

From Thought to Action: How a Hierarchy of Neural Dynamics Supports Language Production

Mingfang (Lucy) Zhang^{2,3}, Jarod Lévy¹, Stéphane d'Ascoli¹, Jérémie Rapin¹, F.-Xavier Alario⁴, Pierre Bourdillon^{2,5}, Svetlana Pinet^{6,7*}, Jean-Rémi King^{1*}

¹Meta AI, ²Hospital Foundation Adolphe de Rothschild, ³École Normale Supérieure, Université PSL, CNRS, ⁴Aix Marseille Université, CNRS, CRPN, ⁵Paris Cité University, ⁶Basque Center on Cognition, Brain and Language, San Sebastián, ⁷Ikerbasque, Basque Foundation for Science, Bilbao

*Equal contribution

Humans effortlessly communicate their thoughts through intricate sequences of motor actions. Yet, the neural processes that coordinate language production remain largely unknown, in part because speech artifacts limit the use of neuroimaging. To elucidate the unfolding of language production in the brain, we investigate with magnetoencephalography (MEG) and electroencephalography (EEG) the neurophysiological activity of 35 skilled typists, while they typed sentences on a keyboard. This approach confirms the hierarchical predictions of linguistic theories: the neural activity preceding the production of each word is marked by the sequential rise and fall of context-, word-, syllable-, and letter-level representations. Remarkably, each of these neural representations is maintained over long time periods within each level of the language hierarchy. This phenomenon results in a superposition of successive representations that is supported by a hierarchy of dynamic neural codes. Overall, these findings provide a precise computational breakdown of the neural dynamics that coordinate the production of language in the human brain.

Date: February 5, 2025

Correspondence: lucy.zhang@psl.eu, s.pinet@cbcl.eu, jeanremi@meta.com



1 Introduction

Humans seamlessly produce intricate sequences of motor actions to communicate their thoughts. Such language production ability systematically requires a complex chaining of motor actions, whose underlying neural mechanisms remain poorly understood.

An extensive cortical network is involved in the production of speech (Indefrey, 2011; Price, 2012), typing (Pinet et al., 2019, 2016; Hu et al., 2023), writing (Planton et al., 2013; Cloutman et al., 2009), and signing (Blanco-Elorrieta et al., 2018; Emmorey et al., 2007). The associative cortices of the temporal and parietal regions primarily build and maintain linguistic features (Matchin et al., 2022; Fedorenko et al., 2024; Chang et al., 2022) in interaction with inferior frontal and supplementary motor areas (Bourguignon, 2014), while also projecting to pre-motor and motor cortices for action planning and execution (Hickok, 2012; Guenther, 2016). This network is thought to enable language production through a hierarchical process. According to this view, the representation of a sentence meaning would trigger an incremental retrieval and encoding of words, syllables, and letters or phonemes (Bock and Levelt, 1994; Dell et al., 1997; Levelt and Meyer, 2000; Logan and Crump, 2011; Roelofs, 1997; Rumelhart and Norman, 1982).

However, two main challenges have hindered the characterization of the timing, content, and coordination of the neural bases of natural language production. First, speech generates auditory feedback, which complicates the discrimination of perception and production processes, and has thus led to using artificial paradigms such as single word production and covert speech tasks (Salmelin et al., 1994; Carota et al., 2022). Second, speech production generates facial artifacts in functional Magnetic Resonance Imaging (fMRI), electro- (EEG) and magneto-encephalography (MEG) (Abbasi et al., 2021; Volfart et al., 2024; Vos et al., 2010). While this issue can be circumvented through the artifact-free yet sparsely-sampled intracranial recordings of epileptic patients, current methods do not readily allow a comprehensive whole-brain study of natural language production.

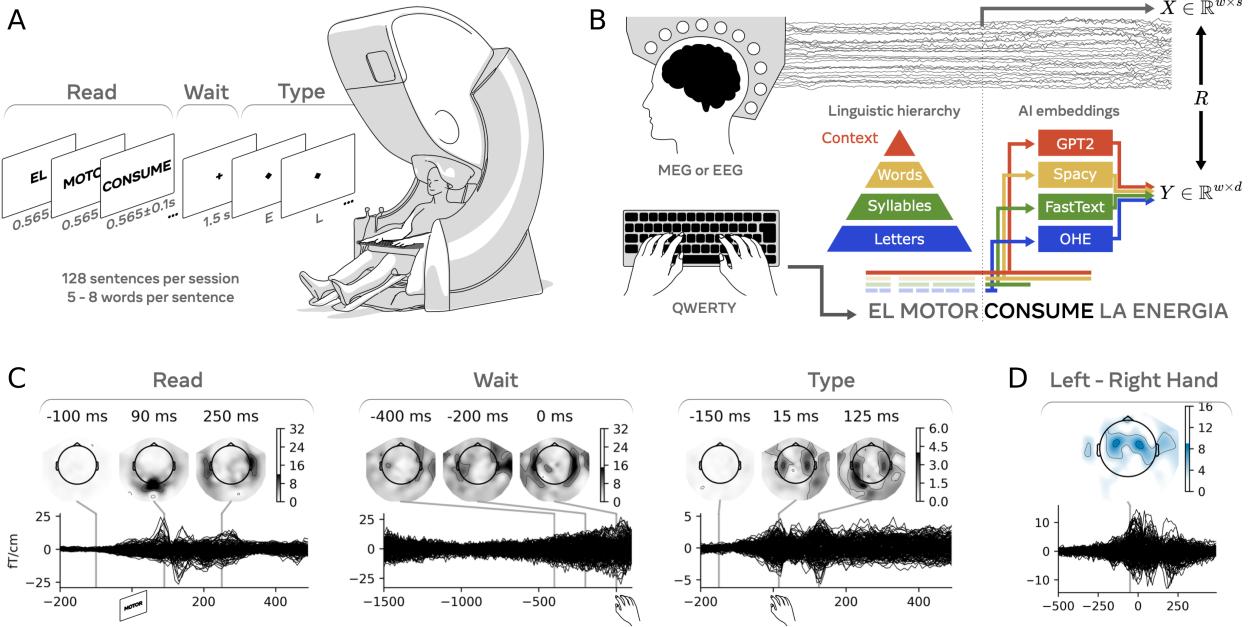


Figure 1 Experimental design and methods. **A.** We recorded the brain activity of healthy volunteers with either MEG or EEG during a read-wait-type task, in which subjects typed briefly memorized sentences. **B.** To investigate when and whether the brain produces a hierarchy of language representations, we linearly decoded, from these signals (X), the vectorial embeddings (Y) of four levels of representations: contextual word embeddings (with GPT-2), de-contextualized word embeddings (with Spacy), syllable embedding (with FastText) and letters (with a One-Hot-Encoder, OHE), and summarize the decoding scores with a Pearson correlation (R). **C.** Median MEG responses time-locked to word presentation (left), first keystroke (middle), and all keystrokes (right). **D.** Difference in MEG responses between keys typed with left vs right hands. A deep learning pipeline optimized to decode sentences from these data is available in our companion paper (Lévy et al., 2025)

To address these challenges, we evaluate here the neural bases of natural language production during typing. This approach presents two major advantages compared to speaking: (i) most of its sensory feedback (the letters on the screen) can be removed and (ii) muscle activity remains distant from M/EEG sensors. Thus, we recorded, with either MEG or EEG, the brain activity of 35 skilled typists while they typed briefly memorized sentences (Fig. 1A). To characterize the hierarchy of language representations generated in the brain during production, we rely on the classic definition of “neural representations” as information linearly decodable from neural activity (DiCarlo and Cox, 2007). To study language representations in the brain, we featurize letters, syllables, words, and contexts into simple vector representations or those from pre-trained language models (Fig. 1B). Finally, we evaluate, with linear decoding, whether and when these hierarchical representations rise and fall during sentence production, and summarize these decoding scores with either accuracy or a Pearson correlation (R). For clarity, and unless stated otherwise, we here focus on the MEG results. We direct the reader to the supplementary materials for the replication of these analyses on EEG data (see Section 6). To explore the possibility of efficiently decoding text from these brain signals, we refer the reader to our companion paper (Lévy et al., 2025).

2 Results

Task-related neural responses. Before decoding language representations from MEG signals, we first examined the average brain responses in the read, wait, and type phases in this task (Figure 1A). During the reading phase, word presentation elicits a posterior MEG response between 100 and 200 ms after their onset, followed by a bilateral and more anterior activation (Figure 1C, see EEG in Figure S2). Next, the wait period is marked by an increase in activation in the lateral MEG channels. Third, the brain response time-locked to keystrokes is marked by a bilateral motor topography, followed by a peak in the left-lateralized sensors.

