Aalto University

School of Science

Degree Programme in Life Science Technologies

Lotta Lammi

Item-level decodability of semantic representations in the brain between individuals

Master's Thesis

Espoo, February 12, 2019

Supervisor:	Professor Riitta Salmelin
Advisor:	Annika Hultén Ph.D

Aalto University

School of Science

Degree Programme in Life Science Technologies

ABSTRACT OF

MASTER'S THESIS

Author:	Lotta Lammi		
Title:			
Item-level decodability of semantic representations in the brain between individuals			
Date:	February 12, 2019	**Pages:**	vi + 76
Major:	Complex Systems	**Code:**	SCI3060
Supervisor:	Professor Riitta Salmelin		
Advisor:	Annika Hultén Ph.D		

Meanings of nouns are widely thought to be grounded in experience. The experience of the world is unique for everyone, but many experiences are shared with other people. This study investigated whether human brain-level semantic representations of items are shared between individuals.

Brain activity of 20 healthy volunteers was measured using magnetoencephalography (MEG), while they were shown black-and-white photographs of concrete nouns. MEG data was analysed on the sensor level. The aim was to predict, for each participant, the presented stimuli based on the participant's brain activation, using a model which was trained on the data of the other participants. The semantic relationship among the stimuli was modelled using a semantic feature set collected through a web-based survey. The correspondence between the stimulus items and the MEG patterns was learned and, subsequently, predicted using machine learning methods, specifically with a zero-shot decoding algorithm.

Cross-subject decoding was successful at the item-level with accuracy significantly above chance level. Most commonalities in semantic representations between subjects were found at 150–250 ms and 350–550 ms after the stimulus onset. Some subjects had more similar brain responses than others, however, similar internal structures of brain responses between individuals did not directly imply good cross-subject decoding performance. Cross-subject decoding did not reach the accuracy of within-subject decoding, suggesting that semantic representations of concrete items are partly individual and partly shared.

More research is needed to uncover the factors that may cause individual differences in semantic representations and in decoding of such representations in the cortex using machine learning methods.

Keywords:	Semantics, word meaning, brain, MEG, decoding
Language:	English

Tekijä:	Lotta Lammi		
Työn nimi:			
Aivovasteiden semanttisten edustumien dekoodaus yli koehenkilöiden			
Päiväys:	12. helmikuuta 2019	**Sivumäärä:**	vi + 76
Pääaine:	Complex Systems	**Koodi:**	SCI3060
Valvoja:	Professori Riitta Salmelin		
Ohjaaja:	Annika Hultén PsT		

Yleisen näkemyksen mukaan sanojen merkitykset muodostuvat kokemuksen kautta. Kokemus ympäröivästä maailmasta on yksilöllinen, mutta monet kokemukset ovat yhteisiä. Tässä diplomityössä tutkittiin merkitysedustumien samankaltaisuutta eri koehenkilöiden aivokuorella.

Koeasetelmassa koehenkilöiden aivoaktivaatiota mitattiin MEG-laitteella henkilöiden katsellessa musta-valkoisia valokuvia, jotka esittivät konkreettisia substantiiveja. MEG-data analysoitiin kanavatasolla, ja aivovasteiden samankaltaisuutta tutkittiin zero-shot -dekoodausalgoritmin avulla. Tavoitteena oli määrittää näytetty ärsykevalokuva koehenkilön aivoaktivaation perusteella käyttäen koneoppimismenetelmää, jossa ennustusmalli opittiin toisen koehenkilön aivoaktivaatiosta. Sanojen merkityksiä mallinnettiin ominaisuusmatriisilla, joka perustui verkkokyselyn tuloksiin.

Koehenkilöiden ajattelemat ärsykesanat pystyttiin tulkitsemaan aivodatasta toiselta koehenkilöltä opitun mallin avulla tilastollisesti merkitsevällä tarkkuudella. Eniten yhtäläisyyksiä aivoaktivaatioissa löydettiin 150–250 ms ja 350–550 ms ärsykkeen esittämisen jälkeen. Havaittiin myös, että joidenkin koehenkilöiden semanttiset merkitysedustumat olivat samankaltaisempia kuin toisten. Kuitenkaan aivokuoren merkitysedustumien etäisyyksien samankaltaisuus eri koehenkilöiden välillä ei suoraan korreloinut hyvän ennustustarkkuuden kanssa. Ennustustarkkuus oli myös huonompi dekoodauksessa koehenkilöiden yli, kuin jos koneoppimismalli opetettiin ja testattiin saman koehenkilön datalla. Tuloksista voidaan päätellä, että osa merkitysten käsittelystä aivokuorella on yksilöllistä ja osa samankaltaista eri ihmisillä.

Näitä yksilöllisiä eroja synnyttävien tekijöiden tunnistaminen vaatii jatkotutkimuksia.

Asiasanat:	semantiikka, sanan merkitys, aivot, MEG, dekoodaus
Kieli:	Englanti

Acknowledgements

Contents

Chapter 1

Introduction

Meanings of concepts are widely thought to be grounded in experience (Binder et al., 2016; Kiefer et al., 2012). The experience of the world is unique for every individual, however we can communicate through our common languages, but sometimes misunderstandings occur. In this study, the aim is to determine how similar the brain activation of different individuals is when processing meanings of concrete nouns.

The term semantic representation refers to how meanings are represented in the brain. The topic is typically studied using tasks that require semantic processing while brain activation is recorded (Rupp et al., 2017). Such tasks are, for example, reading written words (Borghesani et al., 2016) or sentences (Wehbe et al., 2014), listening to spoken words (Correia et al., 2014) or speech (Huth et al., 2016b), naming pictures (Rupp et al., 2017; Clarke et al., 2014) or watching movies (Huth et al., 2016a). Studies of semantic disorders have also yielded much information about the processing of meanings (Harley, 2008).

In semantic disorders the processing of semantic information is disrupted by brain damage (Harley, 2008). There are reports of disorders where processing of some specific semantic category is defective or where the patient has problems to name, for example, some specific nonliving items. In semantic dementia, patients suffer from problems of naming items and forming

meanings. They may have difficulty in listing features related to words or classifying items based on their semantic similarity (Harley, 2008).

Brain decoding methods have enabled researchers to read the mind, in terms of identifying a stimulus based on neural activity (Murphy et al., 2018). Decoding methods are based on machine learning where an algorithm uses training data to learn patterns linked with predictor features in order to make predictions about unseen data (Kotsiantis et al., 2007). These methods are extremely useful in neuroscience as they allow researchers to investigate abstract aspects of brain functions such as semantic processing (Contini et al., 2017; Murphy et al., 2018).

In brain research, most decoding studies are done within-subject, meaning that the training and testing of the model is done on data from the same person. Cross-subject decoding, where the model is trained on data from one person and tested on data from another is much less common. This study investigates the similarity of semantic representations between individuals by performing decoding across subjects. The aim is to predict which individual item the subject has been thinking about, based on the measured brain activity of another subject. The studied brain data is measured with magnetoencephalography (MEG), which is a noninvasive brain imaging method that has an excellent time resolution (1 ms) and a localisation accuracy of 2 to 3 mm in optimal conditions (Lopes da Silva, 2010). Using this method, it is possible to obtain the time course and spatial estimation of brain activity while the subject is performing, for example, a picture naming task. Similar studies have been done using functional magnetic resonance imaging (fMRI) data (Shinkareva et al., 2008; Raizada et al., 2012), but there is a lack of studies that provide accurate time information of neural processes and operate with classification algorithms that are trained and tested on single subjects.

This thesis consists of six chapters. Chapter 2 presents the background of this study, starting with topics of semantic processing and brain decoding and ending with a short theoretical description of MEG and neurophysiology.

Chapter 3 explains the overview of the present study and justifies the chosen methods. Chapter 4 provides a more detailed description of the used methods. The results of the study are reported in the chapter 5 and discussed in chapter 6. Finally, chapter 7 concludes with the main findings of the study.

This study was done as a part of the larger study related to semantic representations of words conducted by Imaging Language group at the Aalto University Department of Neuroscience and Biomedical Engineering.

Chapter 2

Background

2.1 How humans process semantics

Based on a common view of semantics, the meanings of concepts are grounded in experiences of perception and action (Binder et al., 2016; Kiefer et al., 2012). For example, the meaning of 'dog' might be linked to information about the look, sound and feel of a dog. The meanings of concepts are stored as conceptual representations in semantic memory, where the concepts are not linked to any specific event, time or place (Harley, 2008). This conceptual knowledge allows us to understand a word 'dog' even though we haven't met all the dogs in the world.

Decompositional views propose that word meanings can be deconstructed into smaller units called semantic features (Clark, 1973; Vigliocco et al., 2007; Harley, 2008). Semantic features of words form a semantic space. According to Friederici et al. (2015) the semantic space can be thought to be a network where each node represents each feature. The activation of all features is unnecessary for the correct use of words in communication. The number of activated features affect the speed of word recognition, and brain responses depend on how much the features of a word differ from prior context.

