The role of syntax in semantic processing: A study of active and passive sentences

Nicole Rafidi

May 23, 2016

Abstract

Background As you read this sentence, you are performing the complex task of composing the meaning of these words into a whole proposition. A multitude of work in psycholinguistics has shown that humans use both the semantics (the meaning of the words) and the syntax (their order) of a sentence to process its meaning. However, the rules by which the brain combines this information are largely unknown. High temporal resolution non-invasive neuroimaging techniques, such as magnetoencephalography (MEG) have great potential to answer this question, because they measure neural activity on the timescale of natural reading.

Aim Using a machine learning approach, we contrast sentences with the same semantics and different syntax in order to shed light on the processing stages used by the brain when reading English sentences. We examine the robustness of these results to various analysis design choices to maximize sensitivity in future confirmatory analyses.

Data We collected MEG data from 8 participants while they read active- and passive-voiced versions of the same propositions. Each participant read 16 propositions made up of 8 nouns and 4 verbs. The 32 sentences were repeated 15 times and we use an average of some or all of these repetitions for subsequent analysis.

Methods We apply a Naive Bayes classifier at each point in the MEG activity timeseries to generate classification accuracy timeseries for each word in the sentence, cross-validating in a leave-one-proposition-out manner. We seek time windows of above chance accuracy in order to assert when certain words are accessed. To test the robustness of the observed effect, we vary different aspects of the data analysis and observe the effect on decoding accuracy. We also assessed which neural regions contribute most to decoding accuracy at different times.

Results When reading passive-voiced sentences, participants access words relevant to the proposition multiple times throughout the sentence. This draws a sharp distinction with the processing of active-voiced sentences, in which participants seem to access relevant words only once, at word presentation. The result is highly robust to a variety of analysis decisions, and seems, to our surprise, to originate in occipital regions of the brain.

Conclusions This result is consistent with prior work demonstrating reanalysis in syntactically complex sentences, in which participants correct an erroneous initial reading of the sentence. Our work elucidates the timing of reanalysis, which is late in the sentence and is consistent with theory on the mental storage of sentence components by role. Via reanalysis, participants can arrive at the same meaning through different syntactic paths. These results demonstrate the potential of machine learning approaches to reveal neural processing stages in an interpretable manner.

1 Introduction

Understanding language processing has been a long-standing goal of linguistics, psychology, and neuroscience. Crucial to language processing is the reading of a sentence. Sentences capture meaning in many ways, e.g., "The dog found the peach," and "The peach was found by the dog," are the same proposition in different syntactic forms. In English, the former version, called active voice, is the most common. These two versions allow us to test theories of syntactic integration. For example, it is possible that sentences are read without any assumptions of syntax, and the human brain is flexible enough to understand active and passive sentences in the same manner. Alternatively, given the human tendency to make easy assumptions, reading the passive-voiced version may require a correction of the initial reading, which defaults to active voice, since it is more common. Evidence seems largely to point to the second theory of integration, and the correction has been referred to commonly as reanalysis.

While this topic can be studied in many ways, non-invasive neuroimaging is currently the easiest and most direct way to answer questions about the human brain. Of the techniques available, magnetoencephalography (MEG) holds the most promise for examining language. It measures the induced magnetic field produced by concerted neural activity at a rate of 1kHz, which is ample resolution, as humans typically process words on an order of hundreds of milliseconds (Sudre et al. (2012)). Compared to functional magnetic resonance imaging (fMRI), which operates on the order of seconds, MEG is the clear choice. MEG also has an advantage over electroencephalography (EEG) in that magnetic fields, unlike electric fields, are not distorted by the human skull. This improves the localization of MEG signals to actual brain regions (known as source-localization). Analyzing MEG data is difficult from a statistical standpoint: the signal-to-noise ratio of a single MEG trial is quite low, so a stimulus must be presented multiple times and the resultant trials averaged. Given that we can only collect so many trials from a participant in one scan session (dependent on the length of the trials, number of conditions, etc), how can we optimize our scanning protocol to make the most use of the scanner time? Additionally, the high resolution means that there are many highly correlated features to examine. A large number of features but a small number of samples leads most machine learning approaches to overfit, so heavy regularization is typically required in order to get reasonable generalization performance.

In this work we examine data from a pilot study contrasting active and passive sentences using simple classifiers, and find evidence in support of reanalysis. We found that reanalysis occurs with timing that, while varied across individuals, has striking implications for current models of sentence processing. However, this is largely exploratory work, and a confirmatory analysis is needed. In order to maximize sensitivity in subsequent confirmatory analysis, we examine the robustness of the effect found in the pilot data to changes in various experimental parameters such as number of repetitions used and the degree of regularization. This analysis will enable us to conduct a confirmatory analysis using optimal parameterization. We also examine which features contribute most stably to decoding accuracy.

