022).

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

14 Sinai, A. *et al.* Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain* **128**, 1556-1570, doi:10.1093/brain/awh491 (2005).

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

16 Cometa, A. *et al.* Event-related causality in stereo-EEG discriminates syntactic processing of noun phrases and verb phrases. *J Neural Eng* **20**, doi:10.1088/1741-2552/accaa8 (2023).

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

18 Bemis, D. & Pylkkänen, L. Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. . *Cereb Cortex* **23**, 1859-1873 (2013).

19 Bemis, D. K. & Pylkkanen, L. Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J Neurosci* **31**, 2801-2814, doi:10.1523/JNEUROSCI.5003-10.2011 (2011).

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

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

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

23 Castellucci, G. A., Kovach, C. K., Howard, M. A., 3rd, Greenlee, J. D. W. & Long, M. A. A speech planning network for interactive language use. *Nature* **602**, 117-122, doi:10.1038/s41586-021-04270-z (2022).

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

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

26 Forseth, K., Pitkow, X., Fischer-Baum, S. & Tandon, N. What the brain does as we speak. *bioRxiv* **bioRxiv 2021.02.05.429841**, doi:https://doi.org/10.1101/2021.02.05.429841 (2021).

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

28 Bhaya-Grossman, I. & Chang, E. F. Speech Computations of the Human Superior Temporal Gyrus. *Annual review of psychology* **73**, 79-102, doi:10.1146/annurev-psych-022321-035256 (2022).

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

30 Khanna, A. R. *et al.* Single-neuronal elements of speech production in humans. *Nature* **626**, 603-610, doi:10.1038/s41586-023-06982-w (2024).

31 Jamali, M. *et al.* Semantic encoding during language comprehension at single-cell resolution. *Nature*, doi:10.1038/s41586-024-07643-2 (2024).

32 Chomsky, N., Gallego, A. & Ott, D. Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan Journal of Linguisttics*, 226-261 (2019).

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

34 Hagoort, P. The neurobiology of language beyond single-word processing. *Science* **366**, 55-58, doi:10.1126/science.aax0289 (2019).

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

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

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

38 Cai, J., Hadjinicolau, A., Paulik, A., Williams, Z. & Cash, S. Natural language processing models reveal neural dynamics of human communication. *Biorxiv* **2023.03.10.531.095** (2023).

39 Tenney, I., Dipanjan, D. & Pavlick, E. BERT rediscovers the classical NLP pipeline. *arXiv* **1905.05950** (2019).

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

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

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

43 Woolnough, O., Forseth, K. J., Rollo, P. S., Roccaforte, Z. J. & Tandon, N. Event-related phase synchronization propagates rapidly across human ventral visual cortex. *Neuroimage* **256**, 119262, doi:10.1016/j.neuroimage.2022.119262 (2022).

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

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

46 Bansal, A. *et al.* Neural Dynamics Underlying Target Detection in the Human Brain. *Journal of Neuroscience* **34**, 3042-3055, doi: 10.1523/JNEUROSCI.3781-13.2014 (2014).

47 Groppe, D. M. *et al.* iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *J Neurosci Methods* **281**, 40-48, doi:10.1016/j.jneumeth.2017.01.022 (2017).

48 Desikan, R. S. *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).

49 Tager-Flusberg, H. & Seery, A. in *Neural circuit development and function in the brain* (eds JLR Rubenstein & P Rakic) 315-330 (Elsevier, 2013).

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

51 Quian Quiroga, R., Kraskov, A., Koch, C. & Fried, I. Explicit encoding of multimodal percepts by single neurons in the human brain. *Current biology : CB* **19**, 1308-1313, doi:10.1016/j.cub.2009.06.060 (2009).

52 Quian Quiroga, R., Kreiman, G., Koch, C. & Fried, I. Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Science* **12**, 87-91 (2008).

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

54 Wojtasik, M. *et al.* Cytoarchitectonic Characterization and Functional Decoding of Four New Areas in the Human Lateral Orbitofrontal Cortex. *Front Neuroanat* **14**, 2, doi:10.3389/fnana.2020.00002 (2020).