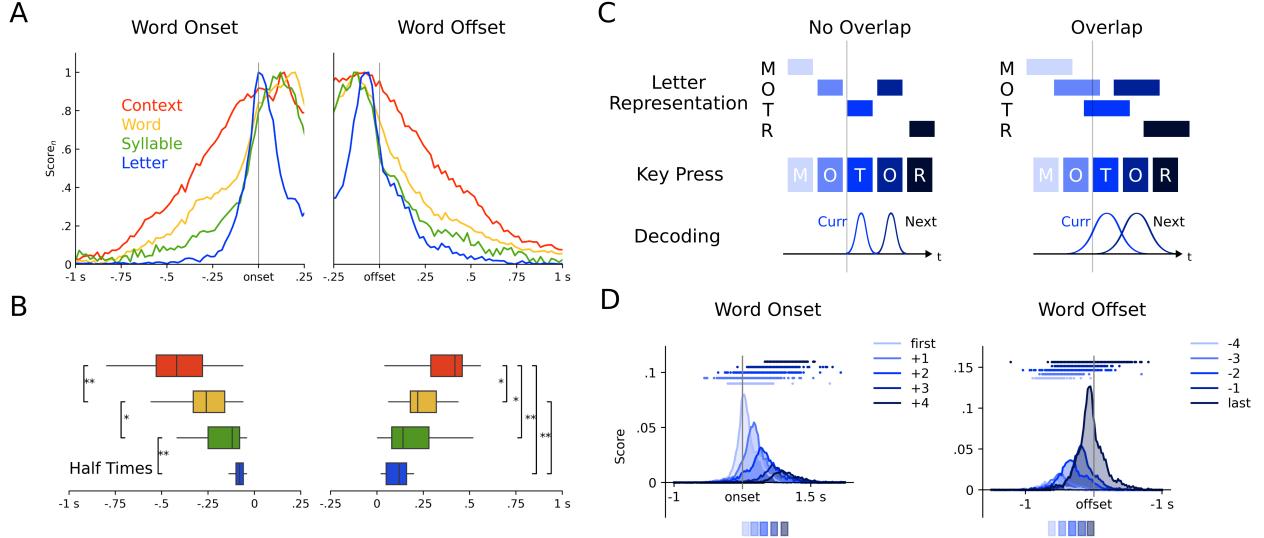


Figure 2 Decoding a hierarchy of linguistic features. **A.** Normalized decoding score of each linguistic feature time-locked to word onset and offset. **B.** Box plot of half-times (when the decoding reaches half of the maximum score) across participants. Asterisks, * and ** indicate significant ($p < 0.05$ and $p < 0.01$, respectively) between the half times of different representations. For clarity, we show skip-level comparisons only if adjacent-level tests are not consistently significant. **C.** Two hypotheses (left/right column) for the neural representations of successive letters. Top: each row corresponds to a putative neural activation of one letter representation. Middle: typing behavior. Bottom: putative decoding scores. **D.** Shifted time-decoding of successive letters relative to word onset or offset. Each curve corresponds to the decoding score at a letter position. Top lines indicate significant decoding scores ($p < 0.05$). Each rectangle below indicates the median start and duration of the corresponding key.

A similar effect can also be seen by comparing left- vs right-hand keystrokes (Figure 1D). Overall, these topographies are consistent with those found during language processing (reading network), the preparation of motor actions, and their actual execution (Munding et al., 2016).

Isolating the neural representations of individual key presses. Current theories predict that language production depends on a hierarchy of language representations. We thus hypothesize that: (1) different levels of representations will be linearly decodable from brain activity, and (2) higher-level representations will be detected before lower-level representations.

To test these hypotheses, we first evaluated whether each letter can be linearly decoded from the MEG activity relative to words. For this, we fitted a series of linear classifiers on all MEG channels at a single time sample relative to word onset and evaluated them on held-out MEG responses. We repeated this process for word-offset responses. Due to the class imbalance induced by the inherent differences in letter frequency, it can be difficult to interpret the chance level performance. For clarity, we thus report the difference in balanced accuracy between this linear decoder and a dummy classifier (chance level of 0%, details see Section 4). Decoding of the first letter of each word peaks around the corresponding key press with a performance of 8.1% (SEM across subjects: $\pm 0.6\%$), but is already significant from -220 ms (Figure S1). Qualitatively similar decoding is observed with EEG, although with a much lower decoding performance: Score=1.6% (SEM across subjects: $\pm 0.5\%$) (Figure S3).

Tracking higher-level representations of language. To extend this approach to higher linguistic features, we adapted this linear decoding analysis to syllables. However, the 288 distinct syllables present in this dataset make classification intractable. We thus embedded each syllable in FastText (Bojanowski et al., 2017), a character-level deep neural network trained on language modeling. This model transforms each syllable into a 300-dimensional vector, whose geometry represents the orthographic regularities of large text corpora. For clarity, we here report the decoding of the first syllable of each word relative to word onset. The resulting decoding shows that these syllable representations can be detected from brain activity up to -520 ms before

word onset, and ultimately peak at +120 ms with decoding of $R=0.10\pm0.01$. The corresponding EEG decoding can be found in [Figure S3](#).

Decomposing the hierarchy of linguistic features. We generalized this embedding approach to two additional levels of the language hierarchy. Word embedding, as provided by Spacy ([Montani et al., 2023](#)), represents words as 300-dimensional vectors such that words used in the same contexts are maximally similar. Contextual word embedding, as extracted from GPT2 ([Radford et al., 2019](#)), follows a similar principle for the *combination* of words and can thus embed context. To get causal representations, we here focused on the context up to the current word. Linear decoding shows that both of these levels of representations are represented in brain activity long before the corresponding word is typed. Decoding of word representations peaks with a $R=0.17\pm0.01\%$, while decoding of context embedding peaks with a $R=0.13\pm0.01\%$. The same results for EEG can be found in [Figure S3](#).

A top-down sequence of activations precedes word production. Theories and empirical findings consistently highlight that language production is supported by a hierarchy of representations. To explore the unfolding of this process, we evaluated whether the rise of each representation depends on its level in the language hierarchy. For this, we estimated, for each subject, the “half-time” of each neural representation – the time point where decoding reaches 50% of its maximal value within the window of focus. The Spearman correlation between half-times and linguistic levels ordered from context to letter is significant across subjects: $R=-0.77\pm0.06$, $p < 10^{-5}$. A similar, although noisier, finding is also replicated with EEG: $R=-0.58\pm0.07$, $p < 10^{-4}$ ([Figure S3](#)). The brain thus appears to produce language in a hierarchical fashion, by first generating representations at the level of context and then sequentially generating the word-, syllable-, and finally letter-level representations ([Figure 2A-B](#)).

Sequential deactivations of the language hierarchy. Although described at the key press level ([Kornysheva et al., 2019](#)), the deactivation of neural representations across the language hierarchy has received less attention. We therefore applied the same analysis presented previously at word offset. We find that the progressive deactivation of representations depends on their level in the language hierarchy ([Figure 2A](#)). Specifically, *after* word offset, the representations of letters go back to chance before those of syllables and words, while context representations are sustained over the longest time period. Overall, this leads to a Spearman correlation between the decay half-times and their linguistic levels of $R=0.64\pm0.10$, $p < 10^{-4}$ ([Figure 2B](#)).

Neural representations sustained for longer than their respective execution. Consistent with the motor execution literature, the decoding of each representation lasts for a relatively long time period ([Figure S1](#)) ([Averbeck et al., 2002](#); [Kornysheva et al., 2019](#); [Pinet et al., 2019](#)). For example, each keypress lasts, on average, for 103 ± 33 ms, but the corresponding neural representation is decodable from -480 ms to +640 ms relative to key press and release, respectively – *i.e.* 10x longer. A similar, although weaker, effect can also be observed for syllables (lasting for 297 ms, decodable from -480 to +900 ms, *i.e.* 4x longer) and words (lasting for 799 ms duration but decodable from -1.5 s to +1.38 s, *i.e.* at least 3x longer). This phenomenon suggests that at all levels of the language hierarchy, *successive* items are *simultaneously* represented at each instant.

Simultaneous representations of multiple keys during word production. To explicitly assess this overlap in terms of letter representations during continuous production, we trained linear classifiers to predict letters at successive positions starting from – or ending at – each word ([Figure 2C](#)). The results show that up to five successive letters can be simultaneously decoded from a single MEG time sample, revealing a superposition of their representations. This phenomenon appears to go over word boundaries: the first three keys can be decoded before word onset and the last three keys can be decoded after word offset ([Figure 2D](#)). Overall, these analyses confirm that multiple *successive* representations are *simultaneously* present in the brain activity, supporting the representational overlap hypothesis (see [Figure 2C](#)). This result generalizes previous work on serial order processes from two-letter words (e.g. [Pinet et al., 2019](#)) to a complex language production setting.

