From single words to sentence production: Shared cortical representations but distinct temporal dynamics

Adam M. Morgan[1*], Orrin Devinsky[2], Werner K. Doyle[1], Patricia Dugan[1], Daniel Friedman[1], Adeen Flinker[1,3]

[1*]Neurology Department, NYU Grossman School of Medicine.
[2]Neurosurgery Department, NYU Grossman School of Medicine.
[3]Biomedical Engineering Department, NYU Tandon School of Engineering.

*Corresponding author(s). E-mail(s): adam.morgan@nyulangone.org;

Abstract

Sentence production is the uniquely human ability to transform complex thoughts into strings of words. Despite the importance of this process, language production research has primarily focused on single words. It remains an untested assumption that insights from this literature generalize to more naturalistic utterances like sentences. Here, we investigate this using high-resolution neurosurgical recordings (ECoG) and an overt production experiment where patients produce six words in isolation (picture naming) and in sentences (scene description). We used machine learning models to identify the unique brain activity pattern for each word during picture naming, and used these patterns to decode which words patients were processing while they produced sentences. In sensorimotor cortex, this procedure predicted each noun in the order it was said in the sentence, confirming that words share cortical representations across tasks. However, in inferior and middle frontal gyri (IFG and MFG), the order in which words were processed depended on the syntactic structure of the sentence. This dynamic interplay between sentence structure and word processing reveals that sentence production is not simply a sequence of single word production tasks. We argue that it is time for the field to leverage the extensive literature on word production for studying more naturalistic linguistic constructs like sentences.

Keywords: ECoG, language, brain decoding, speech production, lexical representation

1 Introduction

Many species in the animal kingdom use symbolic forms of communication: dolphins have names [1], bees dance to signal nectar locations [2], and monkeys and birds use predator-specific calls to warn of threats [3, 4]. While human *lexical* (i.e., word) knowledge is particularly vast, involving tens of thousands of words, the truly remarkable feature of human language is our ability to combine these words into sentences, enabling us to express a limitless number of novel thoughts and ideas.

This communicative ability is central to who we are, but remains poorly understood at the neural level. In particular, the neuroscience of sentence production has been hindered by limitations of traditional noninvasive neural measures, which limit spatial or temporal resolution and are susceptible to motor artifacts, and by the difficulty of experimentally controlling what sentences participants say. Due largely to these challenges, language production research has remained primarily focused on single words. This leaves an important gap in understanding how the brain produces more complex linguistic constructions like sentences. Word production research typically employs *picture naming* paradigms where a participant sees a picture of, e.g., a dog, and says "dog." Among the important insights from this work is that words are not unitary representations. Instead, lexical knowledge involves distinct representations of a word's semantic (i.e., meaning) [5], phonological [6], articulatory [7], and grammatical features [8]. During production, these representational stages are generally thought to come online in a feedforward sequence, starting with meaning and ending with articulation (and perception from sensory feedback) [9–11]. Furthermore, each stage is associated with distinct cortical regions [12, 13], with articulatory planning in inferior frontal gyrus (IFG) [14, 15]; articulation in sensorimotor cortex (SMC); feedback in superior temporal gyrus (and visual cortices for sign languages); grammatical features in middle temporal lobe (MTL) [16], and semantics distributed bilaterally throughout cortex [5].

In contrast to the wealth of knowledge the field has accumulated about single word production, very little is known about the type of speech that is unique to our species: sentences [17–19]. While some work has overcome the obstacles to studying production with non-invasive neural measures [e.g., 15, 17–27], these studies have largely focused on differences between comprehension and production rather than drawing on progress in word production research. One of many open questions regarding sentence production, then, is whether insights from the word production literature can inform our understanding of sentence production. Here, we scale up from the neuroscience of words to the neuroscience of sentences, testing the assumption that what we know about single words generalizes to more complex linguistic constructs.

To do so, we identify the unique cortical activity patterns that encode six particular words during a picture naming task. We then ask whether these cortical representations are the same for words in list and sentence contexts. By recording electrical potentials directly from the cortical surface in ten neurosurgical patients (ECoG), we achieve high spatial and temporal resolution and avoid motor artifacts, bypassing limitations of traditional non-invasive neural measures. We employ a controlled production experiment (Fig. 1A) and sophisticated machine learning techniques, and demonstrate that words' cortical representations are in fact shared across tasks. However, the order

in which these representations are processed by the brain differs depending on the syntactic structure of the sentence, revealing a dynamic interplay between syntax and word production processes.

2 Results

We recorded ECoG from ten neurosurgical patients with electrodes implanted in left peri-Sylvian cortex (Fig. 1B). Patients performed an overt speech production experiment involving the production of the same six words in three tasks: picture naming, list production, and sentence production. In picture naming trials, patients repeatedly saw and named six cartoon characters one at a time. To maximize discriminability, these characters – *chicken*, *dog*, *Dracula*, *Frankenstein*, *ninja*, and *nurse* – differed along a number of dimensions (phonology, number of syllables, proper vs. common noun, etc.). During sentence production, patients overtly described cartoon vignettes depicting transitive actions (e.g., poke, scare, etc.) in response to a preceding question. Questions were constructed using either active syntax ("Who poked whom?") or passive syntax ("Who was poked by whom?"), implicitly priming patients to respond with the same structure ("The chicken poked Dracula" or "Dracula was poked by the chicken") [28]. Finally, patients completed a list production task, where the same vignettes as in the sentence production trials were preceded by an arrow rather than a question, indicating the direction in which participants should list the two characters in the scene: left-to-right (e.g., "chicken Dracula") or right-to-left ("Dracula chicken"). We quantified neural activity as high gamma broadband activity ($70 - 150$ Hz), normalized (z-scored) to each trial's 200 ms pre-stimulus baseline, which correlates with underlying neuronal spiking and BOLD signal [29, 30].

We began by looking at the mean neural activity for each task in seven regions of interest (ROIs) which have previously been implicated in word production [12, 13]. Prior to this and subsequent analyses, we followed previous work [31–36] and temporally warped all trials, setting response times to the median trial duration for each task (-758 ms for picture naming, -1141 ms for sentence production, and -801 ms for list production; see Methods and Supplementary Fig. S1). This boosts signal to noise ratio [34, 36] and tempers extraneous differences. ROIs showed a variety of temporal patterns (Fig. 1B), with the highest levels of activity across tasks achieved in sensorimotor cortex (SMC) during articulation. However, not all of this activity reflects word processing, as various general systems like attention and working memory are also involved in speech production. Notably, many electrodes showed distinct temporal profiles for certain words (Fig. 1C, top). We quantified the amount of evidence for word specificity in each electrode using Bayesian ANOVAs (Fig. 1C, bottom). This information was broadly distributed across cortex, increasing from stimulus onset to speech onset (Fig. 1D). Figure 1E shows the maximum standardized difference (Hedge's g) between each word and all the rest between -750 and 250 ms from speech onset. These networks exclude electrodes with high g-values for more than one word, meaning they are unique to each word.

To more accurately identify word-specific activity patterns, we performed a series of decoding analyses [37, 38] on the picture naming data. In essence, this analysis

