Methods to Identify Semantic Content Differences

R.C.W. Versteeg - 0610097
Radboud University Nijmegen

r.versteeg@student.ru.nl

July 9, 2013

A thesis submitted in partial fulfilment of the requirements for
a degree of Master of Science in Artificial Intelligence.

Academic supervisor:
dr. M.A.J. van Gerven
Department of Artificial Intelligence
Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

Second academic supervisor:
prof.dr.ir. P.W.M. Desain
Department of Artificial Intelligence
Donders Institute for Brain, Cognition and Behaviour
Radboud University Nijmegen

Abstract

In this study we explored two distinct methods to identify semantic content differences using differences in observed brain activity measured by fMRI. The first approach is by analysing functional connectivity and the second approach is based on spatio-temporal activation patterns. The results are not conclusive and more data needs to be gathered to make statements about the neural correlates of semantics in the human brain. However, our results suggest that subtle differences in semantic content can be detected with both functional connectivity methods and by analysing spatio-temporal activation patterns.

Acknowledgements

Contents

1 Introduction

The human brain has powerful computational capabilities that cannot be recreated by artificial intelligence yet. However, the last decade, research has shown improvement in understanding the brain by using better equipment to record brain activity (Yacoub, Harel, & Uğurbil, 2008) and new methods to analyse this brain data (De Martino, De Borst, Valente, Goebel, & Formisano, 2011). By looking at patterns within the brain and the connectivity between brain regions, more can be learned about how the brain processes information and stores it. People are always collecting information when conscious. Surroundings, conversations, but also self-reflective thought elicits semantic knowledge retrieval in the human brain. Semantic knowledge retrieval is an example of something that even happens during resting states (Binder et al., 1999). Semantics is associated with the relationship between signs and symbols and the meaning of these signs and symbols. Semantics can be interpreted as how words or a collection of words map to concepts. How are signs and symbols associated with meaning in the mind? Words can be semantically closer than other words, these words could be grouped in a semantic category. For example a car and a train are both used for transportation and are more alike than a train and a plant. A question that is addressed often is how semantic knowledge is organized in the human brain.

To answer this question, studies explored how the brain represents contrasting categories, for example showing pictures of houses and pictures of faces (Kanwisher, McDermott, & Chun, 1997; Hardoon, Mouro-Miranda, Brammer, & Shawe-Taylor, 2008; Kay, Naselaris, Prenger, & Gallant, 2008). These studies identified brain regions that are sensitive to certain categories. But are these areas only sensitive to these categories? Are there more brain regions that are involved in the representation of a category? Individual differences in brain activity are present when reading or listening to stories (Buchweitz, Mason, Tomitch, & Just, 2009), but of course there are also a lot of resemblances.

Functional magnetic resonance imaging (fMRI) studies, have changed over time. More recent studies use natural stimuli, instead of static stimuli. This is how humans perceive the world in daily life (Zacks, Kurby, Eisenberg, & Haroutunian, 2011). Mitchell and colleagues tried to answer the question how the human brain represents conceptual knowledge, by creating a computational model that predicts the blood oxygenation level dependent (BOLD) response of fMRI (Mitchell et al., 2008). For every noun that is presented visually, the BOLD contrast is predicted by intermediate semantic features of the presented noun. The level of activation of every voxel for every intermediate semantic feature, with respect to the presented noun, creates the BOLD prediction of the presented noun. Mitchell and colleagues used the statistics of word co-occurrence in text associated with these nouns as intermediate semantic features, to predict the neural activation associated with thinking about word meanings. The focus on encoding of more abstract semantic concepts, instead of visual features, is an effective new way of predicting human brain activity.

In the next two paragraphs two state-of-the-art methods of analysing fMRI data will be addressed. The first method is analysing the functional connectivity of the brain during a task and the second method is analysing spatio-temporal activation patterns throughout the brain.

Functional Connectivity Not only the anatomical connections between brain areas group areas together, but brain areas can also be grouped on the basis of their so called functional connectivity (Friston, 1995; Buckner, 2010). Between brain regions that are anatomically distinct, synchronized activity can occur, which means that the regions are functionally connected (Richiardi, Eryilmaz, Schwartz, Vuilleumier, & Van De Ville, 2011). Functional connectivity is about how brain regions interact during a task, instead of analysing single voxels or brain regions in the brain independently. Brain areas can interact in a specific way during a task and this interaction, or functional connectivity, can be significantly different between tasks (Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2012). Functional connectivity can be measured by estimating covariance matrices of the acquired BOLD response during a specific task. Shirer et al. (2012) presented a model, based on such estimated covariance matrices, that could predict which task a participant was executing. This could be a resting-task, an episodic memory recall task, a silently singing of music lyrics task or a subtraction task. In this study an accuracy of 84% was achieved. Even with only one minute of data per task, the model achieved an accuracy of 80%.

Spatio-Temporal Activation Patterns Another method that has been used extensively in the last couple of years is analysing spatio-temporal activation patterns throughout the cortex. Recent brain imaging studies are more focussed on using natural stimuli, for example natural movies instead of static images (Huth, Nishimoto, Vu, & Gallant, 2012) or spoken stories instead of isolated words (Brennan et al., 2012). The classic fixed experiments are easier to conduct and seem to be easy to interpret, but what does a fixed way of presenting pictures or words tell us about the way the human brain works in real-life experiences? This motivated researchers to use natural stimuli in experiments. Brennan et al. (2012) examined the neural basis of language processing with long spoken stories. This study focussed on the syntactic structure and lexical effects. They found brain regions that are involved in syntactic structure building, which are different from studies with isolated words (Caplan, Chen, & Waters, 2008; Rogalsky & Hickok, 2009). A recent study by Huth et al. (2012) used hours of natural movies. By labelling objects in the movie and their time of occurrence, they created a category design matrix. Words were grouped into categories using WordNET (Feinerer & Hornik, 2011; Wallace, 2007; Fellbaum, 1998). Training a category model with hours of BOLD response and regularized linear regression, resulted in a cortical map of a continuous semantic space across the human brain.

In the next chapter we introduce the problem statement by introducing a research question. In Chapter 3 we describe the methods and consider two approaches to identify semantic content differences using brain activity measured by fMRI. Chapter 4 encompasses the results of the different approaches. We conclude this thesis with a discussion, including specific issues of the approaches and improvements of the pipeline for future research.

2 Research question

When listening to short stories, with the same characters and spoken by the same speaker, the stories can still be different from each other. The semantic content is guiding the story. The question is whether it is possible to classify subtle difference in semantic content using differences in observed brain activity. By using two approaches, based on functional connectivity and spatio-temporal activation patterns, we investigate our research question:

- *Can we classify perceived short stories with subtle differences in semantic content using differences in observed brain activity?*

The current study will explore state-of-the-art methods. The first aim is to zoom in on functional connectivity: will subtle differences in semantic content, lead to notable differences in functional connectivity? Shirer et al. (2012) showed that this method works with different tasks, even with one minute of data, but how about subtle differences within a listening task? We hypothesize that this is possible with short stories that are semantically different, with a proper amount of data. One minute of data is probably not enough, but with twice as much data it could work. The second is to extend findings by Huth et al. (2012) into the auditory domain and see if it is possible to classify short stories presented aurally, that are semantically different using spatio-temporal activation patterns. We hypothesize that this classification methods could also be successful, using stories with enough examples of words or categories to train our model.

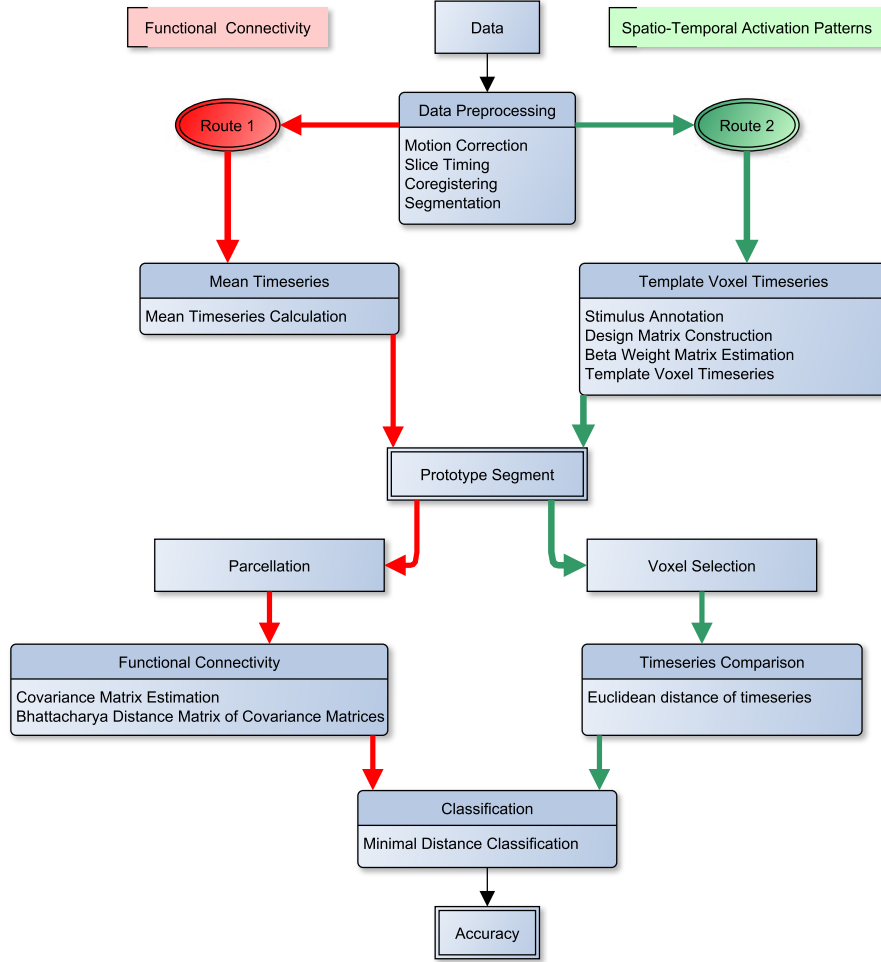# 3 Methods

Figure 1: A graph describing the methods pipeline. A choice can be made to use route 1, based on function connectivity, or route 2, based on spatio-temporal activation patterns.

3.1 Preprocessing

3.1.1 MRI data collection

MRI data were collected on a 3T Siemens Tim-Trio MRI scanner at the Donders Center for Cognitive Neuroimaging Nijmegen using a 32-channel Siemens surface

coil. Functional scans were collected using a parallel-acquired inhomogeneity-desensitized (PAID) sequence (Poser, Versluis, Hoogduin, & Norris, 2006) with repetition time (TR) = 1.650 s, echo times (TE) = 6.9 ms, 16.2 ms, 26 ms and 35 ms, flip angle = 70°, voxel size = 3.0 x 3.0 x 3.0 mm. 39 axial slices covered the entire cortex. A water excitation pulse was used for fat suppression. Anatomical data was collected on the same scanner in the same session using a T1-weighted MP-RAGE sequence with parameters TR = 2300 ms, TE = 3.03 ms, 192 slices with a voxel size of 1 mm^3.

3.1.2 Participants

Two native Dutch speakers (2 females, 24 and 28 years of age, both right-handed) participated in the study. Both participants reported that they did not suffer from any neurological disorders and gave written informed consent prior to the experiment in accordance with the guidelines of the local research ethics committee.

3.1.3 Stimuli

Data was collected with a 4×4 design, using four repetitions of four different 'Jip and Janneke' audio-stories in a pseudo-random order. All stories were spoken in Dutch by the same speaker to avoid leaking discriminative story information by speaker identity, which is not relevant for the semantic content (Charest, Pernet, Latinus, Crabbe, & Belin, 2013). The length of the stories were on average 114.5 seconds (shortest 100 seconds, longest 138 seconds). See Table 1 for more information about the stimuli.

MRI compatible headphones were used to reduce environmental noise. Before the data collection started a test-scan was executed to check the volume level of the stories.

After each story a multiple choice question was shown to encourage the participants to pay attention to the context of the stories. There was no time limit for answering the questions. For every question, four answers were shown. The participants chose the answer by using a button-box. The questions can be found in Appendix B.