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

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

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

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

59 Nogueira, R. *et al.* Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. *Nat Commun* **8**, 14823, doi:10.1038/ncomms14823 (2017).

60 Noonan, M. P. *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, doi:10.1073/pnas.1012246107 (2010).

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

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

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

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

65 Fried, I., Cerf, M., Rutishauser, U. & Kreiman, G. *Single neuron studies of the human brain. Probing cognition.*, 408 (MIT Press, 2014).

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

67 Dijksterhuis, D. E. *et al.* Pronouns reactivate conceptual representations in human hippocampal neurons. *bioRxiv*, 2024.2006.2023.600044, doi:10.1101/2024.06.23.600044 (2024).

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

69 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194 (1999).

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

71 Brainard, D. The Psychophysics Toolbox. *Spatial Vision* **10**, 433-436 (1997).

72 Mitra, P. & Bokil, H. *Observed brain dynamics* (Oxford University Press, 2008).

73 Bansal, A., Golby, A., Madsen, J. & Kreiman, G. in *COSYNE.*

Figure Captions

Figure 1. Task schematic, 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-f.** 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 view (n=1108). **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 auditory (**i**) and visual (**j**) presentations (see color scale on right).

Figure 2. 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. 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). All the 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**). 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. Linear support vector machines separating these electrodes are shown with a thick black line. 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:** top x-axis), 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).

Fig. 3 | Neural signals distinguishing nouns and adjectives in single trials generalize across Word1 and Word2.

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).

Figure 4. Neural signals in left LOF generalize across languages in a bilingual participant and in monolingual participants.

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 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 participants. Electrodes from the bilingual patient are in green with a black arrow indicating the example electrode. Electrodes from one monolingual English participant are in pink and those from 2 monolingual Taiwanese participants are in brown.

Fig. 5 | 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”). Participants 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 electrode 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 nouns and verbs, respectively, in **d,e**.

f. Location of the example electrode in the pars triangularis (see **Figure S11** for electrode coverage).

g. Z-scored β coefficients for the Generalized Linear Model used to predict area under the curve between 200 ms and 800 ms post word onset, using four task predictors: nouns versus verbs, semantic versus non-semantic, 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).