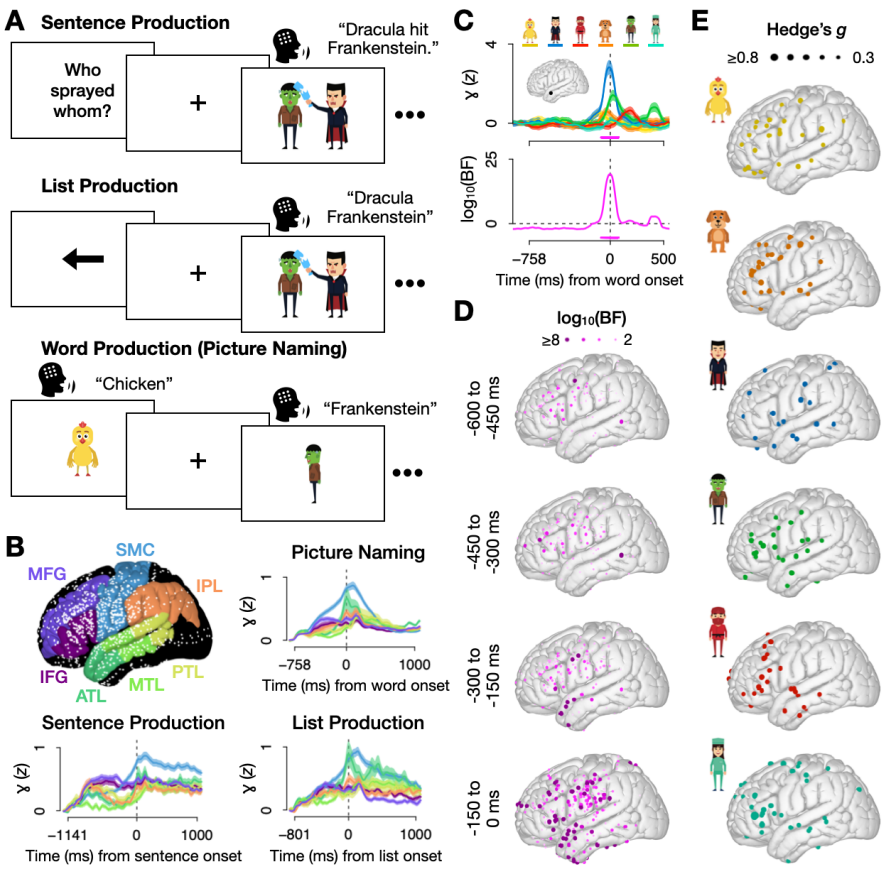

Fig. 1 (A) Task design: In sentence production trials, participants described static cartoon scenes in response to preceding questions. Scenes involved two of the six characters used throughout the experiment (*chicken, dog, Dracula, Frankenstein, ninja, nurse*). Half of the questions were manipulated to appear in active syntax (e.g., "Who sprayed whom?"), implicitly priming active responses ("Dracula hit Frankenstein."). The other half had passive syntax ("Who was sprayed by whom?"), priming passive responses ("Frankenstein was hit by Dracula"). In list production trials, participants saw an arrow pointing to the left, in which case they listed the two characters in the subsequent scene from right to left ("Dracula Frankenstein"), or to the right ("Frankenstein Dracula"). In picture naming trials, the six characters repeatedly appeared one at a time and participants responded with a word (e.g., "chicken"). (B) We recorded electrical potentials from 1256 electrodes (white dots) placed directly on the cortical surface in 10 patients. We identified 7 regions of interest (ROIs) in the word production literature. Line plots show the mean neural activity (z-scored high gamma amplitude) and standard error per task and ROI, locked to speech onset. (C) A sample electrode: mean activity per word during picture naming (top) and the amount of evidence (BF = Bayes Factor) for word-specific information throughout picture naming trials (bottom; calculated with Bayesian ANOVAs where positive values indicate more evidence for word-specificity). (D) The maximum amount of evidence for word specificity during picture naming in each electrode in the four 150 ms windows leading up to speech onset ($t = 0$). (E) Cortical networks for each word. Only showing electrodes for which only one word has a maximum Hedge's g exceeding 0.3. (Hedge's g is a standardized difference metric, interpreted like Cohen's d.)

(schematized in Fig. 2A) learns the unique pattern of activity for each word in a "training" subset of picture naming trials (90% of trials). Then, it predicts which of the six words a patient is saying in each of the withheld "test" trials (10%) by assessing how similar that trial's activity pattern is to each of the six learned patterns. If a test trial's activity pattern is most similar to, e.g., the "chicken" pattern, the classifier will predict that the patient said "chicken." This would be scored a 1 if the patient was in fact saying "chicken," and 0 otherwise. By repeating this process for every time sample in every trial, and for 30 combinations of train and test data parcellations (i.e., 10-fold cross validation repeated three times; see Methods), we were able to calculate the mean prediction accuracy over time.

We repeated this analysis for each patient and for each ROI. Additionally, because words pass through distinct representational stages (e.g., conceptual, phonological, etc.), the time when the training data came from was consequential. For instance, if we trained the model on just data from $t = 0$ (speech onset), we would likely detect articulatory information and miss semantic, phonological, and grammatical aspects of lexical representation. To protect against this, we spanned time, training models on the mean high gamma activity in each 50 ms window from -750ms to 250ms relative to speech onset. In all, this resulted in 1280 classifiers: 10 patients $\times$ 7 ROIs $\times$ 20 training time windows (minus ROIs where patients had insufficient electrode coverage).

Figure 2B shows the prediction accuracies from four sample classifiers. These classifiers predict word identity above chance both before and after their respective training windows (black bars), revealing that whatever information our classifiers encode comes and stays online for longer than 50 ms. In Fig. 2C we plot the maximum accuracy of all classifiers in each region (across patients and training windows), revealing that word identity can be decoded above chance in all seven ROIs (see Supplementary Fig. S2 for all results by ROI). In Fig. 2D, we stacked the time series (like those in Panel B) from all 444 classifiers with significant prediction accuracies (pink highlights). This revealed several patterns. First, the closer a training window (black bar) was to articulation (starting at time 0), the better the decoding, perhaps reflecting higher signal-to-noise ratio for articulatory representations than for earlier stages. Second, training times (black) and significant prediction times (pink) tended to overlap, suggesting that the timecourse of representational stages is relatively consistent across picture naming trials. Finally, almost regardless of training time, most classifiers were able to decode above chance at speech onset ($t = 0$). This suggests that, contrary to the classical view of a series of feed-forward stages that go offline once the next stage has been activated, pre-articulatory representations stay online at least until speech onset, and post-articulatory representations are engaged throughout production.

To assess whether words have the same cortical representations in picture naming and list production, we tested the generalizability of the picture naming classifiers. We followed the same analysis pipeline depicted in Fig. 2A. However, instead of subsets, we use all of the picture naming trials to train classifiers. These classifiers were then used to predict word identity at each time sample during the production of lists like "Dracula Frankenstein." Figure 3A shows the proportion of trials where sensorimotor classifiers predicted the first word ("Dracula," in our example) and the second word ("Frankenstein"). Accuracies are time-locked the onset of the first word in the left