2 Background and Related Work

The question of how the brain handles syntactic complexity and ambiguity, such as that produced by passive-voiced sentences, has been studied, though never with the temporal resolution and machine learning approach that we use here. Early work using EEG recordings during the presentation of German sentences revealed evidence for a reanalysis effect: Bornkessel et al. (2003) found that sentences ordered as the direct object followed by the subject of the verb, the German equivalent of passive sentences, resulted in more positive neural activity in passive sentences during the presentation of the subject than in active sentences. This positivity demonstrates that there is a fundamental contrast between active and passive sentences, and that the timing is very late - at the last necessary moment. However, it is impossible to say what the positivity is. Does it simply reflect processing load? Is it purely syntactic, or is there a semantic component to it? Furthermore, English, unlike German, requires extra words in order to convey an object-first sentence (for example, "the peach was found *by the* dog"), which may change the timing of this effect in English speakers. Nonetheless, this was an early indicator that a fundamental processing difference exists between passive and active sentences. Later work in MEG examined the effect of increasing syntactic complexity in English sentences. Meltzer and Braun (2011) demonstrated a significant power decrease in a broad frequency band after reading syntactically complex sentences. This decrease was not observed during sentence listening, but during the inter-sentence delay period. However, syntactic complexity was induced by adding relative clauses, not by changing the overall structure. This work is important in that it shows that post-sentence neural activity may be relevant in the processing of syntactically complex sentences, although it is unclear what that activation represents.

fMRI studies contrasting active and passive sentences have found increased activation in tempero-parietal (TP) regions, Broca's area, and the inferior frontal gyrus for passive sentences as compared to active sentences (Meyer et al., 2012; Mack et al., 2013; Frankland and Greene, 2015). Because of the low time resolution of fMRI, Meyer et al. (2012) took the analysis one step further and examined EEG data to explain the increased activation, finding activity during verb presentation that traveled from TP regions to Broca's area. Once again, it is difficult to say what this activation actually represents. That is why Frankland and Greene (2015) adopted a multivariate pattern analysis (MVPA) approach. In that study, participants were scanned using fMRI while reading active- and passive-voiced versions of the same propositions, in a paradigm very similar to the one presented in this work, with classifiers applied in a searchlight to the data. The key result is that two proximal yet distinct subregions of left medial superior temporal cortex selectively encode the subject of the sentence and the object of the sentence, respectively. The authors propose that the processing of sentences uses these two regions as registers to store the relevant roles of the nouns of the sentence. What follows naturally from such a hypothesis is that if one of these registers is incorrectly filled, e.g. if a word that was supposedly the subject is found to truly be the object, the brain must perform a transfer operation in order to correctly assign word role. This transfer operation could be a mechanistic explanation for the reanalysis effects seen in the EEG and MEG literature.

In order to determine whether this is indeed the case, we must carefully examine the timecourse of sentence processing. If reanalysis is mechanistically explained by this transfer operation, we should see that the word to be transferred is accessed when the time of the transfer occurs. The prior work discussed here paints an inconsistent picture: we have evidence for active-passive differences during verb presentation, during second noun presentation, and post sentence. Another failing of the prior EEG and MEG work is to determine what exactly is causing this difference. Using the temporal generalization method described in King and Dehaene (2014), which simply consists of training different classifiers over time, we hope to fill this gap in knowledge. From examining this related work, we expect to see a semantic reactivation of the first noun as it is assigned to its correct syntactic role. However, the timing of this reactivation could occur at many candidate times. Based on the Bornkessel et al. (2003) and Frankland and Greene (2015) studies, we expect it to be relatively late, that is we expect reanalysis to occur when the reader has no choice but to transfer the first noun. This could occur as late as the presentation of the second noun but may precede it slightly, due to the particular nature of passive sentences in English.

3 Problem Statement

We hypothesize that when humans read syntactically complex sentences, they initially assign words to roles based on the most simple syntactic reading, and later revise this incorrect reading through the process of reanalysis. Our goal is to demonstrate that reanalysis can be mechanistically explained by the reactivation of the constituent words of the sentence.

4 Method

Data Collection

As their neural activity was recorded with the Elekta Neuromag scanner, 8 participants from Carnegie Mellon University's Machine Learning Department read 32 noun-verb-noun sentences that varied in voicing. The sentences are composed of 8 nouns and 4 verbs. Each of 16 propositions was presented in both active and passive voice 15 times (e.g. "The dog found the peach" and "The peach was found by the dog"). Each word was presented for 300ms with 200ms rest in between, and there was 2s of rest between sentences. Comprehension questions followed 10% of sentences, to ensure semantic engagement. All 8 subjects answered close to all of the questions correctly. If a participant had failed to answer the majority of these very simple questions correctly, he or she would have been excluded from further analysis. See Appendix A for a table of the stimuli used, as well as the comprehension questions. The data were spatially filtered using the temporal extension of the signal space separation (tSSS) algorithm, bandpass filtered from 1 to 150Hz with notch filters applied at 60 and 120Hz, and downsampled to 500Hz. Artifacts from tSSS-filtered same-day empty room measurements, ocular and cardiac artifacts were removed via signal space projection (SSP). Raw and processed versions of the data are available upon request.