Figure 1

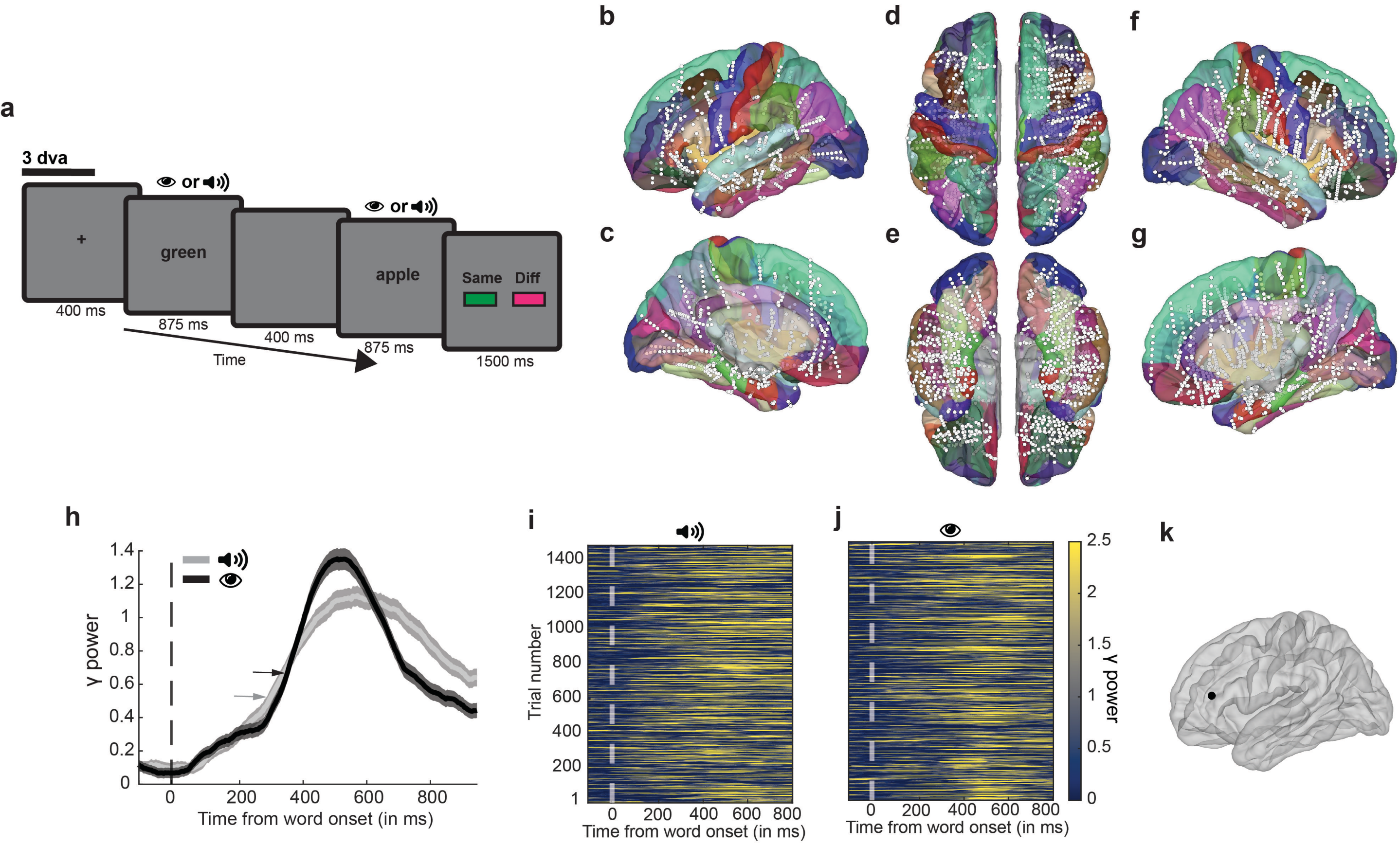

Fig. 1 | Task design, electrode locations and multimodal responses.
a. Task schematic. Two words were sequentially presented either in visual modality or auditory modality. Participants indicated whether the two words were the same (e.g., “apple apple” or “green green”, 8% of trials of each type) or 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 view (n=1108). **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 auditory (**i**) and visual (**j**) presentations (see color scale on right).