A semantic feature space can be utilised in studies of semantic processing (Vigliocco et al., 2007). The feature space can be modelled with several

different methods. A broad representation of features can be obtained, for example, by utilising large text corpora or using feature production norms. In modern linguistics, a corpus database is a large digital collection of authentic texts which are based on written or spoken language (McEnery et al., 2006). A corpus-based feature space could be, for example, a collection of vectors which represent co-occurrences of a target word and neighbouring words (i.e. linguistic context). A vector could contain, for example, 300 most common words which occur with a word 'dog'. Feature production norms can be obtained by asking several people to list important features that describe or define the target word (Cree et al., 2003; McRae et al., 2005). A comprehensive feature space can be used in finding commonalities between words.

Behavioural and electrophysiological studies have suggested that the semantic space of a word and the representation of the word in the brain are organised by similar principles (Friederici et al., 2015). The literature as summarised by Binder et al. (2011) suggest that the semantic space is represented in a distributed fashion all over the cortex.

2.2 Brain decoding

Brain decoding is the ability to predict the presented stimulus based on neuroimaging data using multivariate machine learning methods. Using such methods, it is possible to extract more information from brain activity recordings than with traditional univariate methods that only compare the average activation between different experimental settings (Contini et al., 2017). Brain decoding can enable testing of new theories and examination of how the information is processed in the brain, for example, how the meaning representations of words change in time (Contini et al., 2017).

One way to perform brain decoding is to use a supervised machine learning classifier that aims to learn a set of rules in terms of predictor features from training data to classify new instances in a testing data set (Kotsiantis

et al., 2007). In the training data set, the values of predictor features as well as the class labels are known, whereas the values of predictor features are known in the testing data set. The trained classifier is evaluated by prediction accuracy which tells the percentage of correct classifications with respect to all possible classifications (Kotsiantis et al., 2007). In brain decoding studies, the classification algorithm is typically trained to learn a mapping between stimuli and brain activity. The algorithm is then evaluated on testing data, which consists of brain responses not used during training. The aim is to predict the class (for example the item or category) of a stimulus.

Brain decoding is a powerful tool in basic brain research, but, in addition, it can be applied, for example, to the development of brain-computer interface (BCI) applications (Chan et al., 2011). For example, Collinger et al. (2013) demonstrated that by training a neural decoder, they were able to perform complex movements with a robotic arm, controlled via microelectrodes that were implanted in the motor cortex. There are also promising results about the use of neural decoders to assist in communication (Birbaumer et al., 1999; Birbaumer et al., 2004; Birbaumer, 2006). Researchers have been developing systems where a patient could select letters by regulating their slow cortical potentials. However, the use of these systems requires long training periods, and the selection of the correct letter takes about one minute, which makes the communication slow. In the future, a deeper understanding of how meanings are represented in the brain may also bring advancements in the BCI field, which allow patients to communicate more effectively.

2.3 Decoding of semantic representations and the similarity of activation patterns between individuals

Mitchell et al. (2008) introduced a novel approach to study word meanings utilising the view of the decompositional theory of semantics and machine learning. They predicted brain patterns of unseen stimulus words by combining data from a large text corpus and data from brain activity measurements to train a linear model. The corpus was used to create a feature space which consisted of 25 verbs and the values of frequencies of their co-occurrences with target words. Using the mapping between this feature space and the brain data (recorded using fMRI) the model was able to generate brain patterns of unseen words. Mitchell et al. (2008) employed a method called semantic output classifier, introduced by Palatucci et al. (2009), which utilises zero-shot learning. The zero-shot learning algorithm learns a mapping between semantic features and input brain data and, thereafter, uses the learned mapping to predict the semantic encoding of unseen brain responses.

Brain decoding has been successful in within-subject classification tasks (Mitchell et al., 2008; Chan et al., 2011). There are a few studies where decoding is done across subjects, meaning that the prediction model is learned from one subject or a group of subjects and then tested on an unseen subject's data. This approach is useful to describe commonalities and differences between subjects.

Shinkareva et al. (2008) performed cross-subject decoding on fMRI data. They trained a naïve Bayes classifier with the data of 11 of the 12 subjects and then tested the model with data of the left-out 12th subject. This procedure was done for all subjects in the study. The stimulus set consisted of line-drawings from two categories, tools and dwellings, and both categories had five exemplars. The classifier was able to distinguish between two categories, and for some cases, even to distinguish between exemplars. Prediction ac-

curacy results on item-level were significant for 8 out of 12 left-out subjects, but they were on average lower than when the decoding was done within each subject. These results suggest that subjects have common components of semantic representations, but some components are individual.

Later, the same authors (Shinkareva et al., 2011) explored the commonalities across subjects further and used the same procedure (trained an algorithm on data of 11 subjects and tested on 12th left-out subject) to perform category identification across stimulus modalities (written words and pictures) and across subjects. They had fMRI data of 12 subjects and the stimulus set consisted of line-drawings and written words from two semantic categories, tools and dwellings, and both categories had five exemplars in each modality. When the classifier was trained on picture activation and tested on written word activation, the cross-subject decoding accuracies on category-level were significant for 9 of 12 left-out subjects. In the opposite direction, i.e., when the algorithm was trained on written word activation and tested on picture activation, the cross-subject decoding accuracies on category-level were significant for 8 of 12 left-out subjects. They found out the decoding accuracies were generally higher when identifying the category of a picture stimulus than the category of a written stimulus, both within and across subjects. These results suggest that brain activation is more stable to picture than word stimuli, as the accuracy results of the classification task were more accurate when training on picture activation.

Shinkareva et al. (2012) used the same data set as in 2008 and suggests that most of their subjects shared a similar internal structure of object representations. This was determined by analysing the similarity of the distance matrices of individuals derived from whole-brain fMRI data. However, the authors emphasised that the similarity of distance matrices does not indicate that the object representations would occur in the same spatial locations across individuals.

Just et al. (2010), using Gaussian naïve Bayes and linear regression approaches, performed cross-subject decoding on fMRI data and were able to

identify which of the 60 concrete nouns a subject (1/11) was viewing based on the data of the other 10/11 subjects. All prediction accuracies were significant. The stimulus set consisted of 60 different written words from twelve categories.

Huth et al. (2012) used principal component analysis to compare semantic representation of individuals measured with fMRI. They investigated how much of the variance in the data of one individual can be explained by the principal components of combined data from other individuals or stimuli. Their results suggested that semantic representations include similar components across subjects, but that the finer scale of semantic representations is individual.

All of the results discussed above support the assumption that semantic representations consist of shared components, but some portion of processing semantics is individual. Charest et al. (2014) suggested that individual differences in semantic representations occur because the individual experiences modify the microstructure of the cortex. They showed personally meaningful and unfamiliar images to subjects while fMRI data was recorded. They found that part of the individual perception can be detected in the inferior temporal cortex. The study of Chen et al. (2017) supports this finding as they demonstrated that people with shared memories had a more similar activity structure than people who did not have common memories. This was studied by recording fMRI while subjects were verbally describing the movie they watched before the measurements were performed. The authors found weak but significant correlation between the similarity of activity structure of different subjects and semantic similarity of the verbal description given by subjects during the measurements.

2.4 Magnetoencephalography

Time-series of brain activity can be decoded using data measured with MEG (Contini et al., 2017). Many decoding studies have been done using data measured with fMRI, but those studies give accurate information only about the location of brain activity. Decoding neural measures that also provide millisecond information about the time course of the neural activity, will therefore provide important additional value on the temporal dynamics of semantic processing. Performing decoding on MEG data is a relatively new approach (Contini et al., 2017).

MEG is a noninvasive brain imaging method which measures the weak magnetic fields produced by small changes in the electrical activity of the cerebral cortex (Hämäläinen et al., 1993). During the measurement, the subject is seated in the system with the head covered by a helmet in which the MEG sensors are located. During the measurement, the distance between the sensors and the skull is about 2 to 3 cm. The time-resolution of this imaging method is extremely precise - on the order of 1 ms - and the origin of brain activation can be estimated with an accuracy of 2 to 3 mm, in ideal conditions, by solving the inverse problem (Hämäläinen et al., 1993).

Magnetic fields are measured using flux transformers, magnetometers or gradiometers, where pick-up coils convey the magnetic flux to Superconducting Quantum Interference Devices (SQUID) that convert the flux to a measurable voltage (Parkkonen, 2010). The SQUID sensors are placed in a special container called a dewar which is filled with liquid helium (at temperature of 4 K). Due to the weakness of the magnetic fields generated by neuronal activity, the SQUID sensors need to be extremely sensitive to detect such fields. Therefore, they are also sensitive to artefacts.

Magnetometers are flux transformers that have a pick-up coil with one loop (Hari et al., 2012). They measure magnetic fields that are orthogonal to the surface of the pick-up coil, and also detect magnetic fields that originate outside the head. Gradiometers are flux transformers that are not that sen-

sitive to external interference as their pick-up coils have two loops wound in opposite directions, which cancel out fields from distant sources (Hari et al., 2012). There are two types of gradiometers: axial and planar gradiometers. An axial gradiometer has the two loops arranged vertically one above the other, whereas a planar gradiometer has the two loops placed in the same horizontal plane. Due to the arrangement of the two loops, a planar gradiometer is the most sensitive to neural current flow directly underneath the sensor.

The MEG device is located in a magnetically shielded room to protect the measurements from the magnetic noise and artefacts caused by powerlines, traffic and electrical devices (Hämäläinen et al., 1993). The room is typically isolated from the rest of the building to avoid mechanical vibrations. The subjects should not to have any metal in their body (jewellery or screws from orthopedic surgery etc.) because that could cause artefacts. They should also avoid movement and keep their head in the same position to provide reliable estimates of the head positions. Nevertheless, the recorded MEG data is often noisy and requires preprocessing to clean it before further analysis (Hämäläinen et al., 1993).