Analyses

Assessing the Reanalysis Effect

In the first part of the analysis, we set out to discover when the constituent words of each sentence were decodable the MEG signal. Above chance accuracy at a particular time indicates that the participant is accessing the word at that time. We used a Gaussian Naive Bayes classifier on these data over time, applying it at each point in the timeseries, similar to King and Dehaene (2014). These classifiers are trained on a subset of the propositions to detect certain words, and then are tested on the active and passive forms of the remaining propositions, cross-validating in a leave-one-proposition-out manner (note that this is a leave-two-sentence-out scheme). Each training and test sample is a proposition that is the average of its 15 repetitions. The classifier was trained and tested on the same timepoint across all samples, and used a 200ms window of the neural activity. Since there are 306 sensors, this results in too many features to reasonably use. Therefore only the 500 space-time points with the largest inter-class mean were used to make the classification on the test examples. These parameters were chosen based on prior work by Sudre et al. (2012), which indicated that semantic processing occurs primarily over a 200ms window, and were not cross-validated or modified during this initial analysis.

We train separate classifiers for the first noun, the verb, and the second noun of the sentence. Depending on the type of sentence (active or passive) the first noun could be either the subject (the noun that performs the verb) or the object (the noun acted upon by the verb). To illustrate, in "the dog found the peach," and "the peach was found by the dog," the subject is dog, the verb is found, and the object is peach. This transforms the neural activation timeseries, which are multi-dimensional and non-specific, into single dimensional information timeseries, that track when the participant is accessing the subject, verb, or object.

While generating these timeseries for each participant is simple, aggregating across participants is a non-trivial statistical problem. In particular, establishing which windows of time have significantly above chance classification accuracy is very challenging: standard permutation testing is extremely costly if we perform it over an entire timeseries, and even then we are left with a multiple comparisons problem of many tests that are not independent. For the initial analysis, we use a binomial test across participants and then the cluster permutation test to correct for multiple comparisons (Maris and Oostenveld, 2007). The binomial test generates a timeseries of p-values that indicates the likelihood at that timepoint of the observed number of subjects reporting above chance accuracy (for noun decoding this is 12.5% accuracy). For example, if 5 subjects report above chance accuracy at a timepoint, then the p-value for that timepoint is $\frac{1}{2^5} = 0.03125$. The cluster permutation test corrects this p-value timeseries for multiple comparisons by first thresholding by a chosen p-value (here 0.05). Then it finds contiguous clusters that meet that threshold and a chosen minimum length (here 80ms) and assigns them a score based on the combined t-statistic of the cluster's constituent points. Then the largest cluster undergoes a permutation test wherein the score is recomputed for many random assignments of the classes (here above or below chance accuracy) of the points within the cluster. This generates a histogram of cluster scores, and the corrected p-value for each cluster can be generated by comparing the true cluster score to the histogram (Maris and Oostenveld, 2007)

Robustness Analyses

After conducting this initial analysis we found a significant reanalysis effect in passive sentences. However there were many aspects of the analysis that were chosen without exploration due to computational intractability. In subsequent analyses we assess the robustness of the initial effect by varying these parameters and observing the change in decoding accuracy at the relevant timepoints. The three parameters assessed were the classifier window size (i.e., how much of the timeseries used in classification at a single timepoint), regularization of the Naive Bayes classifier, and the number of repetitions of the stimulus included in the average sample.

Additionally, we attempted to assess which features (sensor-time points) were contributing to the decoding accuracy at various timepoints of interest. For each feature, we calculated the percent of folds (pooled across participants) that that feature appeared in the top 50 features. This yields a heatmap of which sensor-time points occur most frequently in the top 50 overall. In order to summarize, we then collapsed this heatmap over brain regions and over time, to see which brain regions (as indicated by sensor position in the MEG helmet) appeared most often in the top 50 features. It should be stated that these estimates of brain region are very coarse; all MEG sensors receive input from all of the brain, however their strongest input comes from the region closest to them.