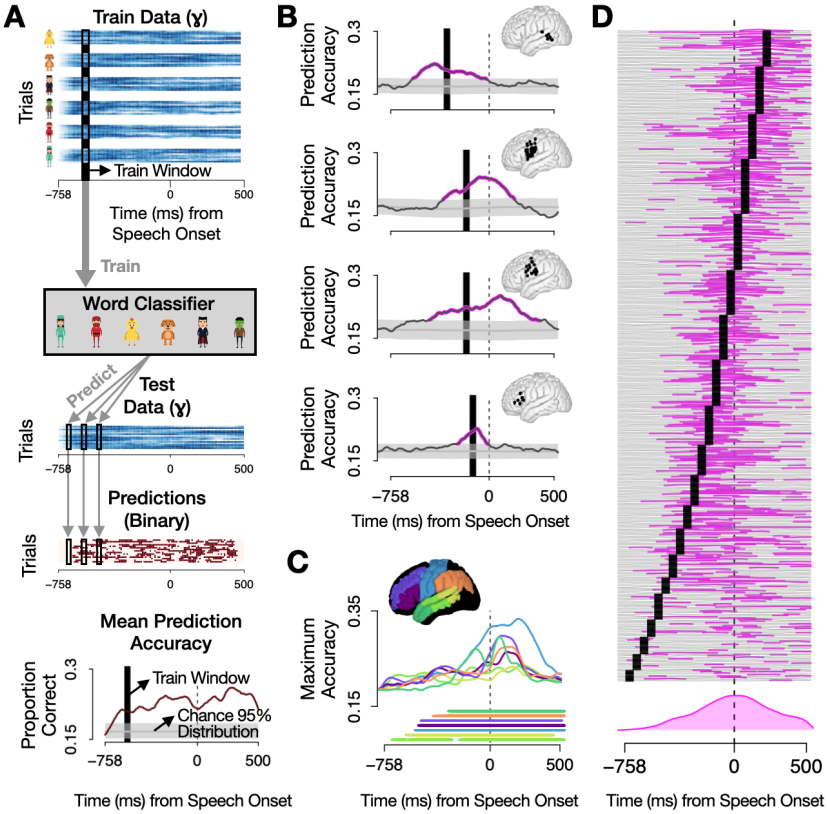

Fig. 2 Word-specific information in picture naming. (A) Simplified schematic of the analysis pipeline with simulated data. We trained a classifier on word identity using the time-averaged picture naming data in the window from -600 to -550 ms (black rectangle). We then predicted word identity in held out "test" trials, generating a binary prediction matrix (maroon): correct (1) or incorrect (0) for each test trial (row) at each time sample (column). We repeated this for 30 different train/test data parcellations (i.e., 10-fold validation, repeated 3 times). Averaging prediction accuracies across trials gives the mean accuracy over time (bottom). A permutation analysis generated 1000 results reflecting chance performance, and significance was determined with respect to this distribution (gray). Not depicted: train and test datasets were in fact 3 dimensional: trial × time × electrode (not shown). This whole process was conducted for each patient, each ROI for which a patient had coverage, and each of the 20 training windows spanning -750 to 250 ms, resulting in a total of 1280 prediction accuracy time series. (B) Prediction accuracies for four sample classifiers that predicted word identity with above-chance accuracy (pink highlights; $p < .05$ for 100 ms or $p < .01$ for 50 ms). From top to bottom, training and test data came from (1) Patient 9, PTL, -350 to -300 ms; (2) Patient 5, SMC, -200 to -150 ms; (3) Patient 6, SMC, -200 to -150 ms; and (4) Patient 5, IFG, -150 to -100 ms. (C) Each ROI's maximum prediction accuracy across classifiers (i.e., across patients and training time windows); significance denoted by bars at bottom ($p < .05$ for 100 ms or $p < .01$ for 50 ms). (See Supplementary Fig. S2 for full results broken down by ROI.) (D) Prediction accuracies from the 444 significant classifiers (across patients, training time windows, and ROIs), stacked vertically to highlight the timecourse of word-specific information in picture naming. Each horizontal line corresponds to one classifier, and pink denotes where accuracy was above chance. Black bars represent the time window the training data came from. The pink curve at the bottom shows the density of the significant predictions, revealing that the most significant decoding occurred at speech onset ($t = 0$).

panel and the second word in the right. These sample results come from classifiers trained on data from SMC between 50 and 100 ms after speech onset during picture naming, and averaged across participants. We accurately predict each word as it is being said – e.g., Dracula when patient says Dracula and then Frankenstein when the patient says Frankenstein. Prediction accuracies from all 97 classifiers that significantly predicted one or both of the two nouns in the list are stacked in Fig. 3B, revealing similar temporal dynamics of lexical information as in picture naming. Overall, there were fewer significant classifiers (97) than for picture naming (444), likely reflecting the lower number of trials analyzed (an average of 359 per patient for picture naming vs. 48 for list production). The three temporal patterns we observed in picture naming were largely preserved, though to a lesser degree. First, the closer to speech onset the training data came from, the more significant detections the classifier made. Second, significant prediction times tended to overlap with training times, though this was less true for classifiers trained on pre-articulatory data. Finally, significant prediction times tended to overlap with speech onset, particularly for the second word in lists, where many classifiers showed above-chance accuracy during articulation even when they failed to do so during their own training time. Overall, the similarity to the picture naming results (Fig. 2) suggests that our list production task may involve similar cognitive processes as picture naming.

There is reason to believe that sentences may behave differently. Whereas word order in lists is linearly structured, word order in sentences is determined by words' syntactic position in a hierarchical structure, which is in turn based on a complex event-semantic representation. It stands to reason that sentence production may involve fundamentally different mechanisms for accessing and producing words. To test this, we used the same classifiers we trained for the list production analysis to predict word identity during sentences. Like lists, each sentence contained two nouns: the subject and the object, and we recorded the proportion of trials where classifiers predicted each of these. We started by analyzing sentences with active syntax, e.g., "Dracula is hitting Frankenstein," which, relative to non-canonical structures like the passive, are easier to process and better preserved in aphasic patients [22, 39–42]. Figure 4A shows the results from the same sensorimotor classifiers previously shown for lists in Fig. 3A. For active sentences, subjects and objects were decoded while each was being said (i.e., Dracula and Frankenstein in the sample sentence, but the particular words in subject/object position varied across trials). Stacking the prediction accuracy time series from all 83 significant classifiers (Fig. 4C) revealed that this was a general pattern: subjects and objects were predicted at their respective production times.

However, in active sentences, the order of the two nouns is confounded with their relative salience in the event. That is, the character performing the action is the subject and comes first, and the character being acted upon is the object and comes second [22]. It is therefore not entirely surprising that the brain processes words in active sentences the way it does in lists: in the order that they are produced. We wondered whether this pattern was generally true of sentence production, or only true when word order aligns with salience. A potentially interesting test case is the English passive (e.g., "Frankenstein is being hit by Dracula"), which involves reversing the

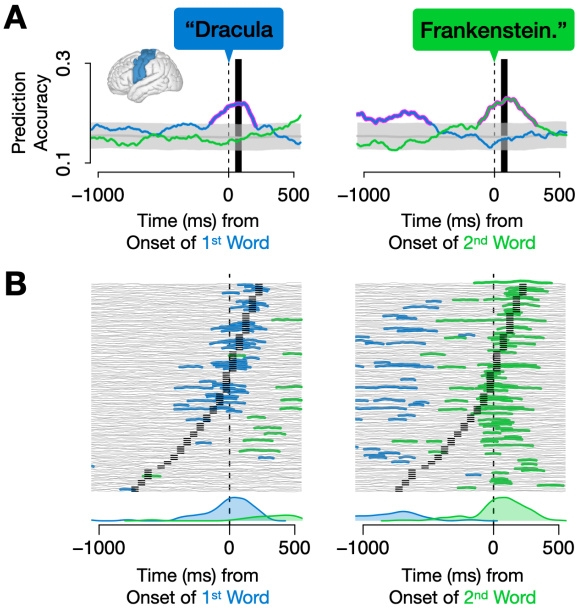

Fig. 3 Predicting words in lists using picture naming data. (A) Sample prediction accuracies for the first and second nouns in lists (e.g., "Dracula" and "Frankenstein" in the example) from classifiers trained on picture naming and tested on lists. The significant detections of the two nouns in the list are evidence of words' common cortical representations across tasks. Training data came from electrodes in SMC (highlighted on brain) between 50 and 100 ms post-speech onset (denoted by black bar) and the resulting prediction accuracies were averaged across patients. Significant prediction accuracy is highlighted in pink ($p < .05$ for 100 ms or $p < .01$ for 50 ms). (B) Prediction accuracy time series (like those in Panel A) from the 97 significant classifiers for lists, stacked vertically to highlight temporal patterns in word-specific information during list production. In this and subsequent stacks of prediction accuracies, each time series (horizontal line) corresponds to the same classifier across the left and right panels. Blue highlights show where a classifier predicted the first noun above chance; green for the second noun ($p < .05$ for 100 ms or $p < .01$ for 50 ms). Black bars denote the time window the training (picture naming) data came from. Blue and green density plots at the bottom summarize the significant predictions. As with picture naming decoding (Fig. 2D), the most significant detections of each word happened at that word's articulation onset.

order of nouns – i.e., producing the character being acted upon first, and the character doing the action second. Passive sentences thus present an opportunity to disentangle serial order and salience, as they convey the same meaning but reverse the order of words. We expected that in sensorimotor cortex, where articulatory information is encoded, words would be decoded in the same order as in speech. Figure 4C shows the predictions of the same SMC classifier from Panel A (and Fig. 3A) for passive sentences, and the results did in fact show this temporal congruence. To look at the overall pattern, we stacked the predictions from all 97 classifiers that made above-chance predictions (Fig. 4D). Despite the temporal congruence in SMC, this analysis revealed an entirely different pattern of results. In passives, both the subject and object remained active throughout the entirety of the sentence, revealing that the brain processes the subject and object simultaneously rather than sequentially. To assess whether this constituted a statistically significant difference from active sentences, we