Letters are represented with a dynamic neural code. How does the brain avoid the interference induced by such overlapping representations? One hypothesis is dynamic neural coding. This coding scheme can be viewed as a positional embedding in the context of artificial neural networks ([Vaswani et al., 2017](#)). In practice,

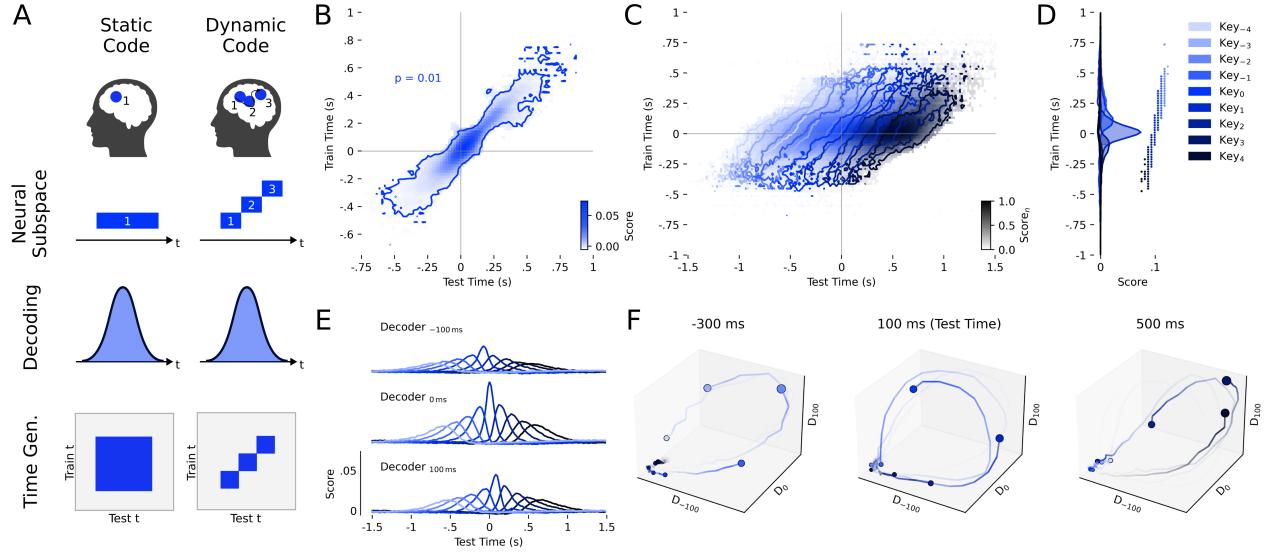


Figure 3 Dynamic code of letter representations. **A.** Static and dynamic codes may present similar decoding curves (middle) but different temporal generalization matrices (bottom). **B.** Temporal generalization of letter decoding. The contour indicates $p < 0.01$ across subjects. **C.** Shifted temporal generalization. Each color indicates a different letter relative to the current letter i . **D.** Scores of each decoder trained at each time t_i evaluated on successive letters (same data as the vertical slice of the temporal generalisation matrices at test time = 0 s in C). Vertical lines indicate statistical significance $p < 0.05$. **E.** Scores of three decoders fitted at -100 ms, 0 ms, and 100 ms on successive letters (i.e. three horizontal slices of the Time Gen. of panel C). **F.** Same as E, plotted as normalized 3D trajectories for three representative test times (-300, 100 and 500 ms). Each axis represents one decoder.

this consists of "moving" the representation across different neural subspaces. Accordingly, the availability of letter information remains constant over time, but *where* this information is represented in neural activity continuously changes (Figure 3A). To test this hypothesis, we implemented temporal generalization analysis (King and Dehaene, 2014) which consists of evaluating whether the neural representations decoded at a time sample effectively generalize to other time samples. If the neural code is stable, then a decoder trained at time t_i should generalize to time t_j . Conversely, if the neural code is dynamic, then the model trained at t_j , but not the model trained at t_i , would be able to decode at t_j . In this view, stable codes lead to square temporal generalization matrices, whereas dynamic codes lead to diagonal matrices (Figure 3A). The results show that letters are represented with a dynamic neural code. The temporal generalization matrix is marked by a diagonal pattern (Figure 3B). We can then evaluate whether each of these decoders can decode, not only the current letter, but also the preceding and the subsequent letters. The results show that successive letters lead to a sequence of "parallel diagonals" (Figure 3C), with multiple letter representations detectable through decoder generalization at the same time sample (Figure 3D). We observed a similar, although noisier, result with EEG signals (Figure S5).

Notably, each of these letter representations moves in a high-dimensional space. For example, we can focus on three decoders trained at -100 ms, 0 ms and +100 ms, respectively (Figure 3E), and visualize them as three dimensions of this space (Figure 3F). The decoding scores of successive letters can then be visualized over time as trajectories in this 3D space. At each instant, each letter is captured differentially by the three decoders, as the representations are positioned in different parts of this space. Consequently, this dynamic coding scheme allows a linear readout of individual letters without being interfered by the others.

A hierarchy of dynamical codes. Do the different levels of language representations follow the same dynamics (Figure 4A)? Here, we hypothesize that the *speed* of the representational dynamics will depend on their level in the language hierarchy. Specifically, context representations would "move" more slowly than letter representations. The temporal generalization matrices of contexts should thus be marked by "wider" diagonal patterns than those of letters (Figure 4A).

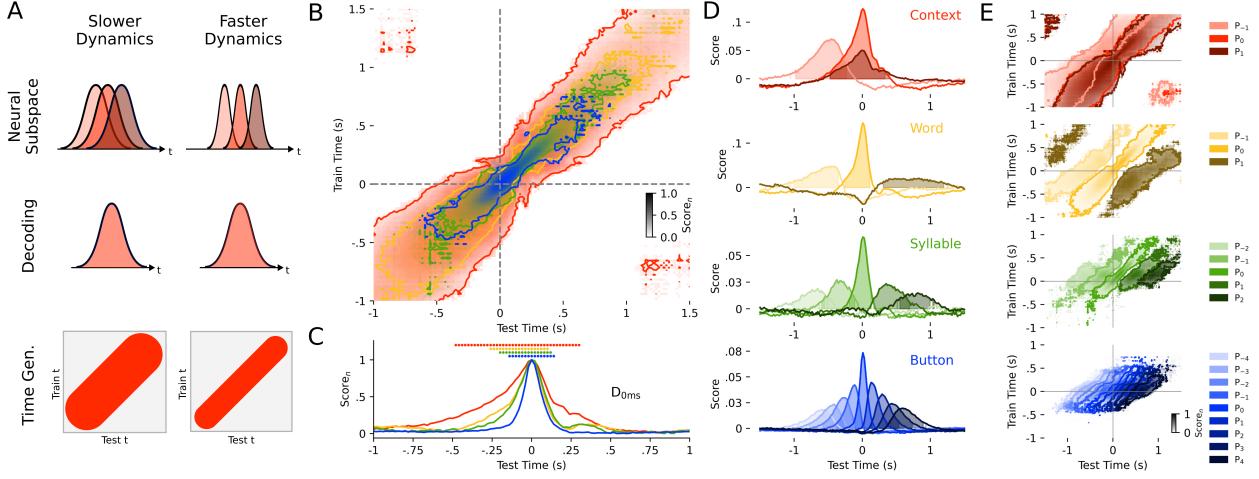


Figure 4 Dynamic neural code across the language hierarchy. **A.** Hypotheses of slow vs fast neural dynamics can lead to similar decoding curves but different temporal generalization matrices (King and Dehaene, 2014). **B.** Normalized temporal generalization of features time-locked at all feature onset. Contours indicate statistical significance across subjects $p < 0.01$. **C.** Normalized generalization scores of decoders from each linguistic level trained at 0 ms and tested over all time samples. Horizontal marks indicate statistical significance $p < 0.05$. **D.** Scores of decoder trained at 0 ms and evaluated across time for each preceding and subsequent feature (i.e. Horizontal slice of panel E). Shaded areas indicate statistical significance $p < 0.05$. **E.** Temporal generalization of previous, current, and next features for each level of the language hierarchy.

All these hypotheses are verified. First, each level of representation is marked by a diagonal pattern: *i.e.* the periods during which letters (1.1 s), syllables (1.4 s), words (2.8 s) and contexts (3 s) are decodable far exceed the periods during which each decoder generalizes (letter : 280 ms, syllable : 320 ms, word : 360 ms, context : 780 ms for decoder trained at 0 ms) (Figure 4B-C). This result confirms that all levels of representation are instantiated by a dynamic neural code. Second, the speed of these representational dynamics depends on their level in the language hierarchy. The generalization duration of the decoder trained at $t=0$ ms, quantified by the difference between the half-times before and after event onset (Section 4), significantly correlates with the level of the language hierarchy: Spearman $R=0.74\pm 0.06$, $p < 10^{-5}$. A similar, although noisier, correlation can be observed from EEG signals: Spearman $R=0.39\pm 0.11$, $p < 10^{-2}$ (Figure S6).