Throughout this thesis, a repetition of a story (one of the sixteen) will be referred to as 'an instance'.

Type	Story 1	Story 2	Story 3	Story 4
Time (s)	138	107	100	113
Words	325	288	258	262

Table 1: The length of every story, expressed by time (seconds) and in number of words.

3.1.4 Stimulus annotation

Four stories were selected from the short audio-book stories 'Jip en Janneke' originally written by Annie M.G. Schmidt and read by Flip van Duijn. The four stories were annotated using PRAAT (Boersma & Weenink, 2010). Every word was indexed with the start time and the end time. The number of words per story varied between 258 words and 325 words per story, with an average of 288 words per story. Speaker identity was identical between stories. The audio volume between the stories is assumed to be similar, as the stories come from a published, and assumed to be mastered, compact disc. This rules out the possibility to detect differences between stories that are not due to the semantic content.

A dictionary was created with all the unique words of the stories. This resulted in a dictionary of 337 words. The dictionary can be found in Appendix C. The transcription of the four stories can be found in Appendix A.

3.1.5 Design matrix construction

For each story a design matrix X was created, in which each TR (1.65 seconds) occupies a row and each entry of the dictionary occupies a column. Resulting in a $M \times N$ matrix, where M are the words and N are the time-points. A value of one was assigned to each entry where that word appeared between this point in time and the previous point in time (one TR back in time), else a zero was assigned to the entry. This resulted in four different design matrices.

Categorization When using all words as regressors, there are not a lot of examples for every word. Instead of words, it could be useful to create design matrices with various categories. Words are grouped together as categories and these can be used as regressors. From previous research it is known that the contrast of inanimate and animate words is clearly represented in the brain (Wiggett, Pritchard, & Downing, 2009). Speer et al. (2009) showed that different brain regions track different aspects of a story, such as the physical location of a character. Categorization can also be done by using a lexical database like WordNET (Huth et al., 2012). A recent study by Çukur et al. (2013) showed that attention can warp the semantic space in the brain. The authors combined words related to people, animals and communication into one category and words related to structure, vehicles and movement into another category.

We created our own four categories based on a mixture of these studies and our own story dictionary: inanimate nouns, animate nouns, action verbs and emotion. Because WordNET is a lexical database for the English language and our stories are in Dutch, we decided to do the categorization manually.

3.1.6 fMRI data preprocessing

Preprocessing of acquired data is very important to draw conclusions about your data and especially for functional connectivity (Gavrilescu et al., 2007).

The multi-echo functional data were first combined using a home-made Matlab script, resulting in a combination of the four echo times. Dummy scans were removed and data were preprocessed using SPM8 (Ashburner et al., 2010). First functional data were motion corrected using SPM8 Realign. Successively, SPM8 Slice Timing Correction was applied and a grey matter mask was created by using SPM8 Segment. A general linear model was applied to filter out scanner drifts and motion of the participant. The Anatomical Automatic Labeling (AAL) (Tzourio-Mazoyer et al., 2002) template was co-registered with the functional mean to use as a template to group voxels together in route 1. The slow hemodynamic response was taken into account by shifting the data three TR's (4.95 s) forward in time, with respect to the story information.

3.1.7 Classification approach

To investigate if it is possible to classify stories with subtle differences in semantic content by using observed brain activity, we will look for similarities and differences between brain activity by using two distinct methods. The first method is based on functional connectivity. Covariance matrices will be estimated and a distance measure will be applied between covariance matrices. The second method is based on spatio-temporal activation patterns and a model that predicts the BOLD response of a story. In this second method a distance measure will be applied to compare timeseries. In the pipeline, two different routes corresponding to the aforementioned approaches can be followed. These will be explained in detail in the upcoming sections.

3.2 Route 1: Functional connectivity

In this thesis there will be two different main classification pipelines that will be explained and reviewed. The first route in the pipeline is based on the correlation structure between various parts in the brain during a story. This is taken to reflect functional connectivity (Buckner, 2010).

3.2.1 Prototype segment: Concatenated timeseries

A prototype segment will be computed in both routes, to compare the actual segment or instance with the prototype segments. In this route the prototype segments will be created by concatenating timeseries of a story type. The concatenated timeseries for every story type is composed for every repetition independently, by leaving this repetition out, to avoid double dipping (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). The BOLD response of the instances from the other repetitions of the same story type are concatenated into story-specific prototype segment. This results in four prototype segments per repetition, that represents prototype BOLD response of a story type.

3.2.2 Grey matter parcellation

To make the acquired data computationally tractable and interpretable, voxels were grouped together by coregegistering the AAL template with the anatomical data of the participant. The AAL template is a broadly used template in neuroimaging to map different brains to the same space. Moreover, it can be used to generalize over multiple participants. All the voxels that are selected for a region are normalized and averaged for an instance, to get an estimate of the region-specific BOLD activation for every instance. The cerebellum and vermis are removed from the template (region 91 up to region 116), because these regions are predominantly related to movement coordination and balance (Rapoport, van Reekum, & Mayberg, 2000).

Y^i is the data matrix of an instance i, with dimensions $N \times V$, where N are the timepoints and V are the voxels. For every instance the mean activation for every AAL region is calculated as follows.

$$M^i = (\mu_1^i, \ldots, \mu_{90}^i)'$$ (1)

where

$$\mu_k^i = \frac{1}{V^k} \sum_{v \in V(k)} y_{:,v(k)}^i$$ (2)

where V^k is the number of voxels in region k and $v(k)$ is the index set of voxels beloning to region k. The average activation of these voxels during instance i are summarized in μ_k^i

3.2.3 Covariance matrix estimation

An empirical covariance matrix was estimated for every repetition of every story (16 covariance matrices). Every matrix contains 90 x 90 cells and represents the covariance between the brain regions defined by by the AAL template, for the particular instance. Covariance reflects how two regions vary together.

The covariance matrix Σ of instance i will be denoted as:

$$\Sigma_i = \begin{pmatrix} \mathrm{Cov}(\mu_1^i, \mu_1^i) & \cdots & \mathrm{Cov}(\mu_1^i, \mu_{90}^i) \\ \vdots & & \vdots \\ \mathrm{Cov}(\mu_{90}^i, \mu_1^i) & \cdots & \mathrm{Cov}(\mu_{90}^i, \mu_{90}^i) \end{pmatrix}.$$ (3)

where the covariance is calculated using the following general covariance formula:

$$\mathrm{Cov}(p, q) = \frac{\sum_{n=1}^{N}(p_n - \bar{p})(q_n - \bar{q})}{N - 1}$$ (4)

where N is the number of observations, p and q are, in our case, datapoints of two regions, $\bar{p}$ and $\bar{q}$ are the sample means of these regions.

The structure in the covariance matrix can teach us about the functional connectivity between brain areas during the story. How the brain areas interact during a story could be different between the different stories or repetitions. If this is the case, this can serve as model to predict which story was presented, by comparing the test-data with a 'prototype' covariance matrix of the different stories by using a similarity measure, like the Euclidean distance (Deza & Deza, 2009), the Kullback-Leibler divergence (Kullback & Leibler, 1951) or the Bhattacharyya measure (Bhattacharyya, 1943).

3.2.4 Similarity measure: Functional connectivity comparison

To measure the similarity between the covariance matrices of the stories, different methods can be used as outlined before. In this case the Bhattacharyya distance will be applied, which is known to operate correctly on multivariate normal distributions, which we assume we are dealing with (Bhattacharyya, 1943).

The Bhattacharyya distance between two stories is computed by the following formula, in which Σ_i and Σ_j are covariance matrices.

$$S_{ij} = \left(\frac{\frac{1}{8} \left(\mu_i - \mu_j \right)^T}{P} \right) \left(\mu_i - \mu_j \right) + \frac{1}{2} \log \left(\frac{\det(P)}{\det(\Sigma_i) \det(\Sigma_j)} \right) \tag{5}$$

with $P = (\Sigma_i + \Sigma_j)/2$. S_{ij} will be the matrix with corresponding distances between Σ_i and Σ_j.

3.2.5 Classification

S_{ij} is the distance matrix which represents which instances are most similar. The stories will be compared by using a minimal distance approach. If the instance with the closest distance to the evaluated instance has the same story type as the evaluated instance, the classification is correct and a one is assigned. Else the classification is false and a zero is assigned. This will result in 16 binary decisions, which will be averaged to get the accuracy of this classification pipeline.

3.2.6 Story Difference Measure

The stories that are presented in the experiment are quite similar. Same characters, similar words and all have a childish feel to it, with a happy ending. To reflect how different the stories are represented in a brain, a story difference measure can be calculated with the following formula.

$$\delta = \bar{S}_i - \bar{S}_j \tag{6}$$

where $\bar{S}_i$ is the mean Bhattacharyya distance between all instances with a different story type and $\bar{S}_j$ is the mean Bhattacharyya distance between instances with the same story type. This results in a story difference measure. The bigger the number, the bigger the semantic difference between the stories. It could be the case that the stories are too similar for a good classification performance for this method.

3.3 Route 2: Spatio-temporal activation patterns

In this second approach a model is trained to predict the BOLD response of a new story. The model is trained by mapping words to BOLD responses from the training set, this mapping is used to predict a BOLD response of a new story. To validate this model, real BOLD response for a new story can be compared to the predicted BOLD response of the model for the same story.

3.3.1 Prototype segment: template voxel timeseries

To train the model, a leave one repetition out method is applied. Three of the four repetitions (12 instances) are used to train the model. First, for each story a mean average of the BOLD response is computed by averaging the BOLD response of the three instances of the story used for training. The formula used to train the model is based on kernel ridge regression (Hoerl & Kennard, 1970; J. H. Friedman, Hastie, & Tibshirani, 2010):

$$B = (YY' + \lambda I_N)^{-1} X'Y \tag{7}$$

Here, Y represents the mean BOLD activation of the stories with dimensions $(N \times V)$, where V are the voxels and N are the timepoints. X represents the concatenated design matrices of the stories $(N \times M)$, I_N the identity matrix of size $N \times N$, B $(M \times V)$ the trained weights between the words and the voxels and λ the regularization parameter.

With the estimated B values predictions can be made for new stories, by multiplying it with a (new) designmatrix.

$$\hat{Y} = X^* \times B, \tag{8}$$

where X^* is the design matrix of the new story and B are the trained weights from Equation 7.

3.3.2 Voxel selection

By using the AAL template in combination with a grey matter mask, a lot of voxels are discarded from the analysis. This still results in a large amount of voxels (¿15.000). Voxel selection could be done more thoroughly. For example areas could be discarded based on the literature, or voxels can be selected on the basis of their explained variance. In this analysis we only use the grey matter mask and the AAL template to select voxels.

3.3.3 Similarity measure: Timeseries comparison

One repetition, consisting of one instance of every story, is left out to use as test data while training the model, this is later cross validated. With the model, BOLD responses are predicted for every story type. This is done for every repetition. The predictions are compared with the actual BOLD response to validate and test the model. The comparison is done by using a distance measure between the predicted BOLD response and the actual BOLD response of each instance. Because the story types do not have the same length, the first 50 timepoints are used for comparison to prevent correct classification based on the length of the story. The distance between the actual and predicted BOLD response is calculated with the following formula:

$$D_{i,t} = \sqrt{\sum_{v=1}^{V} \sum_{n=1}^{N} (\hat{y}_{v,n}^t - y_{v,n}^i)^2}, \tag{9}$$

where $y_{k,n}^i$ is the actual BOLD response of instance i, where v are the voxels that are selected for validation and n are the first 50 timepoints. $\hat{y}_{v,n}^t$ is the predicted BOLD response of story t.

3.3.4 Classification

This results in a comparison of predicted BOLD response of a story type with actual BOLD responses of all four story types. If distance d is smallest for the comparison where the story types of the predicted BOLD response and the actual BOLD response are coherent, the predicted BOLD response of that story type is most similar to the actual BOLD response of the similar story type. In this case a one will be assigned to binary decision value g_i, else g_i will be zero. This will result in 16 binary decisions and can be averaged to get an accuracy value a.

$$a = \frac{1}{I} \sum_{i=1}^{I} g_i \tag{10}$$

where I is the number of instances, g_i the decision value and a the accuracy.