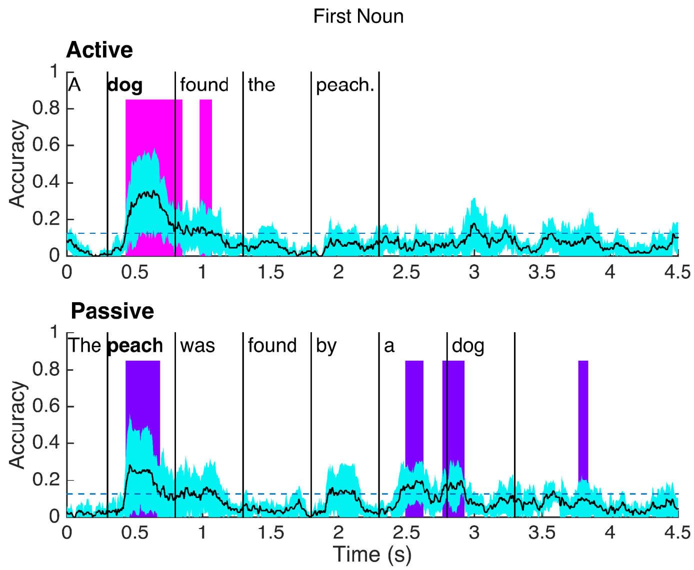

Figure 1: Decoding accuracy over time for classifying the identity of the first noun of the sentence, averaged across participants and over sentences. The black vertical lines show the word presentation times, with the last vertical line indicating the end of the presentation of the sentence. Standard deviation is shown in green, with significant windows highlighted in pink and purple. Accuracies averaged over active sentences are shown above, and accuracies averaged over passive sentences are shown below. Note the repeated peaks in the accuracy timeseries for passive sentences.

5 Results

Assessing the Reanalysis Effect

Figure 1 shows the classification accuracy achieved over time when decoding the first noun of the sentence. Significant windows found via the cluster permutation test are highlighted. The top plot shows the accuracy averaged over active sentences, and the bottom over passive sentences. There is an initial peak in decoding accuracy when the word of interest is presented in both active and passive sentences. Additionally, in passive sentences, there are repeated peaks later in the sentence.

Decoding of the verb was less successful: as seen in Figure 2, left, we are never able to decode verb identity significantly above chance by our cluster permutation criterion. The right half of that figure shows that the second noun in both sentence types is decodable when it is presented, but not during the post-sentence period.

The Effect of Time Window Size

Figure 3 shows the previous sliding window analysis repeated for windows of varying size. The number of features used by the classifier is 500, as in the initial analysis, and all repetitions are used to form the training and test samples. As one would expect, increasing the window size increases the width of accurate windows, smearing the decoding accuracy over time. While window size appears to have little to no effect on decoding accuracy for active sentences, there is a slight

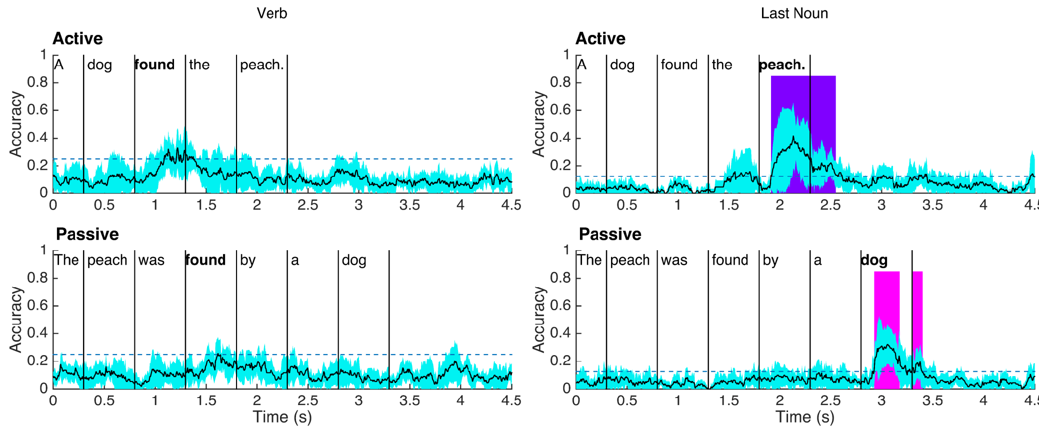

Figure 2: Decoding accuracy over time for classifying the identity of the verb (left) and second noun (right) of the sentence, averaged across participants and over sentences. Format is the same as in Figure 1. Accuracies averaged over active sentences are above, and accuracies averaged over passive sentences are below. We were not able to successfully decode verb identity. However the decoding accuracy pattern over time for the second noun is very similar for active and passive sentences.

decoding accuracy advantage to using smaller windows for the reanalysis period (i.e. the second accuracy peak) in passive sentences.

The Effect of Regularization

Figure 4 shows the initial sliding window analysis repeated such that the classifier is allowed to use different numbers of features in classifying the test example. Window size is 200ms, and all repetitions are used to form the training and test samples. For both active and passive sentences, the trend is that sparser models perform better, but it is a slight advantage.

Figure 5 contains the heatmaps showing the feature score of each sensor for two time periods: first noun presentation, and reanalysis. These time windows were chosen to correspond to peaks in decoding accuracy. The feature score shown is indicative of both the frequency of appearance in the top 50 and the importance of that feature within the classification time window of 200ms. Particularly striking is how heavily occipital these maps are, despite the fact that it is only the left heatmap that was generated from the same time that the word of interest was actually on the screen.