Figure 2.1 presents 10 s of cleaned gradiometer data. The data was processed with the signal space separation method (Taulu et al., 2006) and low-pass filtered to remove artefacts.
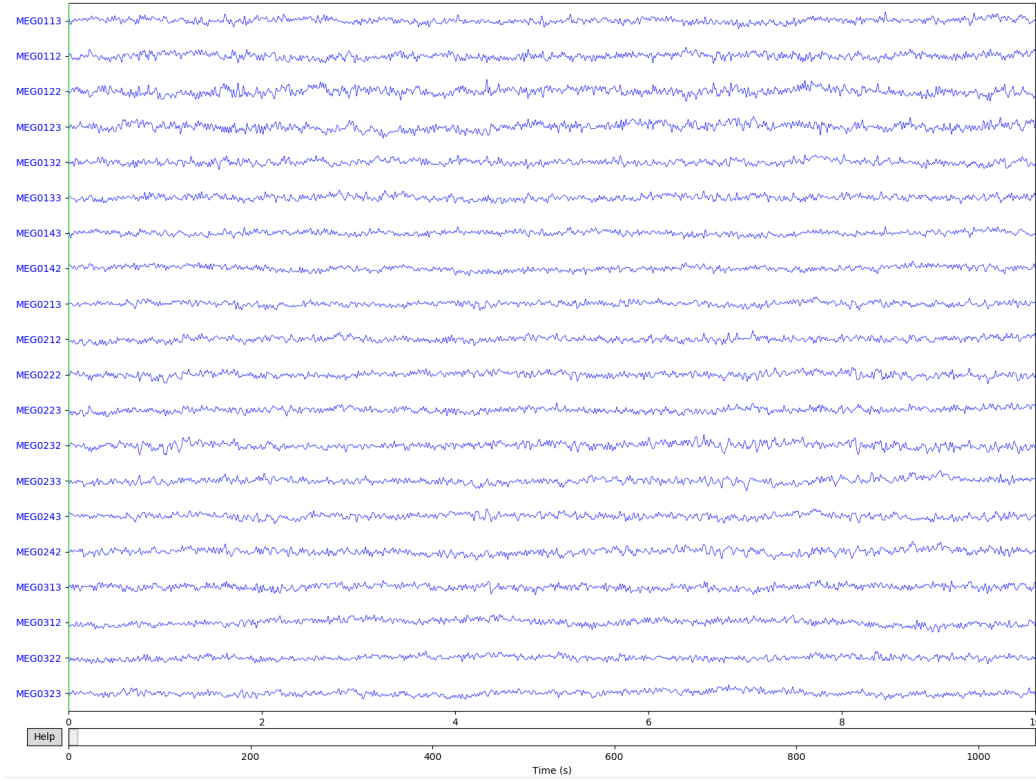
Figure 2.1: An example of MEG data representing a 10 s time window measured by a subset of gradiometers

2.5 Neurophysiology

Neural activity results in flow of electric current, and currents generate magnetic fields (Lopes da Silva, 2010). Neural activation can be divided into two main forms: action potentials and postsynaptic potentials. Action potential refers to the fast depolarisation of the axon membranes, whereas postsynaptic potentials refer to an extended change of membrane potential caused by synaptic activation at the receiving neuron. There are two types of postsynaptic potentials: excitatory and inhibitory potentials. Excitatory potentials are caused by the move of positive ions into the cell, and inhibitory potentials originate from the move of negative ions into the cell or positive ions

out from the cell (Lopes da Silva, 2010).

Action potentials are much faster phenomena (∼1 ms) than postsynaptic potentials (∼10 ms) (Baillet et al., 2001). Furthermore, action potentials are essentially current quadrupoles, as opposed to the dipolar currents of postsynaptic potentials, and their magnetic field thus dies away with distance much more rapidly than magnetic field generated by the postsynaptic potentials. Therefore, MEG signals are primarily generated by postsynaptic activity. Furthermore, the magnetic fields must be strong enough (10fT-1000fT) to be detected by MEG (Hämäläinen et al., 1993). This requires synchronous postsynaptic activity of tens of thousands of cells. Such summation is possible in populations of pyramidal cells that display a parallel alignment.

The folding of the cortex further affects the detection of magnetic field generated by neurons. Some neurons are located in gyri and others are found in the sulcal walls. Pyramidal cells that are located on a wall of a sulcus, with currents oriented parallel to the skull, generate an open magnetic field which can be detected outside the head (Lopes da Silva, 2010).

Chapter 3

Overview of the present study

The purpose of this study was to investigate the similarity of neural semantic processing in time between individuals. Previous fMRI studies have found commonalities in semantic processing (Shinkareva et al., 2008; Shinkareva et al., 2011; Shinkareva et al., 2012; Huth et al., 2012; Just et al., 2010), but the time course of the possible similarity of semantic processing remains unknown.

This master's thesis aims to address three questions:

1. Do different individuals have a common time course of semantic processing?

2. Which time intervals are the most similar between subjects in terms of brain activation when they process item meanings?

3. Are some of the subjects better decoding partners than others?

To address these questions, brain data of 20 individuals was measured with MEG, while they were shown object pictures representing concrete nouns. It was decided to use picture stimuli to activate semantic processing in the brain, as prior studies with the data set used in this study have indicated better decoding performance from brain data for pictures than written or spoken words. The brain data was recorded using MEG to address the

time course of activation. The time course could also be measured using EEG, but then the data would be noisier (Murphy et al., 2018). Here, noise was reduced by averaging across 18 representation times of each stimulus per each subject.

This study employed machine learning, which requires careful experimental design and interpretation of the results to ensure that the desired phenomena are considered as a target of decoding (Murphy et al., 2018). The stimuli were composed of five different picture exemplars of each target word; the stimuli had been tested behaviourally to ensure their quality. By averaging the brain responses across multiple presentations of different exemplars of the same object, the effects of particular visual features in the stimuli were minimised. The data was further cleaned by preprocessing before decoding.

The core idea of this study was to compare semantic processing between individuals using zero-shot decoding (Palatucci et al., 2009). A semantic model was trained on the brain data of one subject and then tested on other subjects' brain data to predict the presented stimulus based on MEG data. The semantic content of each stimulus word was modelled using semantic features derived from a large behavioural study which was based on feature production norms. The semantic features were used as predictor features in decoding.

To perform brain decoding, there must be sufficient amount of variation in the data between different stimulus nouns. Therefore, inter-item variance was first computed. The average signal strength was further calculated to evaluate its effect on decoding accuracy, because the hypothesis was that a lower average signal strength might be linked with decreased prediction accuracy of cross-subject decoding. The similarity of the internal structure of the brain responses was evaluated using distance matrices computed on the MEG data. As a quality check of the data and methods, within-subject decoding was performed before cross-subject decoding. Within- and cross-subject decoding were done at various time intervals of brain data.

If it is possible to decode meanings of individual nouns across subjects, it

means that there are commonalities in the time course of semantic representations between subjects. This will bring new information into group-level studies of brain functions and lead research towards questions about which parts of semantic processing are individually unique vs. shared and why that might be the case.

Chapter 4

Methods

In this chapter the experimental design and analysis methods are reported. The data was collected during the summer of 2015 in Otaniemi as a part of a study conducted by the Imaging Language group at the Department of Neuroscience and Biomedical Engineering at Aalto University. The stimulus set used in the measurements included three kind of stimulus: pictures, written words and spoken words. In this thesis, the focus is on the data collected while pictures were presented.

4.1 Subjects

Twenty healthy Finnish native speaking individuals (10 females, 10 males) with no history of neurological diseases participated in the study. All had normal or corrected to normal vision, and they were strongly right-handed, as indicated by the Edinburgh handedness questionnaire (Oldfield, 1971). Edinburgh handedness questionnaire as well as the other measurement documents are reported in Appendix A. The mean age of volunteers was 22 years (sd. 1.80, range 20–27 years). Individuals with metal or tattoos in their body were excluded from participating. All subjects read the instructions of the study and task before the measurements and filled out a permission form to participate in the study. The study was approved by the research ethics

committee of Aalto University.

The data of three subjects was excluded from the analysis because of poor signal from the coils measuring the head position during the measurements.

4.2   Stimuli

MEG responses were collected to a total of 60 depicted nouns, classified as belonging into one of seven semantic categories: Animals, Body parts, Buildings, Nature, Human characters, Tools and artefacts and Vehicles. However, due to a missing feature vector (see section 4.6.1) one stimulus word was excluded from data analysis; the final set of stimuli was thus 59 nouns. The nouns were common, high-frequency words within the 90th percentile of the corpus distribution derived from Finnish internet pages. There was no statistically significant difference in the lemma frequency of the different categories. Word length was predetermined to be 3-8 letters and did not statistically differ between the categories. The full stimulus list is reported in Appendix B.

