Syntactic Dependencies in Transformers and Their Relation to the Brain

Abraham J. Fresen[*1], *Marianne de Heer Kloots*[2], *Willem Zuidema*[2]

[1]Master's Brain and Cognitive Sciences, [2]Institute for Logic, Language, and Computation

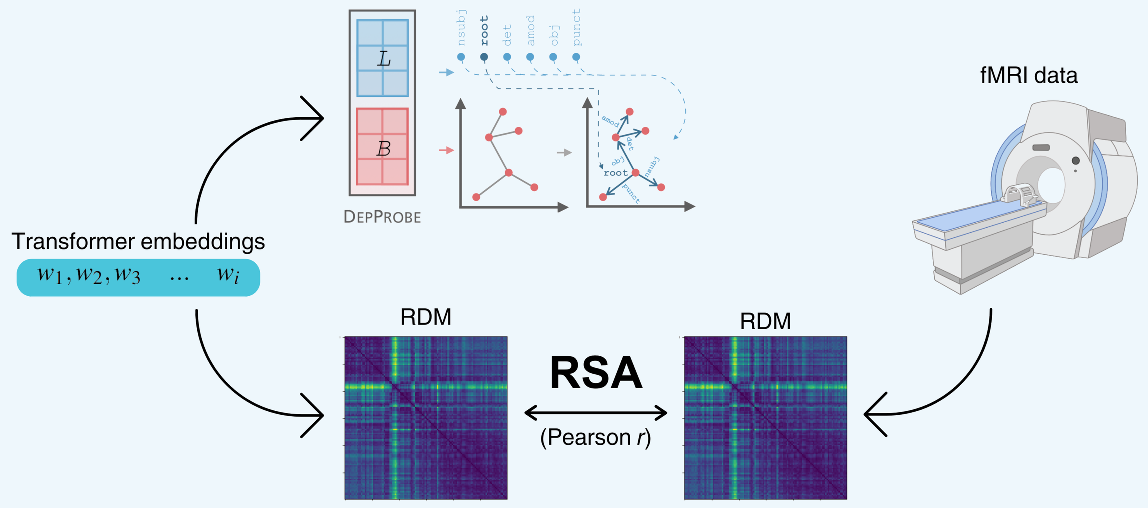
[*]**bram.fresen@student.uva.nl**

Introduction

Language models' internal representations have been shown to correlate with brain activity in language processing (Caucheteux & King, 2022). However, it is still unknown what linguistic features drive this representational similarity (RS). Additionally, it is unclear whether this differs between mono- and multilingual models. Here, we investigated the influence of accurate dependency representations on RS.

Methods

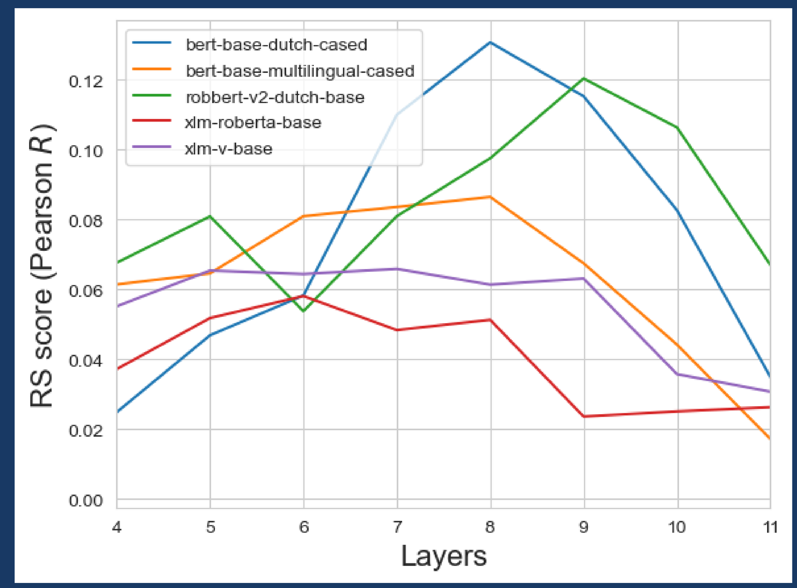
- fMRI data from the MOUS dataset (Schoffelen et al., 2019)
- Preliminary results of four participants
- Signals from the LpMTG, an area associated with syntax processing (Udden et al., 2022)
- Embeddings from monolingual and multilingual transformers
- DepProbe, a method to extract structured syntactic dependency trees from language model representations (Müller-Eberstein et al., 2022)
- Representational similarity analysis (RSA)
- We then compare the RS scores from the RSA with the metric obtained from DepProbe, the labelled attachment score

