communications psychology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s44271-025-00270-1

Decoding words during sentence production with ECoG reveals syntactic role encoding and structure-dependent temporal dynamics

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

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 a largely untested assumption that the principles of word production generalize to more naturalistic utterances like sentences. Here, we investigate this using high-resolution neurosurgical recordings (ECoG) and an overt production experiment where ten patients produced six words in isolation (picture naming) and in sentences (scene description). We trained machine learning classifiers to identify the unique brain activity patterns for each word during picture naming, and used these patterns to decode which words patients were processing while they produced sentences. Our findings confirm that words share cortical representations across tasks, but reveal a division of labor within the language network. In sensorimotor cortex, words were consistently activated in the order in which they were said in the sentence. However, in prefrontal cortex, the order in which words were processed depended on the syntactic structure of the sentence. In non-canonical sentences (passives), we further observed a spatial code for syntactic roles, with subjects selectively encoded in inferior frontal gyrus (IFG) and objects selectively encoded in middle frontal gyrus (MFG). We suggest that these complex dynamics of prefrontal cortex may impose a subtle pressure on language evolution, potentially explaining why nearly all the world's languages position subjects before objects.

Many animals use symbolic forms of communication: dolphins have names [1], bees dance to signal nectar locations[2], and monkeys and birds use predator-specific calls[3,4]. While human word (or *lexical*) 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 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, typically employing *picture naming* paradigms where a participant sees a picture of, e.g., a dog, and says "dog." However, behavioral studies have demonstrated that sentence production is not simply a sequence of single-word production tasks[5,6]. The historical focus on words has left a critical gap in understanding how the brain produces more complex linguistic constructions like sentences.

Among the important insights from research at the word level is that words are not unitary representations. Instead, lexical knowledge involves distinct representations of a word's semantic (i.e., meaning)[7–9], phonological[8–11], articulatory[12], and grammatical features (*lemmas*)[10,13]. Each of these representations is associated with distinct cortical

[1]Neurology Department, NYU Grossman School of Medicine, New York, NY, USA. [2]Neurosurgery Department, NYU Grossman School of Medicine, New York, NY, USA. [3]Biomedical Engineering Department, NYU Tandon School of Engineering, New York, NY, USA. ✉e-mail: Adam.Morgan@NYULangone.org

regions[14,15], with articulatory planning in inferior frontal gyrus (IFG)[16,17]; articulation in sensorimotor cortex (SMC)[18]; feedback in superior temporal gyrus[19] (and visual cortices for sign languages[20]); grammatical features in middle temporal lobe (MTL)[21], and semantics distributed bilaterally throughout cortex[22]. Early production models held that these representational "stages" come online in a strictly feedforward sequence, starting with meaning and ending with articulation (and perception from sensory feedback)[23–25]. However, these models struggled to explain a range of phenomena like speech errors stemming from phonological similarity (e.g., saying *rat* instead of *cat*). Subsequent models, therefore, introduced an interactive architecture, allowing for bidirectional activation between stages of representation[26–28]. More recently, experimental work has revealed that semantic and phonological information come online at roughly the same time[29,30], calling into question the notion of stages and leading to the development of models where representations come online in parallel[31,32].

In contrast to the wealth of knowledge the field has accumulated about single word production, relatively little is known about the type of speech that is unique to our species: sentences[33–35]. Increasingly, researchers are overcoming the obstacles to studying sentence production with non-invasive neural measures (e.g.,[17,33–47]). Many of these studies have focused on the relationship between language production and comprehension, revealing extensive overlap in frontal, temporal, and parietal areas, with additional recruitment of domain-general control regions for production[33,39,40,43,47] (see[48] for a meta-analysis). Others have examined combinatorial processing[10,35,47,49,50], revealing that sentence production engages more of the language network – and to a greater extent – than word production[47]. This research has begun to converge on left anterior temporal lobe (ATL) as a hub for composition[50–52], left posterior temporal lobe (PTL) as encoding hierarchical syntax[35,53,54] and guiding lexical selection[55–57], and left inferior frontal gyrus (IFG) as responsible for coordinating complex linguistic representations to generate a motor code[10,16,35]. Despite these advances, few studies have aimed to directly leverage our understanding of word production to study sentence production. One pressing question, then, is the extent to which the principles of word production generalize to sentences. That is, understanding the processes underlying picture naming only shed light on human cognition more broadly insofar as these processes generalize to more naturalistic language use. However, there is very little evidence supporting this assumption. Here, we address this gap, scaling up from words to sentences by directly testing the hypothesis that the mechanisms underlying single-word production generalize to more complex linguistic behaviors.

To do so, we identified the unique cortical activity patterns that encode six particular words during a picture naming task. We then asked 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 achieved high spatial and temporal resolution and avoided motor artifacts, bypassing limitations of traditional non-invasive neural measures. We employed a controlled production experiment and sophisticated machine learning techniques, and asked whether words' cortical representations are shared across different types of language production behaviors. We hypothesized that word representations would in fact generalize across tasks, but expected that these representations might vary in their temporal dynamics according to task demands. In particular, we expected that in sentences that convey the same event meanings but with different word orders (actives and passives), we should see diverging temporal dynamics in word planning, as event encoding processes remain the same, but grammatical encoding and articulatory processes differ.

Methods

The data in this study were also reported in Morgan et al. (2024), where many of the details below are repeated. The analyses reported in this article were not pre-registered.

Participants

We recorded data from ten neurosurgical patients undergoing evaluation for refractory epilepsy (according to clinical records: 3 women, 7 men, 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, Chanhassen, 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. Participants were compensated at a rate of 20 USD per hour for participation. The research study protocol was approved by the NYU Langone Medical Center Committee on Human Research.

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' voices were 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, Chanhassen, 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 2048 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.

Experimental design

Procedure. The experiment was performed in a single session that lasted approximately 40 minutes. Stimuli were presented in pseudo-random order using PsychoPy[58]. 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, proper vs. common, etc.) to facilitate identification of word-specific information at analysis.

The experiment had a blocked design. Blocks were ordered in terms of importance to ensure that the most valuable data was collected first, in case a patient ceased participation mid-way through (e.g., due to discomfort, fatigue, or seizure activity). 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 a 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, which were demonstrated to reduce task difficulty and facilitate fluent sentence production during pilot testing. 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 the 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. The list production block was designed as a secondary control condition in the original study[59], and as such it was ordered last. 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.

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.

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. (The images of our stimuli in this manuscript were created by the authors for publication purposes, and, while in the same style as the experimental stimuli, are not the images used in the experiment.) 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[60,61]. 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) so that (a) on half of trials the first character to be named appeared on the left, and (b) on half of trials the first character to be named was the agent of the depicted event.

Data coding and inclusion

Speech was manually transcribed and word onset times were manually recorded using Audacity[62] 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.

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[63] (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.

