Top-down information flow drives lexical access when listening to continuous speech

Laura Gwilliams[1] Alec Marantz[2] David Poeppel[2] Jean-Remi King[3]

[1]University of California, San Francisco
laura.gwilliams@ucsf.edu
[2]New York University
{marantz, dp101}@nyu.edu
[3]PSL University, CNRS, Paris, France

Abstract

Speech is noisy, ambiguous and complex. Here we study how the human brain uses high-order linguistic structure to guide comprehension. Twenty-one participants listened to spoken narratives while magneto-encephalography (MEG) was recorded. Stories were annotated for word class (specifically: noun, verb, adjective) under two hypothesised sources of information: (i) 'bottom-up': the most common word class given by the word's phonology; (ii) 'top-down': the true word class given the syntactic context. We trained a classifier on trials where the two hypotheses matched (about 90%), and tested the classifier on trials where they mismatched. The classifier predicted only the syntactic word class labels, in line with the top-down hypothesis. These effects peaked ∼400ms after word offset over frontal MEG sensors. Our results support that when processing continuous speech, lexical representations are quickly built in a context-sensitive manner. We showcase the utility of multivariate analyses in teasing apart subtle representational distinctions from neural time series.

Keywords: MVPA, MEG, language, speech, brain, word class, part of speech, grammatical category

1 Introduction

It is remarkable that, from air particles that vibrate in synchrony, listeners can understand the complex and novel meanings conveyed by their interlocutor. This auditory signal is an *extremely* unreliable indicator of the meaning it contains, even in quiet listening conditions. For instance, speech varies a lot both within and across people, which prohibits any straightforward link between sound and meaning [28]. In addition, speech is often ambiguous – the very same sounds, in the same order, can mean different things depending on context and expectations. Yet, despite these countless challenges, listeners usually understand speech without difficulty.

The saving grace of speech comprehension is context. The higher order structures of language, in the form of syntactic rules and overarching semantic topic, significantly constrains the space of plausible subsequent input [20, 26, 29, 37]. Given the utterance '*MJ looked at the stars using a...*', the upcoming word is more likely to be a noun, given the syntactic structure, and likely to be a word related to astronomy, given the semantic topic.

The brain makes use of these constraints to guide interpretation of the input [7, 9, 38]. For example, violations of semantic or syntactic constraints are known to lead to a reliable increase in brain responses, as measured with electro-encephalography (EEG) [14, 17, 20, 24, 25]. Other forms of expectations likely have an influence, too. For instance, when presented with distorted speech, perception is bolstered by prior written cues [36] and prior exposure to similar distortions [10]. Furthermore, context which arrives *after* the sensory input also influences perception of previous speech sounds [6, 8, 19].

Top-down information may be particularly important in the case of lexical ambiguity. For instance, the grammatical class of some words is ambiguous: Depending on context, the same word may be used as a noun or a verb. The corresponding meaning can be quite different depending on the assigned class, e.g., *lean, spell, watch*, or quite similar, e.g., *disguise, call, step.* Previous work has demonstrated that sentences containing a lot of ambiguous words elicit stronger neural responses in the inferior temporal lobe and inferior frontal gyrus [33], even when comprehension is equivalent to low ambiguity sentences. Context is required to disambiguate between alternative interpretations and understand the correct meaning [12, 15, 32].

While we constantly disambiguate speech using contextual information, when, where and how top-down disambiguation is implemented in the brain remain unknown. In the present study, we use lexical ambiguity to test two competing hypotheses: The first states that word class is first generated bottom-up based on the phonological form of the utterance, and corrected later (if necessary) using top-down information. The second states that the top-down syntactic structure guides word representations directly, without requiring an initial bottom-up parse. For a visual schematic of the predicted results under each hypothesis, see Figure 1. To adjudicate between these alternatives, we recorded magneto-encephalography (MEG) from 21 native English participants while they listened to four short stories. We modelled neural responses as a function of word class (e.g., noun, verb, adjective). Within our ecological task of story listening, we found support for the second hypothesis: Top-down syntactic structure drives lexical recognition directly, with no measurable trace of a bottom-up interpretation preceding it. The brain thus appears to make higher order interpretations accessible as early as possible to form a rapid and coherent understanding of speech in context.