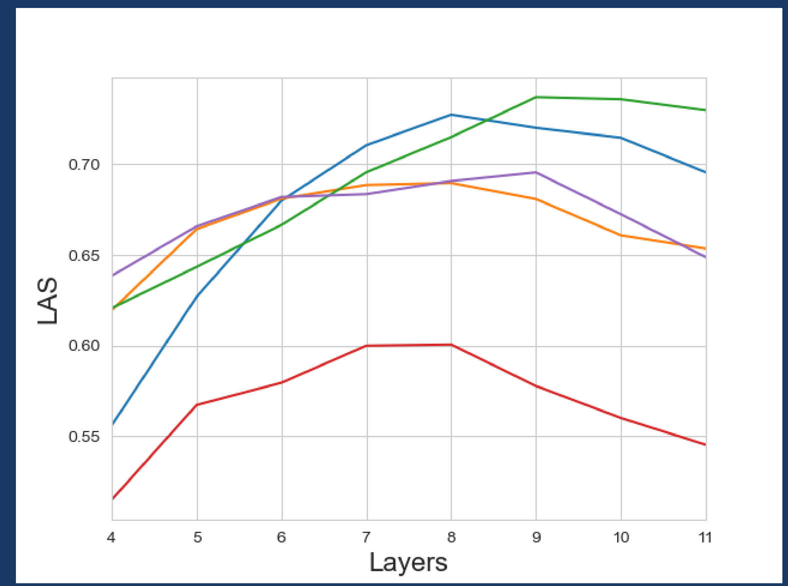


Large language models that represent **syntactic information** accurately are more **brain-like**

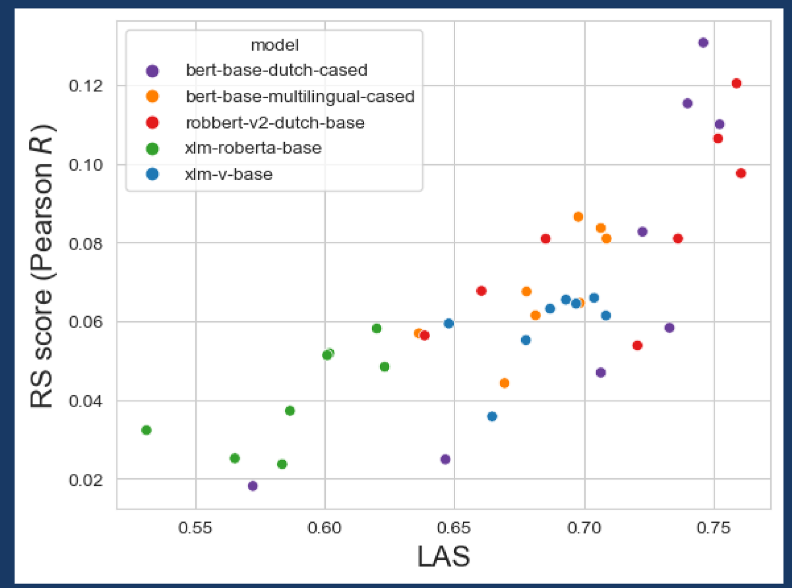
Results

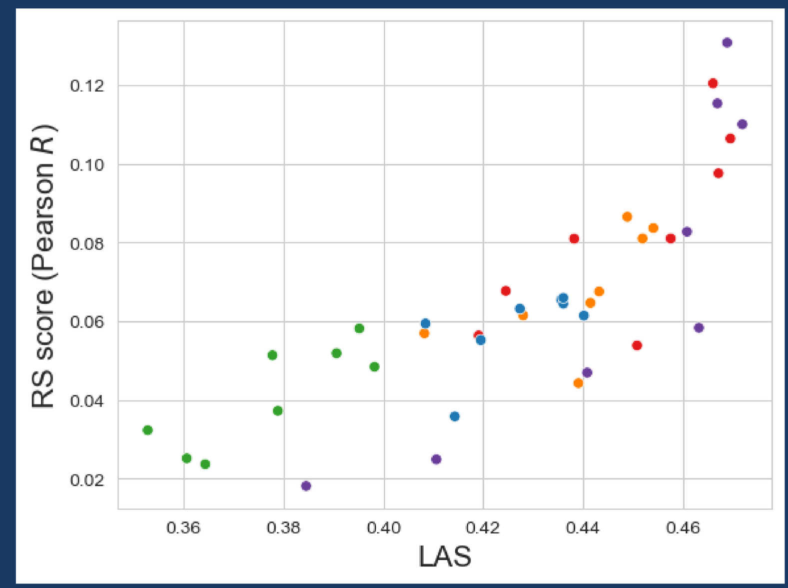
1) Monolingual models outperform multilingual models in both brain similarity and accuracy of dependency representations.