4 Results

4.1 Route 1

The classification accuracies obtained in route 1 of the pipeline are shown in table 2, including the story difference measure.

	Accuracy	*Story Difference Measure*
Participant 1	0.25 (4/16)	0.03
Participant 2	0.31 (5/16)	0.21

Table 2: The results of route 1. *Accuracy* is based on the correct number of classifications. The *Story Difference Measure* is the mean Bhattacharyya distance between instances that have different story types subtracted by the mean Bhattacharyya distance between instances that have the same story type.

According to the binomial test (Howell, 2007), an accuracy of 0.5 is enough to achieve significance in this case (p = 0.05).

The greater part of the instances are classified incorrectly. The congruent colour patterns across the prototype covariance matrices in the design matrices, could be an artefact of time. This does not influence the outcome of the classifier, because the classifier operates on the rows, not the columns. This is exactly the reason why we used the leave-one-out and cross validation method.

Because this method can not predict the right story type from the BOLD response, data will be simulated in the next section, to get a better understanding why this method does not provide a better classification performance.
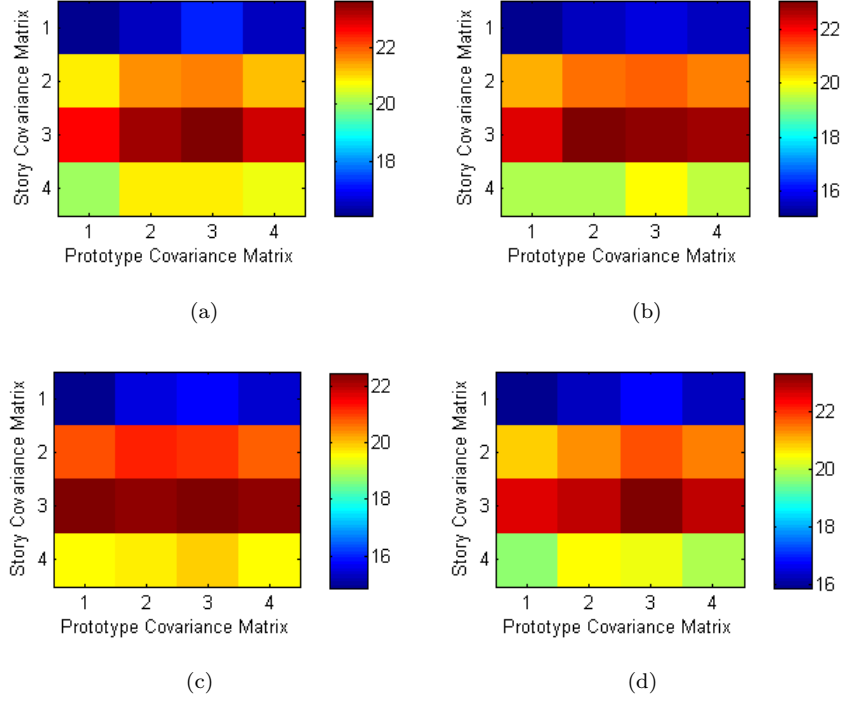
Figure 2: Distance matrices route 1 for Participant 1, expressed in terms of the Bhattacharyya distance. The story covariance matrices are compared with the prototype covariance matrices. If the prototype covariance matrix with the minimal distance to the story covariance matrix correspond to the story type of the story covariance matrix, classification is successful, else it failed. Repetition 1 is represented in subfigure (a), repetition 2 in subfigure (b), repetition 3 in subfigure (c) and repetition 4 in subfigure (d).
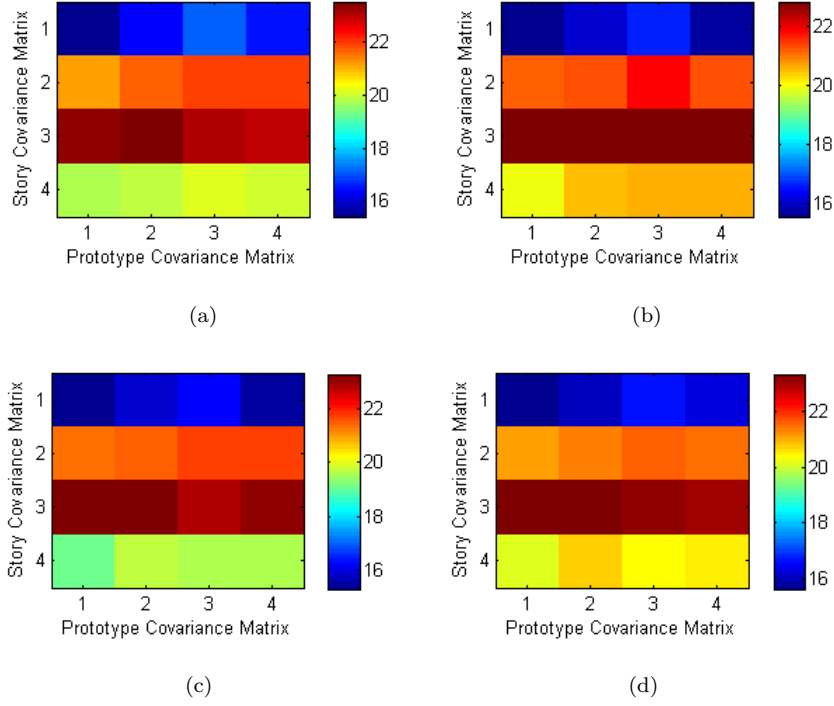
Figure 3: Distance matrices route 1 for Participant 2, expressed in terms of the Bhattacharyya distance. The story covariance matrices are compared with the prototype covariance matrices. If the prototype covariance matrix with the minimal distance to the story covariance matrix correspond to the story type of the story covariance matrix, classification is successful, else it failed. Repetition 1 is represented in subfigure (a), repetition 2 in subfigure (b), repetition 3 in subfigure (c) and repetition 4 in subfigure (d).

14

4.1.1   Simulated data results

To validate the pipeline and see if the obtained results from our participants can be reproduced, additional simulations were executed. Data was sampled out of randomly generated positive definite covariance matrices. Every story type for every sample iteration had its own covariance matrix, which represents the underlying population correlation matrix. The sampled data is used as if it was real data, hence route 1 of the pipeline was followed again. If the stories are close together, more timepoints will be needed to achieve a good performance. The empirical data only contains 50 timepoints, but if the stories are different enough, the amount of samples can be enough.

In comparison to the empirical data results, the simulated data results are coherent. With only 50 samples and a story difference measure of 0.0314 for Participant 1 and 0.2091 for Participant 2 will not lead to a performance that is accurate.

More timepoints will lead to a better estimation of the underlying population correlation matrix, as can be seen in Figure 4(b), Figure 4(c) and Figure 4(d).
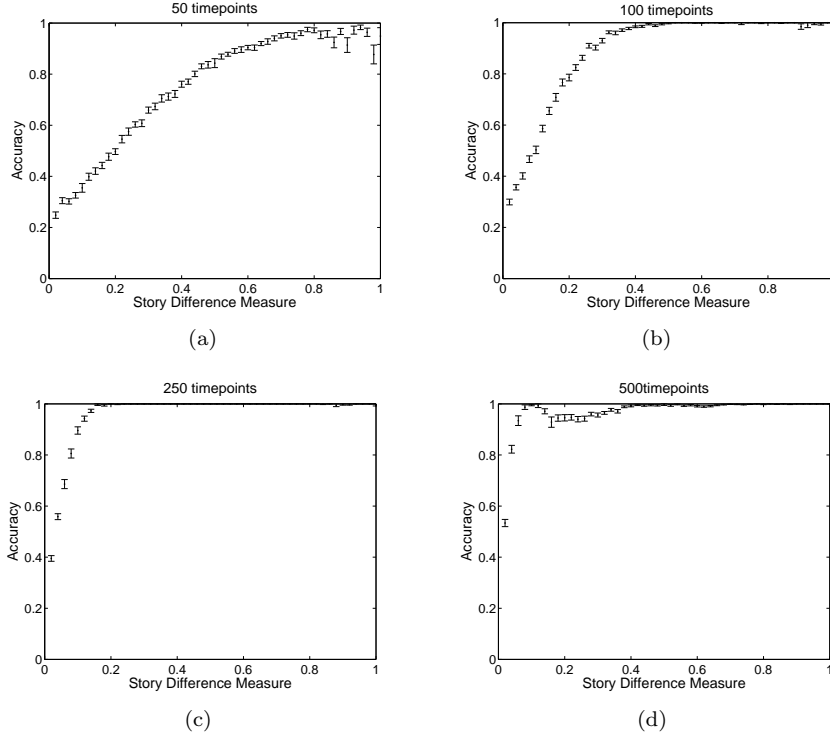
Figure 4: Results of the simulated data. Different number of timepoints are used: 50 (a), 100 (b), 250 (c) and 500 (d) timepoints per instance and different story difference measures are used throughout the samples. Y-axis represents the accuracy and X-axis the story difference measure. A total of 5000 samples in each graph, averaged in bins of 0.02.

16

4.2   Route 2

For this route, different methods were used to examine the classification perfor-
mance. As stated in the methods section, all words could be used as regressors,
but words could also be grouped together to have more examples per regressor
and less regressors. A combination of two categories can be chosen from the
manually composed categories: animate, inanimate, action verbs and emotion.
The categories could also be reviewed on their own. The classification accuracies
obtained with a single category as regressor, with two categories as regressors
and with three and four categories as regressors are shown in Figure 5.
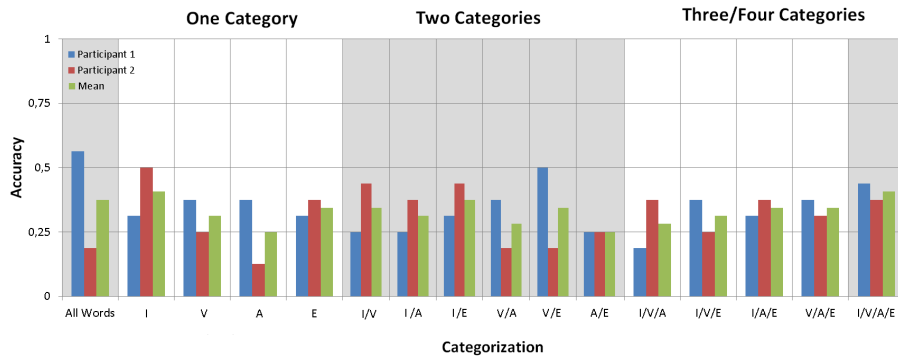
Figure 5: Classification performance, based on the combination category re-
gressors. 'A' is animate, 'I' is inanimate, 'V' is the action verbs category and
'E' is emotion. The category 'All Words' is the initial condition: all words are
regressors.

For Participant 1, the best classification accuracy (0.5625) was obtained in the *All Words* condition. This means that all words from the dictionary were used as regressors, no additional categorization was done. The p-value of this classification accuracy calculated with the binomial test is 0.0075. Note that after applying the conservative Bonferonni correction, to correct for multiple comparisons ($p < 0.003125$), the outcome of this test is statistically not significant (Bland & Altman, 1995).

For Participant 2, the best classification accuracy (0.5) was obtained in the *Inanimate* single category condition. One regressor was used to create the design matrices. This regressor reflected when a inanimate word was presented in the stories. The p-value of this classification accuracy calculated with the binomial test is 0.0271. Note that this performance value is not corrected for multiple comparisons. Correcting for multiple comparisons using the Bonferonni correction ($P < 0.003125$ level), means that this test is statistically not significant.

The mean of Participant 1 and Participant 2 over all conditions, resulted in the best classification accuracy (0.4375) in the *Four Category* condition. Groups of inanimate, animate, emotion and action verbs were created and used as regressors. The p-value of this classification accuracy calculated with the binomial test is 0.0796 for participant one and 0.1897 for participant two. Note that this performance value is not significant when corrected for multiple comparisons, using the conservative Bonferonni correction.

**Explained Variance** Additionally to statistical p-value, the mean explained variance of the estimated timeseries on the real timeseries is plotted for the best conditions for each participant. This analysis can shed light on the brain regions that are responsible for the obtained performances. Only voxels with positive explained variance values are plotted to spot blobs of voxels that are responsible for a high explained variance. The plot for Participant 1 in the best condition *All Words* is shown in Figure 6. The plot for Participant 2 in the best condition *Inanimate* is shown in Figure 9. Plots for both participants in the mean best condition *Four Categories* are shown in Figure 11(a) and Figure 11(b).