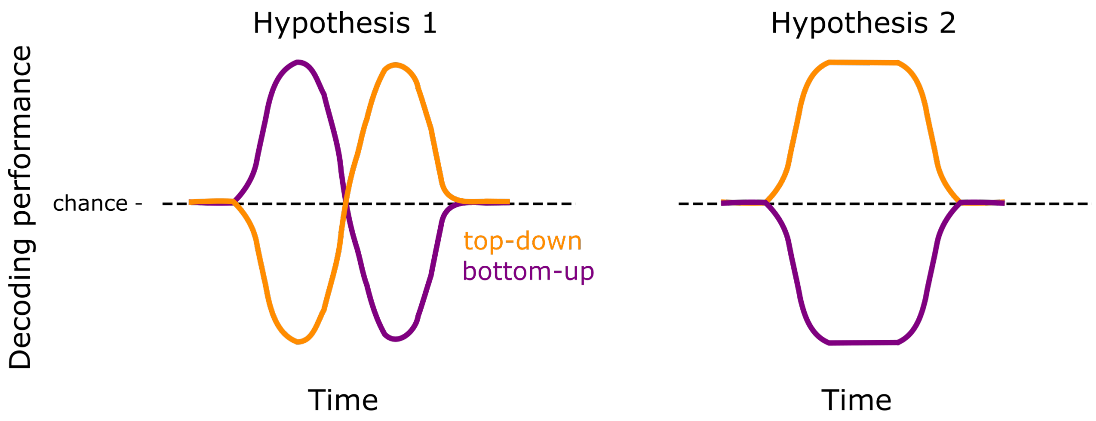

Figure 1: **Analysis predictions.** Schematic of expected results under our two hypotheses. The x-axis represents time relative to word onset. The y-axis represents decoding performance, where the dashed line is chance-level. The purple line represents the decoding of bottom-up labels of word class; the orange line represents decoding the top-down labels. The lines can go above and below chance performance: If the line goes below chance, it means that the classifier is systematically predicting a different class label. Hypothesis 1 predicts that evidence about word class from its phonological form is processed first (bottom-up information), followed by the higher-order syntactic structure (top-down information). Consequently, we would be able to first decode the bottom-up representation of word class, followed by the top-down representation. Hypothesis 2 predicts that top-down information exerts its influence on lexical representations directly. If this is true, we would expect to only be able to decode the top-down labels of word class from neural responses, with no trace of the bottom-up representation being encoded.

2 Methods

2.1 Participants

Twenty-one native English speakers were recruited for the study (13 female; age: M=24.8, SD=6.4). All were right-handed, with normal hearing and no history of neurological disorders. All provided their informed consent and were compensated for their time. The study was approved by the IRB committee at New York University Abu Dhabi, where the study was conducted.

2.2 Stimuli

Four fictional stories were selected from the Manually Annotated Sub-Corpus (MASC), which is a subset of the Open American National Corpus [22] that has been annotated for its syntactic structure using Penn Treebank format.

We synthesised the stories using the Mac-OS text-to-speech application. Three synthetic voices were used (Ava, Samantha, Allison). Story duration ranged from 10-25 minutes. Participants answered a two-choice question via button press on the story content every ∼3 minutes. All participants performed this task at ceiling (98% correct).

2.3 Data acquisition

We used a 208-channel axial gradiometer MEG system (Kanazawa Institute of Technology, Kanazawa, Japan). Data were acquired at a sample rate of 1,000 Hz, with online low-pass filter at 200 Hz and a high-pass filter at 0.03 Hz.

Stimuli were presented to participants though plastic tube earphones placed in each ear (Aero Technologies), at a mean level of 70 dB SPL. Each recording session lasted about one hour. Each participant completed two recording sessions.

2.4 Pre-processing

We removed bad channels from the MEG data using an amplitude threshold cut off of 3SD across all channels within a recording session, and linearly interpolated the bad channels using closest neighbours. We then applied a 1-50 Hz band-pass filter with firwin design [18] and downsampled the data to 250 Hz. The pre-processed continuous MEG data were epoched from -400 to 1,200 ms relative to word onset, and from -400 to 1,200 ms relative to word offset. No baseline correction was applied. All preprocessing was performed using the Python package `mne`, version `0.22.0`.

2.5 Data annotation

We annotated the stories for the identity and timing of the 6,898 words they contained. We were primarily interested in two properties of these words. First is the word class predicted by the *"bottom-up" hypothesis*: *i.e.* the most frequent word class given its phonological form. We obtained these labels by querying the English Lexicon Project [3] for the word class of the words in our stories. The English Lexicon Project derives these labels from a collection of annotated spoken and written corpora. Second is the word class predicted by the *"top-down" hypothesis*, which refers to the word class that a word actually is assigned in the given sentence structure. We obtained these labels from the manual Penn-Treebank syntactic annotation of our stories.