Finally, similarly to letters, these dynamical codes allow successive elements (e.g. syllables, words) to be simultaneously represented in brain activity without sharing the same neural subspaces (Figure 4D-E).

3 Discussion

Results summary. To explore the neural mechanisms coordinating language production, we measured the brain activity of 35 skilled typists with either MEG or EEG while they typed sentences on a keyboard. We find that four levels of language representations can be linearly decoded from these signals. The evolution of these representations follows the predictions of a top-down recruitment of the language hierarchy: before each word is typed, the brain generates context representations, followed by word, syllable, and finally letter representations (Figure 2B). While these representations overlap in time (Figure 2D), we show that they actually exist in separate subspaces of neural activity thanks to a dynamic code (Figure 3A-C, Figure 4).

Top-down activation of the language hierarchy. These results confirm the long-standing predictions that language production requires a hierarchical decomposition of sentence meaning into progressively smaller units that ultimately control motor actions (Chomsky, 1965; Indefrey, 2011; Logan and Crump, 2011). They also extend a growing literature on the temporal characteristics of language representations during production (Munding et al., 2016). For example, MEG results from Carota et al. (2022) suggest a progression from semantic to phonological computations before overt articulation in a picture naming task. Furthermore,

[Khanna et al. \(2024\)](#) recently showed with single-unit neural recordings that several levels of the language hierarchy are activated following a specific order in the posterior frontal gyrus during speech. Our study complements these findings with, here, (i) a complex language production task that reduces sensory confounds, and (ii) a systematic decomposition of the whole-brain activity into a hierarchy of language representations. Interestingly, the reported *top-down* sequence of language representations mirrors the *bottom-up* construction of representations observed with fMRI, M/EEG, and intracranial electrophysiology during language perception ([Millet et al., 2023; Li et al., 2023; Caucheteux and King, 2021; Goldstein et al., 2022](#)). However, the question of how closely these perception and production hierarchies share the same neural codes remains open ([Lin and Hsieh, 2022; Fairs et al., 2022](#)).

A hierarchy of dynamical neural codes. The hierarchy of neural dynamics revealed in this study provides a clear mechanism to simultaneously represent successive features and actions, while avoiding their mutual interference. Our results show that each representation is continuously moving across different subspaces of neural activity – this specific dynamics thus allows different linear read-outs to target a specific representation in each sequence. Critically, we observe that the speed of these dynamics depends on their level in the language hierarchy: at the top level, the representations of context slowly move as the sentence unfolds, whereas at the lowest level, the representations of letters move quickly. This specific neural code thus allows the brain to both (i) represent a rich set of linguistic features and (ii) coordinate their chained execution.

Related works. This dynamic neural code echos recent findings on the neural bases of speech comprehension, which represent *perceived* words and phrases more slowly than those of syllables and phonemes ([Gwilliams et al., 2024](#)). Such temporal characterization of representations enriches previous discussions focused on processing stages in word production ([Munding et al., 2016](#)). Furthermore, this dynamic neural code has been repeatedly observed in the motor literature. For example, [Churchland et al. \(2012\)](#), [Russo et al. \(2020\)](#), and [Crowe et al. \(2014\)](#) showed that the neural activity of the primate motor cortex follows dynamic trajectories during the execution of a simple motor task. Together, these findings thus suggest that the hierarchical organization of neural dynamics may provide a canonical mechanism to coordinate rapid sequences of representations.

Anatomical localization. [Indefrey \(2011\)](#) suggests that word retrieval primarily depends on the middle temporal gyrus (MTG), whereas the preparation of phonological and syllabic representations takes place in posterior temporo-parietal and inferior frontal gyri ([Hickok, 2014; Guenther, 2016](#)). While the present M/EEG study is unsuitable to pinpoint the exact brain location of language representations, it complements these seminal anatomical findings by clarifying the temporal unfolding of the neural representations of natural language production. Nevertheless, the full spatio-temporal properties of this anatomico-functional hierarchy remains to be fully characterized ([Price, 2012; Indefrey, 2011; Hickok, 2012](#)).

The modalities of language production. The present typing paradigm was primarily motivated to circumvent the artifacts and confounds generated by natural speech. However, the architecture of the human brain did not evolve to type on a QWERTY keyboard. This design choice thus leaves several questions open. First, the present study does not address whether the letter representations correspond to abstract linguistic features or rather to keyboard-specific motor actions. Second, whether the neural organization revealed here applies *universally* across the different modalities of human communication remains to be resolved.. Nevertheless, beyond its empirical findings, this study provides the analytical and conceptual tools to unravel the neural and computational mechanisms that orchestrate the transformations of thoughts into actions.

4 Method

4.1 Protocol and preprocessing

Participants. We recruited 35 healthy volunteers at the Basque center on Cognition, Brain and Language (BCBL) (77% female, age 31.6 ± 5.2). Twenty joined MEG sessions and twenty joined EEG sessions (five joined both). To ensure that the participants could type without looking at the keyboard, we only selected native Spanish speakers who scored more than 80% accuracy in a dictation typing task, where the keyboard was covered with a cardboard box. Each participant provided informed consent for this study and received a 12

euro/hour compensation. The study was approved by the local ethics committee. Data from one participant was excluded from the MEG data analysis due to the presence of a metallic component detected during the MEG recording.

Task. Each trial consisted of three phases ([Figure 1A](#)). First, a sentence was presented on the screen through a Rapid Serial Visual Presentation (RSVP) protocol, where words were shown for a random duration between 465-665 ms. Second, participants were tasked to memorize that sentence for a delay period of 1.5 s, during which only a fixation cross was presented on the screen. Finally, participants started typing the sentence from memory once the fixation cross disappeared from the screen. During typing, only a small black square was presented on the screen, rotating 10 degrees clockwise on every keystroke. This helped signal a successful key press without presenting the key input and thus ensured that participants minimized eye movements. Trials were separated by a 1 s interval. Breaks were introduced every 16 trials. Each participant took part in multiple sessions (from 2 to 6), with each session consisting of two blocks of 64 trials. To familiarize participants with the protocol, four training trials were performed at the beginning of each block. Data from these trials was excluded from further analyses. On average, the reading and typing phases lasted for 5.54 s and 4.70 s, respectively.

Stimuli. To balance language complexity and repetition, we designed 128 unique declarative sentences, which varied in length (5 - 8 words), meaning, and syntactic structures. All sentences contained a subject, a verb, and an object. The subject and object were a minimal noun phrase formed by a determiner and a noun in singular or plural form. The noun phrase could be supplemented with an adjective (e.g. "the data"/"the general data") or a prepositional phrase (e.g. "the data"/"the data of the transaction"). All stimuli were presented in black on a 50% gray background. During reading, all letters were capitalized and did not contain any diacritic marks. During typing, participants were instructed to type without diacritic marks as if they were typing in upper case. They were also discouraged from using the backspace key.

Keyboard. Standard keyboards contain electronics and metallic parts that lead to major MEG artifacts. To address this issue, we adapted an MR-compatible keyboard from HybridMojo (LLC) by changing the keys supported by a spring (shift, spacebar, backspace) with a custom non-ferromagnetic silver-spring mechanism. During the task, the keyboard was roughly 70 cm away from the top of the helmet of the MEG scanner.

MEG. MEG recordings were acquired with a 306 MEGIN system (102 magnetometers and 204 planar gradiometers) at a sampling rate of 1 kHz, with an online high-pass filter set at 0.1 Hz and a low-pass filter at 330 Hz. Electrooculograms (EOG) and electrocardiograms (ECG) were recorded through auxiliary channels. Five head position index (HPI) coils were used for tracking head position. The location of each coil relative to the anatomical fiducials (nasion, left and right preauricular points) was defined with a 3D digitizer (Fastrak Polhemus, Colchester, US). HPI were acquired at the beginning of each session. MEG signals were bandpass filtered between 0.5 Hz and 45 Hz with MNE-python's default parameters and downsampled to 50 Hz ([Gramfort, 2013](#)). For [Figure 1C-D](#), we plot the baseline-corrected (-250 ms-0 ms), maxfiltered signal-space-projected (SSP) gradiometers median-averaged across subjects, except for the wait panel where no baseline correction is applied to visualize the activity during this period.

EEG. EEG recordings were acquired with an actiCAP slim system from BrainVision at a sampling rate of 1 kHz, with an online high-pass filter set at 0.02 Hz. A total of 61 EEG channels were recorded together with 3 ocular channels. We applied the same preprocessing methods as described in the MEG section on EEG signals.

Events construction. Language can be described at different levels of representations (letters, syllables, words, sentences). We defined an event's onset as the press of its first key and its offset as the release of its last key (i.e. syllable event "con" starts when "c" is pressed and ends when "n" is released, [Figure 1B](#)). To omit typographical errors from the typed sentences, we computed the Levenshtein distance between the true and typed sentences and kept the subset of keystrokes that matched the true sentence. Boundaries between typed words are given by space keys while boundaries between syllables are defined by manual annotation. Space

keys and mistyped syllables are excluded from subsequent analyses. Epochs are time-locked either to the onset of words or the onset of all features in the language hierarchy.