By averaging and standardizing the explained variance per region (negative and positive values) for each participant in the best condition, the best explaining brain areas can be selected. Figure 7 shows the mean explained variance for Participant 1 in condition *All Words*. Figure 10 shows the mean explained variance for Participant 2 in condition *Inanimate*. Figure 12 shows the cumulative mean explained variance for Participant 1 and Participant 2 in condition *Four Categories*. A complete overview of the brain regions covered by the AAL template can be found in Appendix D.

Figure 6: Brain plot of voxels with positive mean explained variance values, plotted on the corresponding points in the AAL atlas for Participant one in the *All Words* condition. Blobs of voxels can be detected were the explained variance is high. Brain regions that contain these blobs are the inferior frontal gyrus, the superior temporal gyrus, the middle temporal gyrus and the calcarine sulcus.

Figure 7: This graph shows the standardized mean explained variance per brain region. The bars represent the standardized mean explained variance of Participant 1 in the *All Words* condition per AAL brain region. The Y-axis represents brain regions according to the AAL template, the X-axis is the mean standardized explained variance of the corresponding region. The overall best five explaining brain areas for Participant 1 in these conditions are AAL region 43 (left calcarine sulcus), 51 (left middle occipital gyrus), 45 (left cuneus), 81 (left superior temporal gyrus) and 54 (right inferior occipital gyrus). The five worst explaining brain areas are AAL region 5, 69, 10, 26 and 73.

Word	nVox	freq	St1	St2	St3	St4
Aanhollen	1	1				1
Allebei	2	2			2	
Als	2	6	1	4		1
Angstig	1	1				1
Best	1	1		1		
Bij	3	9	9			
Blaast	1	1			1	
Eigenlijk	1	1			1	
Elk	42	1		1		
Etensbak	4	2	2			
Fijne	1	1	1			
Grove	1	1	1			
Heb	1	5	3			2
Hebben	1	3		1		1
Helemaal	1	1			1	
Hij is	1	1	1			
Koel	1	2			2	
Kwaad	1	1	1			
Lange	6	1				1
Meenemen	1	1		1		
Morgen	4	1		1		
Nare	2	1			1	
Pakje	2	1	1			
Pijn	1	1				1
Scherp	2	2				2
Stuiver	1	2		2		
Tis	2	1	1			
Verder	1	1		1		
Voortaan	2	1				1
Wonen	7	1	1			
Zeurt	3	1			1	
+	100	55	22 (11)	13 (8)	9 (7)	11 (9)

Table 3: Words with the largest regressor coefficients of the 100 voxels with the highest explained variance. First column are the words, second column the number of voxels that have this word as highest regressor coefficient, third column the number of occurrences of the word in all stories and the last four columns the distribution of these occurrences over the story types.

Because these results of Participant 1 in the condition *All Words* has voxels with high explained variance, it is interesting to see how these good voxels map to the words that are presented, by examining the regressor coefficients of these voxels. The results of the best 100 voxels can be found in Table 3.