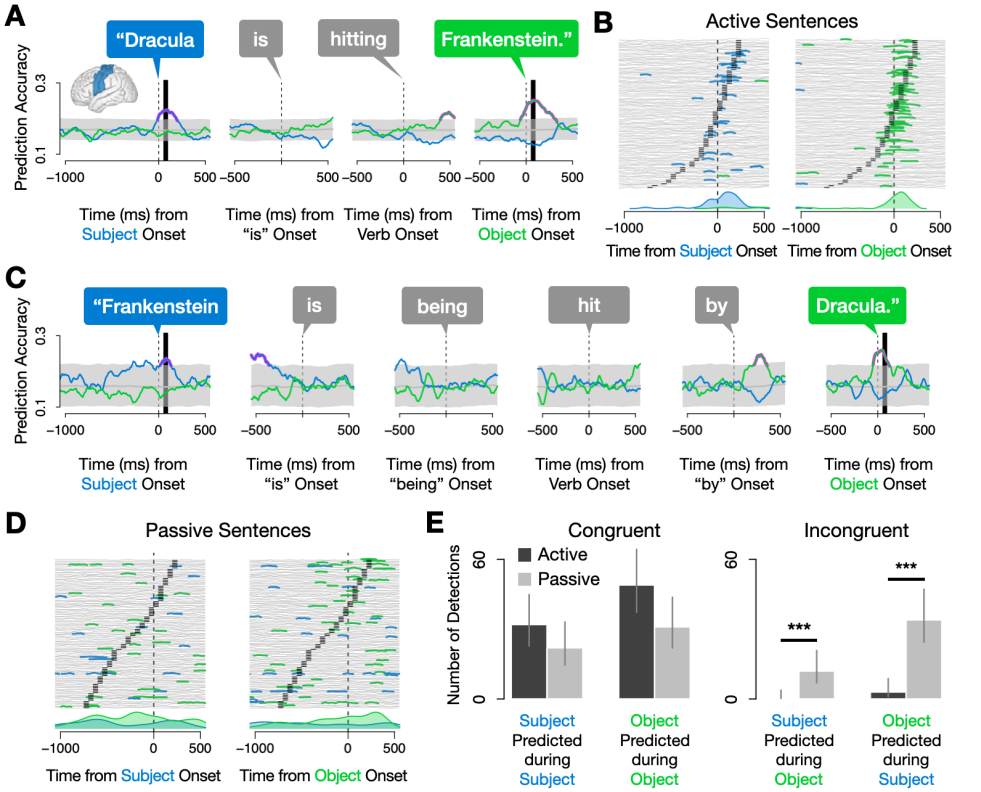

Fig. 4 Predicting words in sentences based on picture naming data. (A) Sample prediction accuracies for the first and second nouns (i.e., the subject and the object) of active sentences, locked to the onset of each word. Classifiers were the same as those used in Fig. 3A, i.e., trained on picture naming data from electrodes in SMC between 50 and 100 ms (shown by black bar) after speech onset; accuracies were averaged across patients. Both nouns were predicted above chance at the time of their respective articulations (pink highlight; $p < .05$ for 100 ms or $p < .01$ for 50 ms). (B) Stacked prediction accuracies from the 83 significant classifiers for active sentences, locked to the onset of each noun. Density plots at bottom show that each word's accuracy peaked during its articulation. (See Supplementary Fig. S3 for density plots broken down by ROI.) (C) Mean prediction accuracies for passive sentences from the same SMC classifiers in Panel A and Fig. 3A. Again, prediction accuracy was above chance for both nouns during their respective articulations. (D) Stacked prediction accuracies from the 97 significant classifiers for passive sentences, locked to the onset of each noun. Unlike in active sentences and lists, there is little correspondence between training time (black bars) and the times when words were detected (green and blue segments). This point is made especially clear by the density plots, which reveal that representations of both the subject and the object nouns were active throughout the production of passive sentences. (E) Number of detections (i.e., classifiers that significantly predicted either word in the sentence) broken down by whether the detection reflected temporally congruent processing (i.e., detection of the subject noun during production of the subject or detection of the object during production of the object) or incongruent (detection of the object noun during production of the subject or vice versa). While both active (black) and passive (gray) sentences involved temporally congruent word processing, passives had significantly more incongruent detections than actives (FDR-corrected $p < .001$ for both incongruent subjects and objects; one-sided tests of equal proportions).

counted the number of classifiers that detected each word during the production of each word (Fig. 4E) – i.e., the number of classifiers that detected the subject when the subject was being said or the object when the object was being said ("congruent" detections), and the number of classifiers that detected the object when the subject was being said or the subject when the object was being said ("incongruent"). In both active and passive sentences we observed congruent (Fig. 4E, left). However, of the incongruent detections we observed, nearly all were in passive trials (Fig. 4E, right). Relative to active sentences, passive sentences involved significantly more incongruent detections of both subjects and objects (for both: FDR-corrected $p < .001$, one-sided test of equal proportions). This pattern was driven by two regions (see Supplementary Fig. S3): IFG, which preferentially encoded subjects throughout passive sentences, and MFG, which preferentially encoded objects. Strikingly, even when patients were articulating the subject of a passive sentence, their brains more strongly encoded the object, as evidenced by the higher number of incongruent detections (predictions of the object) than congruent ones (predictions of the subject) during passive subject production.

3 Discussion

Single word production tasks like picture naming have dominated the neuroscience of language production. Here, we leveraged the unparalleled spatiotemporal precision of ECoG and employed an innovative cross-task classification approach to demonstrate similarities and differences in word processing between single word production and sentence production. We first demonstrated that individual words can be decoded from patterns of activity in picture naming data, verifying that our data contained word-specific information and revealing an unexpected pattern that contradicts the classical depiction of a feedforward sequence of stages in word production. Next, we used the picture naming data to train a series of classifiers on word identity at 20 time points over the course of picture naming and in seven regions of interest. We showed that these classifiers successfully decoded each word during the production of word lists in the order in which words were said. We then applied these same classifiers to sentences to determine whether sentence production could similarly be modeled as a sequence of single word productions. We started with active sentences, which represent the canonical word order in English and involve producing nouns in order from most to least salient. As we observed for lists, classifiers successfully decoded the two words in active sentences in the order in which they were produced. Finally, we used the classifiers to decode word identity during the production of passive sentences, which encode the same events as active sentences but with the reverse word order. Here, we observed a significant departure from the temporal alignment of word processing in the brain and word order in speech: rather than encoding each word as it was being said, the brain encoded both words simultaneously for the duration of the sentence, revealing that sentence production is not simply a sequence of serially executed single-word productions, but involves a dynamic interplay between syntactic structure and word planning.

Our findings validate various aspects of existing cognitive [43–45] and computational [46] models of sentence production. Specifically, these models assume that word representations are invariant across different behaviors. Our finding that the cortical representations of words during picture naming successfully generalize to lists and sentences validate this assumption. Furthermore, these models build in a dependence on syntactic structure during sentence planning, suggesting that the particular dynamics of word representations may vary with syntactic structure. Indeed, our data provided evidence of this in the striking difference in the temporal dynamics of word planning between active and passive sentences.

One area where our findings diverged from cognitive models was in the timecourse of particular word representations during picture naming. According to the classical view of word production [10–13], representational stages (e.g., conceptual, phonological, articulatory) come online in a feedforward sequence, with each stage activating the next before going offline. In our data, however, we observed that once a classifier detected word-specific information, it typically continued to do so until at least the onset of speech. This suggests a slightly different picture, involving a cascade of word representations that, once online, remain online until word articulation is successfully underway.

Finally, we believe that our findings may shed light on a widely noted but poorly understood typological pattern among the worlds' approximately 6,000 languages. Specifically, in sentences like the ones in our experiment which involve a subject, a verb, and an object, there are six possible ways languages can arrange these words: Subject-Verb-Object (as in the English "I eat cake"), Subject-Object-Verb (as in Farsi "*man keik mikhoram*," literally "I cake eat"), and so on. However, of these 6 logically possible word orders, fewer than 5% of languages have a canonical word order where the object precedes the subject [47, 48]. One possible reason for this is that there is a natural tendency in speech to order words from most to least salient, and subjects tend to be more semantically salient than objects. In our experiment, passive sentences provided an opportunity to visualize how the brain processes words when it is required to produce less salient words before more salient ones. Word planning in passive sentences involved a much more complex temporal pattern than in active sentences. Indeed, whereas word planning in actives resembled word planning in picture naming and lists, in passives the brain encoded both the subject and the object for the duration of the sentence. This was driven by patterns in prefrontal cortex, and specifically by IFG, which sustained a representation of the passive subject, and MFG, which sustained the passive object. Furthermore, reaction times, commonly interpreted as an index of processing difficulty, were significantly longer for passives than actives (1,424 ms and 1,164 ms, respectively; $p < .001$, Wilcoxon rank sum test) and lists (801 ms; $p < .001$, Wilcoxon rank sum test). Taken together, these facts point toward a processing-based explanation of the cross-linguistic dominance of subject-before-object word orders. Producing words in least to most salient may simply be harder for the production system. We speculate that, over the course of language evolution, this difficulty may exert a subtle pressure on language change, making it more likely for languages to evolve in the direction of subject-before-object orders.

4 Methods

The data in this study were also reported in Morgan et al. 2024. Details of participants, experimental design, and data collection are repeated below.