4.2 Language representations

We featurized each level of the language hierarchy into target vector representations (Figure 1B). All features are generated from text strings free from diacritic marks.

Context representations: GPT2. We used GPT2’s hidden activations (Radford et al., 2019) to represent the past context of each word. Specifically, for a target i^{th} word in the sentence, we tokenized the text up to and including the current word ($0-i^{th}$ word) with the GPT2 tokenizer and input this context to an inference-only pre-trained GPT2 model as provided by HuggingFace. We defined the context representation as the activation of the i^{th} word in the 8^{th} transformer layer ($d=768$), similar to Caucheteux and King (2022). If a word is represented by multiple tokens, we averaged the activations corresponding to the tokens of the target word.

Word representations: Spacy. To generate word-level representations, we used a pre-trained Spanish language model (es_core_news_lg) from the Spacy library (Montani et al., 2023), leading to a $d = 300$ embedding of each word.

Syllable representations: FastText. To obtain vectorized representations of sublexical strings, we used FastText – a subword embedding model introduced by Bojanowski et al. (2017). We trained a custom FastText model using the Spanish wikidump corpus (date July 19th, 2024)(Joulin et al., 2016; Wikidumps). The corpus was cleaned and preprocessed using the wp2txt pipeline (Hasebe, 2023). The model was trained with default hyperparameters, except for n-gram length which ranged between 1 to 5 and a dimension size set to $d = 300$. Each syllable was then extracted from the library of n-gram vectors.

Letter representations: One-Hot-Encoder. For the letter representations, we simply use a one-hot-embedding of the letter identities.

4.3 Decoding analyses

We deployed four variants of decoding analyses, each addressing a distinct question about the temporal evolution of neural representations.

Time decoding. To assess when a feature is represented, we conducted decoding analyses over time relative to each event. Given $X \in \mathbb{R}^{w \times s \times t}$, with s being the number of M/EEG sensors time-locked to the onset or the offset of w words over t time samples, and their corresponding target vectors y , we fitted and evaluated a linear decoder at each time sample t .

Shifted time decoding. The above analyses decode y_i time-locked to event i . To evaluate the overlap of successive representations, we can extend these analyses to decode y_j time-locked to event i . Concretely, we trained and evaluated a series of decoders to predict y_j relative to t_i .

Temporal generalization. To assess the stability of neural representations, we conducted temporal generalization analyses (King and Dehaene, 2014). This consists of training a decoder $d_{t \rightarrow t'}$ on neural signals X at time t and testing it on t' . If a neural representation is stable, then $d_{t \rightarrow t'}$ should lead to decoding performances similar to $d_{t \rightarrow t}$. By contrast, if the neural code is dynamic, meaning that the neural representation of a feature changes over time, then $d_{t \rightarrow t'}$ should perform worse than $d_{t \rightarrow t}$.

Shifted temporal generalization. Temporal generalization can be further extended to assess the decoder’s generalization of successive features. For this, we evaluated whether each time-decoder (trained with y_i at t_i) effectively decodes, on a test set, the preceding or subsequent features y_j .

Decoding models. We used a ridge linear regression (RidgeCV) as implemented in scikit-learn (Pedregosa et al., 2011), with an alpha regularization (alpha_per_target=True) spanning 50 log-spaced values between 10^{-1} and 10^{10} with default parameters. Neural signals are standardized (zero-meaned and scaled to unit variance) per channel before model fitting.

Train-test procedure. We trained and tested each decoder in a within-subject fashion. We used a group k-fold ($K=5$) cross-validation to assess decoding performance, where the group of each feature is assigned based on unique sentences. Groups are split into a train set (80%) and a test set (20%).

Evaluation metrics. Unless stated otherwise, we evaluated decoding performance with a Pearson correlation between the true and the decoded features averaged across d . For letter decoding, we report the difference in the balanced accuracy score, which is the average of recall measured in each class, between a true model and a dummy model that predicts the most frequent class label. Each representation of language does not have the same space or dimensionality. Consequently, we compared the relative evolution of decoding performances $Score_n$ with a “normalized decoding performance”: *i.e.* $Score_n = \frac{Score - min(Score)}{max(Score) - min(Score)}$, where min and max are applied over time.

Half-times. To assess when each representation rise relative to event onset and offset, we computed the ‘half-time’ of decoding scores, defined as the moment when decoding performance reaches 50% of the maximum decoding score within the window of interest per participant. We focused on the $[-1, 0]$ s window prior to feature onset and $[0, 1]$ s after feature offset.

Significance testing. To test whether the decoding scores are significantly different from chance, we used the Wilcoxon signed rank test implemented in the Scipy library (Virtanen et al., 2020) across subjects. In the case of multiple comparisons, we further adjusted the p-values with a False Detection Rate (FDR) as implemented in MNE (Gramfort, 2013).

Companion paper. We explore the possibility to efficiently decode text from these brain signals in a companion paper (Lévy et al., 2025). Note that the A and D panels of Fig. 1 are shared between these two studies.

5 Acknowledgements

The authors would like to thank Maite Kaltzakorta, Manex Lete, Jessi Jacobsen, Daniel Nieto, Jone Iraeta, Araitz Garnika, Jaione Bengoetxea, Natalia Louleli, Naroa Miralles, Eñaut Zeberio, Craig Richter, Amets Esnal, and Olatz Andonegui. This research is supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033. FXA was supported within the Institute of Convergence ILCB by grants from France 2030 (ANR-16-CONV-0002) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). Parts of this research were carried within the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945304 - Cofund AI4theSciences hosted by PSL University.

References

- Omid Abbasi, Nadine Steingräber, and Joachim Gross. Correcting MEG Artifacts Caused by Overt Speech. *Frontiers in Neuroscience*, 15:682419, June 2021. ISSN 1662-453X. doi: 10.3389/fnins.2021.682419. <https://www.frontiersin.org/articles/10.3389/fnins.2021.682419/full>.
- Bruno B. Averbbeck, Matthew V Chafee, David A. Crowe, and Apostolos P Georgopoulos. Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):13172–13177, 2002. ISSN 00278424. doi: 10.1073/pnas.162485599. ISBN: 0027-8424.
- Esti Blanco-Elorrieta, Karen Emmorey, and Liina Pylkkänen. Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences*, 115(39):9708–9713, September 2018. doi: 10.1073/pnas.1809779115. <https://www.pnas.org/doi/10.1073/pnas.1809779115>. Publisher: Proceedings of the National Academy of Sciences.
- Kathryn Bock and Willem Levelt. Language production: Grammatical encoding. In M.A. Gernsbacher, editor, *Handbook of psycholinguistics*, pages 945–984. Academic Press, San Diego, 1994.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching Word Vectors with Subword Information, June 2017. <http://arxiv.org/abs/1607.04606>. arXiv:1607.04606 [cs].
- Nicolas J. Bourguignon. A rostro-caudal axis for language in the frontal lobe: the role of executive control in speech production. *Neuroscience and Biobehavioral Reviews*, 47:431–444, November 2014. ISSN 1873-7528. doi: 10.1016/j.neubiorev.2014.09.008.
- Francesca Carota, Jan-Mathijs Schoffelen, Robert Oostenveld, and Peter Indefrey. The Time Course of Language Production as Revealed by Pattern Classification of MEG Sensor Data. *The Journal of Neuroscience*, 42(29): 5745–5754, July 2022. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1923-21.2022. <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1923-21.2022>.
- Charlotte Caucheteux and Jean-Rémi King. Language processing in brains and deep neural networks: computational convergence and its limits, January 2021. <https://www.biorxiv.org/content/10.1101/2020.07.03.186288v2>. Pages: 2020.07.03.186288 Section: New Results.
- Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications Biology*, 5(1):1–10, February 2022. ISSN 2399-3642. doi: 10.1038/s42003-022-03036-1. <https://www.nature.com/articles/s42003-022-03036-1>. Publisher: Nature Publishing Group.
- Claire H. C. Chang, Samuel A. Nastase, and Uri Hasson. Information flow across the cortical timescale hierarchy during narrative construction. *Proceedings of the National Academy of Sciences*, 119(51):e2209307119, December 2022. doi: 10.1073/pnas.2209307119. <https://www.pnas.org/doi/full/10.1073/pnas.2209307119>. Publisher: Proceedings of the National Academy of Sciences.
- Noam Chomsky. *Aspects of the theory of syntax*. Aspects of the theory of syntax. M.I.T. Press, Oxford, England, 1965.
- Mark M. Churchland, John P. Cunningham, Matthew T. Kaufman, Justin D. Foster, Paul Nuyujukian, Stephen I. Ryu, and Krishna V. Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, July 2012. ISSN 1476-4687. doi: 10.1038/nature11129. <https://www.nature.com/articles/nature11129>. Publisher: Nature Publishing Group.
- Lauren Cloutman, Leila Gingis, Melissa Newhart, Cameron Davis, Jennifer Heidler-Gary, Jennifer Crinion, and Argye E. Hillis. A Neural Network Critical for Spelling. *Annals of neurology*, 66(2):249–253, August 2009. ISSN 0364-5134. doi: 10.1002/ana.21693. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745512/>.
- David A. Crowe, Wilbert Zarco, Ramon Bartolo, and Hugo Merchant. Dynamic Representation of the Temporal and Sequential Structure of Rhythmic Movements in the Primate Medial Premotor Cortex. *Journal of Neuroscience*, 34(36):11972–11983, September 2014. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2177-14.2014. <https://www.jneurosci.org/content/34/36/11972>. Publisher: Society for Neuroscience Section: Articles.
- Gary S. Dell, Lisa K. Burger, and William R. Svec. Language production and serial order: A functional analysis and a model. *Psychological Review*, 104(1):123–147, January 1997. ISSN 1939-1471, 0033-295X. doi: 10.1037/0033-295X.104.1.123. <https://doi.apa.org/doi/10.1037/0033-295X.104.1.123>.
- James J. DiCarlo and David D. Cox. Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8): 333–341, August 2007. ISSN 13646613. doi: 10.1016/j.tics.2007.06.010. <https://linkinghub.elsevier.com/retrieve/pii/S1364661307001593>.