Before the MEG study, the nouns were tested behaviourally to evaluate their age of acquisition, imageability, emotionality and valence (on a scale 1–7), using a web-based questionnaire with 13 respondents. The pretest revealed statistically significant differences regarding age of acquisition[1 = before school age, 7 = adult] between categories. The nouns in Body parts category [mean rating: 1.02 (sd. 0.03)] were learned earlier than nouns in Buildings category [mean rating: 6.94 (sd.0.09); Wilcoxon rank test W=1.5, $p < 0.01$], Human characters [mean rating: 1.35 (sd. 0.21); (W=1, $p < 0.01$] and Nature [mean rating 1.15 (sd. 1.20); W=4, $p < 0.01$)] categories. All words were judged to be easily visualised [mean rating: 6.73 (sd. 0.32), 1=difficult to visualise, 7=easy to visualise] and emotionally neutral [emotionality: mean rating: 3.50 (sd. 0.93), valence: mean rating: 4.43 (sd. 0.76)]. There were small statistically significant differences in imageability, emotionality and valence between categories when the significance level of $p < 0.05$ was used, but they were not significant on the level of $p < 0.001$.

Each noun was represented by five different exemplars, all of which were black-and-white photographs. The naming agreement of each photograph was estimated from separate pretest with 13 naive respondents naming each item.

The naming agreement based on the statistic H (Snodgrass et al., 1980) was computed using the following function

$$H = \sum_{i=1}^{k} p_i log_2(1/p_i) \tag{4.1}$$

where k is the number of different names given to each picture and p_i is the number of respondents giving each name. The bigger the statistic H is the worse the naming agreement is. In this study, the synonyms and inflected forms of the target word were considered as different names of the picture. The maximum value of the statistic H was 1.7; for each item the naming agreement was > 80%.

4.3 Experimental design

There were altogether three MEG sessions, on separate days for each subject. The subjects had also one MRI session to provide the anatomical data of the brain, but the MRI data was not analysed in this study.

Each noun in a certain modality was repeated 18 times, in total, across the sessions to achieve a sufficient signal-noise ratio. The duration of a MEG session was one hour including two breaks. The stimuli were presented in blocks according to their modality, so that there were six sets of picture objects, six sets of spoken words and six sets of written words on each measurement day. The order of nouns and modalities were randomly generated on each measurement day to avoid bias. Before the first session each subject was asked to name the stimulus pictures to ensure naming agreement.

During the measurements, the stimuli were presented on a screen with a grey background 140 cm away from the subjects' face. The experiments were

performed using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The size of the pictures was 10.6 cm x 10.6 cm, which equals to a visual angle of 4.3 degree. Picture stimuli were displayed for 300 ms, and the inter-stimulus interval was randomised at 700 to 1200 ms. Figure 4.1 illustrates the procedure of the experiment.

Figure 4.1: The procedure of the experiment when 90% of the picture stimuli were displayed (The original picture of a horse: Broster, 2010, Wikimedia Commons)

The task was to think about the item shown in the picture. Silent naming was used to avoid unnecessary muscle activation and, thereby, to improve the data quality. To ensure a subject was focusing on the stimuli, catch trials were shown after 10% of the stimuli. These trials were sentences, and the task was to determine whether the sentence made sense if the last shown stimulus was placed at the beginning of the trial sentence. For example, if the item 'dog' was followed by a phrase 'barks outside' the right answer was 'yes'. Subjects were told to answer 'yes' or 'no' with optical response pads. Half of the subjects answered 'yes' with their right hands and the other half with their left hand. Figure 4.2 illustrates the procedure of the experiment in case the catch trial occurred. The catch trial lasted on the screen until the subject answered.
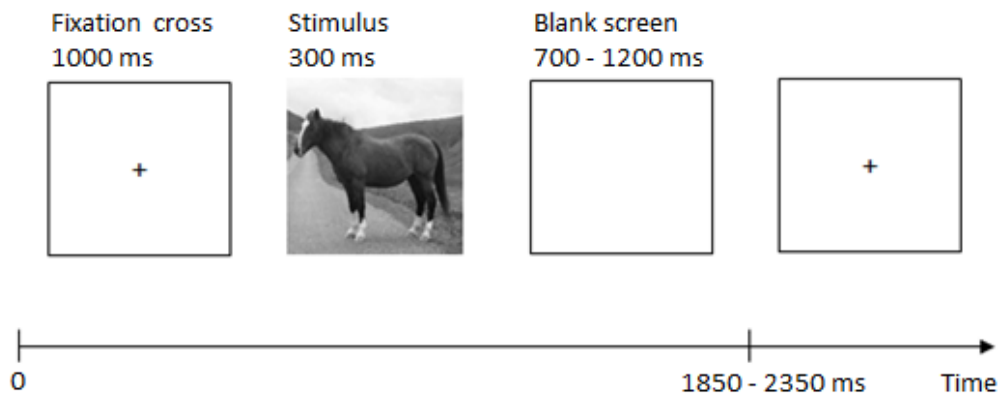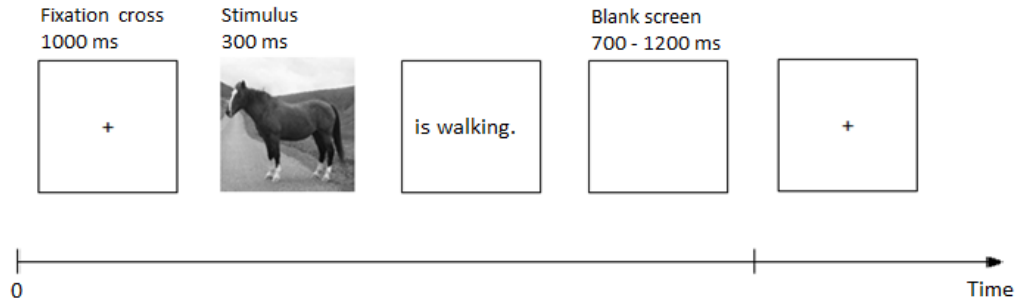
Figure 4.2: The procedure of the experiment when a catch sentence occurred (i.e. the procedure when 10% of the picture stimuli were displayed)(The original picture of a horse: Broster, 2010, Wikimedia Commons)

4.4 MEG measurements

Brain activity was recorded at the Aalto Neuroimaging Infrastructure at Aalto University, in Espoo, Finland using a 306-channel whole-head MEG system (Elekta Neuromag^{TM}, Elekta Oy, Helsinki, Finland) with 204 gradiometers and 102 magnetometers. The pass-band was set to 0.03–330 Hz and sampling rate to 1000 Hz. Subjects were in a seated position during the measurement.

The position of the head within the MEG helmet was determined by using five head position indicator coils whose locations, prior to the MEG recordings, were determined in relation to three readily identifiable points on the head: nasion, right and left preauricular points. The head position was tracked continuously during measurements by energising the coils using a high-frequency signal. Vertical and horizontal eye movements as well as blinks were recorded with two electrode pairs. The first pair was placed above and below the left eye and the second at the outer corners of each eye.

4.5 Data analysis

The MEG data analysis was performed using the Python implementation provided by the MNE software (Gramfort et al., 2014) with the exception of signal space separation (see section 4.5.1). To improve the data quality, only the data measured by the planar gradiometers was used.

4.5.1 Signal space separation and low-pass filtering

To diminish the influence of magnetic fields generated outside the sensor array, the data was first processed using the Signal Space Separation (SSS) method with temporal extension (tSSS) in the Elekta Maxfilter software package (Taulu et al., 2006). Using Maxwell's equations, this method separates the measured MEG signal into three subspaces based on the origin of the signal. The goal is to remove components which originate outside the sensor array or have sources of interference located very close to the sensors as well as noise and artefacts generated by the sensors. The temporal extension detects artefacts by computing correlations between subspaces. If there are high correlations between internal and external subspaces, the subspaces are removed. It has been demonstrated that tSSS improves the signal-to-noise ratio more than the basic SSS approach especially when working with gradiometer data (Haumann et al., 2016).

During this preprocessing phase, the length of the data buffering was set to 16 s, which corresponds to the cut-off frequency of 1/16 Hz. The subspace correlation limit was set to 0.9, and the default settings for automated detection of MEG sensors with poor-quality data were used. In this part of the analysis the data of each subject was transformed to the same reference head position, which was the average position of all subjects across all measurement sessions.

After tSSS processing, a low-pass filter at 40 Hz was applied on MEG data.

4.5.2 Independent component analysis

Eye movements and blinks create artefacts in the MEG signal. In this study, independent component analysis (ICA) was used. ICA expresses the data using projections that are statistically maximally independent and can be very helpful in extracting artefacts to improve data quality (Vigário et al., 2000).

In this analysis, the eye-movements and blinks were detected using EOG-signals (electrooculography signals) measured by two electrode pairs placed close to eyes of a subject. EOG-epochs were created around the detected eye-movements. Thereafter, the EOG-epochs were chosen manually so that only EOG-epochs representing a typical eyeblink were used in the following analysis. After the manual selection, an automatic procedure provided by MNE-Python package was followed to produce the ICA components of the MEG data corresponding the selected EOG-epochs. In this procedure, the 'Fastica' algorithm (Hyvärinen, 1999) was used and 1-3 ICA components that explained 95% of variance were selected. After the automated procedure, the selected ICA components were visualised to ensure that they corresponds to blinks and eye-movements. Thereafter, ICA-projections were created and applied to MEG data to remove the eye-movement and blink components. ICA projections were made for each subject separately using the data measured in the first MEG session of each subject.