4.1 Participants

We recorded data from ten neurosurgical patients undergoing evaluation for refractory epilepsy (3 female, mean age: 30 years, range: 20 to 45). All ten were implanted with electrocorticographic grids and strips. Patients provided informed consent both in writing and then again orally prior to the beginning of the experiment. The implantation and location of electrodes were guided solely by clinical requirements. Eight participants were implanted with standard clinical electrode grid with 10 mm spaced electrodes (Ad-Tech Medical Instrument, Racine, WI). The remaining two participants consented to a research hybrid grid implant (PMT corporation, Chanassen, MN) that included 64 additional electrodes between the standard clinical contacts (with overall 10 mm spacing and interspersed 5 mm spaced electrodes over select regions), providing denser sampling but with positioning based solely on clinical needs. The research study protocol was approved by the NYU Langone Medical Center Committee on Human Research.

4.2 Data collection and preprocessing

Participants were tested while resting in their hospital bed in the NYU Langone epilepsy monitoring unit. Stimuli were presented on a laptop computer screen positioned at a comfortable distance from the participant. Participants' voice was recorded with a cardioid microphone (Shure MX424). The experiment computer generated inaudible TTL pulses marking the onset of a stimulus. These were recorded in auxiliary channels of both the clinical Neuroworks Quantum Amplifier (Natus Biomedical, Appleton, WI), which records ECoG, and the audio recorder (Zoom H1 Handy Recorder). The microphone signal was also fed to the audio recorder and the ECoG amplifier. These redundant recordings were used to sync the speech, experiment, and neural recordings.

The standard implanted ECoG arrays consisted of 64 macro-contacts (2 mm exposed, 10 mm spacing) in an 8×8 grid. Hybrid grids contained 128 electrode channels, including the standard 64 macro-contacts plus 64 additional interspersed smaller electrodes (1 mm exposed) between the macro-contacts (providing 10 mm center-to-center spacing between macro-contacts and 5 mm center-to-center spacing between micro/macro contacts, PMT corporation, Chanassen, MN). The FDA-approved hybrid grids were manufactured for research purposes, which we explained to patients during consent. In all ten patients, ECoG arrays were implanted on the left hemisphere. The location of the grid was solely dictated by clinical needs.

ECoG was recorded at 2,048 Hz, which was decimated to 256 Hz prior to processing and analysis. We excluded electrodes with artifacts (i.e., line noise, poor contact with the cortex, and high amplitude shifts) or with interictal/epileptiform activity prior to subtracting a common average reference (across all valid electrodes and time) from

each individual electrode. We then extracted the envelope of the high gamma component (the average of three evenly log-spaced frequency bands from 70 to 150 Hz) from the raw signal with the Hilbert transform.

The signal was epoched locked to stimulus (i.e., cartoon images) and production onsets for each trial. The 200 ms silent period preceding stimulus onset (during which patients were not speaking and fixating on a cross located at the center of the screen) was used as a baseline, and each epoch for each electrode was z-scored (i.e., normalized) to this baseline's mean and standard deviation.

4.3 Experimental Design

4.3.1 Procedure

The experiment was performed in a single session that lasted approximately 40 minutes. Stimuli were presented in pseudo-random order using PsychoPy [49]. All stimuli were constructed using the same 6 cartoon characters (chicken, dog, Dracula, Frankenstein, ninja, nurse), chosen to vary along many dimensions (e.g., frequency, phonology, number of syllables, animacy, proper vs. common, etc.) to facilitate identification of word-specific information at analysis.

The experiment began with two short familiarization blocks. In the first block (6 trials), participants saw each of the six cartoon characters once with labels (*chicken, dog, Dracula, Frankenstein, ninja, nurse*) written beneath the image. Participants read the labels aloud, after which the experimenter pressed a button to go to the next trial. In the second block, participants saw the same six characters one at a time, twice each with order pseudo-randomized (12 trials), but without labels. Participants were instructed to name the characters out loud. After naming the character, the experimenter pressed a button revealing the target name to ensure patients had learned the correct labels. Participants then completed the first picture naming block (96 trials). Characters were again presented in the center of the screen, one at a time, but no labels were provided.

Next, participants performed a sentence production block (60 trials), which began with two practice trials. Participants were instructed that there were no right or wrong answers, that the goal of the experiment was to understand what the brain is doing when people speak naturally. On each trial, participants saw a 1 s fixation cross followed by a written question, which they were instructed to read aloud, ensuring attention. After another 1 s fixation cross, a static cartoon vignette appeared in the center of the screen depicting two of the six characters engaged in a transitive event (one character acting on the other). Participants were instructed to respond to the question with a description of the vignette. The image remained on the screen until the participant completed their response, at which point the experimenter pressed a button to proceed. After the first 12 trials, the target sentence (i.e., an active sentence after an active question or passive sentence after a passive question) appeared in text on the screen and participants read it aloud. We described these target sentences as "the sentence we expected you to say." The goal of this was to implicitly reinforce the link between the syntax of the question and the target response. If the participant

appeared to interpret these as corrections, the experimenter reminded them that there were no right or wrong answers.

Between each sentence production trial, we interleaved two picture naming trials in order to reduce task difficulty and facilitate fluent sentence production. The picture naming trials showed the two characters that would be engaged in the subsequent vignette, presented in a counterbalanced order such that on half of trials they would appear in the same order as in the target sentence response, and in the opposite order on the other half.

After the sentence block, participants performed the listing block. List production was designed to parallel sentence production. Each trial began with a 1 s fixation cross, followed by an arrow pointing either left or right appeared for 1 s in the center of the screen. After another 1 s fixation cross, a cartoon vignette, taken from the exact same stimuli as in the sentence block, appeared on the screen. Participants named the two characters in the vignette either from left to right or from right to left, according to the direction of the preceding arrow. As in sentence production trials, each list production trial was preceded by two picture naming trials involving the two characters that would appear in the subsequent vignette in counterbalanced order.

Between each block, participants were offered the opportunity to end the experiment if they did not wish to continue. One participant stopped before the list production block, providing only data for picture naming and sentence production. The remaining nine participants completed all three blocks. These nine were also offered the opportunity to complete another picture naming block and another sentence production block. Six consented to an additional picture naming block and two additionally consented to another sentence production block.

4.3.2 Stimulus Design, Randomization, and Counterbalancing

Picture naming stimuli consisted of images of the 6 characters presented in pseudorandom order so that each consecutive set of 6 trials contained all 6 characters in random order. This ensured a relatively even distribution of characters over time, and that no character appeared more than two times in a row. Characters were pseudorandomly depicted in 8 orientations: facing forward, backward, left, right, and at the 45° angle between each of these.

Sentence production stimuli consisted of a written question followed by a static cartoon vignette. Questions were manipulated so half were constructed with passive syntax and the other half with active. All questions had the format: "Who is [verb]-ing whom?" or "Who is being [verb]-ed by whom?". There were 10 verbs: *burn, hit, hypnotize, measure, poke, scare, scrub, spray, tickle, trip*. Each verb was used to create 3 vignettes involving 3 characters in a counterbalanced fashion so that each character was the agent (i.e., active subject) in one vignette and the non-agent (i.e., active object) in one vignette. Each of these three vignettes was shown twice in the sentence production block, once preceded by an active question and once by a passive question, priming active and passive responses [28]. Vignettes were flipped around the vertical axis the second time they appeared so the character that was on the left in the first appearance was on the right in the second appearance. This was also counterbalanced so that on half of the trials in each syntax condition (active/passive) the

subject was on the left. List production stimuli consisted of the same 60 vignettes, also pseudorandomly ordered and counterbalanced across conditions (i.e., arrow direction).

4.4 Data Coding and Inclusion

Speech was manually transcribed and word onset times were manually recorded using Audacity [50] to visualize the waveform and spectrogram of the audio recording. Picture naming trials were excluded if the first word uttered was not the target word (e.g., "Dracula – I mean Frankenstein"). Sentence trials were excluded if the first word was incorrect (i.e., "Dracula" instead of "Frankenstein," regardless of active/passive structure) or if the meaning of the sentence did not match the meaning of the depicted scene; no sentences were excluded because the syntax did not match that of the prime question. Sentences were coded as active or passive depending on the structure the patient used, not the prime structure. Listing trials were excluded if the first word was incorrect ("Dracula" instead of "Frankenstein") or if the order did not match that indicated by the arrow.

In analyses for the three trial types (picture naming, sentence production, and list production), data from all patients who completed trials in that block are included. Data from one patient who did not complete the list production block is not included in the list production analyses, and data from 3 patients who produced 3 or fewer passive sentences during sentence production blocks were not included in the analyses of passive sentences.