A further analysis of the best regressor shows that the BOLD response of the corresponding voxel shows a big increase in activity at the timepoint the word is presented. A graph of this BOLD response is shown in Figure 8.
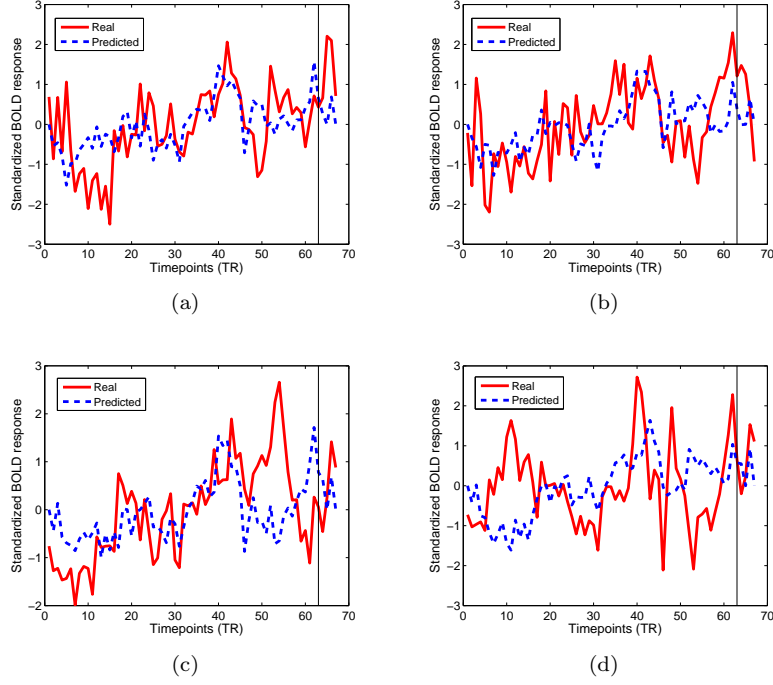
Figure 8: These plots represent the predicted BOLD response of a repetition of story 2, reflected by the real BOLD response of that repetition in the voxel with the highest explained variance. The largest regressor coefficient of this voxel mapped to the word 'elk' and was presented at timepoint 63. Repetition 1 is represented in subfigure (a), repetition 2 in subfigure (b), repetition 3 in subfigure (c) and repetition 4 in subfigure (d).
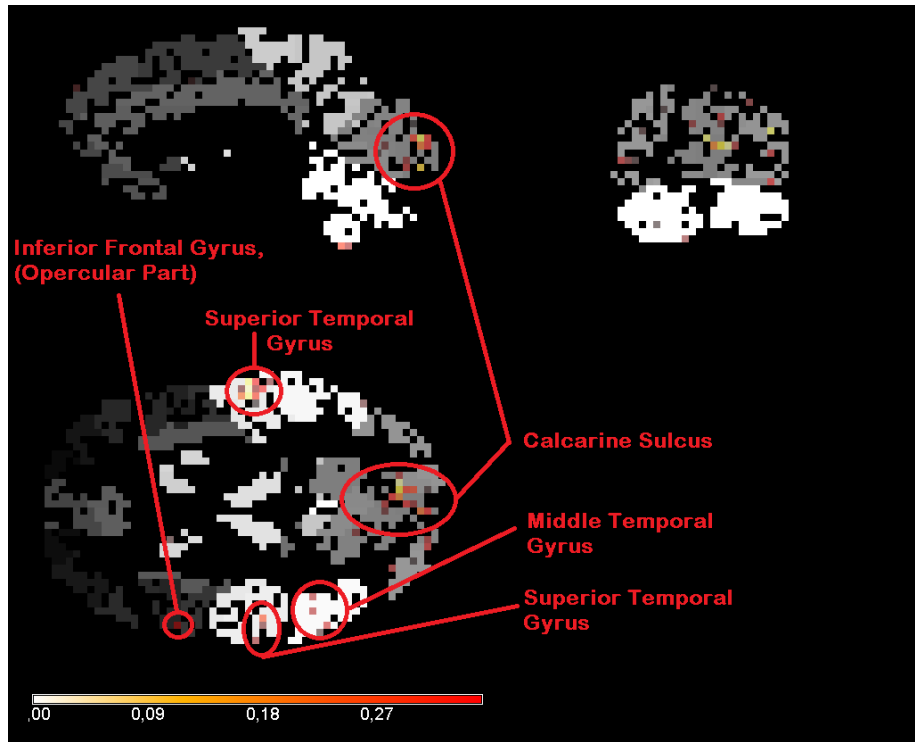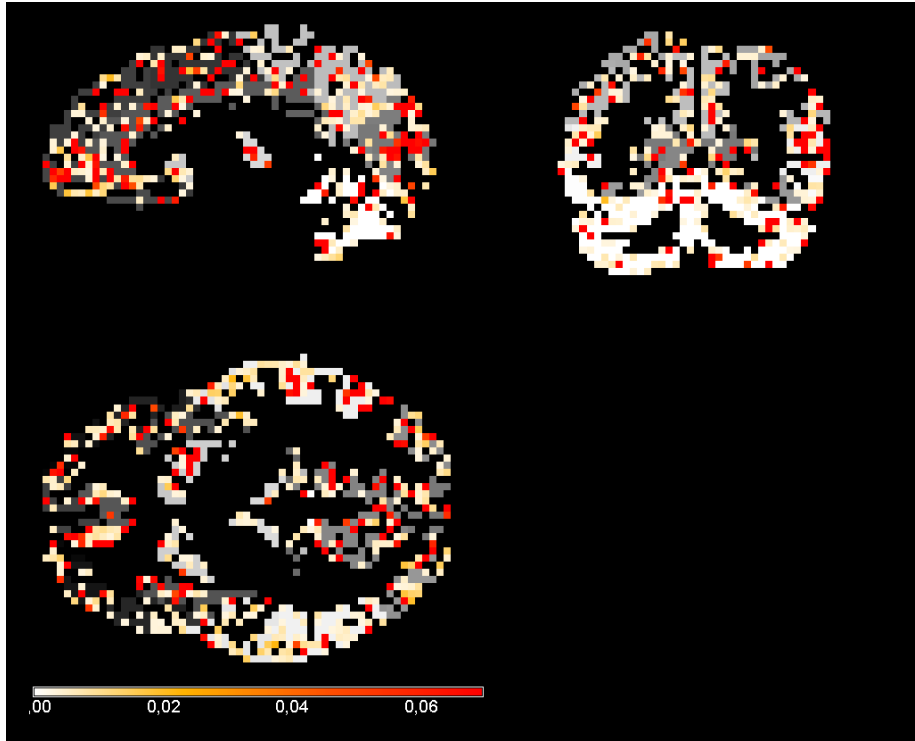
Figure 9: Brain plot of voxels with positive mean explained variance values, plotted on the corresponding points in the AAL atlas for Participant two in the *Inanimate* condition. The voxels that have a high explained variance are widespread throughout the cortex. No big blobs can be detected in the plots and the mean explained variance values of the voxels are not very high.
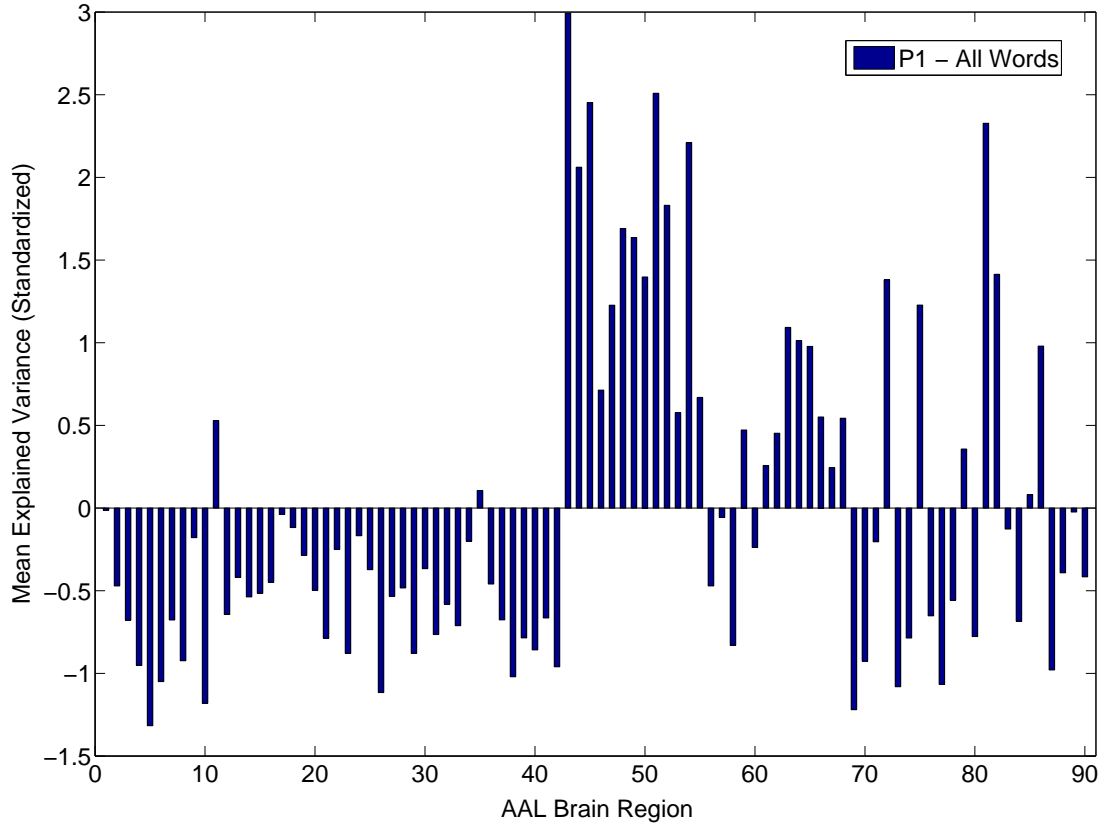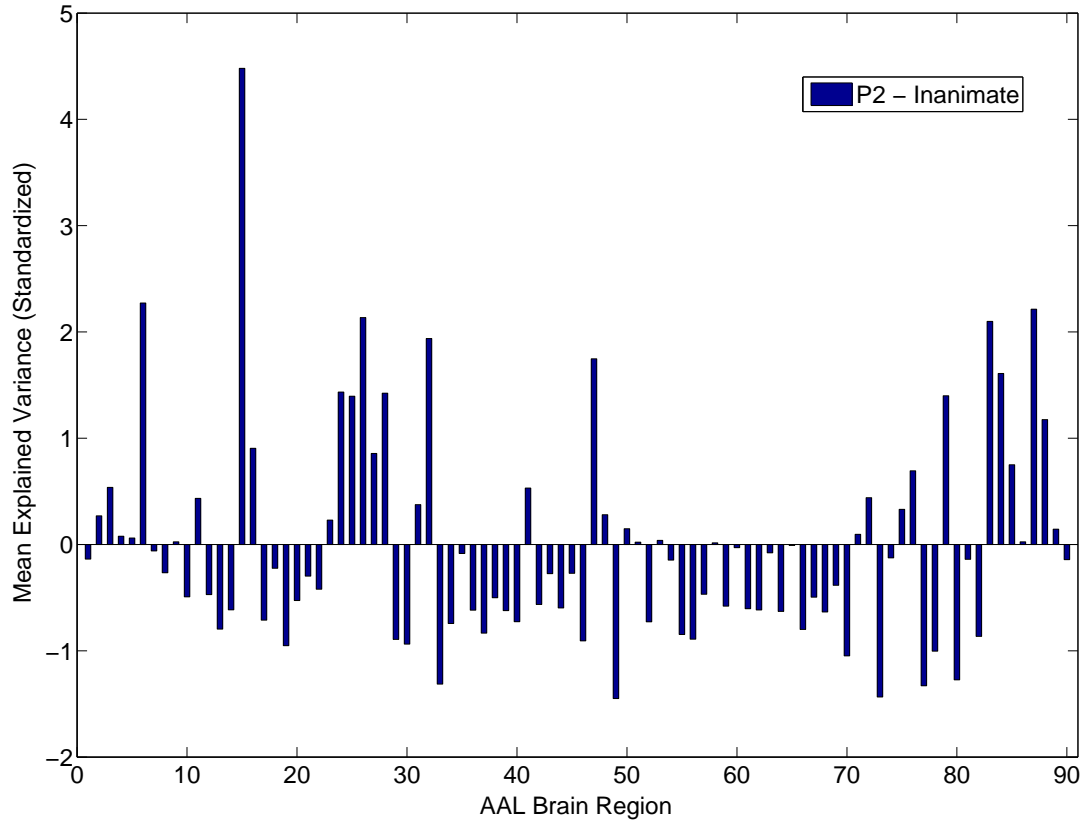
Figure 10: This graph shows the mean standardized explained variance per brain region. The bars represent the standardized mean explained variance of Participant 2 in the *Inanimate* condition per AAL brain region. The Y-axis represents brain regions according to the AAL template, the X-axis is the mean standardized explained variance of the corresponding region. The overall best five explaining brain areas for Participant 1 in these conditions are AAL region 15 (left inferior frontal gyrus, orbital part), 87 (left middle temporal gyrus, temporal pole), 6 (right superior frontal gyrus, orbital part), 26 (right middle frontal gyrus) and 83 (left superior temporal gyrus, temporal pole). The five worst explaining brain areas are AAL region 70, 33, 73, 49 and 78.
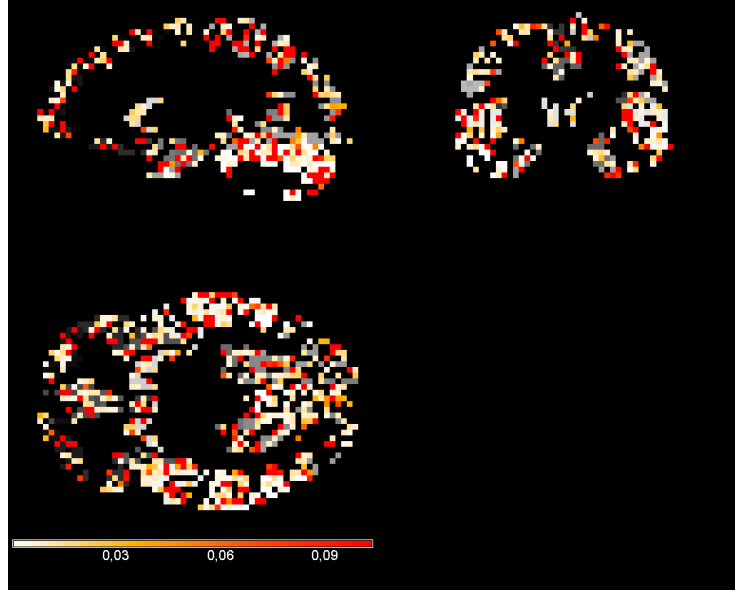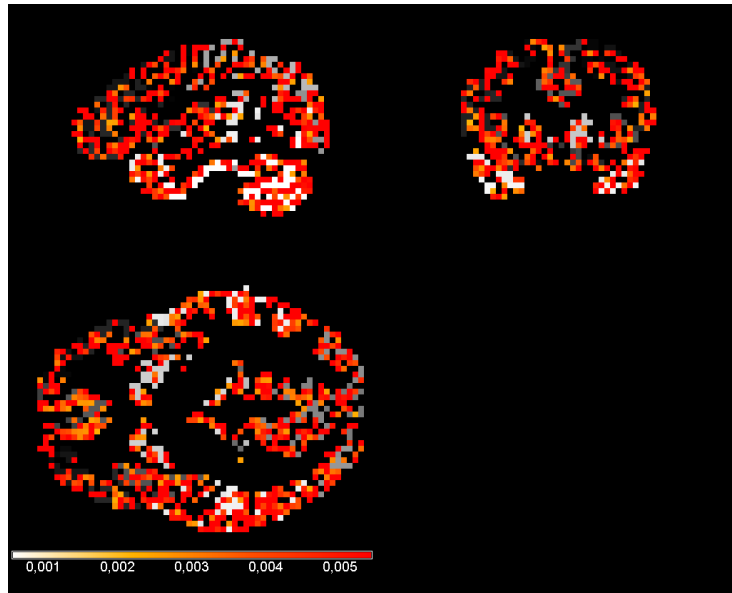
Figure 11: Brain plots of voxels with positive mean explained variance values, plotted on the corresponding points in the AAL atlas for Participant 1 in the four category condition (a) and Participant 2 in the four category condition (b).
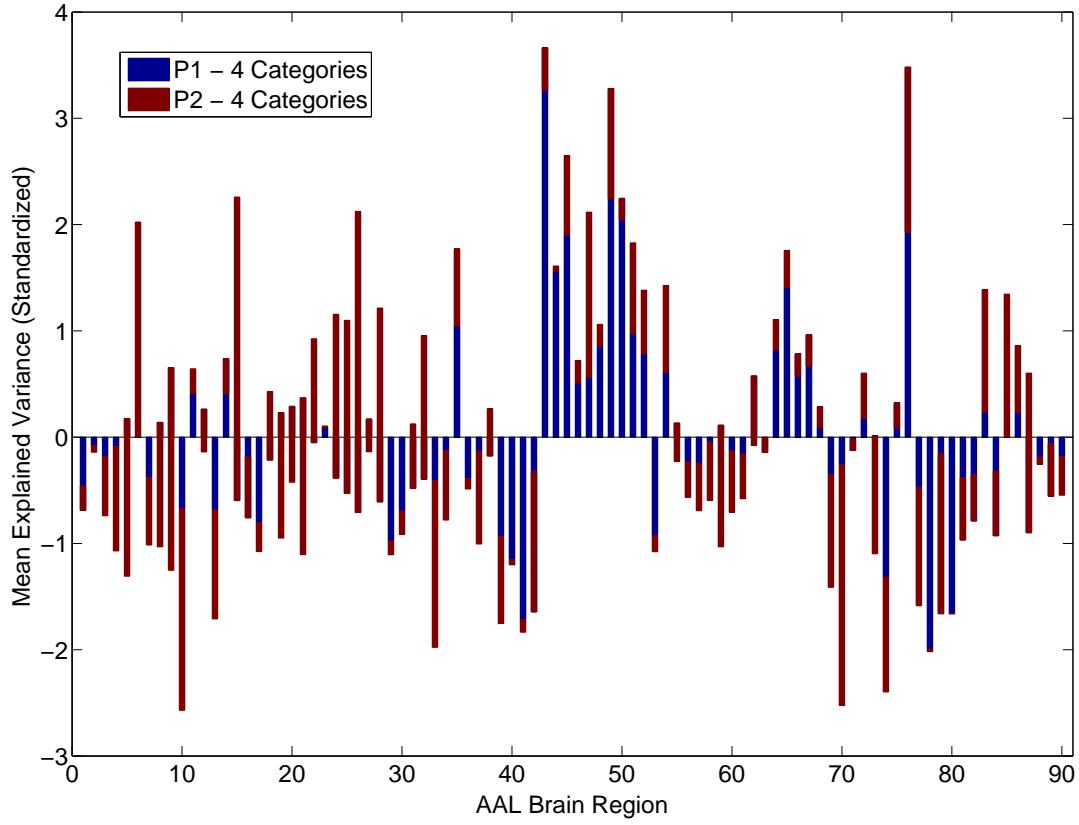
Figure 12: Graph showing the mean standardized explained variance per brain region. The bars are cumulative values of Participant 1 and Participant 2 for condition (*4 Categories*). The Y-axis represents brain regions according to the AAL template, the X-axis is the mean standardized explained variance of the corresponding region. The overall best five explaining brain areas for this condition and these participants are AAL region 43 (left calcarine sulcus), 76 (right pallidum), 49 (left superior occipital gyrus), 45 (left cuneus) and 15 (left inferior frontal gyrus, orbital part). The five worst explaining brain areas are AAL region 10, 70, 74, 33 and 78.

5   Discussion

This study explored the possibilities to identify semantic content differences using differences in observed brain activity by using natural auditory stimuli. The two main approaches were focussed on functional connectivity and spatio-temporal activation patterns. Both routes suggest that it is possible to classify semantic content differences, but more data is needed to prove this conjecture.

Because fMRI can contain a lot of noise, coming from artefacts like respiration or movement, it is difficult to interpret the results of fMRI studies. There are methods that can regress noise out of the signal, like ANATICOR (2010) and CompCor (2007). An fMRI experiment will always benefit from a better signal-to-noise ratio, so this could be a useful improvement for this pipeline. The slow hemodynamic response of fMRI could be modelled more precisely and the data could be shifted by different intervals to find the optimal setting for this shift. Moreover, with only two participants the sample size is very small. More participants would improve the results. More training examples could make a huge difference, this makes the estimation of both models better, in this study we were limited to 12 instances per participant that could be used for training. fMRI technology is improving quickly and it is possible to have a shorter repetition time and still scan the whole brain, which means more observation per minute and therefore again, a better estimation of a model.

In the next sections we will discuss the results and implications of both routes of the introduced pipeline.

5.1   Route 1

The empirical data results are close to chance level. The simulations show that this problem is two-folded: 1. there are not enough timepoints/observations to estimate the underlying population correlation matrix correctly. 2. The presented stories are too similar to estimate the underlying population correlation matrix correctly.