4.5.3 Evoked responses

The data was cropped into segments (epochs) centred around the stimulus triggers, from 200 ms before to 1000 ms after the stimulus onset. The cortical responses to each of the 59 stimuli were obtained, for each subject, by averaging the epochs across the three sessions (number of averaged trials = 18). Averaging improves the signal-to-noise ratio as single trials are typically too noisy to use in the analysis (Salmelin, 2010). The data was baseline corrected using the 200 ms pre-stimulus interval.

4.5.4 MEG data matrices

The data dimensionality was reduced to enable decoding and statistical evaluation. In this step of the analysis, the time line of the responses was corrected for the time-delay of 38 ms caused by the projector which was used to display the stimuli. Three different sized MEG data matrices were created using evoked responses. To evaluate the data quality including inter-item variance and average signal strength the time course from -200 to 1000 ms was used. For this analysis, the evoked responses were downsampled to have 20-ms time windows, resulting in 60 time points. For each subject, a data matrix was formed with the size of 59 (items) x 204 (channels) x 60 (time points).

In decoding phase of the study, shorter time course was used: the responses corresponding to each noun were cropped from 0 ms to 700 ms and downsampled to have 20-ms time windows, resulting in 35 time points. For decoding, a data matrix was formed with the size of 59 (items) x 204 (channels) x 35 (time points) for each subject.

To investigate more the time course of decodability (i.e. what are the time intervals which give the best prediction accuracies), the decoding was done also using cropped evoked responses with length of 100 ms. Therefore, the evoked responses were cropped to have a length of 100 ms in 50 ms time intervals starting from -200 ms to 1000 ms (i.e. the first cropping covered the period from -200 to -100, the second cropping covered the period from -150 to -50 etc. so that the last period was from 900 to 1000 ms). These cropped evoked responses were downsampled to have 10 ms time windows resulting in 10 time points. Therefore, the third size of the MEG data matrices was 59 (items) x 204 (channels) x 10 (time points).

4.5.5 Inter-item variance

Inter-item variance was calculated from downsampled MEG data using matrices of the size of 59 (items) x 204 (channels) x 60 (time points), covering

the time course from -200 ms to 1000 ms. The variance (Mellin, 2010) was calculated for each subject across items at every time point for each channel separately, using the following function

$$\sigma^2 = \frac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{n} \tag{4.2}$$

where n is the number of items (59), x_i is the measured activation on a channel evoked by i^{th} item and μ is the average activation on a channel across all items.

4.5.6 Average signal strength

The amplitude strength of the MEG signal varies between individuals (Salmelin, 2010). In this study, a hypothesis was that lower average signal strength might be linked with decreased prediction accuracy of cross-subject decoding. In order to test the hypothesis, the average signal strength, across items, was calculated using the same matrices that were used to compute the inter-item variance (59 items x 204 channels x 60 time points).

4.5.7 Signal-to-noise ratio

In order to examine the effect of signal-to-noise ratio on decoding performance the amount of noise was estimated by calculating the standard deviation of the average amplitude strength of the MEG signal during the prestimulus interval from -200 to 0 for each of the subject separately. The average signal-to-noise ratio was then determined from downsampled MEG data and time window from -200 to 1000 ms by dividing the average signal strength by the calculated standard deviation.

4.5.8 Dissimilarity matrices of feature space and MEG data

To investigate the category structure of the feature space (see section 4.6.1), the Euclidean distances (Kaufman et al., 2005) between feature vectors corresponding to each pair of stimulus nouns were computed. The following function was used

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (4.3)$$

where $X = (x_1, x_2, ...x_i)$ is a feature vector of one item and $Y = (y_1, y_2, ...y_i)$ is a feature vector of another item. n refers to the length of the feature vector.

In this study, the aim was to investigate the commonalities of semantic representations in the brain across subjects. The main idea was to perform decoding across subjects, but similarity measures were also used to compare the internal structure of representations. Therefore, the euclidean distances between each pair of item representations were calculated for each subject using their MEG data matrices of the size (59 x 204 x 35). Before distance calculation, the MEG data matrices were z-transformed to obtain a vector representation for each of the 59 items and then the dissimilarity matrices were computed.

Thereafter, to compare the internal structure of representations across subjects, the Spearman correlation (Mellin, 2010) between the dissimilarity matrices of each pair of subjects was computed based on the following equation

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{(n^2 - 1)n} \qquad (4.4)$$

where the calculated dissimilarity matrices are first ranked based on the distances to compute the value of d_i which is the difference between ranks of each distance. n refers to number of computed distances (59 x 59). The range

of correlation value ρ is from -1 to 1, where -1 indicates a strong negative correlation and 1 a strong positive correlation.

4.6   Decoding

Decoding was done to compare the semantic representation in the brain across subjects. First the decoding was done within each subject to ensure the reliability of the methods used; subsequently, the decoding was done across subjects. The aim was to predict which stimulus noun a subject was thinking about based on a model which was trained on another subject's data. The zero-shot decoding algorithm (Palatucci et al., 2009) was used as the classifier.

4.6.1   Predictor features

The semantic space of stimulus nouns was modelled using the Aalto-norms feature set. The Aalto-norms set was created in two steps using web-based questionnaires. First, feature production norms were gathered: 241 respondents listed typical features for each stimulus word. In addition to the stimulus words used here, the questionnaire included 241 extra words. The average number of produced unique features for each word was 192 (sd. 43). Sparse feature sets do not work well with machine learning, so the aim was to construct a feature set that could be used to describe all of the stimulus words such that a unique vector would account the degree to which each word is described by each feature.

To optimise the feature set, 73 questions were created based on the most frequent features produced to a subset of 118 words of the total of 300 words used in the first web-based questionnaire. These 118 words included 59 stimulus words used in the present study and as well as 59 words which were used in another study. To ensure that the created questions captured enough of the important features of all the stimulus words, 12 extra questions were added to the question set. Then, this set of 85 questions was used in a

new web-based questionnaire. The task of the 120 novel respondents was to evaluate how much a certain feature relates to a stimulus word on a scale from zero (= not a feature at all) to five (=a very prominent feature). Each respondent evaluated a subset of 20 stimulus words. The set of 85 questions is reported in Appendix C.

The questionnaire responses were averaged across subjects. Thus, a feature space was formed where each of the 59 stimulus words was described by a feature vector with a length of 85. Figure 4.3 illustrates the dissimilarity matrix of this feature space where Euclidean distances between feature vectors were computed. The figure shows that this feature space clusters stimulus words together according to the different categories.
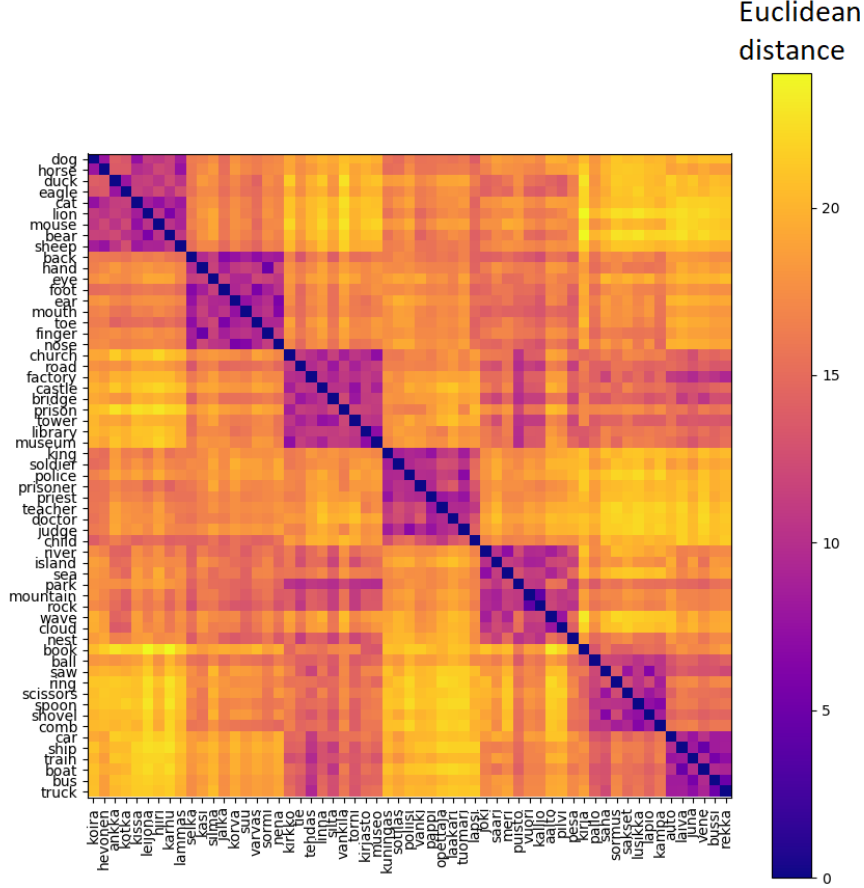
Figure 4.3: Dissimilarity matrix of the feature space

4.6.2 Semantic output code classifier

The aim of the zero-shot decoding algorithm is to learn a semantic output code classifier (SOCC) $f : X \rightarrow Y$, which can predict new values of Y that are not included in the training data set. The procedure used here is the same as that introduced by (Sudre et al., 2012).