4.5 Electrode Localization

Electrode localization in both subject space and MNI space was based on coregistering a preoperative (no electrodes) and postoperative (with electrodes) structural MRI (in some cases, a postoperative CT was employed depending on clinical requirements) using a rigid-body transformation. Electrodes were then projected to the cortical surface (preoperative segmented surface) to correct for edema-induced shifts following previous procedures [51] (registration to MNI space was based on a nonlinear DARTEL algorithm). Based on the subject's preoperative MRI, the automated FreeSurfer segmentation (Destrieux) was used for labeling electrodes' within-subject anatomical locations.

4.6 Significance testing and corrections for multiple comparisons in time series data

Statistical tests on time series data were performed independently at each time sample, producing the same number of p-values as there are samples in the time series. To correct for multiple comparisons we follow [36, 52, 53] and establish a conservative criterion for significance for all time series comparisons: an uncorrected p-value that remains below .05 for at least 100 consecutive milliseconds or below .01 for at least 50 consecutive milliseconds.

4.7 Multi-class classification

For the analyses in Figs. 2–4 we trained multi-class classifiers on word identity using the *caret* and *nnet* packages [54, 55] in R [56]. Classifiers consisted of a series of one-vs-rest logistic regressions (fit as a neural network), which were chosen for their simplicity and interpretability. For the picture naming analyses (Fig. 2), we used a repeated cross-fold validation procedure (3 repeats, 10 folds) to calculate prediction accuracy, and arbitrarily chose a mid-range value of 10^{-3} for decay, the lone hyperparameter in this model (typical values are logarithmically spaced along the range from 10^{-6} to 10^{0}). For the subsequent analyses of list and sentence production data, we first performed repeated cross-validation on the picture naming data again to find the optimal hyperparameters for each individual classifier, and then retrained each model with that hyperparameter and using all of the picture naming trials (rather than a training subset). We then used this model to predict word identity at every time point throughout each trial in the list and sentence production blocks. Prediction accuracy time series like those in Fig. 4A reflect the mean of the binary accuracy scores across sentence production trials separately for the subject and the object (which on different trials were different combinations of the six nouns, e.g., Dracula, dog, etc.), smoothed with a 100 ms boxcar function. To generate the noise distribution (gray shaded area), we performed a permutation analysis, shuffling labels on the test data and repeating the prediction analysis 1000 times. We determined significance by calculating the upper 95th and 99th percentiles of the mean trial accuracies generated by the permutation analysis for each time sample (see Section 4.6 for details on multiple comparisons corrections).

4.8 Temporal warping

The time between stimulus and speech onsets the *planning period*, varied both across and within patients. Consequently, cognitive processes become less temporally aligned across trials the farther one moves away from stimulus onset in stimulus-locked epochs or from speech onset in speech-locked epochs. Temporal warping reduces such misalignments [31–35], which we previously verified in this dataset [36]. Following [34, 36], we linearly interpolated the data in the middle of the planning period (from 150 ms post-stimulus to 150 ms pre-speech) for each trial, setting all trials' planning periods to the same duration (Supplementary Fig. S1): the global median per task (1141 ms for sentences; 801 ms for lists; 758 ms for words). Specifically, for each task we first excluded trials with outlier response times, which we defined as those in the bottom 2.5% or top 5% per participant. We then calculated median response times per task across participants (1,142 ms for sentences, 801 ms for lists, and 758 ms for words), and for each electrode and each trial, concatenated (a) the stimulus-locked data from 150 ms post-stimulus to $\frac{1}{2}$ the median response time with (b) the production-locked data from $-\frac{1}{2}$ median response time to 150 ms pre-speech. We then linearly interpolated this time series to the number of time samples that would, when concatenated between the 150 ms post stimulus (stimulus-locked) and 150 ms pre-speech (speech-locked), result in a time series with the median planning period duration. Finally, we concatenated (a) the unwarped data leading up to 150 ms post-stimulus, (b) the

warped data from the previous step, and (c) the unwarped data starting 150 ms before speech onset, forming the final epochs used in the analyses. We direct the reader to Morgan et al. 2024 for a fuller discussion and demonstration that this temporal warping increased signal to noise ratio in this dataset, as well as analyses of the unwarped high gamma activity across regions of interest.

Acknowledgments

This work was supported by National Institutes of Health grants F32DC019533 (A.M.), R01NS109367 (A.F.), R01NS115929 (A.F.), and R01DC018805 (A.F.).

Data and Code availability

Data will be made available from the authors upon request, provided documentation that the data will be strictly used for research purposes and will comply with the terms of our study IRB. The code is available at https://github.com/flinkerlab/decoding_words_in_sentences.

Declarations

The authors declare that they have no competing interests.

Appendix A Supplementary Information

A.1 Data Warping

Supplementary Fig. S1, reproduced from Morgan et al. 2024, demonstrates the temporal warping procedure for a sample electrode (see Methods Sec. 4.8).

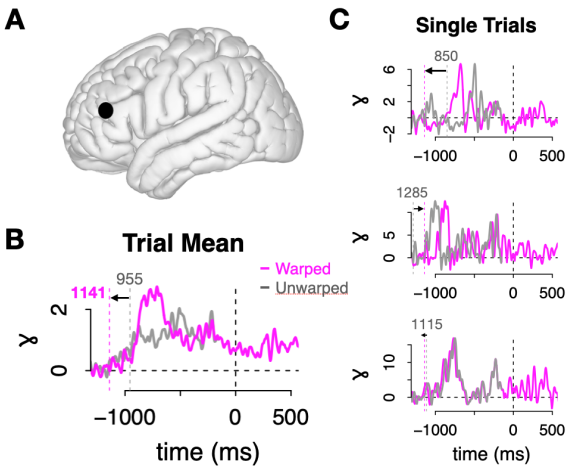

Fig. S1 Warped and unwarped sentence production data from a sample electrode. The data in each trial between 150 ms post stimulus and 150 ms pre-speech were linearly interpolated to set the duration of the planning period to the global median per task (1142 ms for sentence production) [34]. (A) Sample electrode localization in MFG. (B) The mean of this electrode's warped (pink) and unwarped (gray) trials. Prior to warping, this patient's median sentence response time was 995 ms; after warping it was 1,141 ms: the median sentence production response time across patients. The peak of the warped data was higher than the unwarped peak, a sign that warping resulted in better temporal alignment and consequently higher signal-to-noise ratio. (C) Three sample trials: warped (pink) and unwarped (gray) data.

A.2 Picture naming classification results by ROI

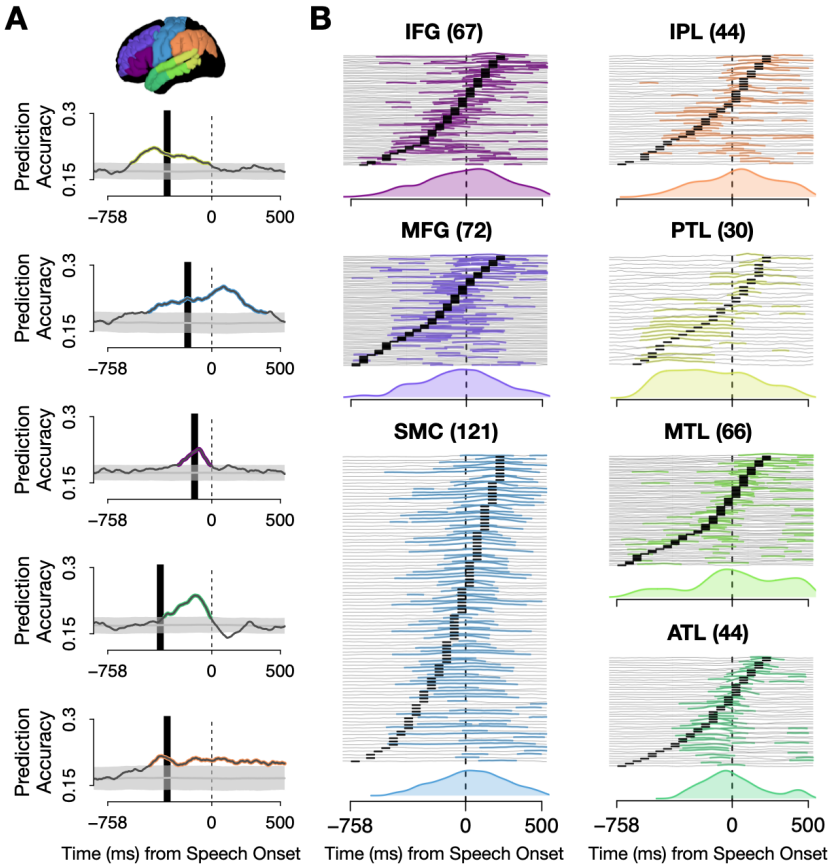

Fig. S2 Picture naming classification results by ROI. (A) Results from five sample classifiers (see Fig. 2 for methods). From top to bottom, the classifiers were trained and tested on picture naming data in (1) Patient 9, PTL, -350 to -300 ms; (2) Patient 6, SMC, -200 to -150 ms; (3) Patient 9, IFG, -150 to -100 ms; (4) Patient 10, ATL, -400 to -350 ms; and (5) Patient 4, IPL, -350 to -300 ms. (B) Stacked picture naming prediction accuracies from each significant classifier by ROI (across patients and training windows). Number of significant classifiers appears in parentheses. Word identity was decodable in all seven ROIs, with the highest decodability in SMC and lowest in PTL, which was also the only ROI to peak during the pre-articulatory period.

