A Data Preprocessing

A.1 Eye-tracking Features

Description	Data source
First fixation duration	Dundee, GECO, Provo, UCL, ZuCo
First pass duration (first fixation duration in the first pass reading)	Dundee, GECO, Provo, UCL, ZuCo
Mean fixation duration	Dundee, GECO, Provo, ZuCo, CFILT-Sarcasm, CFILT-Scanpath
Fixation probability	Dundee
Re-read probability	Dundee
Total fixation duration	Dundee, GECO, Provo, ZuCo
Total duration of all regression going from this word	Dundee
Total duration of all regression going to this word	Dundee
Number of fixations	Dundee, GECO, Provo, ZuCo
Number of long regression (>3 tokens) going from this word	Dundee
Number of long regression (>3 tokens) going to this word	Dundee
Number of refixations	Dundee
Number of regressions going from this word	Dundee, Provo
Number of regressions going to this word	Dundee, Provo
The duration of the last fixation on the current word	GECO
Go-past time	GECO, Provo, UCL, ZuCo
No fixation occurred in first-pass reading	GECO, Provo
Right-bounded reading time	UCL

Table 5: Eye-tracking features provded in the gaze corpora used in this work

A.2 EEG

All four EEG datasets are converted to the EEGLab format[4], if not already provided in this format. The UCL dataset had been preprocessed by the authors. For the other three datasets, bandpass filtering, artifact removal (i.e. removing blinks and other muscle activity) and quality assessment was performed with Automagic[5].

After preprocessing and retaining only the subjects with good data quality, we use the data of 3 subjects from the N400 dataset, 14 subjects from Natural Speech, 12 subject from ZuCo and the same number of subjects as originally from UCL (i.e. 24).

A.3 fMRI

As mentioned in the main paper, we use the preprocessing pipeline from Beinborn et al. (2019) to read the fMRI data, align the scans and select the voxels. We used the Nouns and Pereira readers as is and modified the Harry Potter and Alice readers to extract word-level signals.

---

[4] https://sccn.ucsd.edu/eeglab/index.php
[5] https://github.com/methlabUZH/automagic

B   Detailed Results

embeddings	GECO	ZuCo	Provo	Dundee	Sarcasm	Scanpath	UCL
glove-50	0.010	0.008	0.031	0.010	0.016	0.023	0.044
glove-100	0.018	0.017	0.051	0.014	0.027	0.043	0.054
glove-200	0.026	0.024	0.047	0.021	0.039	0.038	0.054
glove-300	0.020	0.019	0.047	0.016	0.033	0.038	0.059
word2vec	0.015	0.011	0.024	0.014	0.022	0.018	0.028
fasttext-crawl	0.011	0.009	0.017	0.014	0.020	0.010	0.023
fasttext-wikinews	0.010	0.008	0.015	0.014	0.019	0.009	0.019
bert-base	0.007	0.003	0.006	0.008	0.009	0.006	0.003
wordnet2vec	0.018	0.012	0.027	0.016	0.023	0.017	0.040
bert-large	0.008	0.004	0.006	0.008	0.011	0.006	0.003
elmo	0.012	0.009	0.020	0.011	0.014	0.012	0.021

Table 6: Absolute mean squared error averaged over all features for each combination, i.e. averaged error of all eye-tracking features for each dataset.

embeddings	HARRY POTTER	NOUNCS	ALICE	PEREIRA
glove-50	0.005	0.204	0.036	0.044
glove-100	0.015	0.220	0.069	0.055
glove-200	0.007	0.224	0.036	0.050
glove-300	0.008	0.224	0.038	0.050
word2vec	0.005	0.209	0.010	0.044
fasttext-crawl	0.002	0.194	0.009	0.039
fasttext-wikinews	0.001	0.185	0.004	0.037
bert-base	0.001	0.042	0.001	0.012
wordnet2vec	0.015	0.203	0.050	0.046
bert-large	0.001	0.055	0.001	0.013
elmo	0.001	0.135	0.001	0.034

Table 7: Absolute mean squared error averaged over all voxels in each fMRI dataset.

embeddings	N400	NATURAL SPEECH	ZuCo	UCL
glove-50	0.067	0.014	0.009	0.009
glove-100	0.126	0.023	0.011	0.011
glove-200	0.071	0.017	0.013	0.013
glove-300	0.067	0.018	0.014	0.014
word2vec	0.047	0.017	0.012	0.012
fasttext-crawl	0.042	0.013	0.011	0.011
fasttext-wikinews	0.037	0.012	0.010	0.010
bert-base	0.014	0.005	0.006	0.006
wordnet2vec	0.089	0.020	0.015	0.015
bert-large	0.012	0.006	0.006	0.006
elmo	0.024	0.008	0.008	0.008

Table 8: Absolute mean squared error averaged over all electrodes in each EEG dataset.