Figure 2

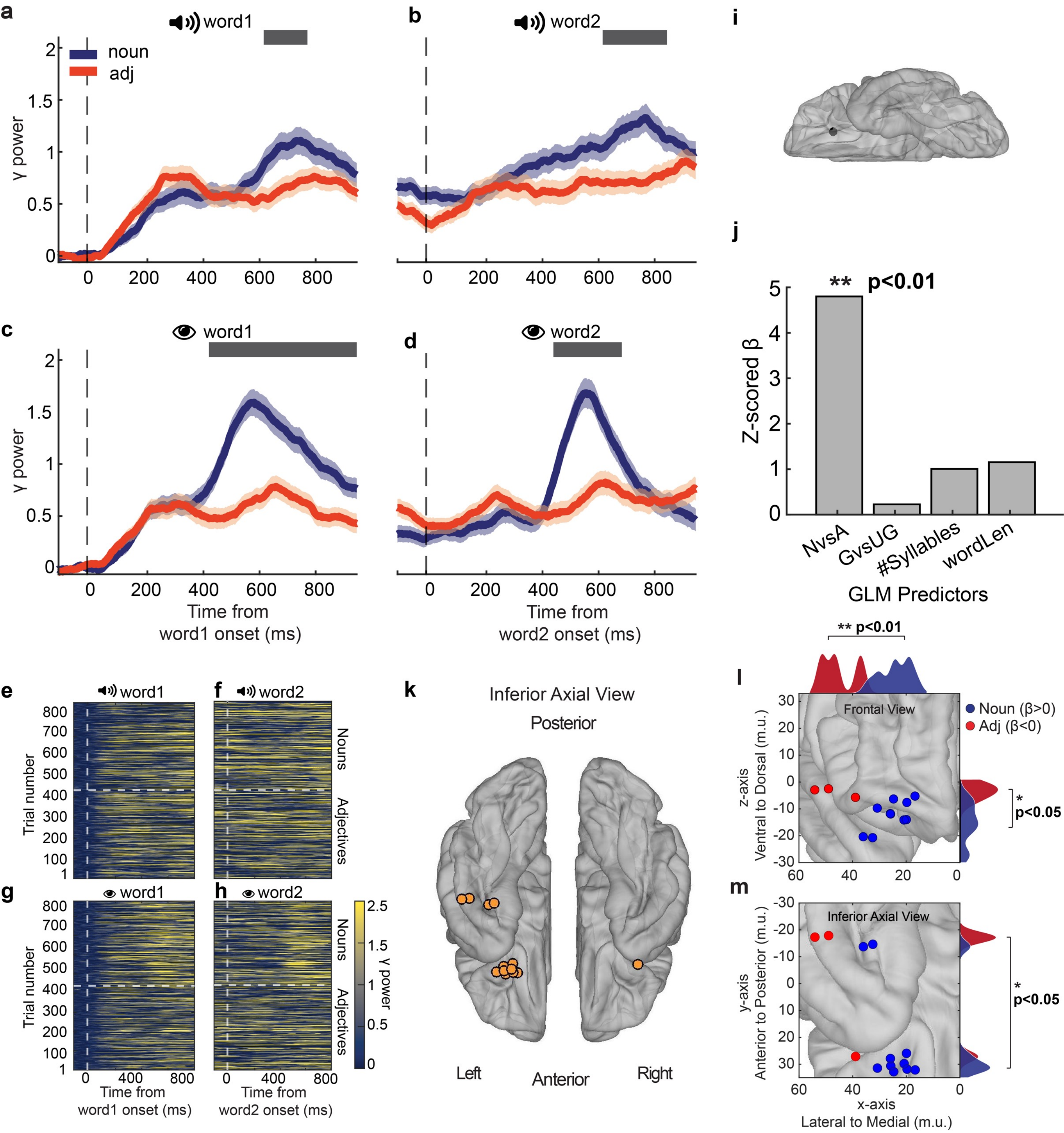
a
word1
noun
adj
γ power
0
0.5
1
1.5
2
0
200
400
600
800
b
word2
c
word1
d
word2
Time from
word1 onset (ms)
Time from
word2 onset (ms)
i
j
Z-scored β
** p<0.01
NvsA
GvsUG
#Syllables
wordLen
GLM Predictors
e
word1
f
word2
g
word1
h
word2
Trial number
Nouns
Adjectives
γ power
2.5
1.5
1
0.5
0
k
Inferior Axial View
Posterior
Left
Anterior
Right
l
** p<0.01
Frontal View
Noun (β>0)
Adj (β<0)
z-axis
Ventral to Dorsal (m.u.)
* p<0.05
m
Inferior Axial View
y-axis
Anterior to Posterior (m.u.)
* p<0.05
x-axis
Lateral to Medial (m.u.)

Fig. 2 | 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. 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). All the 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).