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 test statistics as there are samples in the time series. To correct for multiple comparisons we

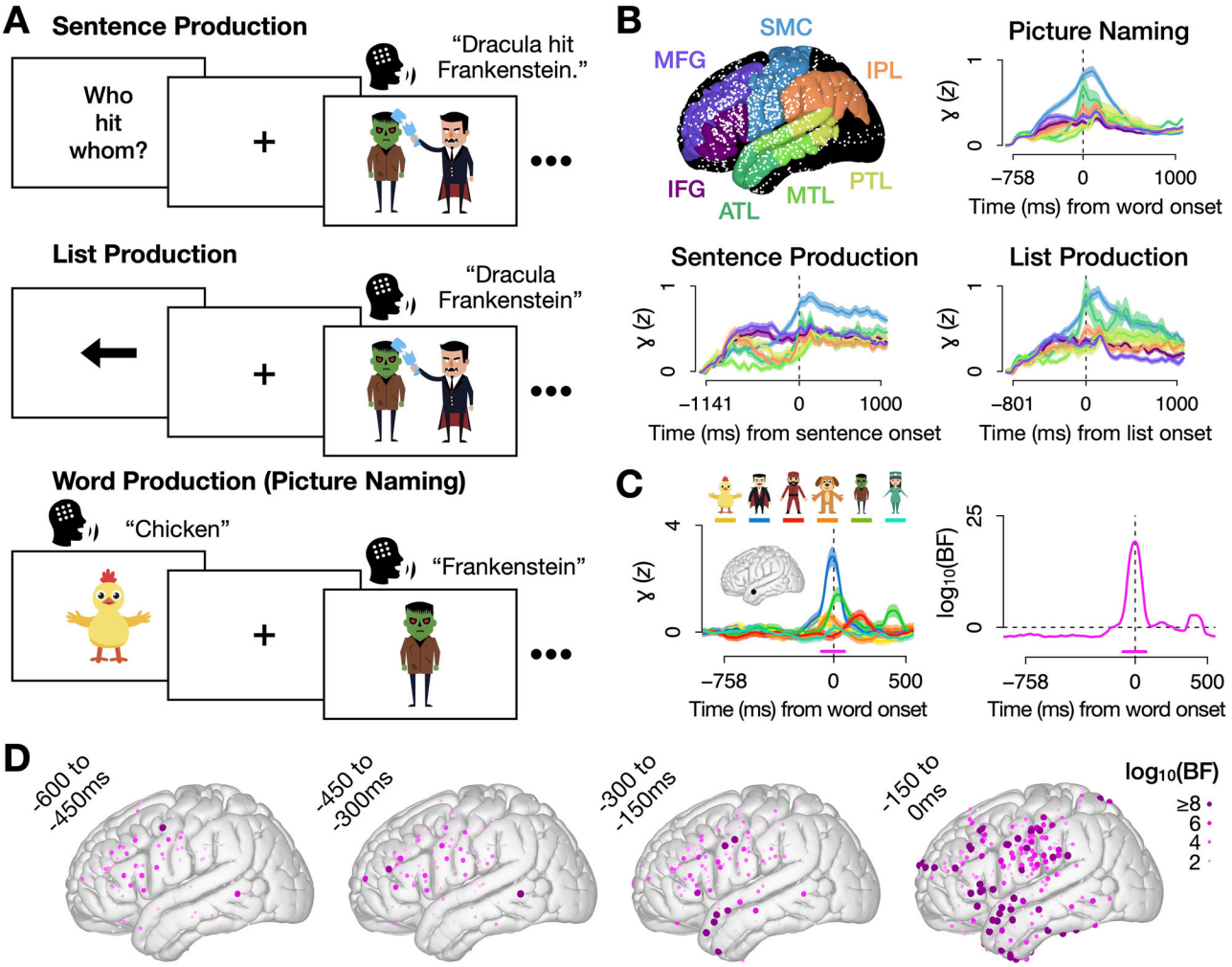

Fig. 1 | Experiment design and task-specific response profiles. A Task design: In sentence production trials, participants described static cartoon scenes in response to preceding questions. Scenes involved two of the six characters used throughout the experiment (*chicken, dog, Dracula, Frankenstein, ninja, nurse*). Half of the questions were manipulated to appear in active syntax (e.g., "Who hit whom?"), implicitly priming active responses ("Dracula hit Frankenstein."). The other half had passive syntax ("Who was hit 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 (number of electrodes per ROI: ATL = 67, IFG = 126, IPL = 75, MFG = 160, MTL = 78, PTL = 47, SMC = 207). **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; pink bar denotes where BF ≥3 for at least 100 consecutive milliseconds). Number of trials per word: chicken = 91, dog = 92, Dracula = 66, Frankenstein = 77, ninja = 84, nurse = 86. **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 ($t = 0$).

follow[59,64,65] 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. For Bayes Factor (BF) analyses of time series (Fig. 1C), bars denote where $BF \geq 3$ for at least 100 consecutive milliseconds (a Bayes Factor of 3, or $\log_{10}(BF)$ of .477, is standardly interpreted as moderate evidence for the alternate hypothesis[66]). All analyses involved accurate assumptions about the data, using non-parametric tests where assumptions of normality were violated or unsupported.

Multi-class classification

For the classification analyses, we trained multi-class classifiers on word identity using the *caret* and *nnet* packages[67,68] in R[69]. 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, 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 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 2.6 for details on multiple comparisons corrections).

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[70–74], which we previously verified in this dataset[59]. Following[59,73], 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. S2): 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 (1142 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.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

We analyzed an existing dataset (first reported in ref. 59), in which ECoG was recorded from ten neurosurgical patients with electrodes implanted in left peri-Sylvian cortex. 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; see Supplementary Fig. S1). 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")[60,61]. 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[75,76].

Neural activity and word specificity

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[14,15]. Prior to this and subsequent analyses, we followed previous work[59,70–74] 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. S2). This boosts signal-to-noise ratio[59,73] 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 (e.g., Fig. 1C, top). We quantified the amount of evidence for word specificity in each electrode using Bayesian ANOVAs (Fig. 1C, bottom); positive values indicate more evidence for word-specificity). This information was broadly distributed across cortex, increasing from stimulus onset to speech onset (Fig. 1D; see Supplementary Fig. S3 for each word's unique network).

Decoding words during picture naming

To more accurately identify word-specific activity patterns, we performed a series of decoding analyses[77,78] on the picture naming data. In essence, this analysis (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 the withheld "test" trials (10%) by assessing how similar each 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 calculated 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.)[14,15,23,25], the time frame of the training data 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 avoid this, we spanned time, training models on the mean high gamma activity in each 50 ms window from −750 ms to 250 ms relative to speech onset. In all, this resulted in 1280 classifiers: 10 patients × 7 ROIs × 20 training time windows (minus ROIs where patients had insufficient electrode coverage).

Figure 2B shows the prediction accuracies from three sample classifiers. These classifiers predicted word identity above chance (pink highlights; Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms) 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. S4 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, possibly reflecting a 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 time course 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 pre-articulatory representations stay online at least until speech onset, and post-articulatory representations are engaged throughout production.