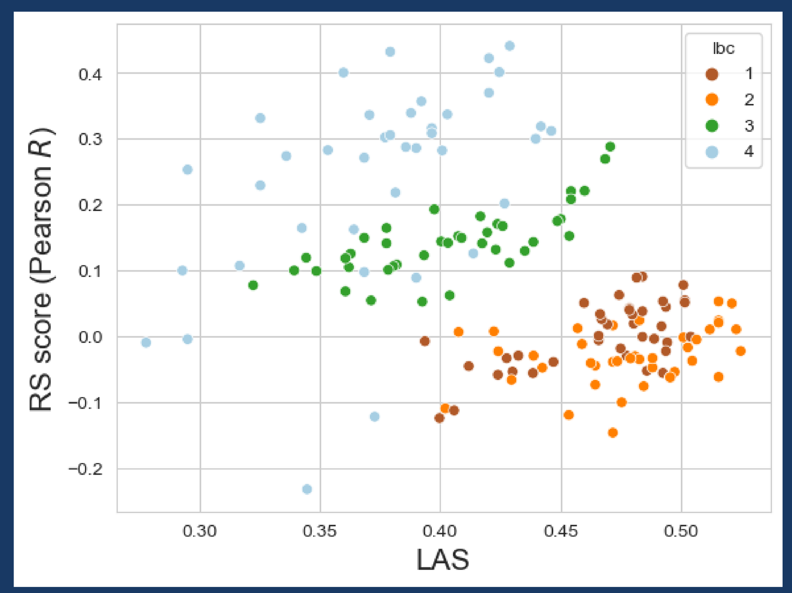


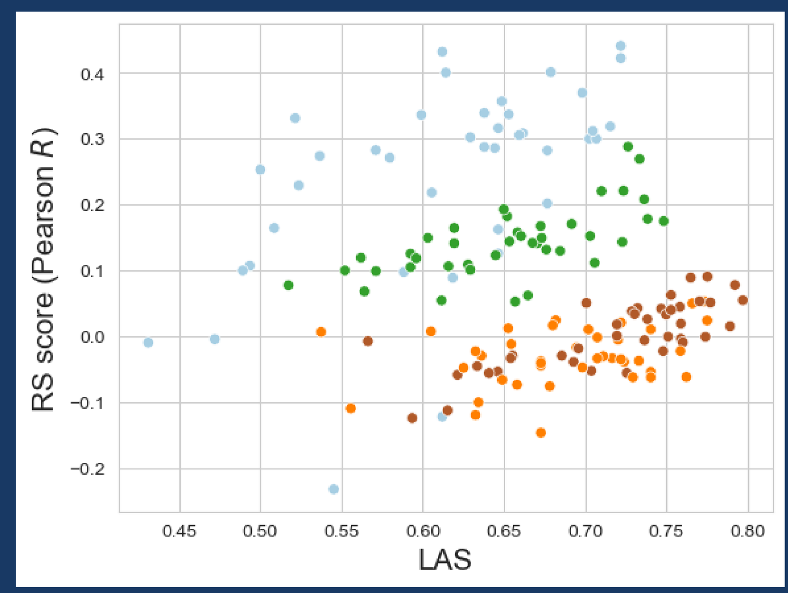
2) Models that accurately represent dependency information show greater similarity to the brain.