Figure 3

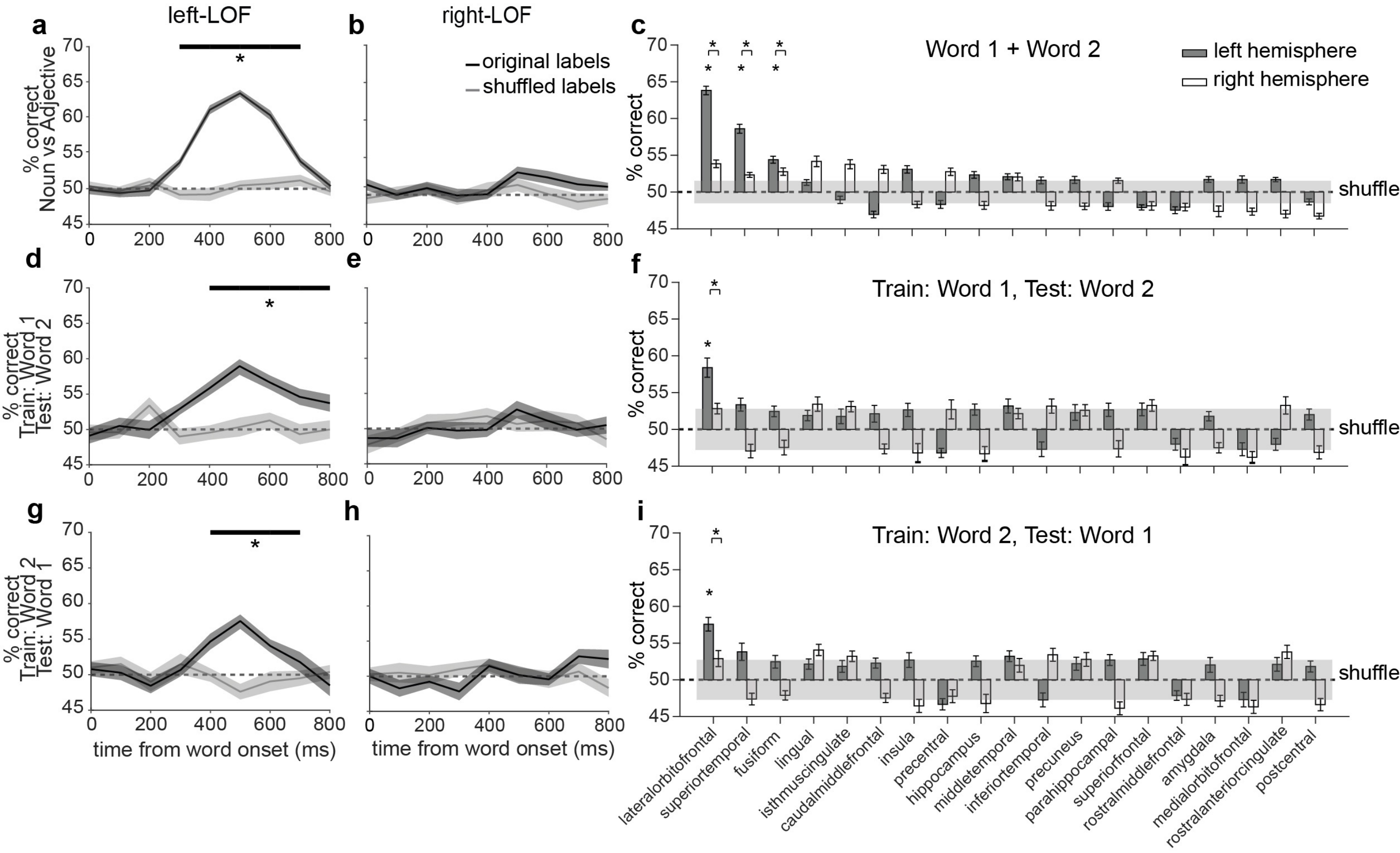

Fig. 3 | Neural signals distinguishing nouns and adjectives in single trials generalize across Word1, and Word2.

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).