Let $X \in R^{N*d}$, where the input MEG data X is given as a matrix with the shape of 59 (items) x timepoints x 204 (channels). Each row of X consists

of the evoked responses activated by each stimulus noun. N is the number of items (59) and d corresponds to the number of dimensions of the neural response (channels*times). The data was z-tranformed and normalised to minimise the effect of differences in signal strength between subjects. Then let $F \in R^{N*p}$ be a matrix of semantic features, where N again is the number of items and p is the length of the feature vector (85). Linear ridge regression was used to learn the weight vector $W \in R^{d*p}$ with the following function

$$W = X^T(XX^T + I_N)^{-1}F \qquad (4.5)$$

where I_N is an identity matrix and is a scalar regularization parameter. The prediction f for neural activity of the semantic features of a previously unseen word is obtained as $f = x * W$.

After the linear mapping has been learned, the SOCC aims to match two previously unseen stimulus words with two unseen MEG data segments. This step is referred to as leave-two-out cross-validation. First, the algorithm translates the MEG segments into two predicted feature vectors using the previously learned linear mapping. Then, the algorithm computes a cosine distance between the predicted feature vector f and the true feature encoding for the two unseen items. The labels (stimulus words) are chosen based on the smallest distance. This process is repeated for all possible leave-two-out combinations; in this study, this analysis was performed for 1711 different combinations.

In the case of cross-subject decoding, separate data matrices were used for training and testing the algorithm. Again, two items were left out from the training data set (of one subject), then the algorithm computed the cosine distance between the predicted feature vector f and the true feature encoding for those two left-out items from the testing data set (of another subject).

In this study, the decoding performance is reported using prediction accuracy, which is an average accuracy across all left-out word pairs. In each distinguishing task, there is a 50% probability of the algorithm choosing correctly between the target words. However, the decoding was performed on

all the possible left-out word pairs, so the chance level of the decoding performance was simulated using permutation tests to get a reliable estimate of the significance (see section 4.6.3). An accuracy of 100% would indicate that the algorithm succeeds in correctly labelling every MEG segment.

4.6.3   Permutation tests

A permutation test was performed to determine the chance level of prediction accuracies to evaluate the performance of the zero-shot classifier. Here, the zero-shot classification was performed 1 000 times on randomly selected subject, whose data was randomised in terms of word labels and MEG data segments. The idea was to find a level where the results are significant and the prediction accuracy is not achieved by coincidence. In case of cross subject decoding, the both training and testing subjects were selected randomly.

4.6.4   Agglomerative hierarchical clustering

Agglomerative hierarchical clustering was done to group the subjects based on similarity of MEG data and cross-subject decoding performance. The clustering was performed, first, based on vectors of correlation coefficients between subjects and, second based on vectors of decoding performance accuracies.

Agglomerative hierarchical clustering is based on the distances between all observations (Kaufman et al., 2005). Here, the euclidean distance metric (see section 4.5.8) was used to determine the distances. First, each observation (here, a subject) is in their own clusters and then the algorithm starts to merge the closest pair of clusters together until there is only one cluster left which covers the entire data set. Here, the Ward variance minimisation algorithm (Müllner, 2011) was used to determine the closest cluster. The Ward variance minimisation algorithm uses the following equation

$$d(u, v) = \sqrt{\frac{|v| + |s|}{T}d(v, s)^2 + \frac{|v| + |t|}{T}d(v, t)^2 - \frac{|v|}{T}d(s, t)^2} \qquad (4.6)$$

where u is a novel cluster consisting of clusters s and t, and v is an unused cluster. T represents the sum of $|v|$, $|s|$ and $|t|$.

Chapter 5

Results

5.1 Task compliance during the MEG measurements

Task compliance during the MEG measurements was evaluated by showing phrases to subjects after 10% of the stimuli. The subject was asked to judge if a sentence made sense, when the previously shown stimulus was placed in front of the phrase. The purpose of this task was to make sure that the subject was concentrating on the stimulus shown and thinking about its meaning. The average percentage of correct answers across subjects was 96.3% (sd. 3.5%). These results confirm that the subjects were thinking about the meanings of the stimuli.

5.2 Inter-item variance of brain activation

Inter-item variance of evoked responses were calculated to ensure the separability of brain responses between items for successful decoding. The variances were computed per channel using the time window of -200 to 1000 ms and are presented in figure 5.1, where each subject is indicated by a separate colour. For all subjects, the variance was largest approximately from 200 to 500 ms, which means that the brain responses to different stimuli varied most during that time interval.

Figure 5.1: A topography of inter-item variance of each MEG channel from -200 to 1000 ms

In figure 5.2, the average variance across channels are presented; each colour refers to a different subject. The figure illustrates that the variance increased as a function of time with a small peak at the time-point 140

ms. The variance differed significantly across subjects (One-way ANOVA, F=235.22 $p < 0.001$). Subject s08 had notably higher variance than the other subjects. During the baseline, the variance was comparably low for each subject, but there were also some differences between subjects. For example, the inter-item variance was high for subject s08 also during the baseline.
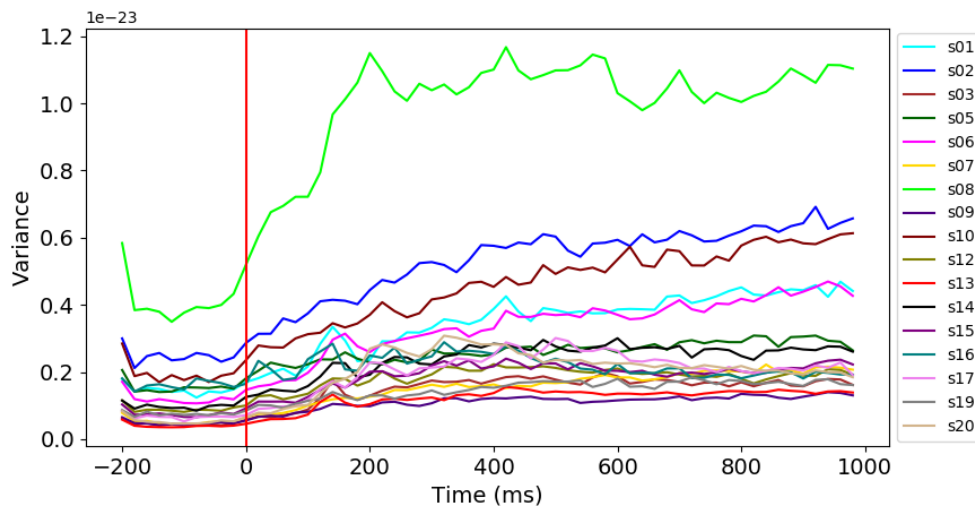

Figure 5.2: The time course of inter-item variance in the different subjects, measured as an average across the MEG channels. The straight vertical red line signifies the stimulus onset.

5.3 Average signal strength and signal-to-noise ratio

The average signal strength was determined to investigate its effect on decoding performance and was computed across the items and MEG channels using the time window of -200 to 100 ms in relation to picture onset. Figure 5.3 shows that the signal strength varied greatly between individuals (One-way ANOVA, F=53.19, $p < 0.001$). The highest signal amplitudes were approximately at 150 ms and at 240 ms after the stimulus onset for most of the subjects (excluding subject s08). The largest amplitudes were detected for subjects s08, s16 and s02. During the baseline, the amplitude of the signal strength was comparably low and after 700 ms there were not any clear peaks in the amplitude strengths.
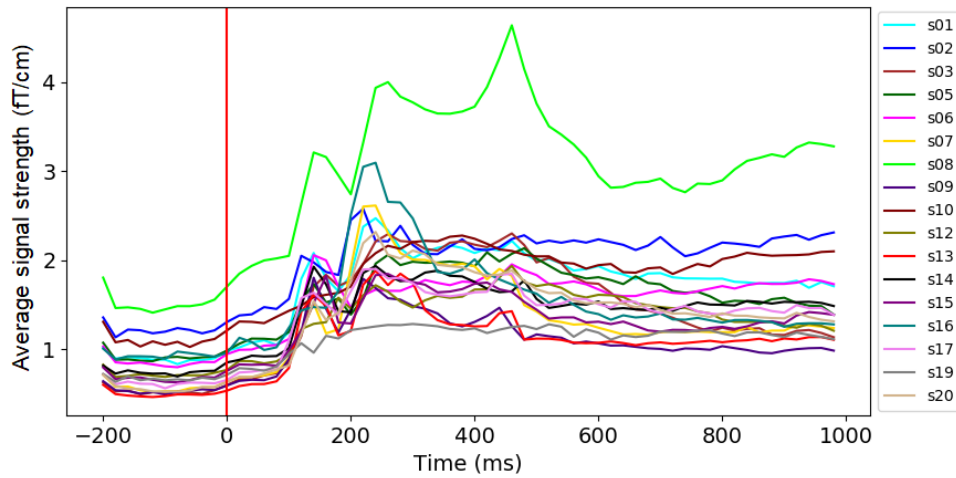
Figure 5.3: The time course of average signal strength. The straight vertical red line signifies the stimulus onset.