The Effect of Repetitions

Figure 6 demonstrates the effect on decoding accuracy as more repetitions are included in creating the training and test samples. That is, row two shows the accuracy when the first two repetitions are averaged, and row 14 shows the accuracy when the first 14 repetitions are averaged (in order of presentation to the participant). For this analysis 200ms windows were used, and 500 features. There are two features of note in this plot: firstly, while accuracy is essentially strictly increasing as a function of repetitions in active sentences, the picture is not so clear for passive sentences.

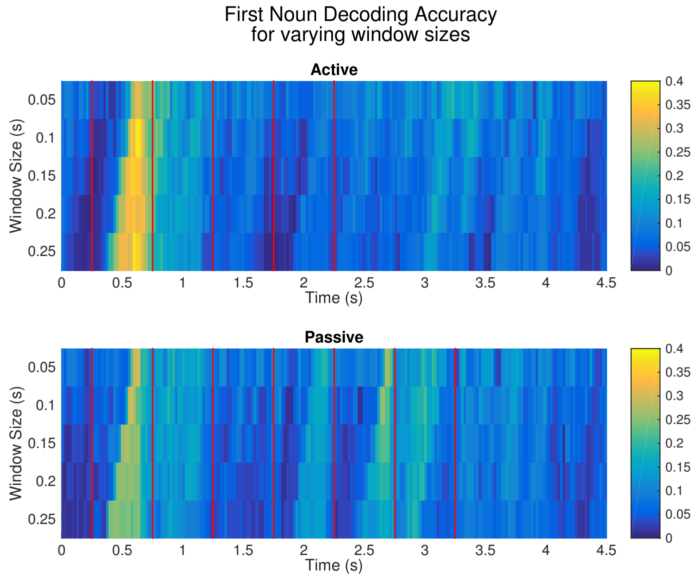

Figure 3: Here we show the analysis from Figure 1 repeated for varying classifier window sizes. The red vertical lines show the word presentation times, with the last vertical line indicating the end of the presentation of the sentence, shifted backwards by the largest possible window size. The effect of window size is more pronounced for the second decoding accuracy peaks in passive sentences than in any other time. Specifically, smaller windows are better for accuracy.

Secondly, accuracy for passive sentences is in general quite low, making it difficult to understand where the different peaks in accuracy across repetitions come from.

In order to better understand how repetitions affect decoding accuracy, we also performed sentence voice classification (i.e., is the sentence active or passive), which is a much easier task simply due to the increase in per-class samples. The result is shown in Figure 7. Accuracy for this task is quite high, even for only two repetitions, and even when there are no words on the screen, indicating that post-sentence activity is indeed fundamentally different for active and passive sentences. Some of this difference can be attributed to the end of the post-sentence period for active sentences and the start of the next sentence or question. A worrisome result is that after 8-10 repetitions, participants appear to be memorizing which sentences are active and which are passive, since we can decode during the presentation of the first word whether the sentence is active or passive. This is also due to a flaw in the stimulus set - some nouns are only ever subjects, and some are only ever objects.

To further investigate the sentence memorization phenomenon revealed in Figure 7, we performed the voice classification task for three groups of repetitions averaged together. The result is shown in Figure 8. What we see is that the blue trace, which is the last 5 repetitions averaged together, is strikingly different from the other two during the first word presentation. The red trace, which is the first 5 repetitions, is strikingly different from the other two shortly after the sentence ends. In both cases this difference is an increase in classification accuracy.

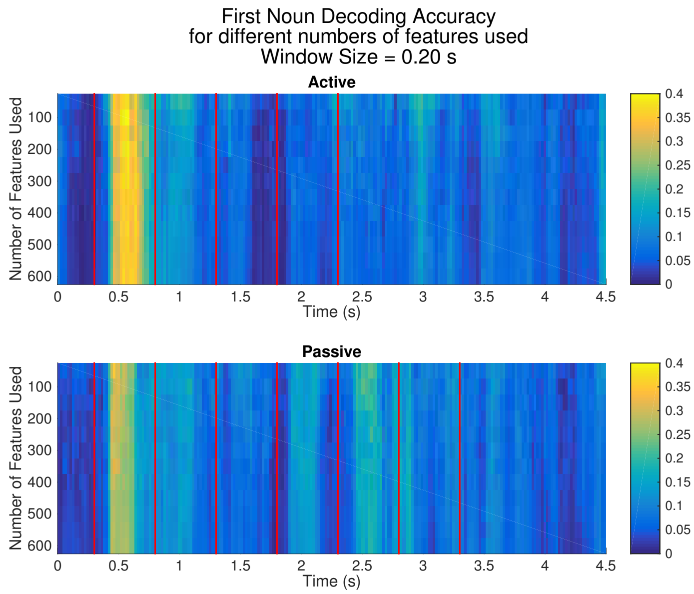

Figure 4: This figure shows the analysis from Figure 1 repeated for varying regularization levels. Here regularization is indicated by the number of features that Naive Bayes is allowed to use. In general, sparser models perform better, but the effect is not strong for the range of values used.