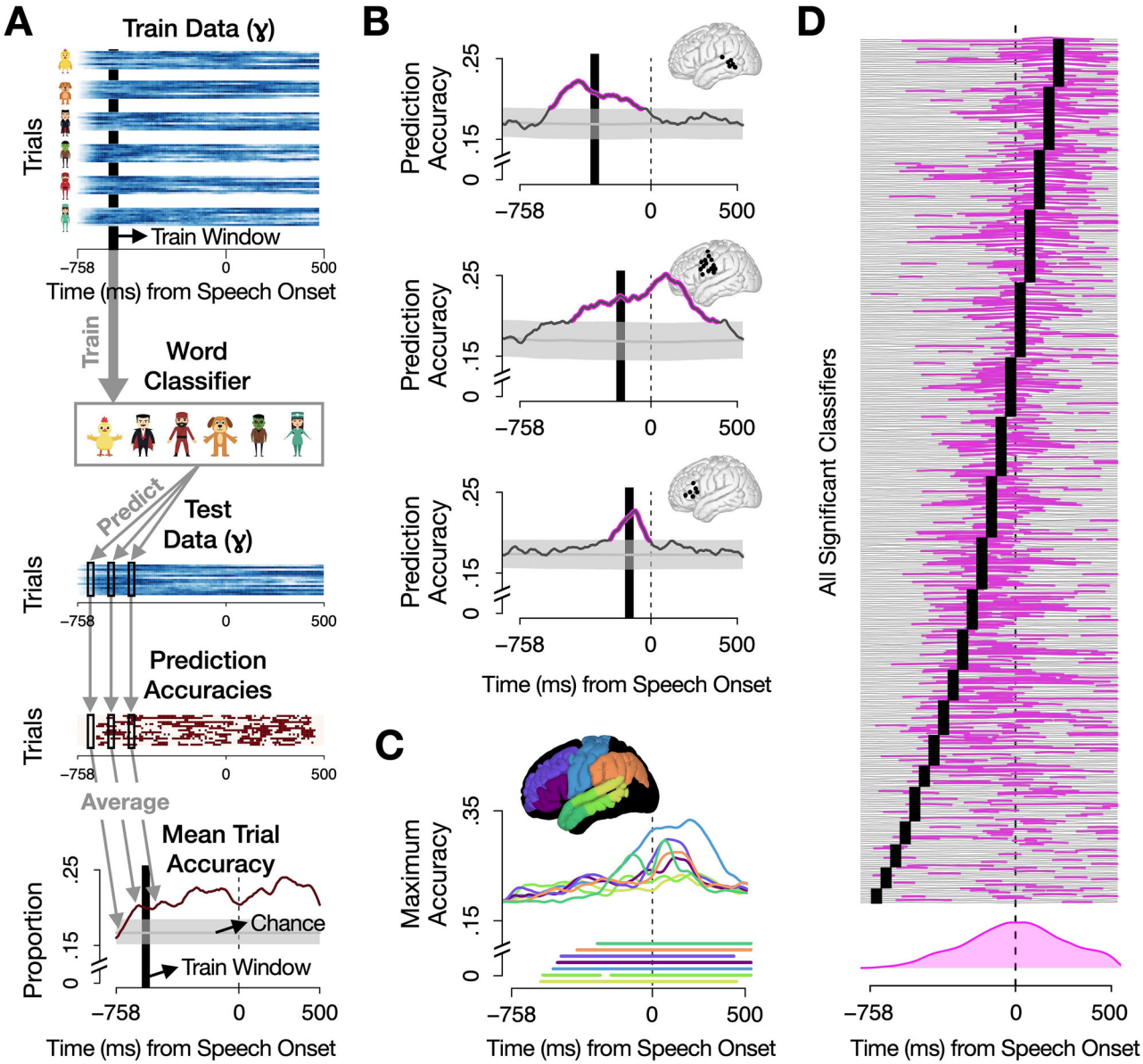

Fig. 2 | Word-specific information in picture naming. A Schematic of the analysis pipeline with simulated data. For each patient and region of interest (ROI), we trained a classifier on word identity using the time-averaged picture naming data from a 50 ms-window (−600 to −550 ms in the example; black rectangle). We then predicted word identity in "test" trials that were not in the training data, coding predictions as correct (1) or incorrect (0) for each test trial (row) at each time sample (column), generating a matrix of prediction accuracies (red). We repeated this for 30 train/test data parcellations (i.e., 10-fold validation, repeated 3 times), and averaged the resulting prediction accuracies across trials to calculate accuracy over time (bottom). A permutation analysis generated 1000 results reflecting chance performance, and significance was determined with respect to this distribution (gray). This whole process was performed 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 classifiers. **B** Prediction accuracies for three sample classifiers with significant predictions (pink highlights; Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms). From top to bottom, training and test data came from (1) Patient 9, PTL, −350 to −300 ms ($n = 513$ trials); (2) Patient 6, SMC, −200 to −150 ms ($n = 331$ trials); and (3) Patient 5, IFG, −150 to −100 ms ($n = 402$ trials). **C** Each ROI's maximum prediction accuracy across classifiers; bars denote significance (Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms). (See Supplementary Fig. S4 for results by ROI.) **D** Prediction accuracies from the 444 classifiers that made significant predictions, stacked vertically to highlight the time course of word-specific information in picture naming. Each horizontal line corresponds to one classifier. Pink denotes where that classifier's accuracy was above chance. Black bars represent the time window of training. The pink curve at the bottom shows the density of significant predictions, revealing that the most significant decoding occurred at speech onset ($t = 0$).

Generalization of word representations from picture naming to word lists

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 2A. However, instead of training and testing on subsets of the picture naming data, we used all of the picture naming data to train classifiers, and then used these classifiers to predict word identity during the production of lists like "Dracula Frankenstein." The plots in Fig. 3A demonstrate results from a sample region, the sensorimotor cortex, showing the proportion of trials where SMC classifiers predicted the first word (left; "Dracula," in our example) and the second word (right; "Frankenstein"). Accuracies are time-locked to the onset of the first word in the left panel and the second word in the right. These sample results

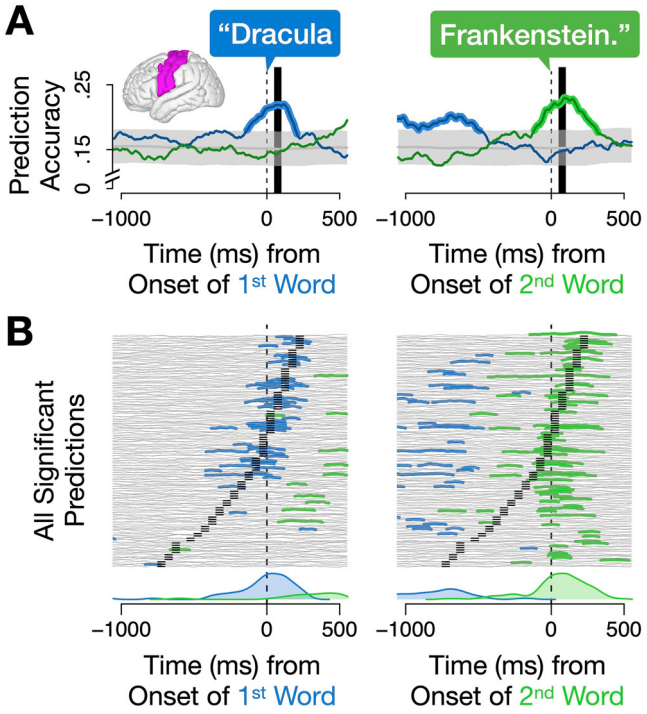

Fig. 3 | Predicting words in lists using picture naming data. A Sample prediction accuracies for the first and second nouns in lists (e.g., "Dracula" and "Frankenstein" in the example) from classifiers trained on picture naming and tested on lists. The significant detections of the two nouns in the list 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 $n = 10$ patients. Significant prediction accuracy is highlighted in blue for the first word and green for the second. **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. 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.

come from classifiers trained on data from SMC between 50 and 100 ms after speech onset during picture naming, and averaged across participants. Blue highlights denote above-chance prediction of the first word and green highlights of the second word (Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms). We accurately predicted each word as it is being said – e.g., Dracula when the patient said Dracula and then Frankenstein when the patient said 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 as in picture naming. 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, these similarities to the picture naming results (Fig. 2) suggest that list production may involve similar processes as picture naming.