Karen Emmorey, Sonya Mehta, and Thomas J. Grabowski. The neural correlates of sign versus word production. *NeuroImage*, 36(1):202–208, May 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.02.040.

Amie Fairs, Raphaël Fargier, and Kristof Strijkers. Shared or different: How linked are word production and comprehension? *TIPA. Travaux interdisciplinaires sur la parole et le langage*, (38), December 2022. ISSN 1621-0360. doi: 10.4000/tipa.4879. <https://journals.openedition.org/tipa/4879#tocto1n3>. Number: 38 Publisher: Laboratoire parole et langage.

Evelina Fedorenko, Anna A. Ivanova, and Tamar I. Regev. The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*, pages 1–24, April 2024. ISSN 1471-0048. doi: 10.1038/s41583-024-00802-4. <https://www.nature.com/articles/s41583-024-00802-4>. Publisher: Nature Publishing Group.

Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A. Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, Aren Jansen, Harshvardhan Gazula, Gina Choe, Aditi Rao, Catherine Kim, Colton Casto, Lora Fanda, Werner Doyle, Daniel Friedman, Patricia Dugan, Lucia Melloni, Roi Reichart, Sasha Devore, Adeen Flinker, Liat Hasenfratz, Omer Levy, Avinatan Hassidim, Michael Brenner, Yossi Matias, Kenneth A. Norman, Orrin Devinsky, and Uri Hasson. Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3):369–380, March 2022. ISSN 1546-1726. doi: 10.1038/s41593-022-01026-4. <https://www.nature.com/articles/s41593-022-01026-4>. Publisher: Nature Publishing Group.

Alexandre Gramfort. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013. ISSN 1662453X. doi: 10.3389/fnins.2013.00267. <http://journal.frontiersin.org/article/10.3389/fnins.2013.00267/abstract>.

Frank H. Guenther. *Neural Control of Speech*. The MIT Press, July 2016. ISBN 978-0-262-33698-7. doi: 10.7551/mitpress/10471.001.0001. <https://direct.mit.edu/books/monograph/4085/Neural-Control-of-Speech>.

Laura Gwilliams, Alec Marantz, David Poeppel, and Jean-Remi King. Hierarchical dynamic coding coordinates speech comprehension in the brain, April 2024. <https://www.biorxiv.org/content/10.1101/2024.04.19.590280v1>. Pages: 2024.04.19.590280 Section: New Results.

Yoichiro Hasebe. WP2TXT: A command-line toolkit to extract text content and category data from Wikipedia dump files, 2023.

Gregory Hickok. Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2):135–145, February 2012. ISSN 1471-0048. doi: 10.1038/nrn3158. <https://www.nature.com/articles/nrn3158>. Publisher: Nature Publishing Group.

Gregory Hickok. The architecture of speech production and the role of the phoneme in speech processing. *Language and cognitive processes*, 29(1):2–20, January 2014. ISSN 0169-0965. doi: 10.1080/01690965.2013.834370. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3904400/>.

Jennifer Hu, Hannah Small, Hope Kean, Atsushi Takahashi, Leo Zekelman, Daniel Kleinman, Elizabeth Ryan, Alfonso Nieto-Castañón, Victor Ferreira, and Evelina Fedorenko. Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex*, 33(8): 4384–4404, April 2023. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/bhac350. <https://academic.oup.com/cercor/article/33/8/4384/6706753>.

Peter Indefrey. The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, 2, 2011. ISSN 1664-1078. doi: 10.3389/fpsyg.2011.00255. <http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00255/abstract>.

Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hervé Jégou, and Tomas Mikolov. FastText.zip: Compressing text classification models, December 2016. <http://arxiv.org/abs/1612.03651>. arXiv:1612.03651 [cs].

Arjun R. Khanna, William Muñoz, Young Joon Kim, Yoav Kfir, Angelique C. Paulk, Mohsen Jamali, Jing Cai, Martina L. Mustroph, Irene Caprara, Richard Hardstone, Mackenna Mejell, Domokos Meszéna, Abigail Zuckerman, Jeffrey Schweitzer, Sydney Cash, and Ziv M. Williams. Single-neuronal elements of speech production in humans. *Nature*, 626(7999):603–610, February 2024. ISSN 1476-4687. doi: 10.1038/s41586-023-06982-w.

J-R. King and S. Dehaene. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*, 18(4):203–210, April 2014. ISSN 1364-6613. doi: 10.1016/j.tics.2014.01.002. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5635958/>.

Katja Kornysheva, Daniel Bush, Sofie S. Meyer, Anna Sadnicka, Gareth Barnes, and Neil Burgess. Neural Competitive Queuing of Ordinal Structure Underlies Skilled Sequential Action. *Neuron*, 101(6):1166–1180.e3, March 2019. ISSN 0896-6273. doi: 10.1016/j.neuron.2019.01.018. [https://www.cell.com/neuron/abstract/S0896-6273\(19\)30045-5](https://www.cell.com/neuron/abstract/S0896-6273(19)30045-5). Publisher: Elsevier.

Willem J. M. Levelt and Antje S. Meyer. Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, December 2000. doi: 10.1080/095414400750050178. <https://www.tandfonline.com/doi/abs/10.1080/095414400750050178>. Publisher: Taylor & Francis Group.

Yuanning Li, Gopala K. Anumanchipalli, Abdelrahman Mohamed, Peili Chen, Laurel H. Carney, Junfeng Lu, Jinsong Wu, and Edward F. Chang. Dissecting neural computations in the human auditory pathway using deep neural networks for speech. *Nature Neuroscience*, 26(12):2213–2225, December 2023. ISSN 1546-1726. doi: 10.1038/s41593-023-01468-4. <https://www.nature.com/articles/s41593-023-01468-4>. Publisher: Nature Publishing Group.

Yi Lin and Po-Jang Hsieh. Neural decoding of speech with semantic-based classification. *Cortex*, 154:231–240, September 2022. ISSN 0010-9452. doi: 10.1016/j.cortex.2022.05.018. <https://www.sciencedirect.com/science/article/pii/S0010945222001666>.

Gordon D. Logan and Matthew J.C. Crump. Hierarchical Control of Cognitive Processes. In *Psychology of Learning and Motivation*, volume 54, pages 1–27. Elsevier, 2011. ISBN 978-0-12-385527-5. doi: 10.1016/B978-0-12-385527-5.00001-2. <https://linkinghub.elsevier.com/retrieve/pii/B9780123855275000012>.

Jarod Lévy, Mingfang(Lucy) Zhang, Svetlana Pinet, Jérémie Rapin, Hubert Banville, Stéphane d’Ascoli, and Jean-Rémi King. Brain-to-Text Decoding: A Non-invasive Approach via Typing. 2025.

William Matchin, Alexandra Basilakos, Dirk-Bart den Ouden, Brielle C. Stark, Gregory Hickok, and Julius Fridriksson. Functional differentiation in the language network revealed by lesion-symptom mapping. *NeuroImage*, 247:118778, February 2022. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2021.118778. <https://www.sciencedirect.com/science/article/pii/S1053811921010508>.

Juliette Millet, Charlotte Caucheteux, Pierre Orhan, Yves Boubenec, Alexandre Gramfort, Ewan Dunbar, Christophe Pallier, and Jean-Remi King. Toward a realistic model of speech processing in the brain with self-supervised learning, March 2023. <http://arxiv.org/abs/2206.01685>. arXiv:2206.01685 [q-bio].