B.1 Correlations between datasets

The following plots show example correlations between the prediction results within one modality, but across datasets. It shows correlations between different stimuli and different recording procedures.
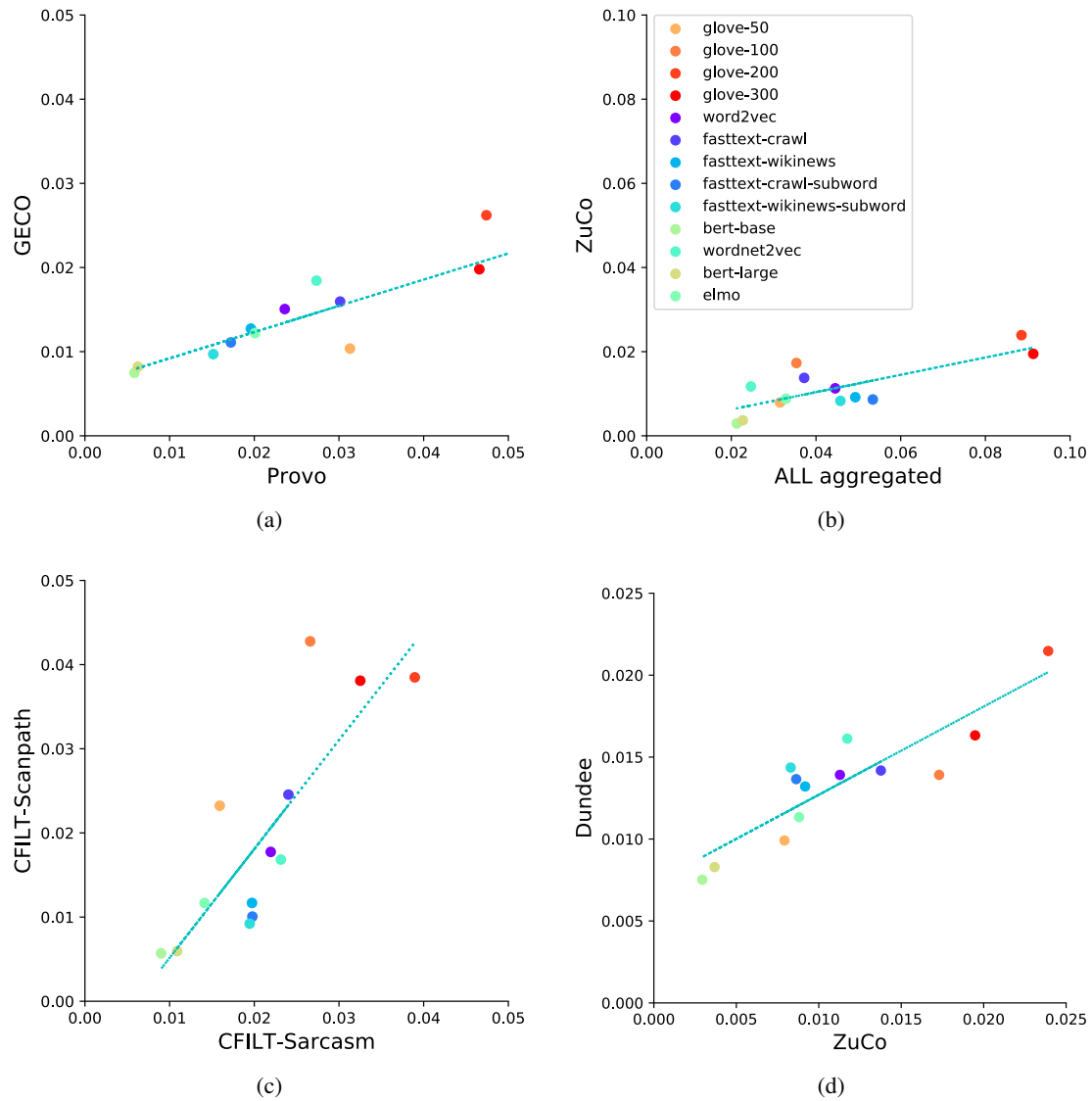
Correlations between eye-tracking datasets

Figure 9: Correlation plots between the prediction results of eye-tracking datasets.

