Table 1: A Comparison of the models explored in this paper, and the data upon which they operate.

Model Name	Section(s)	Text Data	Brain Data	Withheld Data
NNSE(Text)	2, 5	✓	x	-
NNSE(Brain)	2, 5.2.1, 5.3	x	✓	-
JNNSE(Brain+Text)	3, 5	✓	✓	-
JNNSE(Brain+Text): Dropout task	5.2.2	✓	✓	subset of brain data
JNNSE(Brain+Text): Predict corpus	5.3	✓	✓	subset of text data

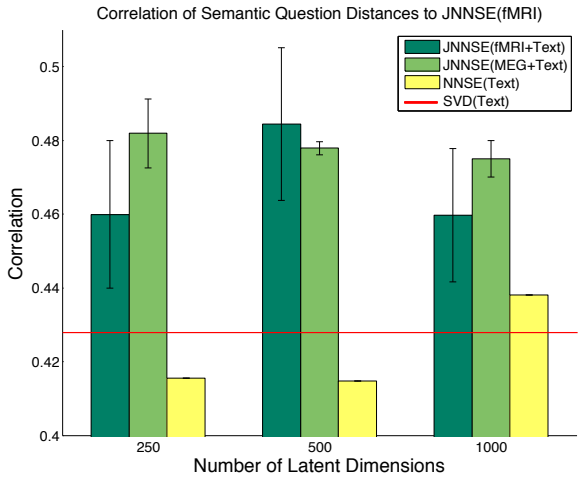

Figure 1: Correlation of JNNSE(Brain+Text) and NNSE(Text) models with the distances in a semantic space constructed from behavioral data. Error bars indicate SEM.

NNSE(Text) algorithm can be used as a VSM, which we use for the task of word prediction from fMRI or MEG recordings. A JNNSE(Brain+Text) created with a particular human subject's data is never used in the prediction framework with that same subject. For example, if we use fMRI data from subject 1 to create a JNNSE(fMRI+Text), we will test it with the remaining 8 fMRI subjects, but all 9 MEG subjects (fMRI and MEG subjects are disjoint).

Let us call the VSM learned with JNNSE(Brain+Text) or NNSE(Text) the *semantic vectors*. We can train a weight matrix W that predicts the semantic vector $\mathbf{a}$ of a word from that word's brain activation vector $\mathbf{x}$: $\mathbf{a} = W\mathbf{x}$. W can be learned with a variety of methods, we will use L_2 regularized regression. One can also train regressors that predict the brain activation data from the semantic vector: $\mathbf{x} = W\mathbf{a}$, but we have found this to give lower predictive accuracy. Note that we must *re-train* our weight matrix W for each subject (instead of re-using $D^{(b)}$ from Equation 4) because testing always occurs on a different subject, and the brain activation data is not inter-subject aligned.

We train ℓ independent L_2 regularized regressors to predict the ℓ-dimensional vectors $\mathbf{a} = \{a_1 \dots a_\ell\}$. The predictions are concatenated to produce a *predicted* semantic vector: $\hat{\mathbf{a}} = \{\hat{a}_1, \dots, \hat{a}_\ell\}$. We assess word prediction performance by testing if the model can differentiate between two unseen words, a task named *2 vs. 2 prediction* (Mitchell et al., 2008; Sudre et al., 2012). We choose the assignment of the two held out semantic vectors $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)})$ to predicted semantic vectors $(\hat{\mathbf{a}}^{(1)}, \hat{\mathbf{a}}^{(2)})$ that minimizes the sum of the two normalized Euclidean distances. 2 vs. 2 accuracy is the percentage of tests where the correct assignment is chosen.

The 60 nouns fall into 12 word categories. Words in the same word category (e.g. screwdriver and hammer) are closer in semantic space than words in different word categories, which makes some 2 vs. 2 tests more difficult than others. We choose 150 random pairs of words (with each word represented equally) to estimate the difficulty of a typical word pair, without having to test all $\binom{60}{2}$ word pairs. The same 150 random pairs are used for all subjects and all VSMs. Expected chance performance on the 2 vs. 2 test is 50%.

Results for testing on fMRI data in the 2 vs. 2 framework appear in Figure 2. JNNSE(fMRI+Text) data performed on average 6% better than the best NNSE(Text), and exceeding even the original SVD corpus representations while maintaining interpretability. These results generalize across brain activity recording types; JNNSE(MEG+Text) performs as well as JNNSE(fMRI+Text) when tested on fMRI data. The results are consistent when testing on MEG data: JNNSE(MEG+Text) or JNNSE(fMRI+Text) outperforms NNSE(Text) (see Figure 3).