From our empirical data of two participants, we see a deviation of the 'story difference measure'. The smaller this measure is, the less accurate the pipeline is. To obtain better results with small story differences, more timepoints are needed. For Participant 1 (story difference measure: 0.0314), almost 10 times the amount of timepoints are needed to achieve a statistically significant correct performance of the classifier. For Participant 2 (story difference measure: 0.2091) a significant correct performance could already be achieved with twice the amount of timepoints. Scanning twice as fast and scanning the whole brain is already possible (Feinberg et al., 2010). Scanning ten times faster is possible with the current technology within fMRI, but still very experimental (Boyacioglu & Barth, 2012). An easier solution is to have longer stories, which consequently means more timepoints and a better estimation of functional connectivity. Presenting stories with the same length, that are semantically more different than

the 'Jip and Janneke' stories is another way to boost the performance. For instance a fairytale, a comedy story, a horror story and a western story. These stories could all be semantically more distant than the stories chosen is this study. The answer to our research question: *Can we classify perceived short stories with subtle differences in semantic content using differences in observed brain activity?*, is that this is not possible in route 1 of the pipeline, although the simulations hint that it is possible with more timepoints/observations.

There are points in this route that can be improved with the right tools, to improve overall performance. As already stated in the previous paragraph, more timepoints will make a big difference, this can be achieved by changing the fMRI protocol to a protocol with a shorter TR or choose longer stories. Moreover, a bigger training set can improve the prototype-story estimate, which could result in better predictions.

In this study we used the AAL throughout route 1. This is a very general template, which roughly divides the brain in 116 distinct areas. Other templates or methods could be used to improve the estimation of brain areas. For instance Diffusion Tensor Imaging could be used to define regions (Hinne, Heskes, Beckmann, & van Gerven, 2013), the brain could be parcellated based on functional localisers (Shirer et al., 2012) or independent component analysis could be used to group voxels in regions (Lahnakoski et al., 2012). These proposed methods are all based on the brain of the participant and not a general template, which could increase the performance dramatically. Alternatively different generic atlasses could be applied to the data instead of the AAL template. For example, the Harvard Oxford atlas (Desikan et al., 2006), the sulci atlas (Varoquaux, Gramfort, Poline, & Thirion, 2010) or the Ncuts atlas (Craddock, James, Holtzheimer 3rd, Hu, & Mayberg, 2011).

To get a better and more informative estimate in comparison to the empirical covariance matrix, a sparse inverse covariance matrix, or precision matrix, can be computed. If an element in this matrix is zero, the brain areas are assumed to be conditionally independent (Whittaker, 2009). This is a great advantage in interpretability of data over a normal covariance matrix. To estimate this sparse inverse covariance matrix, a graphical lasso algorithm (J. Friedman, Hastie, & Tibshirani, 2008), which uses a lasso penalty (L1-regularization) to control the number of zeros in precision matrix, could be used.

5.2 Route 2

The empirical data results of route 2, based on spatio-temporal activation pattern analysis, suggest that it is possible to classify subtle differences in semantic content, using differences in observed brain activity.

How the words are categorized in the design matrix is very important for a good performance, but the best categorization seems to be participant specific. The best mean categorization condition is using all four chosen categories as regressors. Although the accuracy of this mean condition is statistically not significant, it hints that more than three categories provide more stable results.

With more data and more different categorization conditions, this could be explored. The categories that are used in this study are very broad and could be more specific. An automatic categorization algorithm would be very useful, in which a degree of categorization could be selected and used for analysis. This algorithm could be based on WordNET, to automate the (sub)categorization of words.

The in-depth results of Participant one in the best condition *All Words* show that the brain regions that are responsible for the distinction between the story types, are also brain regions that are often reported in literature as regions of interest during processing of spoken word. For example the superior temporal gyrus is reported in the results of this study as an important area, by looking at blobs of voxels with a positive explained variance. The superior temporal gyrus has been reported as an important area for general processing of auditory stimuli (Jobard, Crivello, & Tzourio-Mazoyer, 2003), but this region is also associated with spoken-word recognition tasks (Cabeza & Nyberg, 2000). Moreover, the superior temporal gyrus is reported as a modality specific area for the processing of language and more specifically, phoneme processing of spoken words (Buchweitz et al., 2009). Both left and right superior temporal gyrus showed a relative high mean explained variance as a region.

A blob of voxels with a positive explained variance was also detected in the calcarine sulcus. The left calcarine sulcus was the region with the highest mean explained variance as a region for this participant in this condition. The calcarine sulcus is often associated with vision, the primary visual cortex is concentrated in the calcarine sulcus. A study by Lambert et al. (2002) found activation of the calcarine sulcus by mental imagery, driven by purely verbal cues. It could be due to influences of mental imagery, that this region has a high explained variance for the discrimination between the stories. A different interpretation could be that eye movements evoked by particular words are the basis of the high explained variance in the calcarine sulcus.

A small but interesting blob can be detected in the left frontal inferior gyrus (opercular part). This region is part of 'Broca's area' and is important for language production and verb comprehension (Rogalsky, Matchin, & Hickok, 2008).

The brain regions that are responsible for the discrimination between the different stories for this participant and this condition, do not seem to be random and are often associated to language comprehension.

The words that are important for the discrimination between the stories are often present in one of the four stories and usually only occur one time in the whole story. This is not what we expected, because for a word that occurs in one story and only one time, the model is trained on three examples of this word. Still the model is able to predict when this specific word was presented. We expected that words that occur a lot of times in all stories could improve the model, because the model has a lot of examples in the training set. This does not seem to be the case, looking at the results to which words the best 100

voxels map. The word that shows the biggest explained variance was present in timepoints 63 of story two. Our classifier only takes the first 50 timepoints into account, else it can classify on basis of the length of the story. This implies that the word that produces the largest explained variance was neglected in this approach and the pipeline could be more accurate.

For Participant two, the best categorization condition *Inanimate* showed that voxels with positive explained variance are widespread throughout the cortex. The amount of voxels with a positive explained variance is larger in comparison to the best condition of Participant one, but the values per voxel are a lot lower. That the positive explained variance voxels are widespread, could be due to the fact that there is only one regressor in this model: inanimate words. This category is spreaded out over the cortex. This widespread images of categories has also been reported by Huth et al. (2012).

The brain region that has clearly the highest mean explained variance for this participant in this condition is the left inferior frontal gyrus (orbital part). This region is located next to Broca's area. It could be the case that we are looking at Broca's area here, because these small regions tend to vary in size and shape between participants (Keller et al., 2007).

The best mean condition for Participant one and Participant two was the four category condition. In the plots of the voxels with positive explained variance, no blobs could be detected and the voxels are distributed throughout the cortex. The best mean explanatory regions differed per participant, but a similar distribution of important regions in the other conditions can be detected. For Participant one the left inferior frontal gyrus (orbital part) was again the most important area for the discrimination between the stories and for Participant two the left calcarine sulcus showed the highest mean explained variance. These results show that the areas that are important for the classification are stable for different conditions, but not between participants. Maybe the way these participants listen to the stories are different. Participant one could be visualizing the auditory stimuli vividly and Participant two is more focussed on the comprehension of verbs that are presented. More research is needed to see if these differences between participants occur more often and which areas are important for semantic processing of auditory stimuli. It has been shown that gender has an effect on the processing of higher order semantic in the brain (Wirth et al., 2007), in our study both participants were women, but the study shows that there can be individual differences between semantic processing.

Also for this route, more data has to be acquired to acknowledge the findings in this study. As stated before, different categorization protocols can be used, to explore how semantics are organized in the human brain. Moreover, different voxel selection could increase the results in this pipeline. In this study we selected voxels by using a grey matter parcellation and the AAL template.

Within this selection, voxels can be discarded on basis of their explained variance or a different criterion.

5.3 Conclusion

In this study we explored two distinct methods to identify subtle semantic content differences using differences in observed brain activity. The results are not conclusive and more data needs to be gathered to make statements about the neural correlates of semantics in the human brain. However, our results suggest that subtle differences in semantic content can be detected with both functional connectivity methods and by analysing spatio-temporal activation patterns.

Our results tend to advocate a inter-subject variability of the processing of semantic content, but this could be simply contradicted by the amount of data that has been gathered. Moreover, fMRI is an indirect measure of neuronal response, which makes the acquired data sensible to noise.

Improvements can be made by parcellating the brain in subject-specific areas, by using DTI data, functional localizers or ICA. More training data per subject would increase the estimation of the models. This could be achieved by more repetitions of the same story, but also more observations per story should increase the accuracy of the model. Especially for the functional connectivity method, more observations are crucial to identify subtle differences in semantic content.

References

Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., ... Phillips, C. (2010). SPM8 manual [Computer software manual]. Retrieved from http://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*, 90–101.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99-109.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state. a functional MRI study. *Journal of Cognitive Neuroscience*, *11*, 80-95.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal*, *310*(6973), 170.

Boersma, P., & Weenink, D. (2010). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org

Boyacioglu, R., & Barth, M. (2012). Generalized iNverse imaging (GIN): Ultrafast fMRI with physiological noise correction. *Magnetic Resonance in Medicine*.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkanen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163-73.

Buchweitz, A., Mason, R., Tomitch, L., & Just, M. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology and Neuroscience*, *2*(2), 111-123.

Buckner, R. L. (2010). Human functional connectivity: new tools, unresolved questions. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10769-10770.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, *12*(1), 1–47.

Caplan, D., Chen, E., & Waters, G. (2008). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, *44*(3), 257-75.

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex*, *23*(4), 958-966.

Craddock, R., James, G., Holtzheimer 3rd, P., Hu, X., & Mayberg, H. (2011). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, *33*(8), 1914-1928.

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human

brain. *Nature Neuroscience*, *16*(6), 763-770.

De Martino, F., De Borst, A. W., Valente, G., Goebel, R., & Formisano, E. (2011). Predicting eeg single trial responses with simultaneous fmri and relevance vector machine regression. *NeuroImage*, *56*(2), 826–836.

Desikan, R. S., Sgonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968 - 980.

Deza, M. M., & Deza, E. (2009). *Encyclopedia of Distances.* Springer Berlin Heidelberg.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., ... Yacoub, E. (2010). Multiplexed echo planar imaging for subsecond whole brain fMRI and fast diffusion imaging. *PLoS ONE*, *5*(12).

Feinerer, I., & Hornik, K. (2011). *wordnet: WordNet Interface* [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=wordnet

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication).* The MIT Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432-441.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1-22.

Friston, K. (1995). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, *2*, 56-78.

Gavrilescu, M., Stuart, G. W., Rossell, S., Henshall, K., McKay, C., Sergejew, A. A., ... Egan, G. F. (2007). Functional connectivity estimation in fMRI data: Influence of preprocessing and time course selection. *Human Brain Mapping*, *29*(9), 1040-1052.

Hardoon, D., Mouro-Miranda, J., Brammer, M., & Shawe-Taylor, J. (2008). Using image stimuli to drive fMRI analysis. In M. Ishikawa, K. Doya, H. Miyamoto, & T. Yamakawa (Eds.), *Neural Information Processing* (Vol. 4984, p. 477-486).

Hinne, M., Heskes, T., Beckmann, C. F., & van Gerven, M. A. (2013). Bayesian inference of structural brain networks. *NeuroImage*, *66*(0), 543-552.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Howell, D. (2007). *Statistical Methods for Psychology.* Thomson Wadsworth.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210-1224.

Jo, H. J. J., Saad, Z. S., Simmons, W. K., Milbury, L. A., & Cox, R. W. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, *52*(2), 571–582.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies.

Neuroimage, *20*(2), 693-712.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302-4311.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–5.

Keller, S. S., Highley, J. R., Garcia-Finana, M., Sluming, V., Rezaie, R., & Roberts, N. (2007). Sulcal variability, stereological measurement and asymmetry of Broca's area on MR images. *Journal of Anatomy*, *211*(4), 534-555.

Kriegeskorte, N., Simmons, K. W., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 49-86.

