From Single Words to Sentence Production: Shared Cortical Representations but Distinct Temporal Dynamics

Preprint: https://doi.org/10.1101/2024.10.30.621177

Adam Morgan[1], Orrin Devinsky[1], Werner Doyle[1], Patricia Dugan[1], Daniel Friedman[1], Adeen Flinker[1,2] *NYU Grossman School of Medicine[1], NYU Tandon School of Engineering[3]*

Language production research has primarily focused on single-word tasks, such as picture naming. This work has converged on models that describe a cascade of distinct representations – conceptual, grammatical (lemma), phonological, articulatory, etc. – each associated with particular brain regions [1-5]. In contrast, the neural underpinnings of sentence production remain less understood, and it is unclear whether insights from word-level studies fully generalize to more naturalistic utterances like sentences. This gap arises from challenges in experimentally controlling what sentences participants produce and limitations of traditional neural measures like fMRI, MEG, and EEG, which are highly sensitive to motor artifacts and provide limited spatial or temporal resolution [6]. To address this, we used electrocorticography (ECoG) to record electrical potentials directly from the surface of the brain in 10 awake neurosurgical patients as they performed an overt production experiment (Fig. 1). During a picture naming block, participants repeatedly named six characters (e.g., chicken, nurse, Dracula), chosen to vary along multiple dimensions like frequency and phonology to maximize discriminability. Participants then completed a sentence production block, describing cartoon scenes in response to questions manipulated to have active (Who hit whom?) or passive (Who was hit by whom?) structure. This stochastically elicited active ("Dracula hit Frankenstein") and passive ("Frankenstein was hit by Dracula") sentence responses.

Using machine learning classifiers, we identified the unique neural activity patterns associated with each word during picture naming. These models were trained on data from each of seven regions of interest (ROIs), for each of the 20 consecutive 50ms windows from -750 to 250ms from speech onset, and for each of the 10 patients, resulting in ~1400 models. We first used these models to predict which word patients were processing during held-out trials in picture naming, successfully decoding word identity in 444 models (Fig. 2). Consistent with (interactive) feedforward models of word production, prediction accuracy for each model tended to peak in the time window that the model's training data came from. Our data further reveal that representations tend to stay online until speech onset, something models leave unspecified.

Next, to test generalizability, we used these same classifiers — trained on picture naming data — to predict word identity throughout sentence production (Fig. 3A,B). For active sentences, we again successfully decoded each word in the order of production – subject then object. As with picture naming, decoding accuracy tended to correspond to the time the training data came from, with models trained at earlier times detecting words farther in advance of their production in sentences. These findings are evidence for shared representations and processes across picture naming and active sentence production.

Intriguingly, applying this same procedure to passive sentences revealed a different pattern. While we again successfully decoded word identity with above chance accuracy (Fig 3C), across models both the subject and the object remained active throughout the entire sentence (Fig. 3D). This was driven by sustained representations in prefrontal cortex, which encoded not only words but also their position in the sentence, with inferior frontal gyrus (IFG) selectively encoding the subject and middle frontal gyrus (MFG) the object (Fig. 3E). In contrast, sensorimotor cortex, which encodes articulatory information, behaved as in active sentences, encoding words in the time window where the training data came from (Fig 3C).

Our findings reveal a previously uncharacterized division of labor within the language network. Sensorimotor cortex encodes lexical information robustly and in a task-agnostic way, while prefrontal regions are sensitive to syntactic structure, likely reflecting flexibility under varying task demands. Furthermore, we uncovered a spatial code within prefrontal cortex for syntactic role, with subjects encoded by IFG and objects encoded in MFG. We propose that the complex temporal dynamics of word processing in prefrontal cortex may impose a subtle processing pressure over the course of language evolution, offering a possible explanation for why nearly all the world's languages place subjects before objects [7,8].

Fig 1. Experimental procedure

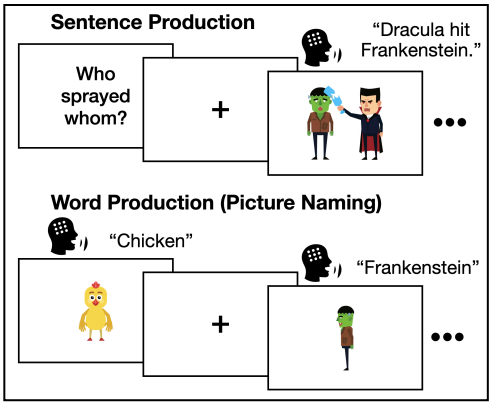
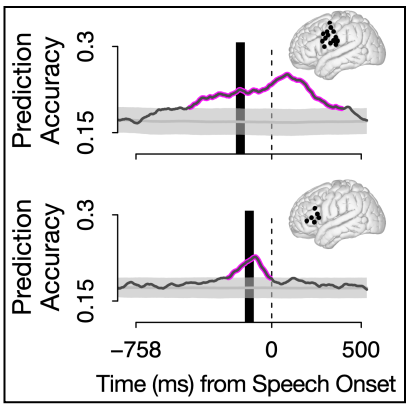


Fig 2. Results from two sample classifiers trained and tested on picture naming data. Brains show training electrodes (top: sensorimotor; bottom: inferior frontal). Grey distribution is chance accuracy (permutation). Pink highlights denote above-chance prediction accuracy. Black bar is the training data time window.