The signal-to-noise ratio was estimated by dividing the average signal strength in every time point by standard deviation of the signal strength during the baseline. The standard deviations of the signal strength during the baseline are illustrated in figure 5.4. The figure shows the subject s08 and s10 had the largest variation in signal strength during the baseline. High variation during the baseline indicate there are still some noise left despite the noise-reduction techniques used.
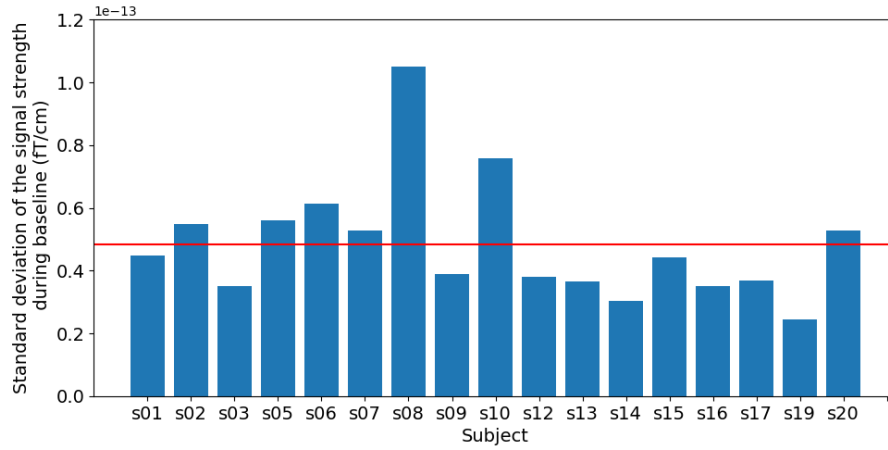
Figure 5.4: Standard deviation of average signal strength across channels during the baseline. The horizontal red line illustrates the mean of standard deviations.

Figure 5.5 illustrates the estimated signal-to-noise ratio. The subjects s03, s14 and s16 seem to have the best ratio as their have the strongest effects whereas the subject s10 has the worst ratio. The signal-to-noise ratio varied greatly between individuals (One-way Anova, F=31.81, $p < 0.001$).
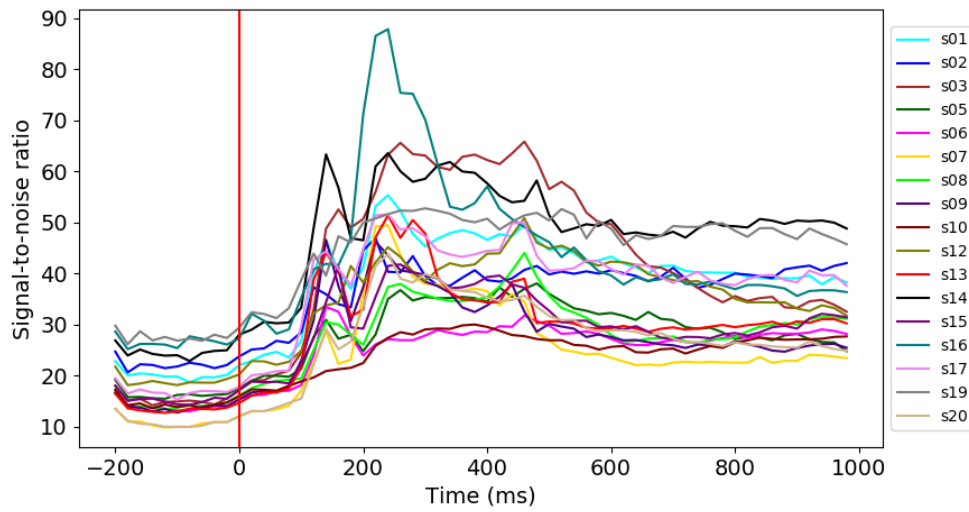
Figure 5.5: The time course of the estimated signal-to-noise ratio. The straight vertical red line signifies the stimulus onset.

5.4   Item-level within-subject decoding

To establish the baseline level of decoding accuracy with respect to this data set, decoding was first done within each subject, that is, the model was trained and tested on the same subject's data.

In figure 5.6, the within-subject decoding accuracies are presented. Here, the data from 0 to 700 ms was used. The mean accuracy across subjects was 86.7% (sd. 4.6%). A permutation test indicated that prediction accuracies above 62.0% were significantly above chance level at $p < 0.05$. Therefore, all the within-subject decoding results were significant, indicating that it was possible to find a linear mapping between the semantic features and the brain data so that the classification algorithm could succeed in labelling the brain data correctly in most of the leave-two-out tasks.
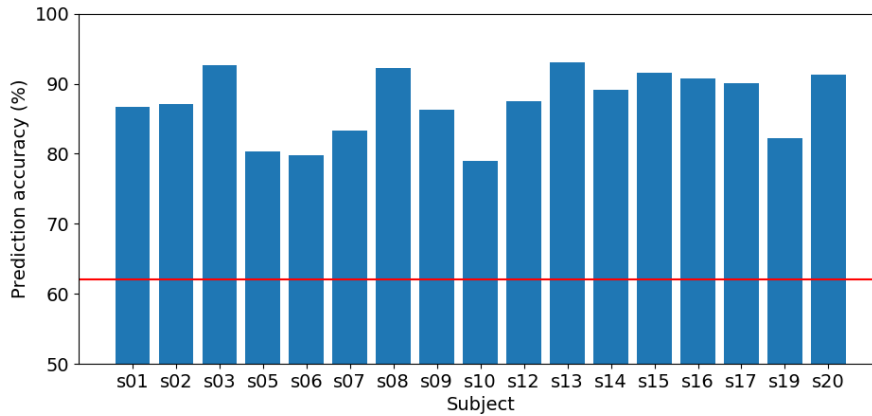
Figure 5.6: Within-subject decoding accuracies. The significance level is marked as a red line.

Figure 5.7 presents a time course of prediction accuracies for each subject. The data was cropped to 100-ms bins, and the algorithm was trained and tested for each data bin. The model succeeded in distinguishing between two target words best during the time window from approximately 150 ms to 500 ms. It indicates that most of the semantic information was processed during

that time. For some subjects, the algorithm was able to label the MEG data correctly even when the data from 700 to 1000 ms was used, which suggests that semantic processing can last longer for some individuals. During the baseline (-200-0 ms), the prediction accuracies were below the significance level.
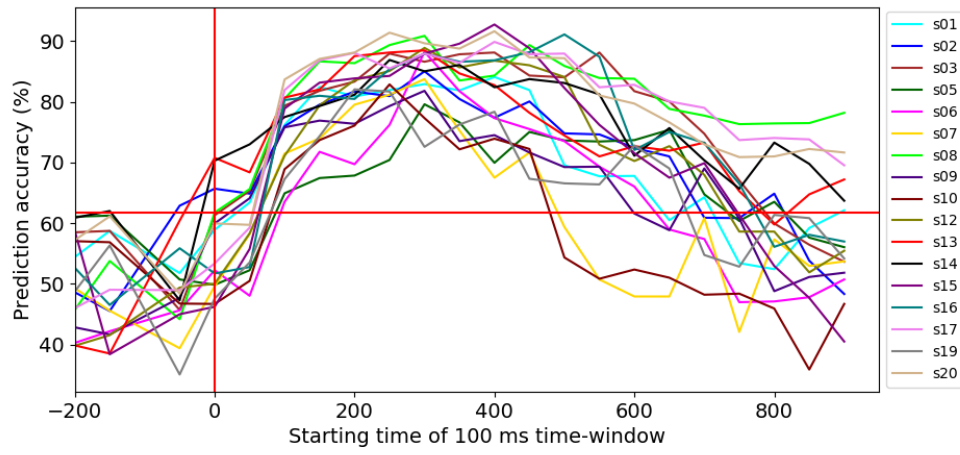
Figure 5.7: Within-subject decoding accuracies for 100 ms time windows, every 50 ms. The subjects are presented in different colours and the straight horizontal red line indicates the significance level, whereas the vertical straight line indicates the stimulus onset.