We focus on trials where the word class is an adjective, a noun or a verb in both the top-down *and* bottom-up definition of word class. This sub-selection yielded 3,941 epochs per subject per run.

Some of the words whose class differ across these two definitions were polysemous (had related meanings, and likely etymologically related) while others were homographs (had unrelated meanings, and likely not etymologically related). Unfortunately we did not have enough trials to separate the analysis by this factor, but it would be an interesting avenue for future work to explore.

2.6 Analysis implementation

The majority of decoding analyses used the python package `scikit-learn`, version `0.24.1`. This includes the functions `LogisticRegressionCV`, `StandardScaler`, and `ShuffleSplit`.

Trial-type sets We organised trials into two sets. "Match trials" refer to trials where the word class was identical across the bottom-up and top-down hypotheses. This encompassed 3,662 trials per subject, per run (93%). "Mismatch trials" refers to trials where the word class was different depending on how the word class was defined (279 trials, 7%). Our primary question was whether the bottom-up or top-down definition of word class best explains neural activity when the definitions conflict. For this, match trials were used to train the classifier, and mismatch trials were used to test the classifier.

Optimisation We used a logistic regression trained to perform a one-versus-all classification on the 3-class problem (noun, verb, adjective). The model was fit on each time sample independently, and no sliding window was used. We optimised the regularisation parameter at each time sample, selecting the best model fit on ten log-spaced alpha parameters from 1e-4 to 1e+4 (using `LogisticRegressionCV`).

We used the classifier to estimate the probabilistic class prediction for the held out test trials: *i.e* the soft-maxed distance of that trial from the hyper-plane distinguishing one word-class category from another.

Because the test set contained relatively few trials, we stabilised the performance estimates by allocating different shuffled subsets of training trials with a 20-split ensemble scheme. The subsets comprised

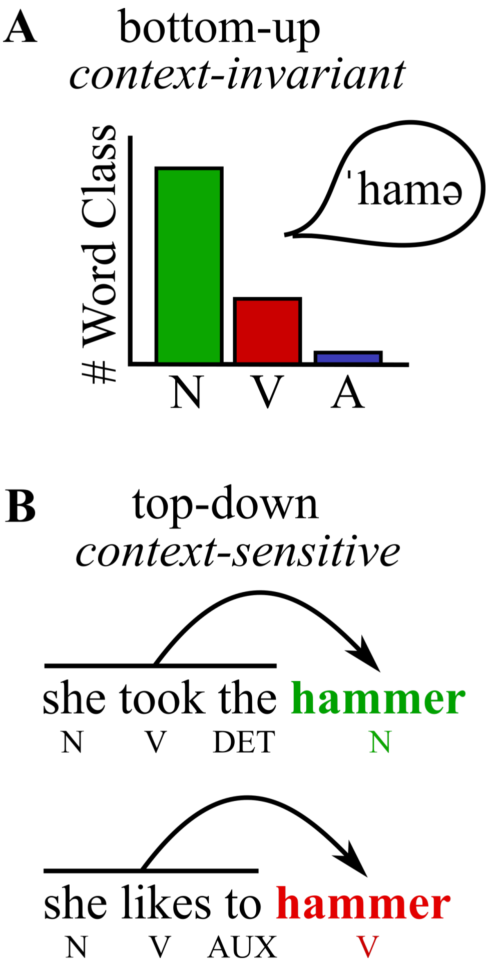

Figure 2: **Definitions of word class.** A: Bottom-up word class corresponds to the most frequent word class category assigned to a particular phonological form. In this case, the word 'hammer' is more often used as a noun than a verb; therefore, the bottom-up definition of word class for this word is noun. B: Top-down word class corresponds to the syntactic word class the item is being used in within the sentence structure. The two sentences provide examples where the same word, 'hammer' is being used as a noun (the sentence above) and as a verb (the sentence below).

a random 75% of trials from the train set (2,746 trials). These random subsets were allocated 20 times, once for each split. Each time, we tested the model using the same 279 trials from the test set.

Evaluation To evaluate decoding performance we compared the model's probabilistic predictions (`ypred`) separately to the bottom-up labels (y_{bu}) and to the top-down labels (y_{td}). We used the receiver operating characteristic (ROC) area under the curve (AUC) to summarise the likelihood that brain activity responded similarly to either hypothesis. This evaluation was repeated for each of the 20-splits within the cross-validation loop. Classifier performance was then averaged across the 20 splits for each subject.

