Word Salience during Sentence Reading

Sharmistha Jat, Partha Talukdar, Tom Mitchell
Organization for Human Brain Mapping (oHBM) 2021
{sharmisthaj,ppt}@iisc.ac.in, tom.mitchell@cs.cmu.edu

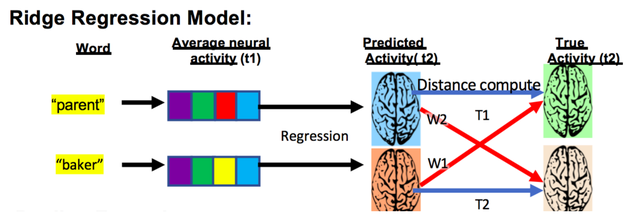
Contributions

1. Attention plays an important role in human language comprehension, where the mechanisms of syntactic and semantic integration access information presented earlier in the sentence and combine it with each newly presented word [Grace 2020]
2. The salience of a past word when encountering later parts of the same sentence is often described as a combination of that past word's degree of activation in working memory and manipulations of the focus of attention [Myachykov 2005]
3. In the work presented here, we define the saliency of earlier word w during a later time t in the sentence as the degree to which the identity of word w modulates observed MEG activity at time t.
4. In MEG data collected while human subjects read simple active and passive voice sentences, we measure and report here the saliency of nouns and verbs at each point in the sentence.
5. We consider two approaches to studying the salience of a word w presented at time t1 on the neural activity observed at later time point t2 in the same sentence.
6. We use Simple Sentence Corpus (SSC), consisting of a mix of 256k simple active and passive sentences of the form: "*the woman encouraged the girl"* and *"the woman was encouraged by the boy"* [Jat et al. 2019]

Dynamic Word Representation

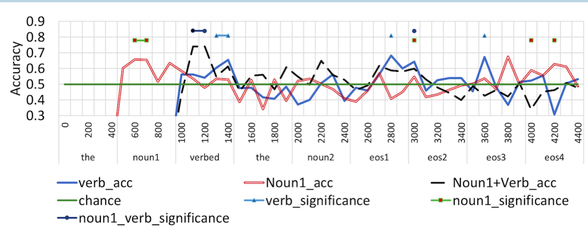
1. Through our method-1 experiments we find that earlier brain representations of the words are unable to locate word related activity later during sentence comprehension.
2. This result may indicate a dynamic representation of a word during the sentence reading
3. For example in the sentence " The apple crumble was delicious", the word "apple" may be represented by a mental image of a "red round fruit", while the same "apple" after the phrase "apple crumble" would be represented by a mental image of a brown pulp

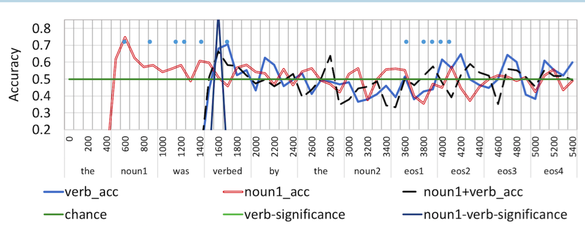
Method-1



This approach examines the degree to which the neural activity during word w at time t1 can predict aspects of observed MEG activity at the later time t2 (Figure 1). In this scheme, we represent the stimuli using the averaged neural activity (t1) for each distinct noun1, verb, and noun1+verb pair. We use the stimuli representation to predict the brain activity at each later time t2, using a linear ridge regression (Golub et al. 1979). If such prediction is successful (accuracy $a >$ chance), we deem the earlier word(s) to be salient at time t with accuracy a. In a 2-by-2 test of model performance, the classification accuracy determines whether earlier words w1 and w2 align with later observed MEG activity b1 and b2, or with b2 and b1, respectively (Mitchell 2008). We use a modification of this test by pooling three examples together to form one sample, similar to Leila et al. 2015.

Result-1

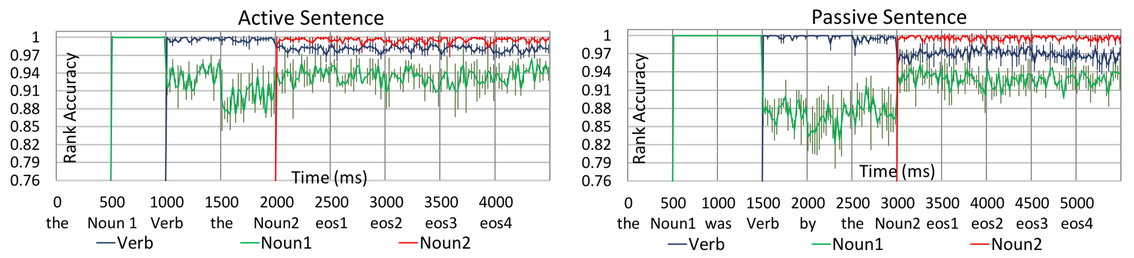




Method-2

The second approach is to use a state of the art language model, BERT (Devlin 2019), to create an evolving representation of the meaning of the sentence as new words arrive. BERT representation therefore captures the salience of word w appearing at time t1, on the evolving sentence meaning at time t2 (Figure 2)(Jat 2019). We then predict the brain activity at time t2 using the BERT models stimuli representation. We perform a rank test to probe for past word reference. More details of this method can be found in Jat et al. 2020.

Result-2



These graphs show the salience of past stimuli (noun1, verb and noun2) based on the BERT model's internal representations (Layer 18 activations)

BERT Model's word Salience

the	noun1	verb	the	noun2		
0	1	0.95	0.95	0.99		
the	noun1	was	verb	by	the	noun2
0	1	1	0.86	0.88	0.88	0.99

Input: current representation; Output: GloVe representation of past words. 6 in 1 rank accuracy measure reported; Observations: BERT model's internal context representation has varied salience of past stimuli in it's activations.

Conclusion

We propose two frameworks to detect the word salience during sentence comprehension. In our first approach we use brain activity during stimulus word to detect the saliency in later parts of the sentence. This approach discovers differences in active versus passive sentence comprehension, but might have missed other effects due to poor accuracy. We experiment with improved stimulus representation from the Deep learning model BERT. However, the BERT model's context has varied salience over the past words (Figure 2, result D), therefore the final word salience conclusion is riddled with confounds of the salience of BERT's representation. In our future work, we hope to improve the framework to discover the saliency result more robustly.

References

1. Tom Mitchell et al. (2008), 'Predicting Human Brain Activity Associated with the Meanings of Nouns', Science, vol 320,
2. Myachykov, Andriy and Posner, Michael. (2005). Attention in Language.
3. Grace W. Lindsay (2020), Attention in Psychology, Neuroscience, and Machine Learning, Frontiers in Computational Neuroscience
4. Sharmistha Jat et al. (2020), BioArxiv, Spatio-temporal Characteristics of Noun and Verb Processing during Sentence Comprehension in the Brain
5. Sharmistha Jat et al. (2019), ACL