Generalization to sentences: Word processing in actives follows order of articulation

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[45,79–82]. Figure 4A shows the results from the same sensorimotor classifiers previously shown for lists in Fig. 3A. For active sentences, we decoded subjects and objects – Dracula and Frankenstein in the sample sentence (but the particular words in subject/object position varied across trials) – while each was being said (Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms). 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, the order of the two nouns is confounded with their relative salience in the event in active sentences. 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. It may, therefore, not be entirely surprising that the brain processes words in active sentences 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 order of nouns – i.e., producing the character being acted upon first, and the character doing the action second. On half of the sentence production trials, we primed patients with questions formed with passive syntax ("Who was hit by whom?") rather than active syntax. This successfully elicited a mix of active and passive utterances from patients. Consistent with previous findings[61], active primes were more successful in eliciting actives (96.93% of trials) than passive primes were in eliciting passives (49.15%). A mixed-effects logistic regression modeling passive production as a function of prime syntax revealed that passive primes did indeed result in more passive responses than active primes ($\beta = 4.242$, $z = 9.506$, $p < 0.001$, effect size (log-odds) = 4.242, 95% Confidence Interval (CI) = [3.367, 5.117]; model converged with a random intercept for patient).

Sustained, simultaneous encoding of words in passive sentences

Because passive sentences convey the same meanings as actives but with the order of the nouns reversed, they present an opportunity to disentangle serial order and salience. Specifically, if the reason we decoded subjects before objects in active sentences was solely because subjects were more agentive, then we should expect to see the temporal dynamics of word processing reverse in passives. We started by examining prefrontal regions. Prefrontal cortex has previously been shown to involve higher activity for passives than actives[83–88], but the precise reason for this remains debated[45]. Figure 5A shows the mean decoding accuracy in MFG during passive sentences, and indeed reveals a distinct temporal profile from what we observed for actives: sustained encoding of the object throughout the duration of the sentence (Permutation Tests, $p < 0.05$ for 100 ms or $p < 0.01$ for 50 ms), starting even before the onset of the sentence. However, this was not true everywhere in cortex during passive sentences. Figure 5B shows the predictions of the same sensorimotor classifier from Fig. 4A. In this region, responsible for articulation and sensory feedback, words were decoded as they were in actives: in the order in which they were produced. To look at the overall pattern, we stacked the predictions from all 97 classifiers that made above-chance predictions (Fig. 5C; for a breakdown by ROI see

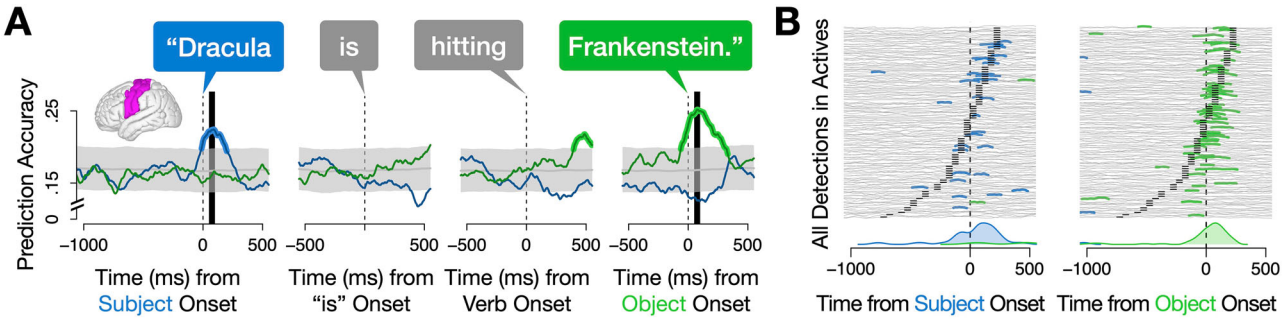

Fig. 4 | Decoding words in active sentences using picture naming data. A Sample prediction accuracies for the first and second nouns (i.e., subject and object) of active sentences, locked to the onset of each word. This was the same classifier as in Fig. 3A, i.e., trained on picture naming data from electrodes in SMC between 50 and 100 ms (black bar) after speech onset (averaged across all $n = 10$ patients). Both nouns were predicted above chance at the time of their respective articulations (blue/green highlights). **B** Stacked prediction accuracies from the 83 significant classifiers for active sentences, locked to the onset of both nouns. Density plots at the bottom show that each word's accuracy peaked during its articulation. (See Supplementary Fig. S5 for density plots broken down by ROI).