Statistical assessment To evaluate the reliability of decoding over time, we used a non-parametric temporal permutation cluster test across participants. First, we compute a t-value at each time-point by submitting the distribution of decoding accuracy across subjects to a one-sample t-test against chance performance. Second, we identified putative clusters by grouping consecutive t-values that exceeded a t > 1.96 ($p < 0.05$) threshold. Third, the mean t-value within the cluster is compared to a null distribution of t-values, formed by randomly flipping the sign of the distance from chance level, re-running the cluster forming step, and collecting the average t-value. This was performed 1,000 times. We consider clusters significant when their mean t-values exceeds 950 of the lowest values in the distribution ($p < 0.05$).

Word-class decoding To evaluate the overall decoding of word class, we also used a 20-split shuffle ensemble cross-validation scheme on the match trials. This means that on each of the 20 splits, we randomly shuffle the data, and then separate trials into train and test partitions. The classifier was trained on 75% of trials and tested on the held-out 25%. Performance was evaluated on the 20-split average similarly to the above procedure.

3 Results

Our goal was to understand the contribution of bottom-up and top-down computations during naturalistic listening, focusing on the lexical representation of word class (e.g., noun, verb, adjective). We defined word class in two different ways. First, a bottom-up definition: The most frequent word class ascribed to that phonological form (Figure 2A). For instance, the phoneme sequence 'hammer' is most often used as a noun, and so, this would be the bottom-up definition regardless of the context it was being used in. Second, a top-down definition: The true word class assigned given the syntactic structure it is occurring within (Figure 2B).

First, we observed that word class was decodable from the neural responses to the spoken narratives. For this, we subset all words that were either a noun, verb, or adjective, and whose bottom-up and top-down definitions gave the same word class label (3,662 trials). We then used logistic regression to distinguish the three classes, and the Area Under the Curve (AUC) to summarise decoding performance. Using a temporal permutation cluster test, we found that nouns were decodable during the entire epoch, timelocked both to word onset (average $t = 8.1$, $p < .001$) and word offset (average $t = 9.2$, $p < .001$). Verbs were decodable from -40 to 1050 ms from word onset (average $t = 7.3$, $p < .001$) and from -400 to 1080 ms relative to word offset (average $t = 8.2$, $p < .001$). Although adjectives were not significantly decodable relative to word onset after correction for multiple comparison, they were decodable relative to word offset from 210-320 ms (average $t = 3.6$, $p = .004$) and from 370-810 ms (average $t = 4.4$, $p < .001$).

Next, we averaged decoding performance over the three classes (black trace in Figure 3B, to assess the time-course of word class encoding, more broadly. We find that decoding performance is higher relative to word offset than word onset (mean AUC from 0-1000 ms; word onset = 0.515; word offset = 0.52; t value = 2.7; $p = 0.02$). Furthermore, we find two reliable peaks in decoding performance. Relative to word onset, averaged over subjects, the peaks occur at around 110 ms and 680 ms. Relative to word offset, they occur at 390 ms and 600 ms.

Overall, this first set of analyses confirms that word class is encoded in neural activity, and is maximally decodable around 400-600 ms after word offset.

Second, and critically for the aims of the current study, we assessed whether the neural representation of word class is built using a "bottom-up → top-down" sequence of representation, or whether top-down processes influence word class generation directly. This is the analysis for which we plot predictions in Figure 1. We trained a logistic regression classifier on trials where the two definitions matched (same as above), and examined the predictions of trials where they diverged (see Methods for details). We found significant decoding of top-down labels relative to word onset, from 30-260 ms (average $t = 2.8$, $p = .03$) and 650-1100 ms (average $t = 2.7$, $p = .005$). They were also decodable relative to word offset (average $t = 3.7$, $p < .001$). We found that decoding of bottom-up labels were significantly worse than chance in all cases, relative to word onset (0-190 ms; average $t = -3.0$, $p = .02$) and relative to word offset (-260-840 ms; average $t = -3.1$, $p < .001$). This demonstrates that model predictions significantly