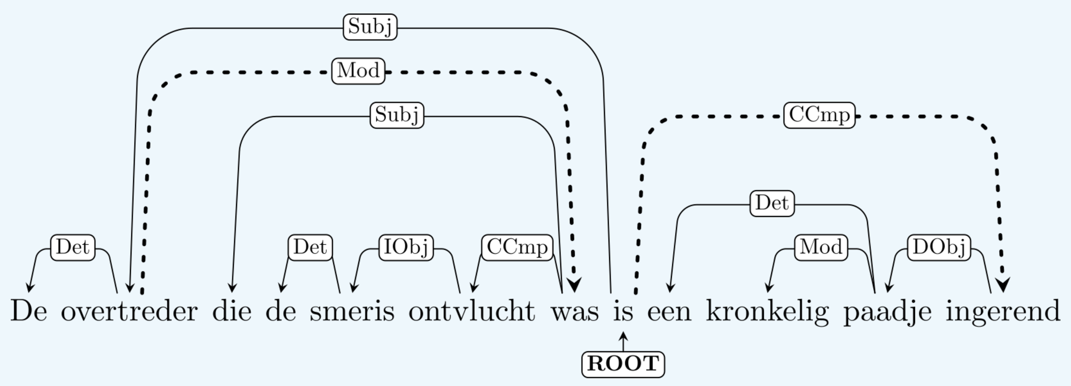




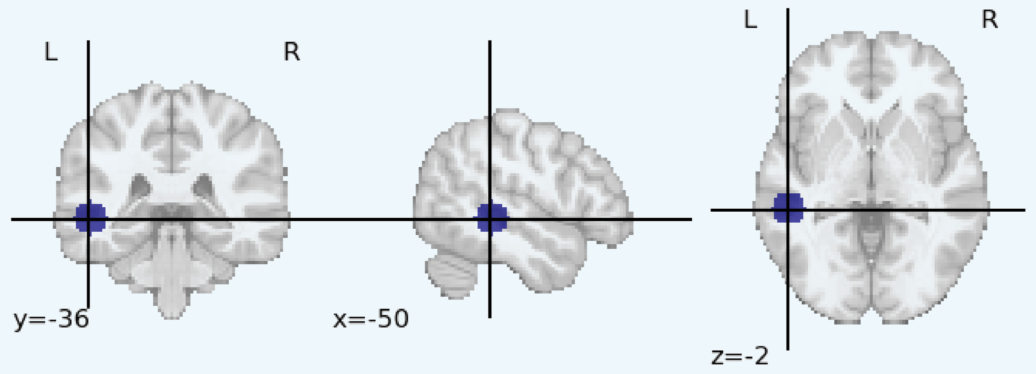
3) The relationship between brain similarity and accuracy of dependency representations is mediated by syntactic complexity.







Example sentence from the MOUS. If we parse this sentence from left to right, the maximum number of open verbal dependencies is 4 during the retrieval of the word "ontvlucht" as "overtreder" is dependent on the verb "is", "die" is dependent on the verb "was", "smeris"is dependent on the verb "ontvlucht", and "ontvlucht is dependent on the verb "was".



Region of interest used for all analyses, the left posterior middle temporal gyrus (LpMTG) .

Gold parse	Slope	Intercept	*t*	*p*	*SE*
Udden et al.	0.836	-0.288	7.870	< 0.001	0.106
Biaffine	0.425	0.037	7.543	< 0.001	0.056

Mixed linear model estimates. IV : labelled attachment score, DV : Representational similarity score, RX : model.

Model	Slope	Intercept	*r*	*p*	*SE*
BERTje	-0.03	0.744	-0.491	< 0.01	0.01
mBERT	-0.043	0.762	-0.891	< 0.001	0.004
XLM-r	-0.031	0.761	-0.625	< 0.001	0.007
RobBERT	-0.048	0.678	-0.84	< 0.001	0.006
XLM-v	-0.038	0.759	-0.856	< 0.001	0.004

Linear model estimates (biaffine parse): IV : Left branching complexity, DV : Labelled attachment score.

Model	Slope	Intercept	*r*	*p*	*SE*
BERTje	0.085	-0.1	0.775	< 0.001	0.013
mBERT	0.063	-0.097	0.518	< 0.01	0.019
XLM-r	0.121	-0.18	0.817	< 0.001	0.016
RobBERT	0.081	-0.162	0.779	< 0.001	0.012
XLM-v	0.095	-0.13	0.84	< 0.001	0.011

Linear model estimates: IV : Left branching compexity, DV : Representational similarity score.

Updates + references