A.3 Sentence classification results by ROI

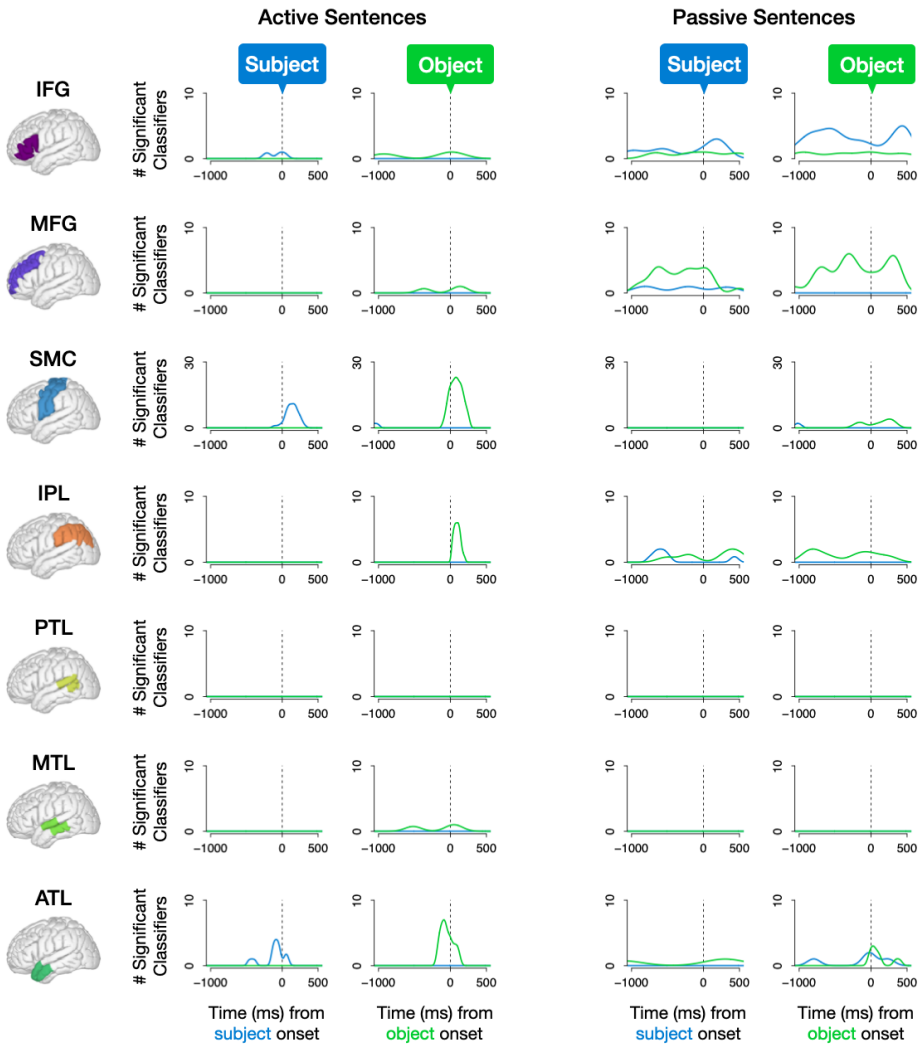

Fig. S3 Sentence production classification results by ROI: Density of significant prediction accuracies of subject and object in active (left) and passive (right) sentences. The y-axis is scaled to the number of significant classifiers in each region and for both syntactic structures. (NB: due to high decodability in SMC it was scaled to 30; all other ROIs go from 0 to 10.) Notably, incongruent detections (i.e., predicting the subject during production of the object or predicting the object during production of the subject), which nearly exclusively occurred in passive sentences (Fig. 4E), appear to have been driven by processing in IFG, which preferentially encoded passive subjects, and MFG, which preferentially encoded passive objects.

References

[1] Janik, V.M., Sayigh, L.S., Wells, R.S.: Signature whistle shape conveys identity information to bottlenose dolphins. Proceedings of the National Academy of Sciences **103**(21), 8293–8297 (2006) https://doi.org/10.1073/pnas.0509918103

[2] Wenner, A.M., Wells, P.H., Rohlf, F.J.: An analysis of the waggle dance and recruitment in honey bees. Physiological Zoology **40**(4), 317–344 (1967) https://doi.org/10.1086/physzool.40.4.30158452

[3] Seyfarth, R.M., Cheney, D.L., Marler, P.: Vervet monkey alarm calls: semantic communication in a free-ranging primate. Animal Behaviour **28**(4), 1070–1094 (1980) https://doi.org/10.1016/S0003-3472(80)80097-2

[4] Gill, S.A., Sealy, S.G.: Functional reference in an alarm signal given during nest defence: seet calls of yellow warblers denote brood-parasitic brown-headed cowbirds. Behavioral Ecology and Sociobiology **56**, 71–80 (2004) https://doi.org/10.1007/s00265-003-0736-7

[5] Huth, A.G., De Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L.: Natural speech reveals the semantic maps that tile human cerebral cortex. Nature **532**(7600), 453–458 (2016) https://doi.org/10.1038/nature17637

[6] Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F.: Phonetic feature encoding in human superior temporal gyrus. Science **343**(6174), 1006–1010 (2014) https://doi.org/10.1038/nature11020

[7] Moses, D.A., Metzger, S.L., Liu, J.R., Anumanchipalli, G.K., Makin, J.G., Sun, P.F., Chartier, J., Dougherty, M.E., Liu, P.M., Abrams, G.M., *et al.*: Neuroprosthesis for decoding speech in a paralyzed person with anarthria. New England Journal of Medicine **385**(3), 217–227 (2021) https://doi.org/10.1056/NEJMoa2027540

[8] Badecker, W., Miozzo, M., Zanuttini, R.: The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. Cognition **57**(2), 193–216 (1995) https://doi.org/10.1016/0010-0277(95)00663-J

[9] Dell, G.S.: A spreading-activation theory of retrieval in sentence production. Psychological review **93**(3), 283 (1986)

[10] Butterworth, B.: Lexical access in speech production. In: Lexical Representation and Process, pp. 108–135 (1989). https://doi.org/10.5555/86149.86157

[11] Levelt, W.J., Roelofs, A., Meyer, A.S.: A theory of lexical access in speech production. Behavioral and brain sciences **22**(1), 1–38 (1999) https://doi.org/10.1017/S0140525X99001776

[12] Indefrey, P., Levelt, W.J.: The spatial and temporal signatures of word production components. Cognition **92**(1-2), 101–144 (2004) https://doi.org/10.1016/j.cognition.2002.06.001

[13] Indefrey, P.: The spatial and temporal signatures of word production components: a critical update. Frontiers in psychology **2**, 255 (2011) https://doi.org/10.3389/fpsyg.2011.00255

[14] Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T., Crone, N.E.: Redefining the role of broca's area in speech. Proceedings of the National Academy of Sciences **112**(9), 2871–2875 (2015) https://doi.org/10.1073/pnas.1414491112

[15] Tremblay, P., Small, S.L.: Motor response selection in overt sentence production: a functional mri study. Frontiers in psychology **2**, 253 (2011) https://doi.org/10.3389/fpsyg.2011.00253

[16] Schwartz, M.F., Kimberg, D.Y., Walker, G.M., Faseyitan, O., Brecher, A., Dell, G.S., Coslett, H.B.: Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. Brain **132**(12), 3411–3427 (2009) https://doi.org/10.1093/brain/awp284

[17] Menenti, L., Segaert, K., Hagoort, P.: The neuronal infrastructure of speaking. Brain and language **122**(2), 71–80 (2012) https://doi.org/10.1016/j.bandl.2012.04.012