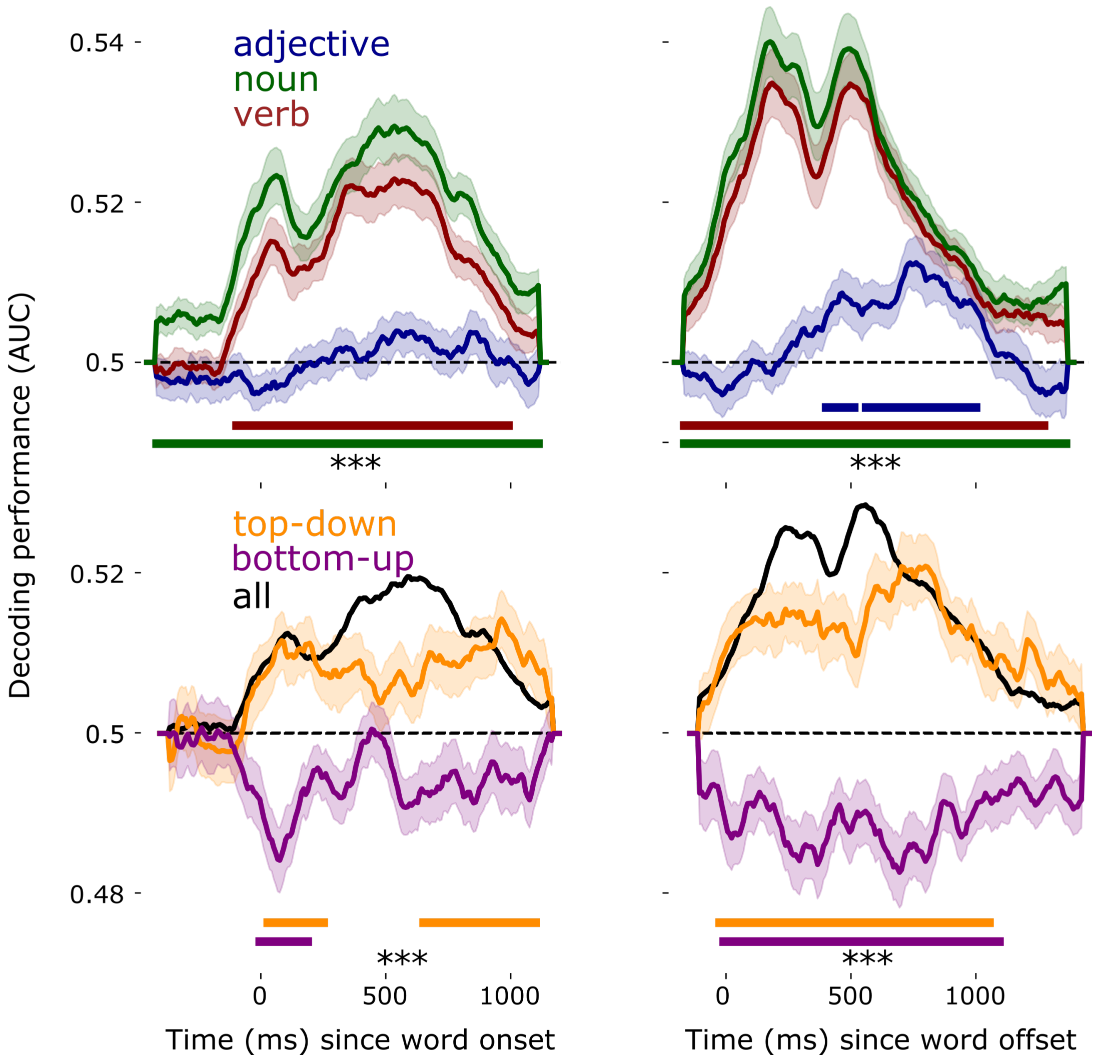

Figure 3: **Timecourse of word class decoding.** Above: Result of decoding the three word classes time-locked to word onset (left) and word offset (right). Below: Comparing bottom-up and top-down labels to the decoding model's predictions on the mismatch trials. Horizontal lines below the timecourse represent the significant temporal clusters resulting from the permutation test. *** = p < .001. Shading represents standard error of the mean across subjects.

correlate with the top-down definition of word class, and significantly anti-correlate with the bottom-up definition of word class. Our results therefore align with Hypothesis 2 shown in Figure 1: Top-down labels are predicted above chance throughout the timecourse of reliable decoding; bottom-up labels are predicted below chance throughout.

Finally, we sought to investigate the spatial patterns with which word class is encoded. For this, we focused on the encoding of nouns and verbs relative to word offset, given these contrasts and timing yielded the strongest decoding performance. We applied a logistic regression decoding model to trials where the top-down and bottom-up labels matched, withholding regularisation so that the spatial patterns remained interpretable. We found that the coefficients that discriminate these word classes are largest at sensors over frontal/temporal cortex, evolving from a right-lateralised topography to a bilateral topography over time. The sensors with highest absolute model weight are shown in Figure 4.