Figure 4

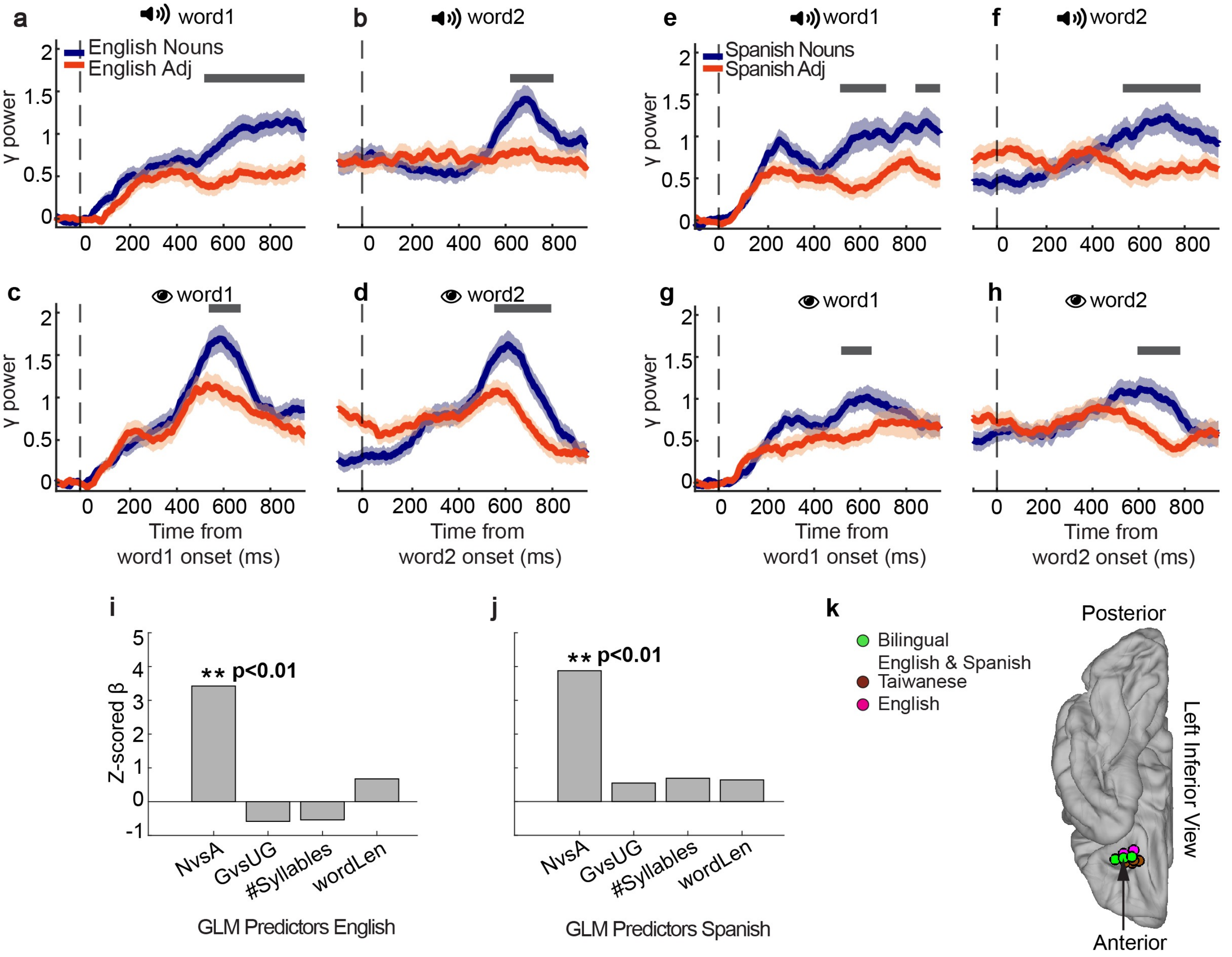

Fig. 4 | Neural signals in left LOF generalize across languages in a bilingual participant and in monolingual participants. a-h. Trial averaged responses of an electrode in the left lateral orbitofrontal cortex from a bilingual patient. The format follows **Fig. 2a-d**. (**a-d**) English words (audio: 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 participants. Electrodes from the bilingual patient are in green with a black arrow indicating the example electrode. Electrodes from one monolingual English participant are in pink and those from 2 monolingual Taiwanese participants are in brown.