6 Discussion

Figure 1 paints an interesting picture of what the reanalysis effect is. As hypothesized, there is a significant reactivation of the first noun in passive sentences that does not happen in active sentences. While the timing of this reactivation is highly varied across people, there is a strong indication that reanalysis is occurring when the sentence is known to be passive-voiced (i.e., during ”by the”). This ties in nicely with both the Bornkessel et al. (2003) and the Frankland and Greene (2015) results. The first study reported a difference in active and passive EEG signals during second noun presentation in German. In English, the voice of the sentence is noticeable earlier, and thus we observe the effect earlier. In addition to simply illustrating a difference between active and passive sentence processing, our analysis also uncovers the information content of the reanalysis signal: the neural signature of the first noun, which is the word requiring assignment to a new role. This the first clear demonstration of this fact. It is as if we have captured the moment at which the first noun switches from the ”subject register” to the ”object register” as described in Frankland and Greene (2015). Also notable is that verb and second noun processing show no difference across the two types of sentences.

It could be argued that in reading a passive sentence, the reader should know even earlier what the voice of the sentence is. Specifically, in the stimuli in this experiment, the presence of the verb ”was” is a clear indicator that the sentence is passive. These results support a lazier hypothesis of reanalysis, in which the reactivation occurs only when necessary to make way for the second noun. For example, the data would indicate ”peach” would not require reanalysis in ”The peach was found.”, since we do not observe any reactivation that early in the sentence. However, the

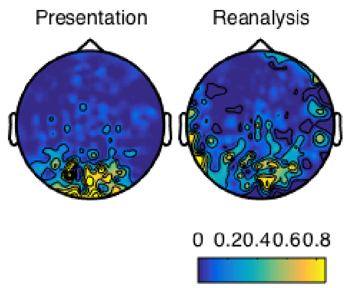

Figure 5: Helmet plots of the frequency each sensor appeared in the top 50 over folds and subjects, shown for the two accuracy peaks in Figure 1, bottom.

variation across individuals (note the three peaks in Figure 1) indicates that individual differences potentially play a role in the exact timing of reanalysis. This could explain why prior results, such as (Meyer et al., 2012) and (Bornkessel et al., 2003) disagree so strongly on the effect timing. Further information about participants will need to be collected to ascertain the reason for these differences.

The results from the robustness analyses are fairly straightforward. We hope to find hyperparameters that maximize the decoding accuracy at the time of reanalysis. We can think of decoding accuracy as the sensitivity of our analysis to the effect of interest. Ideally, we would choose both this and the classifier window size via cross-validation. However, training a classifier at each timepoint is already quite expensive, and cross-validating two hyperparameters would make the problem much less tractable.

While the window size used by the classifier does not strongly affect decoding at the time of word presentation (see Figure 3), it does appear to affect the decoding accuracy achieved during the reanalysis period. Specifically, the classifier is more sensitive to the reanalysis effect if it uses a short time window. This implies a rapidly changing signal that does not remain stable over longer time windows. Further analysis of classifier similarity across time windows, as demonstrated in King and Dehaene (2014), will be required to test this implication more thoroughly. As predicted, stronger regularization improved decoding accuracy (shown in Figure 4). This is to be expected, since we are working with high dimensional data (306 sensors $\times$ the number of time points) but very few examples (32 total).

The result of varying the number of repetitions, shown in Figure 6, were more difficult to decipher. This is because there are two issues at play when more repetitions are used: firstly, using more repetitions typically improves the signal-to-noise ratio and thus improves decoding accuracy. Secondly, as the participant reads the same sentences over and over, they become desensitized to the

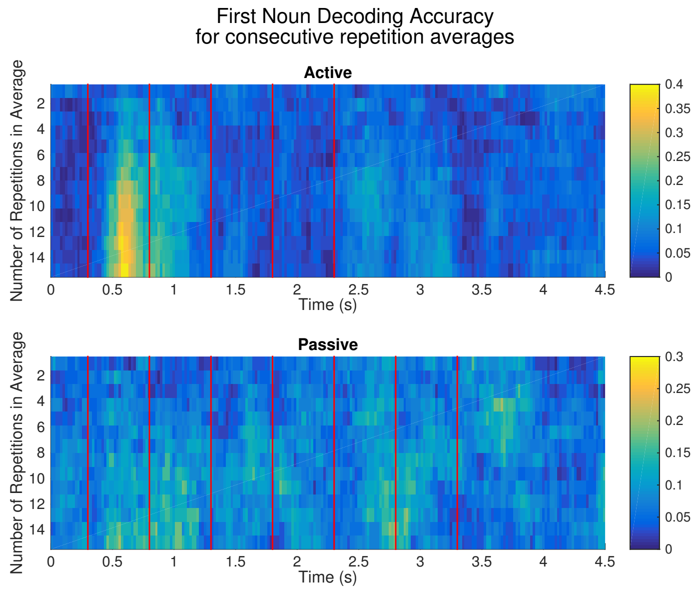

Figure 6: This figure shows the analysis from Figure 1 repeated as repetitions from the experiment are included in the average. While more repetitions lead to superior decoding for active sentences, the picture is less clear for passive sentences, particularly post-sentence.