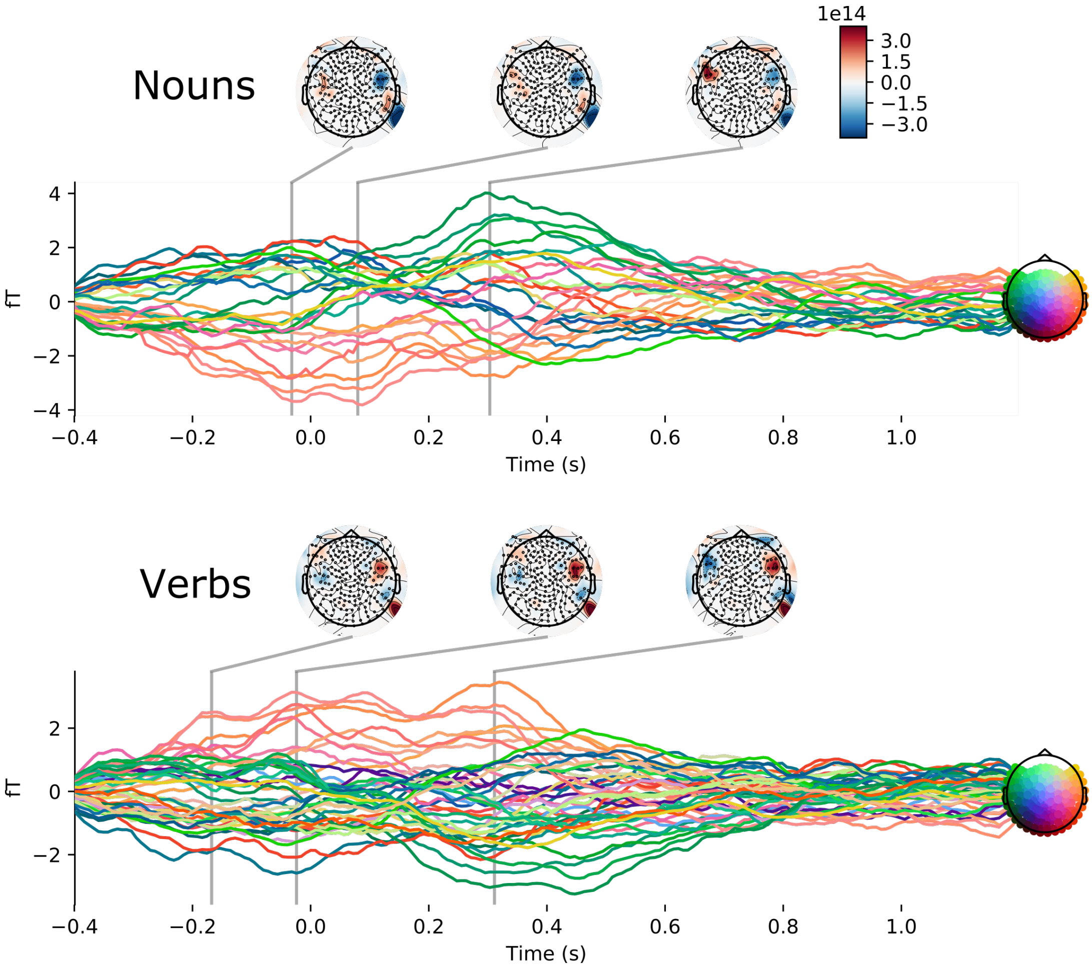

Figure 4: **Sensor weights of decoder over time.** Visualising the decoding model coefficients when discriminating nouns and verbs. Topographies are depicted at three peaks in the sensor magnitudes. Time-courses are displayed for sensors whose absolute max coefficient magnitude exceeds the 80th percentile over sensors. Each line in the timecourse corresponds to a single MEG sensor, coloured by its x/y position in the helmet. Time is plotted relative to word offset.

4 Discussion

Deriving meaning from speech requires overcoming a multitude of sensory and structural ambiguities. We propose that the brain uses high-order linguistic structure to guide the correct interpretation of what is being said. Here, we test a specific prediction of this proposition: Lexical representations are context sensitive, and are built from the global structure that they occur within. The lexical feature we test is word class (noun, verb, adjective), and the type of global structure we test is syntactic. Our main results show that, in line with our predictions, the representation of a word is formed based upon the words which precede it. This suggests that hierarchical linguistic structure primarily exerts influence from the top down, altering how lower order structure is built and interpreted.