Ines Montani, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, and Henning Peters. explosion/spaCy: v3.7.2: Fixes for APIs and requirements, October 2023. <https://zenodo.org/doi/10.5281/zenodo.1212303>.

Dashiel Mundig, Anne-Sophie Dubarry, and F.-Xavier Alario. On the cortical dynamics of word production: a review of the MEG evidence. *Language, Cognition and Neuroscience*, 31(4):441–462, April 2016. ISSN 2327-3798. doi: 10.1080/23273798.2015.1071857. <https://www.tandfonline.com/doi/full/10.1080/23273798.2015.1071857>. Publisher: Routledge.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. ISSN 1533-7928. <http://jmlr.org/papers/v12/pedregosa11a.html>.

Svetlana Pinet, Anne-Sophie Dubarry, and F.-Xavier Alario. Response retrieval and motor planning during typing. *Brain and Language*, 159:74–83, August 2016. ISSN 0093934X. doi: 10.1016/j.bandl.2016.05.012. <https://linkinghub.elsevier.com/retrieve/pii/S0093934X15301802>.

Svetlana Pinet, Gary S. Dell, and F.-Xavier Alario. Tracking Keystroke Sequences at the Cortical Level Reveals the Dynamics of Serial Order Production. *Journal of Cognitive Neuroscience*, 31(7):1030–1043, July 2019. ISSN 0898-929X. doi: 10.1162/jocn_a_01401. https://doi.org/10.1162/jocn_a_01401.

Samuel Planton, Mélanie Jucla, Franck-Emmanuel Roux, and Jean-François Démonet. The “handwriting brain”: A meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex*, 49(10):2772–2787, November 2013. ISSN 0010-9452. doi: 10.1016/j.cortex.2013.05.011. <https://www.sciencedirect.com/science/article/pii/S0010945213001391>.

Cathy J. Price. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2):816–847, August 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.04.062. <https://www.sciencedirect.com/science/article/pii/S1053811912004703>.

Alec Radford, Jeff Wu, R. Child, D. Luan, Dario Amodei, and I. Sutskever. Language Models are Unsupervised Multitask Learners. OpenAI blog, 2019.

Ardi Roelofs. The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3):249–284, September 1997. ISSN 00100277. doi: 10.1016/S0010-0277(97)00027-9. <https://linkinghub.elsevier.com/retrieve/pii/S0010027797000279>.

David E. Rumelhart and Donald A. Norman. Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance. *Cognitive Science*, 6(1):1–36, 1982. ISSN 1551-6709. doi: 10.1207/s15516709cog0601_1. https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0601_1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0601_1.

Abigail A. Russo, Ramin Khajeh, Sean R. Bittner, Sean M. Perkins, John P. Cunningham, L. F. Abbott, and Mark M. Churchland. Neural Trajectories in the Supplementary Motor Area and Motor Cortex Exhibit Distinct Geometries, Compatible with Different Classes of Computation. *Neuron*, 107(4):745–758.e6, August 2020. ISSN 0896-6273. doi: 10.1016/j.neuron.2020.05.020. <https://www.sciencedirect.com/science/article/pii/S0896627320303664>.

Riitta Salmelin, R Hari, O V Loumasmaa, and Mikko Sams. Dynamics of brain activation during picture naming. *Nature*, 368:463–465, 1994.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fb0d053c1c4a845aa-Abstract.html.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. Van Der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul Van Mulbregt, SciPy 1.0 Contributors, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T Webber, Janko Slavić, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius De Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmeler, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0686-2. <https://www.nature.com/articles/s41592-019-0686-2>.

Angelique Volfart, Katie L. McMahon, and Greig I. De Zubicaray. A Comparison of Denoising Approaches for Spoken Word Production Related Artefacts in Continuous Multiband fMRI Data. *Neurobiology of Language*, 5(4):901–921, September 2024. ISSN 2641-4368. doi: 10.1162/nol_a_00151. <https://direct.mit.edu/nol/article/5/4/901/122510/A-Comparison-of-Denoising-Approaches-for-Spoken>.

De Maarten Vos, Stephanie Riès, Katrien Vanderperren, Bart Vanrumste, Francois-Xavier Alario, Van Sabine Huffel, and Boris Burle. Removal of Muscle Artifacts from EEG Recordings of Spoken Language Production. *Neuroinformatics*, 8(2):135–150, June 2010. ISSN 1539-2791, 1559-0089. doi: 10.1007/s12021-010-9071-0. <http://link.springer.com/10.1007/s12021-010-9071-0>.

Wikidumps. Wikimedia Downloads. <https://dumps.wikimedia.org/>.

6 Supplementary Information

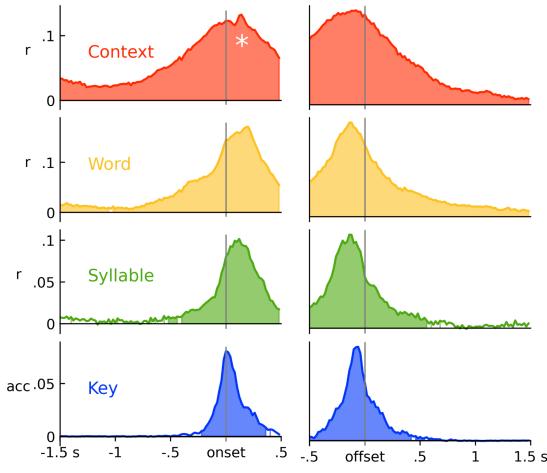


Figure S1 MEG raw decoding score of the language hierarchy in [Figure 2B](#). Left column time-locked on word onset and right column to word offset. Shaded area indicate significant decoding score $p < 0.05$. r stands for Pearson correlation and acc stands for balanced accuracy.

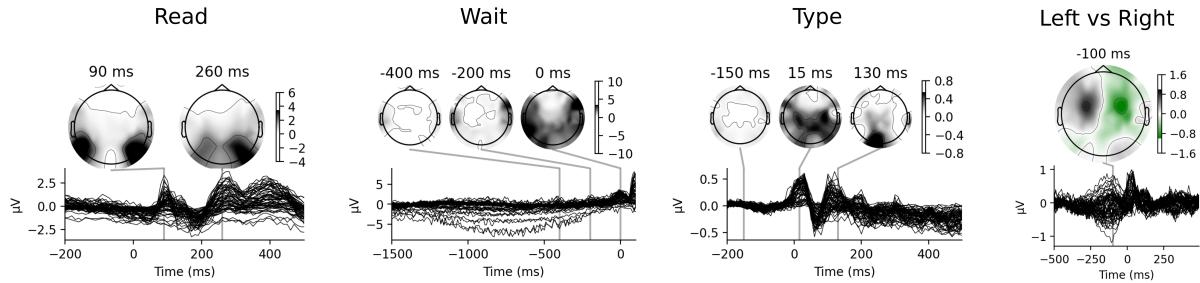


Figure S2 Evoked response of EEG for three phases of the typing tasking (same as [Figure 1C-D](#)). Type response plotted with baseline correction between -250 ms and -0.05 ms. No baseline correction applied otherwise.

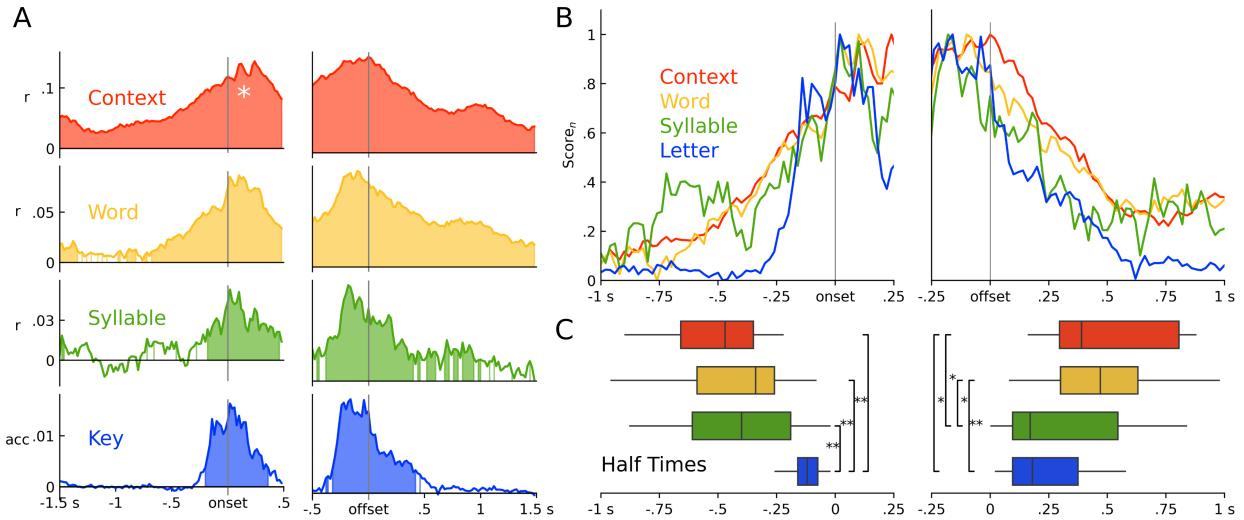


Figure S3 Decoding of the language hierarchy with EEG, replication of Figure 2A-B. **A.** Raw score. **B.** Normalized score. Significant decoding starts at -1.5 s for contexts, -1.5 s for words, -180 ms for syllables, and -200 ms before word onset. After word offset, contexts can be decoded till 1.46 s, words till 1.46 s, syllables till 940 ms, and letters till 420 ms. **C.** Half time across subjects. Significant pair-wise comparisons in half-time are indicated by brackets between feature levels (*: $p < 0.05$, **: $p < 0.01$). Significant comparisons among all pair-wise tests are shown unless adjacent levels are consistently different across the hierarchy. Peak decoding scores at word onset: context 0.14 ± 0.01 , word 0.09 ± 0.01 , syllable 0.04 ± 0.01 , letter $1.6\% \pm 0.2\%$.