stimuli and can even memorize them. While more repetitions is better for maximizing sensitivity to the reanalysis effect, Figure 6 also shows that a post-sentence effect seems to disappear with more repetitions. This could be some kind of habituation effect: either the neural representation observed there disappears or it occurs earlier as the participants become accustomed to the sentences. It is difficult to distinguish these two forms of habituation with only this analysis. To further examine the question of memorization, we conducted the same analysis but for an easier classification problem, the binary choice of whether the sentence was active or passive. Figure 7 shows that decoding accuracy is very high for this task even for low signal-to-noise samples, but that the ability to distinguish the sentences early (during the first noun) appears only after 8 to 10 repetitions. This early accuracy could either be attributed to memorization or to the task of distinguishing inanimate objects from animate ones.

In order to try to tease apart these explanations, we subsequently ran the analysis for groups of repetitions, dividing the total of 15 into three blocks of five repetitions each. The result is shown in Figure 8. These blocks differed at two key times: during the first noun, presentation, when the last group elicits much higher accuracy, and briefly after the end of the sentence, when the first group elicits much higher accuracy. The first phenomenon is evidence of some degree of memorization of the sentences. The second appears to be related again to the habituation observed in Figure 6.

7 Conclusions

In sum, we were able to characterize the reanalysis effect as occurring late during the sentence. Individual variation in the timing of the effect can potentially explain inconsistencies in prior work.

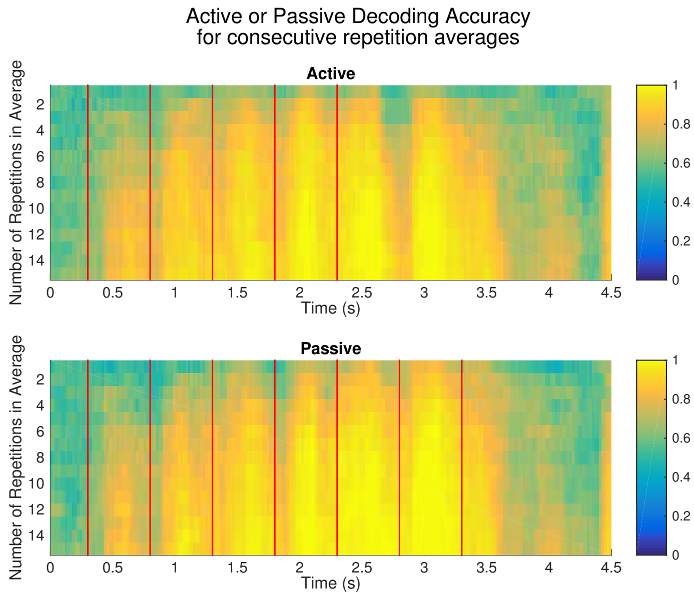

Figure 7: Here we show the results from the voice classification task, which identifies whether the sentence is active or passive, repeated as repetitions from the experiment are included in the average. Accuracy increases with the number of repetitions used, including early in the sentence, indicating potential memorization.

The reanalysis effect appears to be best captured by examining a short period of time and using a sparse representation. While memorization and habituation were both consequences of the large number of repetitions, the signal-to-noise ratio improvement was very valuable. This settles a long-standing debate about the informational content of the reanalysis signal, that it in fact can be explained by a reactivation of the first noun of the sentence.

Limitations

This work is largely exploratory. It should be noted that given the battery of analyses these data have been subjected to, the only analysis appropriate for consideration of effect size is the initial one, shown in Figures 1 and 2. The true test of the parameters examined here will be to apply them to an untouched data set in order to estimate the effect size directly.

Throughout we only decode a word if it is timelocked to the stimulus in a consistent way across repetitions and sentences. It is possible that more complex temporal patterns of reactivation exist, but we fail to detect them due to temporal variation across sentences and/or repetitions.

Lastly, the stimuli are confounded: only inanimate nouns take the role of objects, and only animate nouns are subjects. In future data collection we will fix this confound by balancing the nouns used in the various roles. Previous work largely does not suffer from this exact confound.