Previous work investigating top-down processing has primarily used adverse listening conditions. For instance, making comprehension more difficult by adjusting the signal-to-noise ratio by adding noise on top of the speech signal [11, 38] or by degrading the quality of the speech signal [21, 35]. Other studies have engineered language stimuli to include systematic ambiguity at the phonetic [19, 27] and lexical levels [31, 33], or provided top down information in a different modality, such as hearing speech while reading text [36]. All these studies demonstrate, in different ways, that top down information serves to resolve noise and ambiguity in the speech stimulus. Our work highlights that top-down processing is not only recruited when speech is particularly difficult to understand. Rather, even in ideal listening conditions, the predominant direction of information flow remains top-down.

The major implication of our results is that, in the ecological task of story listening, the neural representation of a lexical item encompasses not just information at t_0, but also information provided in previous time-steps. This result is consistent with the *Syntax First* [13] model of language processing, which posits that even though syntactic structure is a very complex property of language, it is one of the first to come online during processing. Our results also aid interpretation of recent studies that use neural networks to model language processing [5, 30, 34]. One key result of this previous work is that artificial models do better at predicting neural responses when they make use of longer context windows [4, 5, 16, 23]. Our findings anchor a concrete interpretation of this result: The boost in explanatory power is caused by incorporating higher order syntactic information into the lexical representations that are being cross-correlated.

Taken together, our findings are inconsistent with the classic view of language processing that assumes that lower order properties of speech, closer to the sensory signal, are processed first, and serially composed into more complex abstract features over time. Here, we show that during continuous speech processing, higher order structures of speech precede lower levels, more akin to a reverse hierarchy [1, 2]. The reverse hierarchy theory, as put forward for visual processing, suggests that because higher order structure is more robust to noise in the sensory signal, it is used to guide processing and interpretation at the lower levels. We posit that a similar process is happening here for the case of speech processing: Higher order information, in the form of syntactic structure and semantic content, serves to inform interpretations at lower levels. This top-down process may aid the processing of the low-level, and generally noisy, representations of speech signals.

More broadly, our results confirm that linguistic processing is context sensitive [5, 23]. This means that experimenters should be cautious when presenting human subjects with isolated syllables or words because the computations applied to these speech units in isolation may not be the same when they are embedded in continuous speech. The effect we report here is only observable because we presented subjects with language in context, where the language input has an overarching semantic topic and syntactic frame. Prior work that presents subjects with de-contextualised language input (i.e. isolated words or phrases) may have overestimated the role of bottom-up processing for everyday speech comprehension.

Acknowledgements This work was supported by ANR-17-EURE-0017, the Fyssen Foundation and the Bettencourt Foundation to JRK for his work at PSL and NYU Abu Dhabi Institute (G1001) awarded to AM. We thank Arianna Zuanazzi for helpful discussions on an earlier version of this manuscript.

References

[1] Merav Ahissar and Shaul Hochstein. The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, 8(10):457–464, 2004.

[2] Merav Ahissar, Mor Nahum, Israel Nelken, and Shaul Hochstein. Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515):285–299, 2009.

[3] David A Balota, Melvin J Yap, Keith A Hutchison, Michael J Cortese, Brett Kessler, Bjorn Loftis, James H Neely, Douglas L Nelson, Greg B Simpson, and Rebecca Treiman. The english lexicon project. *Behavior research methods*, 39(3):445–459, 2007.

[4] Charlotte Caucheteux and Jean-Rémi King. Language processing in brains and deep neural networks: computational convergence and its limits. *BioRxiv*, pages 2020–07, 2021.

[5] Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications biology*, 5(1):1–10, 2022.

[6] Cynthia M Connine, Dawn G Blasko, and Michael Hall. Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *Journal of Memory and Language*, 30(2):234–250, 1991.

[7] Thomas E Cope, E Sohoglu, W Sedley, Karalyn Patterson, PS Jones, Julie Wiggins, C Dawson, M Grube, RP Carlyon, TD Griffiths, et al. Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nature communications*, 8(1):1–16, 2017.

[8] Delphine Dahan. The time course of interpretation in speech comprehension. *Current Directions in Psychological Science*, 19(2):121–126, 2010.

[9] Matthew H Davis and Ingrid S Johnsrude. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing research*, 229(1-2):132–147, 2007.

[10] Matthew H Davis, Ingrid S Johnsrude, Alexis Hervais-Adelman, Karen Taylor, and Carolyn McGettigan. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2):222, 2005.