[18] Lukic, S., Thompson, C.K., Barbieri, E., Chiappetta, B., Bonakdarpour, B., Kiran, S., Rapp, B., Parrish, T.B., Caplan, D.: Common and distinct neural substrates of sentence production and comprehension. NeuroImage **224**, 117374 (2021) https://doi.org/10.1016/j.neuroimage.2020.117374

[19] Giglio, L., Ostarek, M., Sharoh, D., Hagoort, P.: Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. Proceedings of the National Academy of Sciences **121**(11), 2310766121 (2024) https://doi.org/10.1073/pnas.2310766121

[20] Braun, A.R., Guillemin, A., Hosey, L., Varga, M.: The neural organization of discourse: An h215o-pet study of narrative production in english and american sign language. Brain **124**(10), 2028–2044 (2001) https://doi.org/10.1093/brain/124.10.2028

[21] Menenti, L., Gierhan, S.M., Segaert, K., Hagoort, P.: Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional mri. Psychological science **22**(9), 1173–1182 (2011) https://doi.org/10.1177/0956797611418347

[22] Walenski, M., Europa, E., Caplan, D., Thompson, C.K.: Neural networks for sentence comprehension and production: An ale-based meta-analysis of neuroimaging studies. Human brain mapping **40**(8), 2275–2304 (2019) https://doi.org/10.1002/hbm.24523

[23] Haller, S., Radue, E.-W., Erb, M., Grodd, W., Kircher, T.: Overt sentence production in event-related fmri. Neuropsychologia **43**(5), 807–814 (2005) https://doi.org/10.1016/j.neuropsychologia.2004.09.007

[24] Grande, M., Meffert, E., Schoenberger, E., Jung, S., Frauenrath, T., Huber, W., Hussmann, K., Moormann, M., Heim, S.: From a concept to a word in a syntactically complete sentence: an fmri study on spontaneous language production in an overt picture description task. Neuroimage **61**(3), 702–714 (2012) https://doi.org/10.1016/j.neuroimage.2012.03.087

[25] Schönberger, E., Heim, S., Meffert, E., Pieperhoff, P., Costa Avelar, P., Huber, W., Binkofski, F., Grande, M.: The neural correlates of agrammatism: Evidence from aphasic and healthy speakers performing an overt picture description task. Frontiers in Psychology **5**, 246 (2014) https://doi.org/10.3389/fpsyg.2014.00246

[26] Geranmayeh, F., Brownsett, S.L., Leech, R., Beckmann, C.F., Woodhead, Z., Wise, R.J.: The contribution of the inferior parietal cortex to spoken language production. Brain and language **121**(1), 47–57 (2012) https://doi.org/10.1016/j.bandl.2012.02.005

[27] Geranmayeh, F., Wise, R.J., Mehta, A., Leech, R.: Overlapping networks engaged during spoken language production and its cognitive control. Journal of Neuroscience **34**(26), 8728–8740 (2014) https://doi.org/10.1523/JNEUROSCI.0428-14.2014

[28] Bock, J.K.: Syntactic persistence in language production. Cognitive psychology **18**(3), 355–387 (1986) https://doi.org/10.1016/0010-0285(86)90004-6

[29] Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., Malach, R.: Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. Science **309**(5736), 951–954 (2005) https://doi.org/10.1126/science.1110913

[30] Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., Malach, R.: Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Current biology **17**(15), 1275–1285 (2007) https://doi.org/10.1016/j.cub.2007.06.066

[31] Williams, A.H., Poole, B., Maheswaranathan, N., Dhawale, A.K., Fisher, T., Wilson, C.D., Brann, D.H., Trautmann, E.M., Ryu, S., Shusterman, R., *et al.*: Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. Neuron **105**(2), 246–259 (2020) https://doi.org/10.1016/j.neuron.2019.10.020

[32] Wang, K., Begleiter, H., Porjesz, B.: Warp-averaging event-related potentials. Clinical Neurophysiology **112**(10), 1917–1924 (2001) https://doi.org/10.1016/S1388-2457(01)00640-X

[33] Picton, T.W., Lins, O.G., Scherg, M.: The recording and analysis of event-related potentials. Handbook of neuropsychology **10**, 3–3 (1995)

[34] Edwards, E., Nagarajan, S.S., Dalal, S.S., Canolty, R.T., Kirsch, H.E., Barbaro, N.M., Knight, R.T.: Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage **50**(1), 291–301 (2010)

[35] Molina, M., Tardón, L.J., Barbancho, A.M., De-Torres, I., Barbancho, I.: Enhanced average for event-related potential analysis using dynamic time warping. Biomedical Signal Processing and Control **87**, 105531 (2024) https://doi.org/10.1016/j.bspc.2023.105531

[36] Morgan, A.M., Devinsky, O., Doyle, W., Dugan, P., Friedman, D., Flinker, A.: A low-activity cortical network selectively encodes syntax. bioRxiv, 2024–06 (2024) https://doi.org/10.1101/2024.06.20.599931

[37] Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L.: Encoding and decoding in fMRI. Neuroimage **56**(2), 400–410 (2011) https://doi.org/10.1016/j.neuroimage.2010.07.073

[38] Holdgraf, C.R., Rieger, J.W., Micheli, C., Martin, S., Knight, R.T., Theunissen, F.E.: Encoding and decoding models in cognitive electrophysiology. Frontiers in systems neuroscience **11**, 61 (2017) https://doi.org/10.3389/fnsys.2017.00061

[39] Caramazza, A., Zurif, E.B.: Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain and language **3**(4), 572–582 (1976) https://doi.org/10.1016/0093-934X(76)90048-1

[40] Grodzinsky, Y.: The neurology of syntax: Language use without broca's area. Behavioral and brain sciences **23**(1), 1–21 (2000) https://doi.org/10.1017/S0140525X00002399

[41] Love, T., Swinney, D., Walenski, M., Zurif, E.: How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. Brain and Language **107**(3), 203–219 (2008) https://doi.org/10.1016/j.bandl.2007.11.004

[42] Thompson, C.K., Choy, J.J.: Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. Journal of psycholinguistic research **38**, 255–283 (2009) https://doi.org/10.1007/s10936-009-9105-7

[43] Ferreira, V.S., Morgan, A.M., Slevc, L.R.: Grammatical encoding. The Oxford Handbook of Psycholinguistics, 453–469 (2018)

[44] Bock, K., Ferreira, V.S.: Syntactically speaking. The Oxford handbook of language production, 21–46 (2014)

[45] Slevc, L.R.: Grammatical encoding. In: Language Production, pp. 4–31. Routledge, ??? (2023). https://doi.org/10.4324/9781003145790-2

[46] Chang, F., Dell, G.S., Bock, K.: Becoming syntactic. Psychological review **113**(2), 234 (2006) https://doi.org/10.1037/0033-295X.113.2.234

[47] Dryer, M.S.: Determining dominant word order. The world atlas of language structures online (2013)

[48] Hammarström, H.: Linguistic diversity and language evolution. Journal of Language Evolution **1**(1), 19–29 (2016) https://doi.org/10.1093/jole/lzw002

[49] Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K.: Psychopy2: Experiments in behavior made easy. Behavior research methods **51**, 195–203 (2019) https://doi.org/10.3758/s13428-018-01193-y

[50] Audacity Team: Audacity (R). http://audacity.sourceforge.net/

[51] Yang, A.I., Wang, X., Doyle, W.K., Halgren, E., Carlson, C., Belcher, T.L., Cash, S.S., Devinsky, O., Thesen, T.: Localization of dense intracranial electrode arrays using magnetic resonance imaging. Neuroimage **63**(1), 157–165 (2012) https://doi.org/10.1016/j.neuroimage.2012.06.039

[52] Maris, E., Oostenveld, R.: Nonparametric statistical testing of eeg-and meg-data. Journal of neuroscience methods **164**(1), 177–190 (2007) https://doi.org/10.1016/j.jneumeth.2007.03.024

[53] Flinker, A., Doyle, W.K., Mehta, A.D., Devinsky, O., Poeppel, D.: Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. Nature human behaviour **3**(4), 393–405 (2019) https://doi.org/10.1038/s41562-019-0548-z

[54] Kuhn, Max: Building predictive models in r using the caret package. Journal of Statistical Software **28**(5), 1–26 (2008) https://doi.org/10.18637/jss.v028.i05

[55] Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S, 4th edn. Springer, New York (2002). ISBN 0-387-95457-0. https://www.stats.ox.ac.uk/pub/MASS4/

[56] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2021). R Foundation for Statistical Computing. https://www.R-project.org/