

## A RAW PREDICTION RESULTS

Figure 5 shows the prediction results obtained using each feature group. To be able to better judge different levels of accuracy, instead of looking at the  $R^2$  scores, we compute  $R^{2+}$ , in which we replace the positive  $R^2$  values by their squared root, making them easier to resolve visually, and the negative ones with 0.

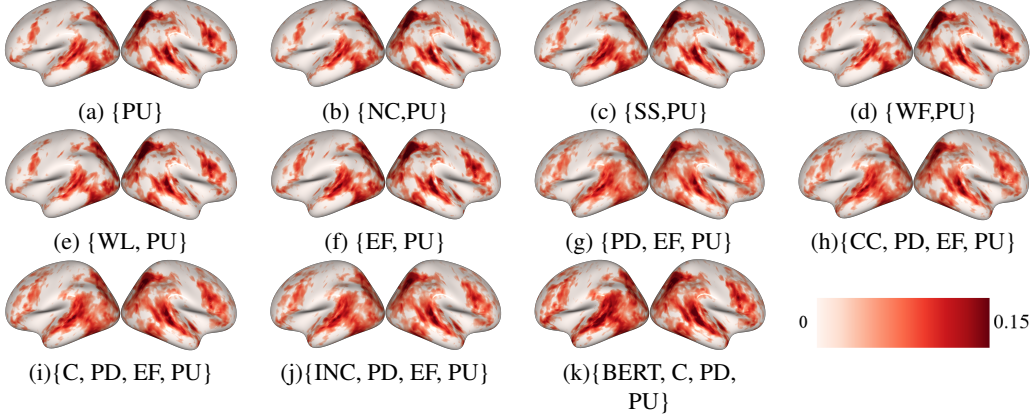


Figure 5: Cross-subject prediction performance of all syntactic feature groups. The figures show cross-subject average  $R^{2+}$  scores. Here, PU = Punctuation, NC = Node Count, SS = Syntactic Surprisal, WF = Word Frequency, WL = Word Length, EF = All effort-based metrics, PD = POS and DEP Tags, CC = ConTreGE Comp, C = ConTreGE, INC = InConTreGE, BERT = BERT embeddings and '{,}' indicates that these features were concatenated in order to make the predictions.

## B ACQUIRING AND PREPROCESSING THE fMRI DATA

We obtained the raw data from Wehbe et al. (2014). This fMRI data is acquired at a rate of 2s per image and comprise  $3 \times 3 \times 3mm$  voxels. The data for each subject is slice-time and motion corrected using SPM8 (Ashburner et al., 2008), then detrended and smoothed with an isotropic spherical Gaussian kernel with a standard deviation of  $3mm$ . The brain surface of each subject is reconstructed using Freesurfer (Fischl, 2012) and a grey matter mask is obtained. Pycortex (Gao et al., 2015) is used to handle and plot the data. All subject results are converted to MNI space using pycortex.