[11] Matthew H Davis, Michael A Ford, Ferath Kherif, and Ingrid S Johnsrude. Does semantic context benefit speech understanding through "top–down" processes? evidence from time-resolved sparse fmri. *Journal of cognitive neuroscience*, 23(12):3914–3932, 2011.

[12] Miriam Faust and Christine Chiarello. Sentence context and lexical ambiguity resolution by the two hemispheres. *Neuropsychologia*, 36(9):827–835, 1998.

[13] Angela D Friederici. Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, 6(2):78–84, 2002.

[14] Angela D Friederici, Erdmut Pfeifer, and Anja Hahne. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive brain research*, 1(3):183–192, 1993.

[15] Edward Gibson. The interaction of top–down and bottom–up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54(3):363–388, 2006.

[16] Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, et al. Shared computational principles for language processing in humans and deep language models. *Nature neuroscience*, 25(3):369–380, 2022.

[17] Ana C Gouvea, Colin Phillips, Nina Kazanina, and David Poeppel. The linguistic processes underlying the p600. *Language and cognitive processes*, 25(2):149–188, 2010.

[18] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *Neuroimage*, 86:446–460, 2014.

[19] Laura Gwilliams, Tal Linzen, David Poeppel, and Alec Marantz. In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, 38(35):7585–7599, 2018.

[20] Peter Hagoort, Lea Hald, Marcel Bastiaansen, and Karl Magnus Petersson. Integration of word meaning and world knowledge in language comprehension. *science*, 304(5669):438–441, 2004.

[21] Ronny Hannemann, Jonas Obleser, and Carsten Eulitz. Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain research*, 1153:134–143, 2007.

[22] Nancy Ide and Catherine Macleod. The american national corpus: A standardized resource of american english. In *Proceedings of corpus linguistics*, volume 3, pages 1–7. Lancaster University Centre for Computer Corpus Research on Language . . . , 2001.

[23] Shailee Jain and Alexander Huth. Incorporating context into language encoding models for fmri. *Advances in neural information processing systems*, 31, 2018.

[24] Edith Kaan, Anthony Harris, Edward Gibson, and Phillip Holcomb. The p600 as an index of syntactic integration difficulty. *Language and cognitive processes*, 15(2):159–201, 2000.

[25] Marta Kutas and Steven A Hillyard. Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947):161–163, 1984.

[26] Chia-Lin Lee and Kara D Federmeier. Wave-ering: An erp study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*, 61 (4):538–555, 2009.

[27] Yune-Sang Lee, Peter Turkeltaub, Richard Granger, and Rajeev DS Raizada. Categorical speech processing in broca's area: an fmri study using multivariate pattern-based analysis. *Journal of Neuroscience*, 32(11):3942–3948, 2012.

[28] Alvin M Liberman, Franklin S Cooper, Donald P Shankweiler, and Michael Studdert-Kennedy. Perception of the speech code. *Psychological review*, 74(6):431, 1967.

[29] George A Miller and Stephen Isard. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2(3):217–228, 1963.

[30] Peng Qian, Xipeng Qiu, and Xuanjing Huang. Bridging lstm architecture and the neural dynamics during reading. *arXiv preprint arXiv:1604.06635*, 2016.

[31] Jennifer Rodd, Gareth Gaskell, and William Marslen-Wilson. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language*, 46(2):245–266, 2002.

[32] Jennifer M Rodd, M Gareth Gaskell, and William D Marslen-Wilson. Modelling the effects of semantic ambiguity in word recognition. *Cognitive science*, 28(1):89–104, 2004.

[33] Jennifer M Rodd, Matthew H Davis, and Ingrid S Johnsrude. The neural mechanisms of speech comprehension: fmri studies of semantic ambiguity. *Cerebral Cortex*, 15(8):1261–1269, 2005.

[34] Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A Hosseini, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45), 2021.

[35] Ediz Sohoglu, Jonathan E Peelle, Robert P Carlyon, and Matthew H Davis. Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25):8443–8453, 2012.

[36] Ediz Sohoglu, Jonathan E Peelle, Robert P Carlyon, and Matthew H Davis. Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):186, 2014.

[37] Michael J Spivey-Knowlton, John C Trueswell, and Michael K Tanenhaus. Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47 (2):276, 1993.

[38] Adriana A Zekveld, Dirk J Heslenfeld, Joost M Festen, and Ruurd Schoonhoven. Top–down and bottom–up processes in speech comprehension. *Neuroimage*, 32(4):1826–1836, 2006.