Lahnakoski, J., Salmi, J., Jskelinen, I., Lampinen, J., Glerean, E., Tikka, P., & Sams, M. (2012). Stimulus-related independent component and voxel-wise analysis of human brain activity during free viewing of a feature film. *PLoS One*, *7*(4), e35215.

Lambert, S., Sampaio, E., Scheiber, C., & Mauss, Y. (2002). Neural substrates of animal mental imagery: calcarine sulcus and dorsal pathway involvement–an fMRI study. *Brain Research*, *924*(2), 176-83.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M. M., Malave, V. L., Mason, R. A., & Just, M. A. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191-1195.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227-1235.

Rapoport, M., van Reekum, R., & Mayberg, H. (2000). The role of the cerebellum in cognition and behavior: a selective review. *Journal of Neuropsychiatry and Clinical Neurosciences*, *12*(2), 193-198.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., & Van De Ville, D. (2011). Decoding brain states from fmri connectivity graphs. *NeuroImage*, *56*(2), 616-626.

Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *19*(4), 786-96.

Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. *Frontiers in Human Neuroscience*, *2*(14).

Shirer, W. R., Ryali, S., Rykhlevskaia, E. E., Menon, V. V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral cortex*, *22*(1), 158-165.

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences.

Psychological Science, *20*(8), 989-99.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273-289.

Varoquaux, G., Gramfort, A., Poline, J.-B., & Thirion, B. (2010). Brain covariance selection: better individual functional connectivity models using population prior. In *NIPS'10* (p. 2334-2342).

Wallace, M. (2007). Jawbone Java WordNet API [Computer software manual]. Retrieved from http://mfwallace.googlepages.com/jawbone.html

Whittaker, J. (2009). *Graphical Models in Applied Multivariate Statistics*. Wiley Publishing.

Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. *Neuropsychologia*, *47*(14), 3111-3117.

Wirth, M., Horn, H., Koenig, T., Stein, M., Federspiel, A., Meier, B., ... Strik, W. (2007). Sex differences in semantic processing: Event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cerebral Cortex*, *17*(9), 1987–1997.

Yacoub, E., Harel, N., & Uğurbil, K. A. (2008). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, *105*(30), 10607–10612.

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066.

Appendix A Stories

These are the transcriptions of the four stories that were used in the experimen. All stories in Dutch.

Ze komen thuis met een hondje (story 1) 'Janneke,' zegt moeder, 'weet je de winkel van Smit? De winkel van lapjes en van wol?' 'Ja,' zegt Janneke. 'Ga daar eens een pakje naalden halen,' zegt moeder. 'Fijne naalden, geen grove.' 'Goed,' zegt Janneke. En ze roept Jip. 'Ga je mee naalden kopen?' Daar gaan ze samen. Het is mooi weer. En er is veel te zien op straat. Er is een vrouw met bloemen. En een kar met pruimen. Jip en Janneke krijgen ieder een pruim van de pruimenman. 'Kijk, dat hondje loopt met ons mee,' zegt Jip. 'Ja,' zegt Janneke, wat een gek hondje. Wat een gek lang hondje. Zijn achterpootjes zitten zo ver van zijn voorpootjes. 'Hij is vies,' zegt jip. 'Maar hij is lief.' 'Ga maar naar ja baas,' zegt Janneke, 'ga maar terug.' Maar het hondje wil niet terug. Hij wil mee. 'Arm hondje,' zegt Janneke, 'heb je geen baas?' Het hondje kijkt haar verdrietig aan. 'Heb je geen huis?' Het hondje kan nit praten, maar zijn oogjes zeggen: 'Nee, ik heb geen huis.' 'Ga dan maar met ons mee,' zegt Janneke. 'Ja hoor,' zegt Jip. 'Bij mij thuis mag je wonen.' 'Nee, bij mij thuis,' zegt Janneke. 'Nee, bij mij thuis,' schreeuwt Jip. Ze krijgen echt ruzie. En ze vergeten de naalden. En ze komen kwaad bij Jannekes moeder, het hondje achter hen aan. 'Wat is dat?' vraagt moeder. 'Waar zijn de naalden? En wat is dat voor een vies beest.' ''t Is geen vies beest,' zegt Jip. ''t Is mijn hond!' 'Nee, mijn hodn,' gilt Janneke. Dan komt Jips moeder er ook bij. En eindelijk mag het hondje bij Jip thuis in de tobbe. En als het schoon is, mag het bij Janneke in een mandje slapen. En het krijgt een etensbak bij Jip. En ook een etensbak bij Janneke. Nu is het hondje van hen samen. En hoe het heet? Het heet Takkie. En de moeders zeggen: 'Die kinderen toch. Je stuurt ze om naalden en ze komen met een hond thuis.'

Jip zingt op straat (Story 2) 'Moeder,' roept Jip. 'Er staat een man op straat. Hij zingt.' 'Ik hoor het,' zegt moeder. 'Hier, je mag hem wat geld brengen. Hij heeft het verdiend.' Jip geeft de man het geld en dan gaat hij naar Janneke. 'Ga je mee geld verdienen?' vraagt hij. 'Op straat met zingen?' 'Kan dat dan,' zegt Janneke. 'Ja, dat kan. Ik zal wel zingen, dan moet jij die grote hoed van je vader meenemen.' Daar staan ze in het straatje. Jip zingt zo hard als hij kan. hij wordt er schor van. En Janneke staat daar met de hoed. Maar ze krijgen niets. Niemand let erop. 'Je moet het echt vragen,' zegt Jip. 'Daar komt een mevrouw aan. Vraag maar wat geld voor de zanger.' Janneke is erg verlegen. Maar ze doet het toch. 'Wel, wel,' zegt mevrouw. 'Halen jullie geld op? Daar dan,' en ze geeft Janneke twee muntjes. Maar verder komt er niemand meer. 'Ik schei uit,' zegt Jip, 'ik kan niet meer. Er komt geen geluid meer uit.' 'Kijk eens, moeder, we hebben geld verdiend,' zegt Jip als ze weer thuis zijn. 'Hoe dan?' vraagt moeder. 'Met zingen,' zegt Jip. 'Net als die man van straks.' 'O,' zegt moeder, 'maar dat moet je niet doen. Je mag best zingen, maar niet voor geld.' 'Maar we hebben toch lekker wat verdiend,' zegt Jip. 'Dat geld geven we morgen aan die man,' zegt moeder. 'En jullie krijgen een sprits van me, als jullie nog een keer ook voor mijn zingen.' Dat doen ze. En ze krijgen elk een grote sprits.

Het is zo warm (Story 3) Het is zooooo warm! Janneke wil niet meer spelen. Ze ligt op het gras. En ze heeft alleen haar badpakje aan. 'Wacht,' zegt Jip. 'Ik ga je natspuiten.' 'Nee,' zegt Janneke. 'Ja,' zegt Jip. 'Ja, ja, ik ga je helemaal natspuiten.'

En Jip neemt de tuinspuit. Hij zet het kraantje aan. Janneke rent heel hard weg. Maar het helpt niet. Ze krijgt een heel straaltje water over zich heen. 'Oe! Oe!!' brult Janneke. 'Niet doen!' Maar Jip gaat door en Janneke wordt klets-klets-nat. En eigenlijk is het wel fijn. Ze is ineens zo lekker koel. 'Nou jij,' zegt ze. En ze pakt Jip de tuinspuit af. 'Nee! Ik wil niet!' huilt Jip. Hij is zelf heus bang. Maar Janneke heeft echt geen medelijden met hem. En Jip krijgt zo'n grote straal dat het water uit zijn haren druipt. Er staat een heel plasje om hem heen. Nu zijn ze allebei nat. En ze zijn allebei lekker koel. 'En nou de poes,' zegt Jip. 'Poezen mogen niet nat worden,' zegt Janneke. 'Dat is zielig. Poes kan er niet tegen.' 'H toe, een klein beetje,' zeurt Jip. En hij loopt achter poes aan. Die arme poes krijgt een heel straaltje water. Ze zegt heel hard mauw en is ineens weg. Daar zit poes. In de boom. Erg nat. 'Nare jongen,' zegt Janneke. 'Ze wordt wel weer droog,' zegt Jip. 'In de zon.' En poes wordt wel weer droog. Maar ze blijft wel drie dagen boos op Jip. En blaast tegen hem. En dat heeft hij echt verdiend.

Jip snijdt zich (Story 4)   'Ga je mee? Ik ga in de tuin spelen,' zegt Jip. 'Nee,' zegt Janneke. 'Ik moet appels schillen. Een hele emmer vol appels. Voor de appelmoes.' 'Ik ga ook appels schillen,' zegt Jip. 'Ik heb maar n mes,' zegt Janneke. 'Dan ga ik eerst een mes halen.' Jip gaat naar huis en haalt een mes. Uit de keuken. En hij holt met het mes naar Janneke. Nu gaat hij apples schillen. 'Kijk,' zegt Janneke. 'Ik maak heel lange kronkelschillen. Dat is erg moeilijk. Maar ik kan het al.' Jip probeert het ook. Hij krijgt er een kleur van. Maar de schil breekt af. 'Zo moet je doen,' zegt Janneke. 'Kijk, zo.' Jip doet het zo. Maar O! Jee! Au! Daar snijdt hij in zijn duim! Er komt bloed uit. Wat een schrik! Jip staat met zijn duim naar boven en kijkt zo angstig. 'Moeder!' roept Janneke. Daar komt haar moeder aanhollen. 'Wat is er?' roept ze, roept ze, maar ze ziet het al. 'Wacht,' zegt ze. 'Ik zal de duim verbinden. Ik heb er een mooi lapje voor. Kom maar mee. Doet het pijn?' 'Ja,' zegt Jip. 'Maar ik huil niet.' Nee, Jip is heel dapper en hij huilt niet. Hij krijgt een lap om de duim en Janneke staat erbij te kijken. 'Hij wou appels schillen,' zegt ze tegen moeder. 'En toen is hij thuis een mes gaan halen.' 'Ja, dat zie ik,' zegt Jannekes moeder. 'Maar dat mes is ook zo scherp. Veel te scherp. Voortaan moet je het eerst vragen hoor Jip, als je een mes wil hebben. En nu, weet je wat, Janneke mag schillen. En Jip mag eten. Omdat Jip gewond is mag hij appeltjes eten.' Dat is natuurlijk fijn. Want eten is niet moedlijk. Schillen wel.

Appendix B   Questions

These are the questions that were used in the experiments in between the instances. All questions and answers were visually presented and are in Dutch.

- Wat gingen Jip en Janneke kopen?

    1. Naalden
    2. Pennen
    3. Een hondje
    4. Vis

- Hoeveel cent gaf Jip aan de straatmuzikant?

    1. 2 cent
    2. 5 cent
    3. 10 cent

    4. 25 cent

- Hoeveel dagen bleef poes boos op Jip?

  1. En dag
  2. Twee dagen
  3. Drie dagen
  4. Vier dagen

- Wie verbindt Jip? 1. De moeder van Jip

  1. De moeder van Jip
  2. De vader van Jip
  3. De moeder van Janneke
  4. De vader van Janneke

- Wat krijgen Jip en Janneke op de markt?

  1. Een appel
  2. Een banaan
  3. Een peer
  4. Een pruim

- Hoeveel verdiende Jip en Janneke als straatmuzikanten?

  1. 2 cent
  2. 5 cent
  3. 10 cent
  4. 25 cent

- Waar speelden Jip en Janneke?

  1. Op het terras
  2. In huis
  3. Op het gras
  4. Op straat

- Waar haalde Jip het mes vandaan

  1. Bij Janneke thuis
  2. Bij Jip thuis
  3. Het mes lag er al
  4. Van Janneke

- Hoe heet de winkel waar Janneke naalden moet kopen?

  1. De lapjes winkel
  2. De winkel van Smit
  3. De Textielwinkel
  4. De winkel van Naalden

- Wat moet Janneke meenemen?

  1. Teksten van liedjes
  2. Centen
  3. Een grote hoed
  4. Een trommel

- Waarom mag de poes niet nat worden volgens Janneke?

  1. Omdat het slecht voor de haren is
  2. Omdat het dan gaat stinken
  3. Omdat het zielig is
  4. Omdat poes het dan koud krijgt

- Waarom schilt Janneke appels?