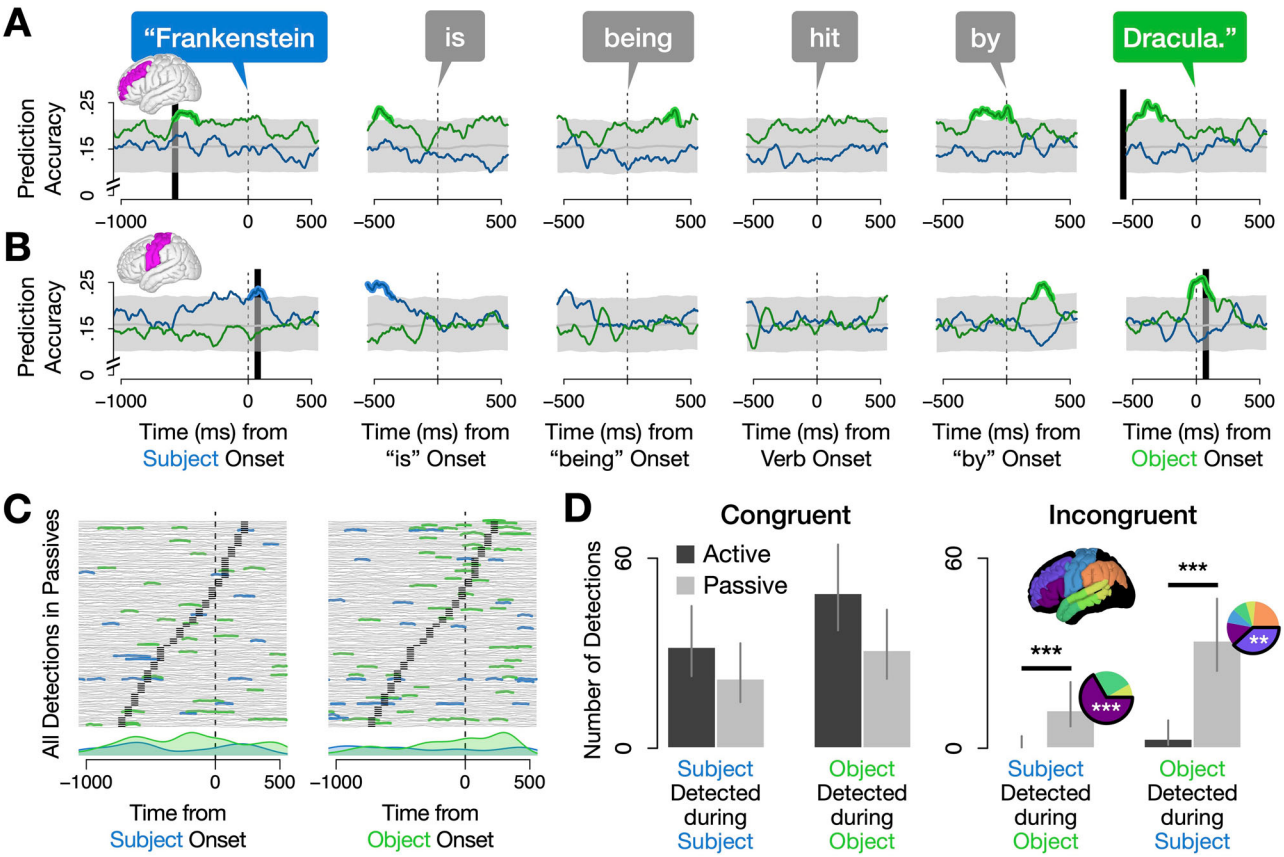

Fig. 5 | Word decoding results for passive sentences. A Mean prediction accuracies for passive sentences from MFG classifiers trained on picture naming data from −600 to −550 ms ($n = 10$ patients). This prefrontal region sustained a representation of the object throughout the sentence. **B** Mean prediction accuracies for passive sentences from the same SMC classifier in Fig. 4A (and Fig. 3A) ($n = 10$ patients). As in these previous analyses, this classifier detected each noun during its respective articulation. **C** Stacked prediction accuracies from the 97 significant classifiers for passive sentences, locked to each noun's onset. Unlike in active sentences, there is little correspondence between training time (black bars) and when words were detected (green and blue segments). This point is made especially clear by the density plots, which reveal that both subjects and objects were active throughout passive sentence production. **D** Number of classifiers that significantly predicted either word in the sentence, broken down by whether the prediction revealed temporally congruent processing (i.e., detection of the subject during production of the subject or detection of the object during the object) or incongruent (detection of the object during the subject or vice versa). Error bars reflect 95% CIs on detection counts, computed by applying the Wilson score interval to the underlying binomial proportions [107] and scaling the resulting bounds by the number of classifiers ($n = 1280$). While both active (black) and passive (gray) sentences involved temporally congruent word processing, passives had significantly more incongruent detections than actives for both incongruent subjects and incongruent objects. Pie charts show where these incongruent detections were made: incongruent passive subjects were detected in IFG more than any other ROI (8 out of $n = 12$ classifiers), whereas incongruent passive objects were found in MFG (13 out of $n = 34$ classifiers; see Supplementary Fig. S5 for more detail).

Supplementary Fig. S5). This analysis revealed that in passives, both the subject and object remained active throughout the entirety of the sentence, indicating that the brain processes both characters in the sentence simultaneously rather than sequentially. To assess whether this constituted a statistically significant difference from active sentences, we counted the number of classifiers that detected each word during the production of each word (Fig. 5D) – 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" detections. In both active and passive sentences we observed many congruent detections (Fig. 5D, left).

However, of the incongruent detections (Fig. 5D, right), nearly all were in passive sentences. Relative to active sentences, passive sentences involved significantly more incongruent detections of both subjects (one-sided Test of Equal Proportions, $\chi^2(1) = 10.132$, FDR-corrected $p < 0.001$, difference in proportions $= 0.010$, 95% CI $= [0.004, 1]$) and objects (one-sided Test of Equal Proportions, $\chi^2(1) = 24.687$, FDR-corrected $p < 0.001$, difference in proportions $= 0.025$, 95% CI $= [0.016, 1]$). (The Test of Equal Proportions was chosen due to the bounded nature of the counts, which were between 0 and 1280, or the total number of classifiers. Note that pairwise tests were only performed where the number of detections was higher for passives than actives because the reverse pattern is uninterpretable: active analyses having roughly three times more data than passives, statistical power was higher for actives, meaning that the higher number of active detections is likely trivial. On the other hand, the significantly higher number of incongruent detections for passives than actives, in spite of passives' lower power, lends extra credibility to those effects.)

This incongruent noun representation in passive sentences was driven by two regions (see pie charts and Supplementary Fig. S5): IFG, which preferentially encoded subjects more than any other region (one-sided Test of Equal Proportions, $\chi^2(1) = 23.000$, FDR-corrected $p < 0.001$, difference in proportions $= 0.041$, 95% CI $= [0.012, 1]$), and MFG, which preferentially encoded objects (one-sided Test of Equal Proportions, $\chi^2(1) = 11.421$, FDR-corrected $p < 0.001$, difference in proportions $= 0.045$, 95% CI $= [0.013, 1]$). Notably, no such regional specificity was observed in active sentences. A Bayesian Contingency Analysis[89] revealed substantial evidence against role-specific encoding during active sentences in prefrontal cortex, with a Bayes Factor of 0.029 for IFG (i.e., 34.126 times more evidence for the null hypothesis that there is no role-specific encoding) and 0.042 in MFG (23.672 times more evidence for the null).

Strikingly, even when patients were producing the subject of a passive sentence, we detected numerically more objects (i.e., incongruent detections) than subjects (congruent), although this difference was only marginally significant ($\chi^2(1) = 2.210$, $p = 0.069$, difference in proportions $= 0.010$, 95% CI $= [-0.001, 1]$, one-sided Test of Equal Proportions). In summary, our decoding analysis revealed a marked difference in lexical processing between active and passive sentences. While actives exhibit a pattern similar to that in list and single word production, where words are activated in the order they are produced, passive sentences involve sustained, simultaneous encoding of both the subject and object nouns. This difference was region-specific, with sensorimotor cortex consistently representing lexical information in task-agnostic ways, but IFG and MFG showing sensitivity to syntactic structure, highlighting a division of labor within the language network where prefrontal regions support structure-dependent processing demands.

Discussion

Single word production tasks like picture naming have dominated the neuroscience of language production, but there is little direct evidence for the critical assumption that what is true of words in isolation remains true in more naturalistic utterances like sentences. In this study, we leveraged the unparalleled spatiotemporal precision of ECoG and employed an innovative cross-task classification approach to assess the similarities and differences in the production of words and sentences. We first demonstrated that individual words can be decoded from patterns of neural activity in picture naming data, confirming that our data contained word-specific information. We then trained classifiers to identify words in seven regions of interest and at 20 time points spanning the picture naming epoch. Applying these classifiers to sentence production data revealed three key findings. First, we successfully decoded nouns during sentence production using picture naming data, validating the assumption that word representations are shared across linguistic behaviors. Second, by comparing word decoding results between sentences with active vs. passive syntax, which convey the same event meanings but with nouns in reverse orders, we demonstrated that the temporal dynamics of word processing depend on syntactic structure. In active sentences, which represent the canonical word order in English and typically involve producing nouns in order of agency (from most agentive to least), there was a tight correspondence between word processing in the brain and word order in speech: classifiers decoded the subject and object nouns in the order they were produced. In contrast, passive sentences showed a significant departure from this temporal alignment: rather than encoding each word as it was said, the brain encoded both words simultaneously for the duration of the sentence. Third and finally, our data revealed a spatial code in prefrontal cortex for words' syntactic position, with subjects preferentially encoded in IFG and objects in MFG.

Syntax-dependent dynamics of word processing

Our findings validate various aspects of cognitive[90–92] and computational[93,94] models of sentence production. These models assume that word representations are invariant across different behaviors, which is reflected in our finding that the cortical representation of words during picture naming generalizes to both list and sentence contexts. Furthermore, these models build in a dependence on syntactic structure during sentence planning, predicting that the dynamics of word processing may vary with syntactic structure. Consistent with this, we observed striking differences between active and passive sentences: in actives, the subject and object were decoded sequentially in the order in which they were produced, whereas in passives we continuously decoded both nouns throughout the duration of the sentence. This sustained representation was driven by activity in IFG and MFG, aligning with prior experimental[83–85], stimulation[86], and lesion-symptom mapping studies[87,88] implicating these prefrontal regions in the processing of noncanonical structures like passives.

The role of prefrontal cortex in sentence production

There is ongoing debate regarding the specific functions of IFG and MFG in passives. While we have attributed these differences to "syntax," they could in principle derive from other differences between actives and passives. The literature identifies a number of such extraneous differences: working memory[95], thematic role assignment[84], and syntactic movement (a syntactic operation associated with certain complex structures[83,85,87]). However, much of this prior work relies on measures like "activity," which do not directly map onto specific psychological processes or representations, complicating efforts to discern among these possibilities. By tracking word-specific information, our findings reveal at least one function of IFG and MFG in passives: the sustained encoding of words and their syntactic roles. This finding poses a challenge for the syntactic movement account, as movement pertains to abstract structure rather than specific words[96,97]. Similarly, the thematic role assignment hypothesis is inconsistent with our results, as thematic roles did not differ between actives and passives.