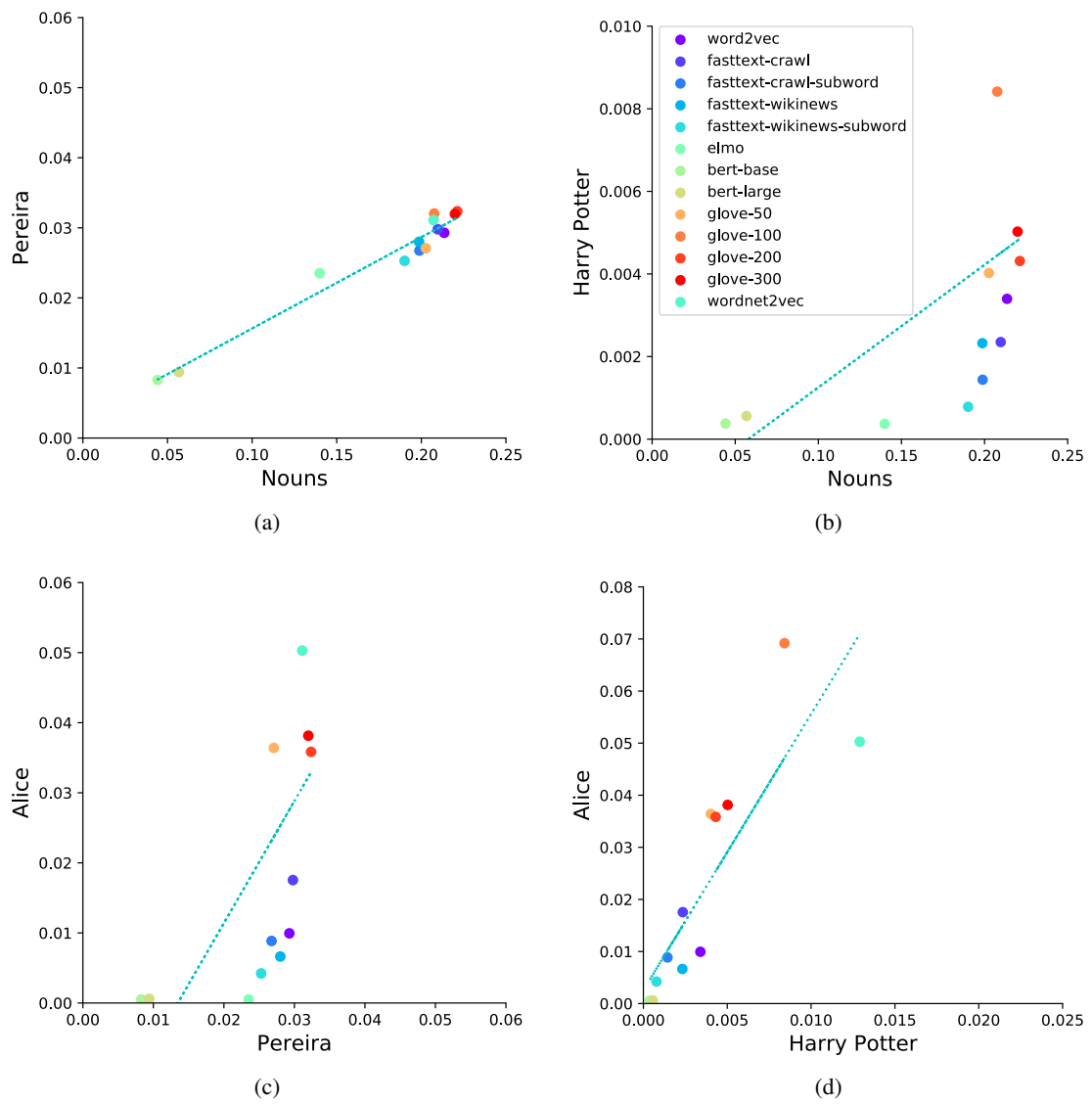**Correlations between fMRI datasets**

(a)

(b)

(c)

(d)

Figure 10: Correlation plots between the prediction results of fMRI datasets.