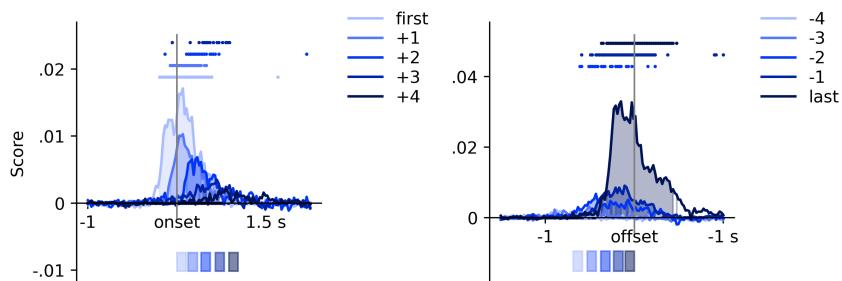


Figure S4 EEG replication of Figure 2D. Scattered dots and shading indicate time steps with significant decoding scores ($p < 0.05$). Rectangles below the x-axis indicate the median start and duration of a key at positions i relative to the first or last keys of the words.

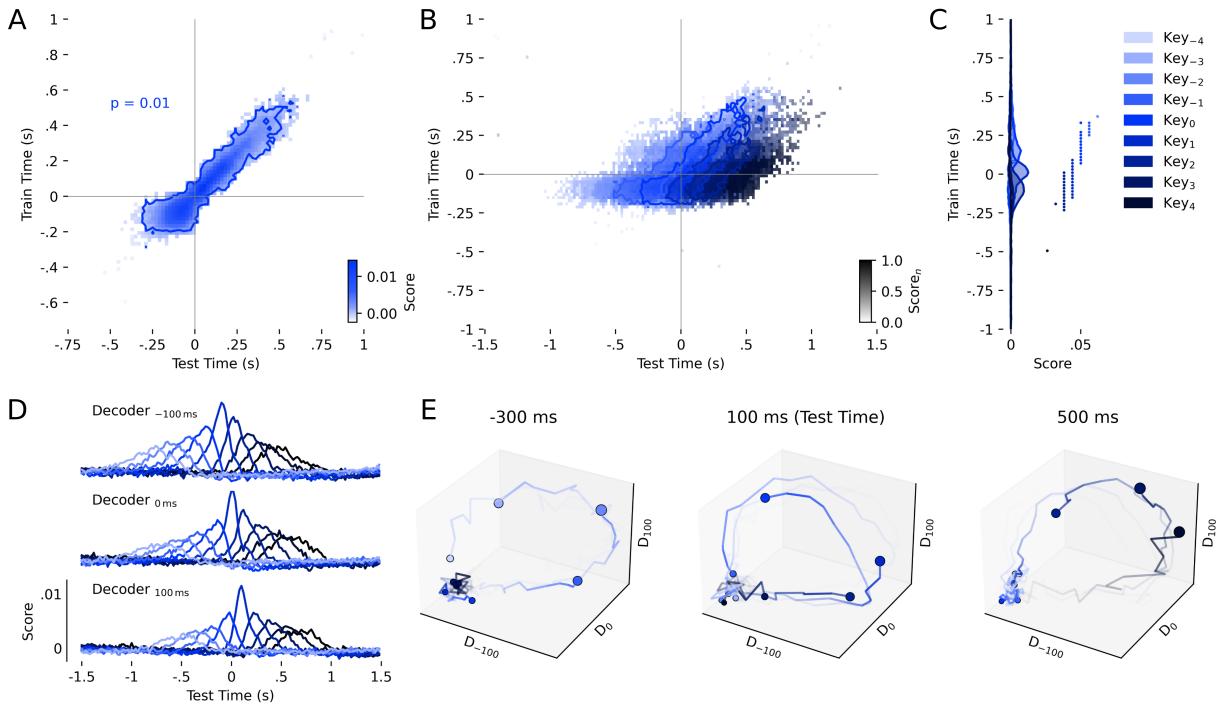


Figure S5 EEG replication of Figure 3B-F. **A.** Temporal generalization of letter decoding. The contour indicates $p < 0.01$ across subjects. **B.** Shifted temporal generalization. **C.** Scores of each decoder trained at each time t_i evaluated on successive letters (same data as the vertical slice of the temporal generalisation matrices at test time = 0 s in C). Vertical lines indicate statistical significance $p < 0.05$. **D.** Scores of three decoders fitted at -100 ms, 0 ms, and 100 ms on successive letters (i.e. three horizontal slices of the Time Gen. of panel B). **E.** Same as D, plotted as normalized 3D trajectories for three representative test times (-300, 100 and 500 ms). Each axis represents one decoder.

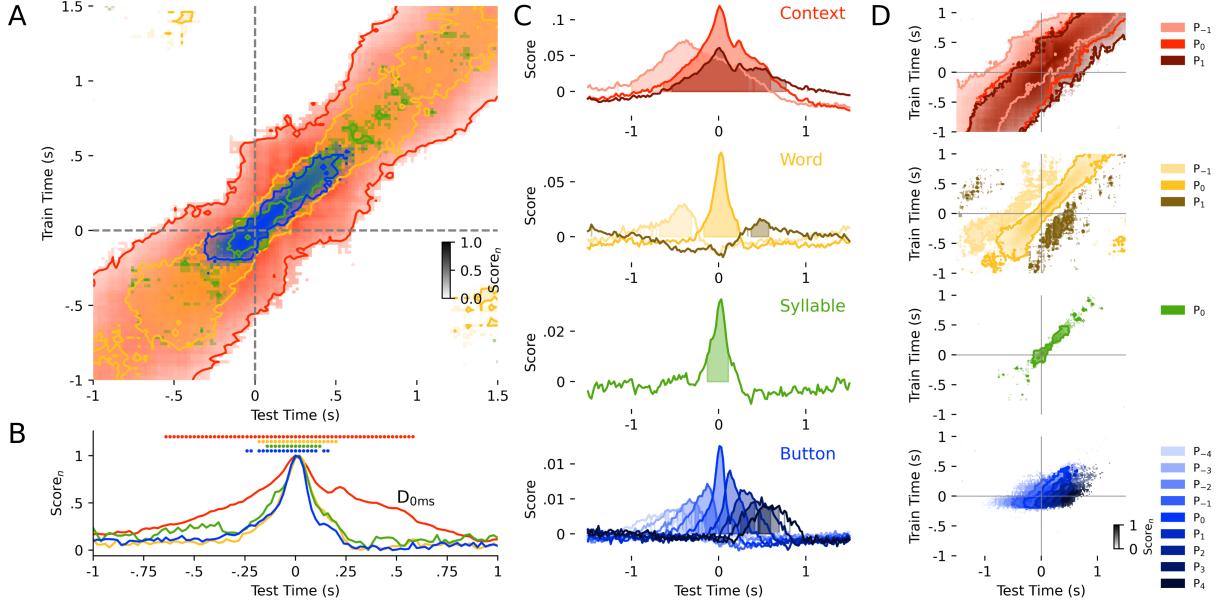


Figure S6 EEG replication of Figure 4B-E. **A.** Normalized temporal generalization of features time-locked at all feature onset. Contours indicate statistical significance across subjects $p < 0.01$. **B.** Normalized generalization scores of decoders from each linguistic level trained at 0 ms and tested over all time samples. Horizontal marks indicate statistical significance $p < 0.05$. **C.** Scores of decoder trained at 0 ms and evaluated across time for each preceding and subsequent feature (i.e. Horizontal slice of panel D). Shaded areas indicate statistical significance $p < 0.05$. **D.** Temporal generalization of previous, current, and next features for each level of the language hierarchy.