Of these three alternative possibilities, our results align most strongly with a working memory (or perhaps attention) account. Passives likely engage such cognitive resources to a higher degree, as they are less common than actives, involve planning more words, and, in our stimuli, require the reversal of the canonical agent-first ordering. We suspect that this latter property is most likely to drive the effect in our data given that IFG and MFG encoded words according to their syntactic role in the sentence. Ongoing access to information about which noun occupies which syntactic position may be needed to override default syntactic processes, which would

presumably favor mapping agents to subject, as in active sentences. Thus, the sustained encoding of nouns in IFG and MFG may constitute a core neural mechanism by which the brain exerts top-down control of speech, enabling flexible sentence production to meet situational or task demands. A working memory account further aligns with the fact that IFG and MFG sit squarely within the *multiple-demand network*, a domain-general system that supports cognitive resources like attention and working memory[98–100]. While this explanation fits our findings better than the thematic role assignment or syntactic movement accounts, it does not rule out the possibility that these regions additionally support other such functions. This is particularly true in light of the fact that the relative contributions of these functions may differ between production and comprehension, meaning that our results may not be directly comparable to previous comprehension work.

A spatial code for syntactic roles

Syntactic roles, or structural positions like subject and object, are key parts of all models of language processing. They are necessary for mapping words to semantic/thematic roles (i.e., who performs an action vs. who it is done to). In production, these roles must be quickly and flexibly linked to specific words so that, for example, "Dracula" can be the subject of one sentence and the object of another in rapid succession without causing confusion (see discussion of the "fast-changing weights" in Chang et al.'s 2006 computational model). Theoretical accounts often point to coherence as a possible neural mechanism for this type of binding[94,101], but there remains little empirical evidence for this (or any) particular implementation. Here, we uncovered a spatial code for syntactic roles during passives, with subjects encoded in IFG and objects in MFG.

Two caveats, however, warrant consideration. First, while active sentences also have subjects and objects, we found no evidence for role-specific encoding in prefrontal cortex during actives (see Supplementary Fig. S5), despite higher statistical power (patients produced approximately three times more actives than passives). Indeed, Bayesian analyses revealed substantial evidence for a lack of regional specificity for syntactic roles in actives. This suggests that the spatial code we observed is not the general neural mechanism for encoding syntactic role (or linear position, whichever it may be). However, an alternative explanation is that the brain does not rely on syntactic roles during the production of actives. Speakers may develop a heuristic strategy for producing very common structures[102], for instance, something like "produce the more agentive noun first." Indeed, previous work has shown that sentence production is not always driven by bottom-up syntactic encoding in production, and can instead be driven by attention[103] (see ref. 91 for discussion). The second caveat is that the absence of evidence for position-specific encoding in actives limits our ability to disentangle whether this feature of passives encodes syntactic roles or linear position. Prior work has linked IFG to linear rather than hierarchical-syntactic representations[54], potentially indicating a linear position interpretation of our results is a better fit. Further work is needed to conclusively tease apart these possibilities.

Regardless, the pattern we observed for passives constitutes a clear demonstration of a neural code for a noun's position in a sentence, be it linear or hierarchical. This important finding shows how a spatial organization of lexical information can and does encode positional information for words in sentences.

Limitations

Future work is needed to identify *which* lexical representations we detected. Our stimuli, designed for different purposes, do not readily distinguish between different types of lexical information. For example, an electrode that selectively responds to "Dracula" and "Frankenstein" might reflect semantic information (*monsterhood*) or form-level information, as both words have stress on the first of three syllables. One approach to isolating lexical representations could involve examining the spatial or temporal distribution of lexical information. For instance, if different lexical representations correspond to specific ROIs and come online in a feedforward sequence during picture naming, then one would expect word-specific information to first emerge in conceptual regions, then in lemma regions, and so on. However, it remains an open question whether either of these assumptions are valid. Spatially, there is growing consensus that at least semantic knowledge is broadly distributed across cortex[22], which, if true, would mean that decoding in any ROI could be driven by semantic information. Indeed, this could be true of other types of lexical representations as well, where evidence is stronger for a more spatially concentrated code, but still not a matter of certainty. Temporally, if parallel models of word production are correct and lexical representations come online contemporaneously[30–32], then temporal patterns are also of limited use in distinguishing between representations. But even under strictly feedforward circumstances, our data reveal another complication: the signal-to-noise ratio of different representations appears to change over time, leading to potentially sizable differences between when a representation comes online and when it is detected. Evidence for this is clearest in the picture naming data (Fig. 1D), where words tended to be decoded around the onset of articulation regardless of when the training data came from. This was particularly true for earlier training times (the lower part of the "stack" plot), where many classifiers successfully decoded words around time 0 even when they were unable to do so at the time they were trained on. This points toward a signal-to-noise ratio increase leading up to articulation, making information easier to decode at articulation even if it came online much earlier. Thus, we caution against directly interpreting the lack of decoding in classifiers trained at earlier times, as false negatives are likely to become more frequent the earlier the training data come from.

Interestingly, there was one clear exception to this trend. Lexical information in posterior temporal lobe (PTL) peaked far earlier than in other ROIs (Supplementary Fig. S4B), consistent with evidence implicating this region in phonological encoding, one of the earlier stages in feedforward models of word production (see ref. 15 for a review). However, other work has associated PTL with more general processes such as lexical access and coordinating different types of information[55–57], which would implicate multiple levels of lexical representation. Thus, while there are suggestive features of our data, ongoing disagreements in the field about the timecourse and localization of lexical representations, combined with limitations of our stimulus design, prevent us from answering these questions definitively. Nonetheless, this study introduces an innovative approach to investigating these issues – one that we anticipate will be instrumental in resolving these questions in future research.

Word order typology aligns with neural costliness

Finally, we suggest that our findings may shed light on a widely noted but poorly understood pattern among the world's 6000 languages. Specifically, there are six possible orders in which a language can arrange subjects, verbs, and objects: 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 place objects before subjects (e.g., Object-Subject-Verb)[104,105]. One possible reason for this is that subjects tend to be more agentive (semantically salient) than objects, and there is a natural tendency in speech to order words from most to least agentive (a bias apparently preserved across hominids[106]). In our experiment, passive sentences provided an opportunity to visualize how the brain processes words when producing less agentive words before more agentive ones. In these cases, word planning involved a much more complex temporal pattern. Indeed, whereas word planning in actives resembled 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 sustained activation of both nouns in prefrontal cortex. Specifically, IFG sustained a representation of the subject, and MFG sustained a representation of the object. Furthermore, reaction times, commonly interpreted as an index of processing difficulty, were significantly longer for passives than both actives and lists (see Section A.2.4 in Supplementary Information). Taken together, these facts point toward a processing-based explanation of the cross-linguistic dominance of subject-before-object word

orders like those in English and Farsi. Producing words in order from least to most salient may simply be harder for the production system. We speculate that, over the course of language evolution, this difficulty exerts a subtle pressure on language change, making it more likely for languages to evolve in the direction of subject-before-object word orders.

Data availability

Data will be made available from the authors upon request to the corresponding author (Adam.Morgan@NYULangone.org), provided documentation that the data will be strictly used for research purposes and will comply with the terms of our study IRB. Numerical data underlying the manuscript figures are available on https://doi.org/10.17605/OSF.IO/GEUTY.