  1. Om ze daarna op te eten

2. Voor in de salade
3. Om het te leren
4. Voor in de appelmoes

- Waar mag Takkie uiteindelijk slapen?

  1. Bij Jip
  2. Bij Janneke
  3. Buiten
  4. Bij de baas

- Wat moeten Jip en Janneke doen voor een grote sprits?

  1. Zingen voor de moeder van Jip
  2. Zingen voor de moeder van Janneke
  3. Het geld aan de straatmuzikant geven
  4. Afwassen

- Waar speelden Jip en Janneke?

  1. Op het terras
  2. In huis
  3. Op het gras
  4. Op straat

- Hoeveel appels moet Janneke schillen?

  1. Vijf stuks
  2. Een emmer vol
  3. Voor iedereen en
  4. Twaalf stuks

Appendix C   Dictionary

Word	Frequency	Animate	Inanimate	(Action) Verbs	Emotion
'jip'	41	1			
'en'	40				
'zegt'	40			1	
'janneke'	33	1			
'ze'	30	1			
'maar'	27				
'is'	26				
'een'	23				
'de'	22				
'het'	20				
'ik'	19	1			
'je'	19	1			
'hij'	17	1			
'moeder'	15	1			
'dat'	14		1		
'niet'	14				
'er'	13				
'''t'	12				
'ga'	12			1	
'met'	11				
'zo'	11				
'hondje'	10	1			
'bij'	9				
'van'	9				
'daar'	8				
'dan'	8				
'mag'	8			1	
'nee'	8				
'op'	8				
'thuis'	8		1		
'zijn'	8				
'aan'	7				
'geen'	7				
'in'	7				
'ja'	7				
'kan'	7				
'mee'	7			1	
'mes'	7		1		
'voor'	7				
'wel'	7				
'als'	6				
'die'	6		1		
'heel'	6				
'komt'	6				
'krijgt'	6				
'moet'	6			1	
'naalden'	6		1		
'nat'	6				
'ook'	6				

Word	Count				
'poes'	6	1			
'wat'	6				
'zingen'	6			1	
'appels'	5		1		
'heb'	5				
'krijgen'	5				
'man'	5	1			
'naar'	5				
'roept'	5			1	
'schillen'	5		1		
'staat'	5				
'uit'	5				
'wil'	5			1	
'''ie'	4				
'doen'	4				
'duim'	4	1			
'echt'	4				
'gaat'	4			1	
'halen'	4			1	
'heeft'	4			1	
'hem'	4	1			
'jullie'	4				
'kijk'	4			1	
'meer'	4				
'mij'	4	1			
'nu'	4				
'om'	4				
'straat'	4		1		
'te'	4				
'weer'	4				
'wordt'	4			1	
'centen'	3		1		
'doet'	3			1	
'erg'	3				
'grote'	3				
'haar'	3				
'hard'	3				
'hebben'	3				
'hond'	3	1			
'huis'	3		1		
'komen'	3			1	
'lekker'	3				1
'n'	3				
'tegen'	3				
'toch'	3				
'twee'	3				
'verdiend'	3			1	
'vies'	3				1
'vraagt'	3			1	
'water'	3				

'we'	3	1			
'zingt'	3				
'''n'	2				
'achter'	2				
'af'	2				
'al'	2				
'allebei'	2	1			
'baas'	2	1			
'beest'	2	1			
'cent'	2		1		
'droog'	2				
'eerst'	2				
'eten'	2		1		
'etensbak'	2		1		
'fijn'	2				
'gaan'	2			1	
'geeft'	2			1	
'gek'	2				
'geld'	2		1		
'gilt'	2			1	
'heen'	2				
'heet'	2				
'hele'	2				
'hen'	2				
'hoe'	2				
'hoed'	2		1		
'hoor'	2				
'huilt'	2			1	1
'ineens'	2				
'jannekes'	2	1			
'jij'	2	1			
'kijkt'	2				
'klets'	2				
'koel'	2				
'loopt'	2				
'mevrouw'	2	1			
'mijn'	2				
'mooi'	2				
'niemand'	2				
'nou'	2				
'samen'	2	1			
'scherp'	2				
'snijdt'	2			1	1
'spelen'	2				
'sprits'	2		1		
'spuiten'	2			1	
'straaltje'	2		1		
'stuiver'	2		1		
'terug'	2				
'tuinspuit'	2		1		

Word					
'vak'	2		1		
'veel'	2				
'vragen'	2			1	
'wacht'	2			1	
'warm'	2				
'wat een'	2				
'weg'	2				
'winkel'	2		1		
'z''n'	2				
'zal'	2				
'zeggen'	2			1	
'zich'	2	1			
'''m'	1				
'aanhollen'	1			1	
'achterpoo	1	1			
'ahh'	1				
'alleen'	1				
'angstig'	1				1
'appelmoe	1		1		
'appeltjes'	1		1		
'arm'	1				
'arme'	1				
'auw'	1				1
'badpakje'	1		1		
'bang'	1				
'beetje'	1				
'best'	1				
'blaast'	1			1	
'blijft'	1				
'bloed'	1		1		1
'bloemen'	1		1		
'boom'	1		1		
'boos'	1				
'boven'	1				
'breekt'	1				
'brengen'	1			1	
'brult'	1			1	
'centje'	1		1		
'd''r'	1				
'dagen'	1				
'dapper'	1				1
'deze'	1				
'door'	1				
'drie'	1				
'druipt'	1			1	
'eigenlijk'	1				
'eindelijk'	1				
'elk'	1				
'emmer'	1		1		
'en een'	1				

'erbij'	1					
'eriseen'	1					
'fijne'	1					
'geef'	1					
'geluid'	1		1			
'geven'	1					
'gewond'	1					1
'goed'	1					
'gras'	1		1			
'grove'	1					
'haalt'	1					
'haren'	1		1			
'he'	1					
'helemaal'	1					
'helpt'	1					
'het hondje	1	1				
'het is'	1					
'heus'	1					
'hier'	1					
'hij is'	1	1				
'holt'	1					
'huil'	1				1	1
'huuuoooo	1					1
'ieder'	1	1				
'jaa'	1					
'jahoor'	1					
'jeej'	1					1
'jips'	1	1				
'jongen'	1	1				
'kar'	1		1			
'keer'	1					
'keuken'	1		1			
'kijken'	1					
'kinderen'	1	1				
'klein'	1					
'kleur'	1		1			
'kom'	1				1	
'kopen'	1				1	
'kraantje'	1		1			
'kronkelsch	1		1			
'kwaad'	1					
'lang'	1					
'lange'	1					
'lap'	1		1			
'lapje'	1		1			
'lapjes'	1		1			
'let'	1					
'lief'	1					1
'ligt'	1				1	
'maak'	1				1	

'mandje'	1		1		
'mauw'	1		1		
'me'	1				
'medelijde	1				1
'meeneme	1			1	
'met ons'	1	1			
'moeders'	1	1			
'moeilijk'	1				
'mogen'	1				
'morgen'	1				
'nare'	1				1
'natuurlijk'	1				
'neemt'	1				
'net'	1				
'niets'	1				
'nog'	1				
'omdat'	1				
'ons'	1				
'oogjes'	1	1			
'ooooh'	1				
'over'	1				
'pakje'	1		1		
'pakt'	1				
'pijn'	1				1
'plasje'	1		1		
'poezen'	1	1			
'praten'	1			1	
'probeert'	1			1	
'pruim'	1		1		
'pruimen'	1		1		
'pruimen-n	1	1			
'rent'	1				
'ruzie'	1				1
'schei'	1				
'schil'	1		1		
'schoon'	1				
'schor'	1				
'schreeuwt	1				1
'schrik'	1				
'slapen'	1			1	
'smit'	1	1			
'staan'	1			1	
'straal'	1		1		
'straatje'	1		1		
'straks'	1				
'stuurt'	1				
't'	1				
'takkie'	1	1			
'tis'	1				
'tob'	1		1		

Word					
'toe'	1				
'toen'	1				
'tuin'	1		1		
'vader'	1	1			
'van de'	1				
'van zijn'	1				
'ver'	1				
'verbinden	1			1	
'verder'	1				
'verdienen	1			1	
'verdrietig'	1				1
'vergeten'	1			1	
'verlegen'	1				1
'vol'	1				
'voorpootje	1	1			
'voortaan'	1				
'vraag'	1			1	
'vrouw'	1				
'waar'	1				
'weet'	1			1	
'weet je'	1			1	
'wol'	1		1		
'wonen'	1			1	
'worden'	1			1	
'wou'	1			1	
'zanger'	1	1			
'zelf'	1				
'zet'	1				
'zeurt'	1			1	
'zie'	1			1	
'zielig'	1				1
'zien'	1			1	
'ziet'	1			1	
'zit'	1			1	
'zitten'	1			1	
'zon'	1		1		
'één'	1		1		
		40	54	56	21
#Words	1158	246	116	169	27

Appendix D   AAL regions

AAL region	Brain Area	AAL region	Area
1	Precentral gyrus (L)	46	Cuneus (R)
2	Precentral gyrus (R)	47	Lingual gyrus (L)
3	Superior frontal gyrus (L)	48	Lingual gyrus (R)
4	Superior frontal gyrus (R)	49	Superior occipital gyrus (L)
5	Superior frontal gyrus, orb. (L)	50	Superior occipital gyrus (R)
6	Superior frontal gyrus, orb. (R)	51	Middle occipital gyrus (R)
7	Middle frontal gyrus (L)	52	Middle occipital gyrus (L)
8	Middle frontal gyrus (R)	53	Inferior occipital gyrus (L)
9	Middle frontal gyrus, orb. (L)	54	Inferior occipital gyrus (R)
10	Middle frontal gyrus, orb. (R)	55	Fusiform gyrus (L)
11	Frontal inferior gyrus, orb. (L)	56	Fusiform gyrus (R)
12	Frontal inferior gyrus, orb. (R)	57	Postcentral gyrus (L)
13	Inferior frontal gyrus, tri. (L)	58	Postcentral gyrus (R)
14	Inferior frontal gyrus, tri. (R)	59	Superior parietal cortex (L)
15	Inferior frontal gyrus, orb. (L)	60	Superior parietal cortex (R)
16	Inferior frontal gyrus, orb. (R)	61	Inferior parietal cortex (L)
17	Rolandic operculum (L)	62	Inferior parietal cortex (R)
18	Rolandic operculum (R)	63	Supramarginal gyrus (L)
19	Supplementary motor area (L)	64	Supramarginal gyrus (R)
20	Supplementary motor area (R)	65	Angular gyrus (L)
21	Olfactory cortex (L)	66	Angular gyrus (R)
22	Olfactory cortex (R)	67	Precuneus (L)
23	Superior frontal gyrus, med. (L)	68	Precuneus (R)
24	Superior frontal gyrus, med. (R)	69	Paracentral lobule (L)
25	Middle frontal gyrus, orb. (L)	70	Paracentral lobule (R)
26	Middle frontal gyrus, orb. (R)	71	Caudate nucleus (L)
27	Gyrus rectus (L)	72	Caudate nucleus (R)
28	Gyrus rectus (R)	73	Putamen (L)
29	Insula (L)	74	Putamen (R)
30	Insula (R)	75	Pallidum (L)
31	Anterior cingulum (L)	76	Pallidum (R)
32	Anterior cingulum R)	77	Thalamus (L)
33	Median cingulum (L)	78	Thalamus (R)
34	Median cingulum (R)	79	Heschl gyrus (L)
35	Posterior cingulate gyrus (L)	80	Heschl gyrus (R)
36	Posterior cingulate gyrus (R)	81	Superior temporal gyrus (L)
37	Hippocampus (L)	82	Superior temporal gyrus (R)
38	Hippocampus (R)	83	Temporal pole, sup. temp. (L)
39	Parahippocampal gyrus (L)	84	Temporal pole, sup. temp. (R)
40	Parahippocampal gyrus (R)	85	Middle temporal gyrus (L)
41	Amygdala (L)	86	Middle temporal gyrus (R)
42	Amygdala (R)	87	Temporal pole, mid. temp. (L)
43	Calcarine sulcus (L)	88	Temporal pole, mid. temp. (R)
44	Calcarine sulcus (R)	89	Inferior temporal gyrus (L)
45	Cuneus (L)	90	Inferior temporal gyrus (R)