Figure 5

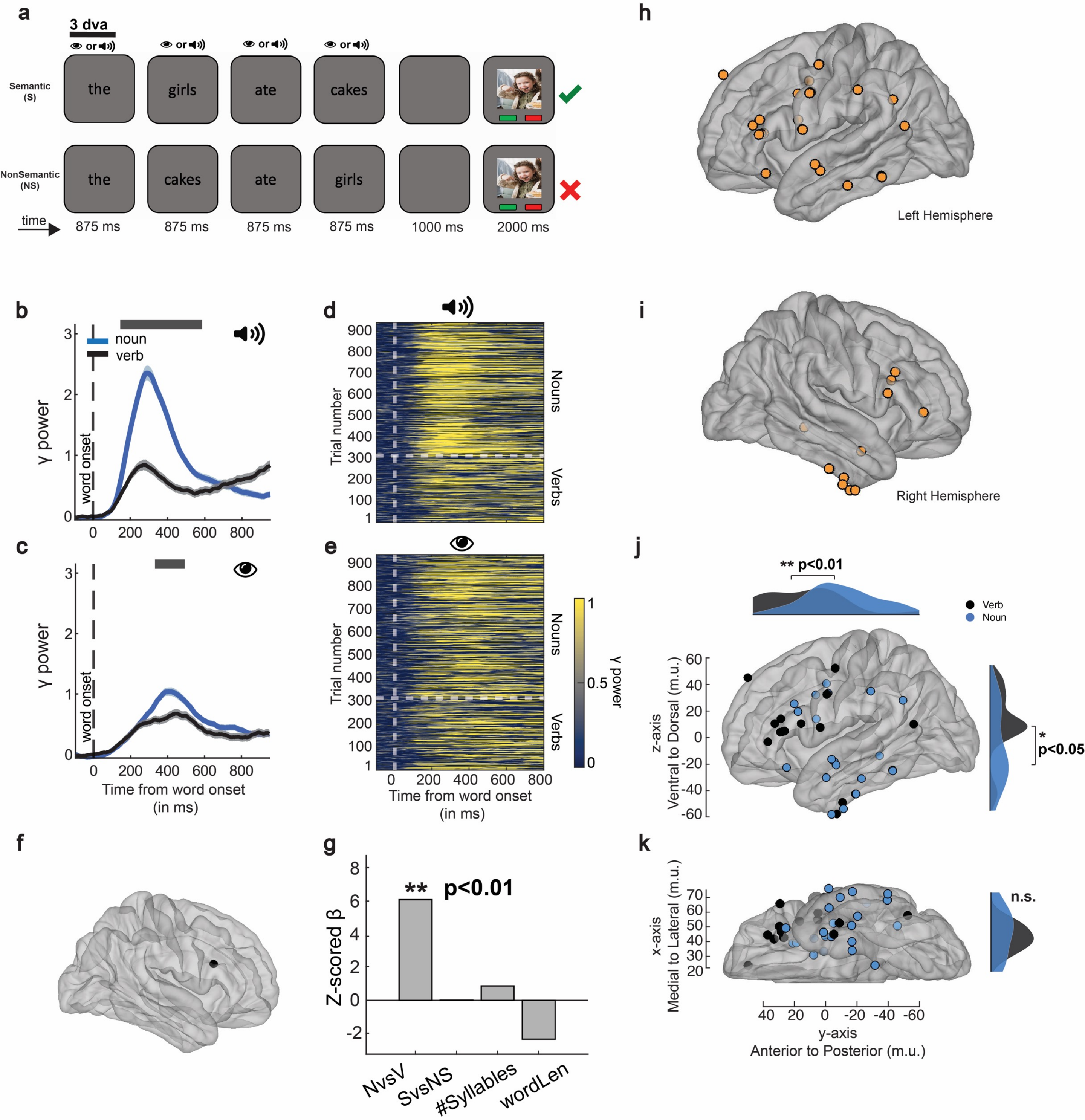
a
3 dva
or
Semantic (S)
the
girls
ate
cakes
NonSemantic (NS)
the
cakes
ate
girls
time
875 ms
875 ms
875 ms
875 ms
1000 ms
2000 ms
b
noun
verb
γ power
word onset
0
1
2
3
0
200
400
600
800
c
γ power
word onset
Time from word onset (in ms)
d
Trial number
900
800
700
600
500
400
300
200
100
1
Nouns
Verbs
e
Trial number
Time from word onset (in ms)
γ power
1
0.5
0
f
g
** p<0.01
Z-scored β
8
6
4
2
0
-2
NvsV
SvsNS
#Syllables
wordLen
h
Left Hemisphere
i
Right Hemisphere
j
** p<0.01
Verb
Noun
z-axis
Ventral to Dorsal (m.u.)
60
40
20
0
-20
-40
-60
*
p<0.05
k
x-axis
Medial to Lateral (m.u.)
70
60
50
40
30
20
n.s.
40
20
0
-20
-40
-60
y-axis
Anterior to Posterior (m.u.)

Fig. 5 | 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”). Participants 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 (see **Figure S11** for electrode coverage).

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).