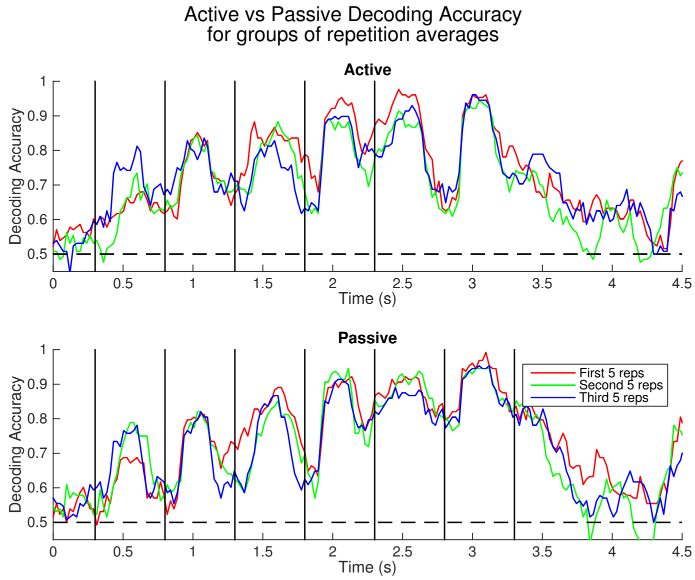

Figure 8: Accuracy for the voice classification task for three groups of repetitions averaged together. In red is the accuracy for the average of the first 5 repetitions, in green the second 5 repetitions, and in blue the third 5 repetitions. Note the discrepancies between the third group and the other two groups, particularly during the presentation of the first noun.

Future Work

These results form a clear recommendation for subsequent confirmatory analysis. The reanalysis effect is best observed with short time windows, and the neural representation at the relevant timepoints is very sparse. Furthermore, excessive numbers of repetitions can lead to participant memorization of the stimuli, therefore fewer repetitions will need to be collected. However, signal-to-noise ratio is an important factor for decoding accuracy across the board, and so we are motivated to try to collect as many repetitions as possible before reaching the point of memorization. Currently a confirmatory data set is being collected that contrasts active and passive sentences and also contains shortened sentences that consist of a single noun and verb, to test whether the reanalysis effect requires the presence of the second noun to occur. We hope to take the lessons learned from this work to confirm and better characterize the reanalysis effect.

8 Acknowledgments

This research has been supported in part by NICHD under award number R01HD075328, in part by the NSF Graduate Research Fellowship Program. Special thanks to Dan S. and Mariya for watching me give my DAP presentation twice, and for always being there to provide thoughtful feedback.

References

Bornkessel, I., Schlesewsky, M., and Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, 18(3):269–298.

Frankland, S. M. and Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(37).

King, J. R. and Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, 18(4):203–210.

Mack, J. E., Meltzer-Asscher, A., Barbieri, E., and Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain sciences*, 3(3):1198–1214.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–90.

Meltzer, J. A. and Braun, A. R. (2011). An EEG-MEG Dissociation between Online Syntactic Comprehension and Post Hoc Reanalysis. *Frontiers in human neuroscience*, 5:10.

Meyer, L., Obleser, J., Kiebel, S. J., and Friederici, A. D. (2012). Spatiotemporal dynamics of argument retrieval and reordering: an FMRI and EEG study on sentence processing. *Frontiers in psychology*, 3(December):523.

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., and Mitchell, T. (2012). Tracking Neural Coding of Perceptual and Semantic Features of Concrete Nouns. *NeuroImage*, 62(1):451–463.

Appendix A Stimuli

Active				
Art-1	Subject	Verb	Art-2	Object
A	dog	found	the	peach
The	dog	kicked	a	school
A	dog	inspected	a	hammer
The	dog	touched	the	door
The	doctor	found	a	school
A	doctor	kicked	the	peach
A	doctor	inspected	a	door
The	doctor	touched	the	hammer
The	student	found	a	door
A	student	kicked	the	hammer
The	student	inspected	the	school
A	student	touched	a	peach
A	monkey	found	the	hammer
The	monkey	kicked	a	door
The	monkey	inspected	the	peach
A	monkey	touched	a	school

Table 1: Active sentences used in the experiment.

Passive						
Art-1	Subject	Verb-Aux	Verb	Prep	Art-2	Object
The	peach	was	found	by	a	dog
A	school	was	kicked	by	the	dog
A	hammer	was	inspected	by	a	dog
The	door	was	touched	by	the	dog
A	school	was	found	by	the	doctor
The	peach	was	kicked	by	a	doctor
A	door	was	inspected	by	a	doctor
The	hammer	was	touched	by	the	doctor
A	door	was	found	by	the	student
The	hammer	was	kicked	by	a	student
The	school	was	inspected	by	the	student
A	peach	was	touched	by	a	student
The	hammer	was	found	by	a	monkey
A	door	was	kicked	by	the	monkey
The	peach	was	inspected	by	the	monkey
A	school	was	touched	by	a	monkey

Table 2: Passive sentences used in the experiment.

Engagement Question	
Question	Answer
Was there a vegetable?	N
Could you get hurt doing this?	Y
Was a tool seen?	Y
Was it bouncy?	N
Could this item be put in a pocket?	N
Was fruit damaged?	Y
Was it a hard surface?	Y
Was it bendy?	N
Was this item bigger than a whale?	N
Was it a soft item?	N
Did this involve a building?	Y
Was there something fuzzy?	Y
Was this item smaller than an elephant?	Y
Is this something you could do?	Y
Was something blue seen?	N
Was this item squishy?	N

Table 3: Comprehension questions used in the experiment.