Code availability

The code is available at https://doi.org/10.17605/OSF.IO/GEUTY.

Received: 31 January 2025; Accepted: 19 May 2025;
Published online: 03 June 2025

References

1. Janik, V. M., Sayigh, L. S. & Wells, R. S. Signature whistle shape conveys identity information to bottlenose dolphins. *Proc. Natl Acad. Sci.* **103**, 8293–8297 (2006).
2. Wenner, A. M., Wells, P. H. & Rohlf, F. J. An analysis of the waggle dance and recruitment in honey bees. *Physiol. Zool.* **40**, 317–344 (1967).
3. Seyfarth, R. M., Cheney, D. L. & Marler, P. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094 (1980).
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. *Behav. Ecol. Sociobiol.* **56**, 71–80 (2004).
5. Momma, S., Slevc, L. R. & Phillips, C. The timing of verb selection in Japanese sentence production. *J. Exp. Psychol.: Learn., Mem., Cogn.* **42**, 813 (2016).
6. Momma, S. & Ferreira, V. Beyond linear order: The role of argument structure in speaking. *Cogn. Psychol.* **128**, 101397 (2021).
7. Riès, S. K. et al. Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proc. Natl Acad. Sci.* **114**, 4530–4538 (2017).
8. Graves, W. W., Grabowski, T. J., Mehta, S. & Gordon, J. K. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *J. Cogn. Neurosci.* **19**, 617–631 (2007).
9. Carota, F., Schoffelen, J.-M., Oostenveld, R. & Indefrey, P. The time course of language production as revealed by pattern classification of meg sensor data. *J. Neurosci.* **42**, 5745–5754 (2022).
10. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* **326**, 445–449 (2009).
11. Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P. & Salmelin, R. An MEG study of picture naming. *J. Cogn. Neurosci.* **10**, 553–567 (1998).
12. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).
13. 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**, 193–216 (1995).
14. Indefrey, P. & Levelt, W. J. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
15. Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).
16. Flinker, A. et al. Redefining the role of Broca's area in speech. *Proc. Natl Acad. Sci.* **112**, 2871–2875 (2015).
17. Tremblay, P. & Small, S. L. Motor response selection in overt sentence production: a functional MRI study. *Front. Psychol.* **2**, 253 (2011).
18. Bouchard, K. E. & Chang, E. F. Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J. Neurosci.* **34**, 12662–12677 (2014).
19. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S. & Houde, J. F. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl Acad. Sci.* **110**, 2653–2658 (2013).
20. Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S. & Casasanto, D. Visual cortex entrains to sign language. *Proc. Natl Acad. Sci.* **114**, 6352–6357 (2017).
21. Schwartz, M. F. et al. Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain* **132**, 3411–3427 (2009).
22. 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**, 453–458 (2016).
23. Butterworth, B.: Lexical access in speech production. In: Lexical Representation and Process, pp. 108–135 (1989).
24. Caramazza, A. How many levels of processing are there in lexical access? *Cogn. Neuropsychol.* **14**, 177–208 (1997).
25. Levelt, W. J., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–38 (1999).
26. Dell, G. S. A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **93**, 283 (1986).
27. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801 (1997).
28. Roelofs, A. The weaver model of word-form encoding in speech production. *Cognition* **64**, 249–284 (1997).
29. Carota, F., Schoffelen, J.-M., Oostenveld, R. & Indefrey, P. Parallel or sequential? decoding conceptual and phonological/phonetic information from meg signals during language production. *Cogn. Neuropsychol.* **40**, 298–317 (2023).
30. Strijkers, K., Costa, A. & Pulvermüller, F. The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage* **163**, 206–219 (2017).
31. Strijkers, K. & Costa, A. The cortical dynamics of speaking: Present shortcomings and future avenues. *Lang., Cogn. Neurosci.* **31**, 484–503 (2016).
32. Pickering, M. J. & Strijkers, K. Language production and prediction in a parallel activation model. *Topics in cognitive science*, (2024).
33. Menenti, L., Segaert, K. & Hagoort, P. The neuronal infrastructure of speaking. *Brain Lang.* **122**, 71–80 (2012).
34. Lukic, S. et al. Common and distinct neural substrates of sentence production and comprehension. *NeuroImage* **224**, 117374 (2021).
35. Giglio, L., Ostarek, M., Sharoh, D. & Hagoort, P. Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. *Proc. Natl Acad. Sci.* **121**, 2310766121 (2024).
36. Salmelin, R., Hari, R., Lounasmaa, O. & Sams, M. Dynamics of brain activation during picture naming. *Nature* **368**, 463–465 (1994).
37. 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**, 2028–2044 (2001).
38. Haller, S., Radue, E.-W., Erb, M., Grodd, W. & Kircher, T. Overt sentence production in event-related fMRI. *Neuropsychologia* **43**, 807–814 (2005).
39. 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. *Psychol. Sci.* **22**, 1173–1182 (2011).
40. 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 Lang.* **121**, 47–57 (2012).
41. Grande, M. et al. 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**, 702–714 (2012).
42. Schönberger, E. et al. The neural correlates of agrammatism: Evidence from aphasic and healthy speakers performing an overt picture description task. *Front. Psychol.* **5**, 246 (2014).
43. Geranmayeh, F., Wise, R. J., Mehta, A. & Leech, R. Overlapping networks engaged during spoken language production and its cognitive control. *J. Neurosci.* **34**, 8728–8740 (2014).
44. Blanco-Elorrieta, E., Kastner, I., Emmorey, K. & Pylkkänen, L. Shared neural correlates for building phrases in signed and spoken language. *Sci. Rep.* **8**, 5492 (2018).
45. Walenski, M., Europa, E., Caplan, D. & Thompson, C. K. Neural networks for sentence comprehension and production: An ale-based meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* **40**, 2275–2304 (2019).
46. Giglio, L., Ostarek, M., Weber, K. & Hagoort, P. Commonalities and asymmetries in the neurobiological infrastructure for language production and comprehension. *Cereb. Cortex* **32**, 1405–1418 (2022).
47. Hu, J. et al. Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cereb. Cortex* **33**, 4384–4404 (2023).
48. Hagoort, P. & Indefrey, P. The neurobiology of language beyond single words. *Annu. Rev. Neurosci.* **37**, 347–362 (2014).
49. Pylkkänen, L., Bemis, D. K. & Elorrieta, E. B. Building phrases in language production: An MEG study of simple composition. *Cognition* **133**, 371–384 (2014).
50. Westerlund, M. & Pylkkänen, L. The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia* **57**, 59–70 (2014).
51. Pylkkänen, L. The neural basis of combinatory syntax and semantics. *Science* **366**, 62–66 (2019).
52. Pylkkänen, L. Neural basis of basic composition: what we have learned from the red–boat studies and their extensions. *Philos. Trans. R. Soc. B* **375**, 20190299 (2020).
53. Matchin, W. et al. Agrammatism and paragrammatism: a cortical double dissociation revealed by lesion-symptom mapping. *Neurobiol. Lang.* **1**, 208–225 (2020).
54. Matchin, W. & Hickok, G. The cortical organization of syntax. *Cereb. Cortex* **30**, 1481–1498 (2020).
55. Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. Converging language streams in the human temporal lobe. *J. Neurosci.* **26**, 7328–7336 (2006).
56. Choi, Y.-H., Park, H. K. & Paik, N.-J. Role of the posterior temporal lobe during language tasks: a virtual lesion study using repetitive transcranial magnetic stimulation. *Neuroreport* **26**, 314–319 (2015).
57. Hope, T. M. & Price, C. J. Why the left posterior inferior temporal lobe is needed for word finding. *Brain* **139**, 2823–2826 (2016).
58. Peirce, J. et al. Psychopy2: Experiments in behavior made easy. *Behav. Res. methods* **51**, 195–203 (2019).
59. Morgan, A.M. et al. A low-activity cortical network selectively encodes syntax. bioRxiv, 2024–06 (2024)
60. Bock, J. K. Syntactic persistence in language production. *Cogn. Psychol.* **18**, 355–387 (1986).
61. Pickering, M. J. & Ferreira, V. S. Structural priming: a critical review. *Psychol.Bull.* **134**, 427 (2008).
62. Audacity Team: Audacity (R). http://audacity.sourceforge.net/
63. Yang, A. I. et al. Localization of dense intracranial electrode arrays using magnetic resonance imaging. *Neuroimage* **63**, 157–165 (2012).
64. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG and MEG data. *J. Neurosci. Methods* **164**, 177–190 (2007).
65. Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O. & Poeppel, D. Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Hum. Behav.* **3**, 393–405 (2019).
66. Jeffreys, H.: The Theory of Probability. OUP Oxford, (1998)
67. Kuhn & Max Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
68. Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S, 4th edn., New York ISBN 0-387-95457-0. https://www.stats.ox.ac.uk/pub/MASS4/ (2002).
69. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. https://www.R-project.org/ (2021)
70. Williams, A. H. et al. Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. *Neuron* **105**, 246–259 (2020).
71. Wang, K., Begleiter, H. & Porjesz, B. Warp-averaging event-related potentials. *Clin. Neurophysiol.* **112**, 1917–1924 (2001).
72. Picton, T. W., Lins, O. G. & Scherg, M. The recording and analysis of event-related potentials. *Handb. Neuropsychol.* **10**, 3–3 (1995).
73. Edwards, E. et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* **50**, 291–301 (2010).
74. Molina, M., Tardón, L. J., Barbancho, A. M., De-Torres, I. & Barbancho, I. Enhanced average for event-related potential analysis using dynamic time warping. *Biomed. Signal Process. Control* **87**, 105531 (2024).
75. Mukamel, R. et al. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* **309**, 951–954 (2005).
76. Nir, Y. et al. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr. Biol.* **17**, 1275–1285 (2007).
77. Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. Encoding and decoding in fMRI. *Neuroimage* **56**, 400–410 (2011).
78. Holdgraf, C. R. et al. Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* **11**, 61 (2017).
79. Caramazza, A. & Zurif, E. B. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain Lang.* **3**, 572–582 (1976).
80. Grodzinsky, Y. The neurology of syntax: Language use without Broca's area. *Behav. Brain Sci.* **23**, 1–21 (2000).
81. Love, T., Swinney, D., Walenski, M. & Zurif, E. How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. *Brain Lang.* **107**, 203–219 (2008).
82. Thompson, C. K. & Choy, J. J. Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *J. Psycholinguist. Res.* **38**, 255–283 (2009).
83. Mack, J. E., Meltzer-Asscher, A., Barbieri, E. & Thompson, C. K. Neural correlates of processing passive sentences. *Brain Sci.* **3**, 1198–1214 (2013).
84. Yokoyama, S. et al. Cortical activation in the processing of passive sentences in L1 and l2: An fMRI study. *Neuroimage* **30**, 570–579 (2006).