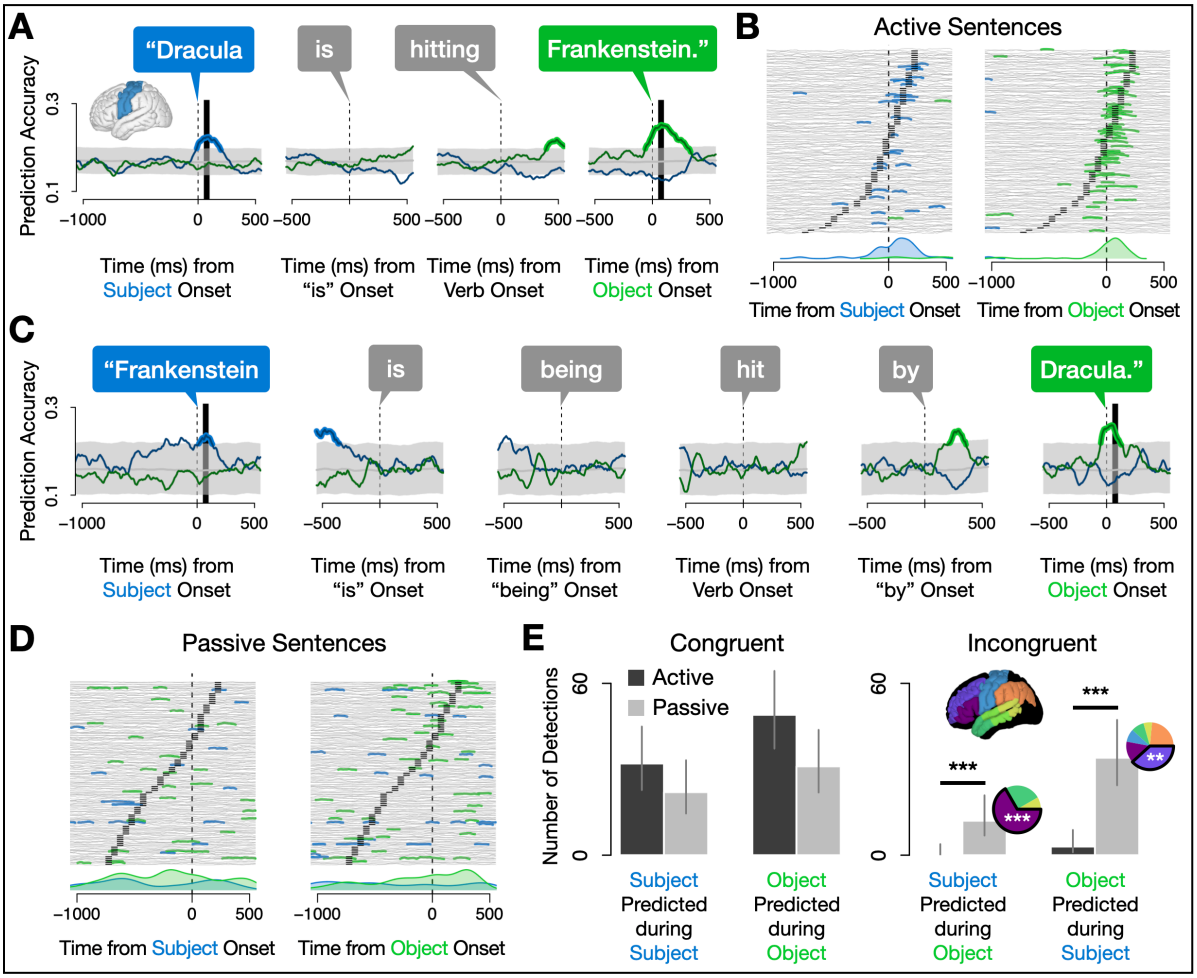


Fig 3. Results of decoding word identity during active (A-B) and passive (C-D) sentence production using classifiers trained on picture naming. (A,C) Sample decoding results from sensorimotor cortex. Blue/green highlights = significant subject/object decoding. (B,D) Prediction accuracies like in A and C, but for all significant classifiers, stacked vertically. Density plots of sig. windows are beneath each stack, revealing congruent temporal encoding of nouns in actives but sustained encoding throughout passives. (E) Sig. detections by structure and detection time. Passives had more incongruent detections than actives ($p<.001$), driven by inferior (violet) & middle (indigo) gyri.

References

[1] Dell, 1986 [2] Butterworth, 1989 [3] Levelt, Roelofs, Meyer, 1999 [4] Indefrey, Levelt, 2004 [5] Indefrey, 2011 [6] Flinker, Piai, Knight, 2018 [7] Dryer, 2005 [8] Hammarström, 2016