85. Feng, S. et al. Differences in grammatical processing strategies for active and passive sentences: An fMRI study. *J. Neurolinguist.* **33**, 104–117 (2015).
86. Riva, M. et al. Evaluating syntactic comprehension during awake intraoperative cortical stimulation mapping. *J. Neurosurg.* **138**, 1403–1410 (2022).
87. Tyler, L. K. et al. Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* **134**, 415–431 (2011).
88. Kinno, R. et al. Agrammatic comprehension caused by a glioma in the left frontal cortex. *Brain Lang.* **110**, 71–80 (2009).
89. Morey, R.D., Rouder, J.N., Jamil, T.: BayesFactor: Computation of Bayes Factors for Common Designs. (2023). R package version 0.9.12-4.6
90. Ferreira, V.S., Morgan, A.M., Slevc, L.R.: Grammatical encoding. The Oxford Handbook of Psycholinguistics, 453–469 (2018).
91. Bock, K., Ferreira, V.S.: Syntactically speaking. The Oxford handbook of language production, 21–46 (2014).
92. Slevc, L.R.: Grammatical encoding. In: Language Production, pp. 4–31. Routledge, (2023).
93. Chang, F., Dell, G. S. & Bock, K. Becoming syntactic. *Psychol. Rev.* **113**, 234 (2006).
94. Murphy, E. Rose: A neurocomputational architecture for syntax. *J. Neurolinguist.* **70**, 101180 (2024).
95. Fiebach, C. J., Schlesewsky, M., Lohmann, G., Von Cramon, D. Y. & Friederici, A. D. Revisiting the role of broca's area in sentence processing: syntactic integration versus syntactic working memory. *Hum. Brain Mapp.* **24**, 79–91 (2005).
96. Chomsky, N.: Aspects of the Theory of Syntax vol. 11. MIT Press, (2014).
97. Chomsky, N.: The Minimalist Program. MIT Press, (2014).
98. Duncan, J. The multiple-demand (md) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010).
99. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl Acad. Sci.* **110**, 16616–16621 (2013).
100. Hugdahl, K., Raichle, M. E., Mitra, A. & Specht, K. On the existence of a generalized non-specific task-dependent network. *Front. Hum. Neurosci.* **9**, 430 (2015).
101. Martin, A. E. & Doumas, L. A. A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biol.* **15**, 2000663 (2017).
102. Goldberg, A. E. & Ferreira, F. Good-enough language production. *Trends Cogn. Sci.* **26**, 300–311 (2022).
103. Kuchinsky, S.E.: From Seeing to Saying: Perceiving, Planning, Producing. University of Illinois at Urbana-Champaign, (2009)
104. Dryer, M.S. Determining Dominant Word Order. (eds Dryer, M. S. & Martin, H.). WALS Online (v2020.4) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591 (2013).
105. Hammarström, H. Linguistic diversity and language evolution. *J. Lang. Evol.* **1**, 19–29 (2016).
106. Brocard, S., Wilson, V.A., Berton, C., Zuberbühler, K., Bickel, B.: A universal preference for animate agents in hominids. iScience **27**(6) (2024)
107. Dorai-Raj, S.: Binom: Binomial Confidence Intervals For Several Parameterizations. R package version 1.1-1.1. https://CRAN.R-project.org/package=binom (2014).

Acknowledgements

We thank the entire Flinker Lab for feedback on this project and Yasamin Esmaeili for the Farsi example in the text. This work was supported by National Institutes of Health grants F32DC019533 (A.M.), R01NS109367 (A.F.), R01NS115929 (A.F.), and R01DC018805 (A.F.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Adam M. Morgan was responsible for project conceptualization, experimental design, data collection, data processing and analysis, funding acquisition, and manuscript preparation. Orrin Devinsky, Werner K. Doyle, Patricia Dugan, and Daniel Friedman were responsible for the clinical aspects of the project, including electrophysiological recordings and electrode localization. Adeen Flinker oversaw the project and was involved in conceptualization, design, data analysis, funding acquisition, and manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-025-00270-1.

Correspondence and requests for materials should be addressed to Adam M. Morgan.

Peer review information *Communications Psychology* thanks Constantijn van der Burght, Kristof Strijkers, and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Troby Ka-Yan Lui. A peer review file is available.