5.5 Correlation between MEG dissimilarity matrices

To investigate the similarity of the internal structure of semantic representations across subjects and to verify the results of cross-subject decoding, the correlation between distance matrices of the MEG data were computed. The results are presented in figure 5.8. The significant ($p < 0.05$) Spearman correlation coefficients between different subjects varied in the range from 0.05 to 0.46. The largest correlation was between subjects s16 and s20, and for most of the subjects, the correlation was significant. Hierarchical clustering shows how the subjects were clustered based on correlations between the distance matrices of the MEG data. Subjects close to each other had more similar internal structure than those further away from each other.
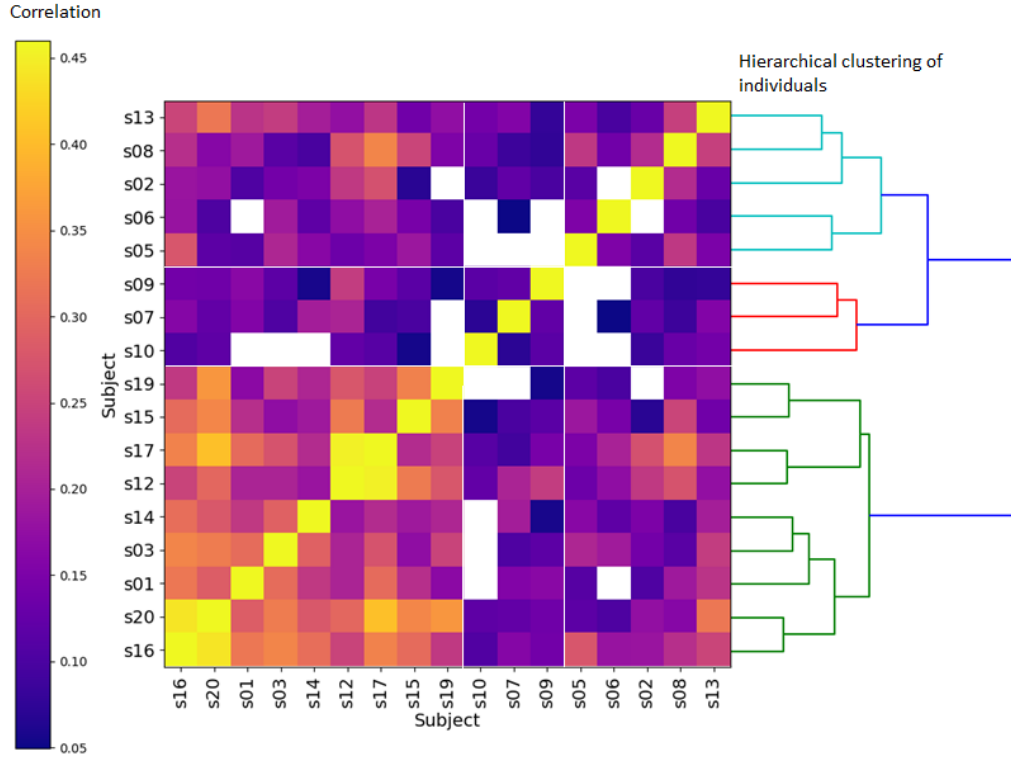
Figure 5.8: Correlations between MEG dissimilarity matrices of each subject. The white colour indicates insignificant correlation, yellow indicates relatively high correlation and blue indicates low correlation.

5.6   Item-level decoding across subjects

In this study the aim was to examine similarities in semantic representation across subjects. Similarity was investigated by performing decoding across subjects. The model was trained with the data of one subject and tested with the data of another subject; here this subject combination is called a testing pair ($N = 17 * 17 - 17$). The significance level determined by a permutation test was 58.1% ($p < 0.05$). Of all the testing pairs, 95.2% demonstrated significant results, thus, successful decoding (mean 66.5%, sd. 7.1%). The best prediction accuracy was 85.6% and was achieved when the model was trained with data of s20 and tested with the data of s01.

Figure 5.9 displays the prediction accuracies of all the subject combinations. The within-cluster decoding accuracies were higher than the average across all the subjects, which supports the idea that some of the subjects have more similar brain responses than others.
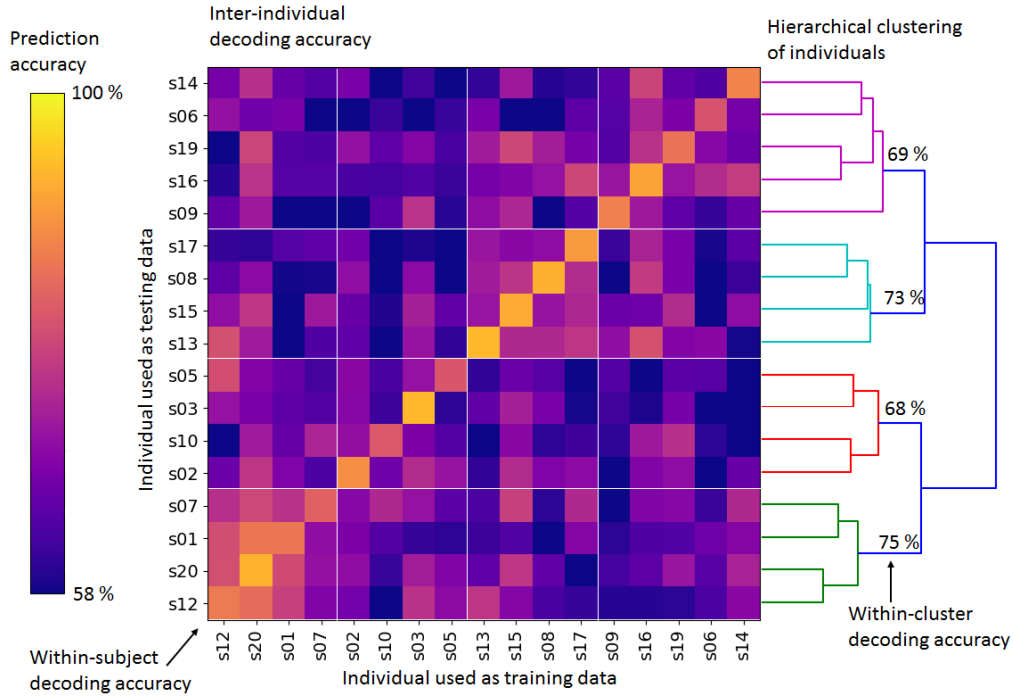
Figure 5.9: Cross-subject decoding accuracies. Yellow indicates high prediction accuracy and blue low accuracy. The yellowish diagonal line presents the within-subject decoding results. On the right side of the figure, the hierarchical clustering is presented.

The subjects that performed well as testing pairs seemed to also have a more similar internal structure of brain responses reported in section 5.5. For example, the subjects s01, s12 and s20 were clustered together in both cases. On the other hand, the subjects s16 and s20 seemed to have the most similar internal structure of brain responses, but in decoding task the prediction accuracies of this pair of subjects were not remarkably precise compared to other testing pairs and they occurred in different clusters.

Figure 5.10 illustrates the percentage of significant decoding results ($p < 0.05$) as a function of time. Here, the data was cropped to 100-ms bins, at 50-ms time intervals. The model was tested on every testing pair, with the

testing and training data taken from the same time intervals. The results show that during the time from 100 ms to 600 ms more than half of the decoding results were significant, so most of the similarities across subjects occurred during that time. More precisely, when the time-window of 150 to 250 ms was used for training and testing the algorithm 80% of the decoding results were significant. Another peak in decoding occurred in the time window from 350 to 500 ms. This indicates that those time windows are the most similar across subjects, as regards semantic processing.
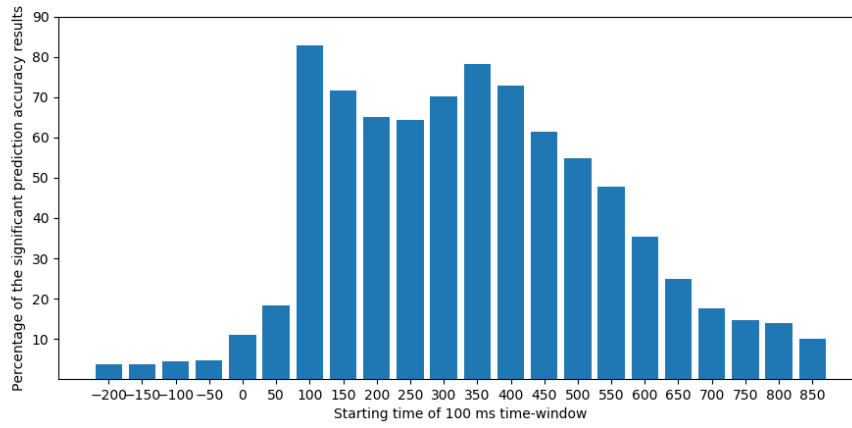
Figure 5.10: Percentages of significant cross-subject decoding results per 100 ms time windows. The percentage shows the portion of all the testing pairs that reached a significant prediction accuracy (>58.1%, ($p < 0.05$), per time window. Therefore, if 100% of the decoding results were significant, all the testing pairs (N=272) would have resulted with significant prediction accuracy.

Chapter 6

Discussion

In this study, the semantic representations of individual words were investigated. The aim was to study similarities in the timing of representations across subjects using zero-shot decoding across subjects. It was possible to predict individual words from MEG responses by training the algorithm on the data of one subject (with the to-be-tested items removed) and testing on another subject's data. This indicates that there are commonalities between individuals in how words are represented in the brain.

6.1   Quality of the data

To test the quality of the data, inter-item variance, signal-to-noise ratio and average signal strength were computed and item-level decoding was performed within each subject. During the decoding process, the semantic content of the stimulus items was modelled using the Aalto-norms as a feature set. The dissimilarity matrix of this feature set displayed clear clusters which followed the pre-defined categorical structure (Animals, Body parts, Human characters, Buildings, Nature, Tools and artefacts and Vehicles). When the feature set was used, the within-subject decoding results were significant for all the subjects. This shows the representations of meanings were modelled successfully as such correlations were found between brain activity and the

feature set.

The brain responses to different items varied for all the subjects. There was a peak in inter-item variances for all the subjects at 140 to 160 ms after picture onset. Previous picture naming studies have suggested that this is part of the time window where visual object recognition happens (Indefrey, 2011; Levelt et al., 1998; Salmelin, 2007). The large inter-item variance in this time window is, therefore, likely to reflect real item-level differences in the visual properties of the presented pictures.

The average signal strength varied greatly for different subjects and the subject (s08) who had the largest inter-item variance had also the highest amplitude of the signal strength. For some of the subjects (s02, s08 and s10) the inter-item variance was fairly high also during the 200-ms prestimulus interval, which may indicate noise in the brain signal and that is why the signal-to-noise ratio was estimated.

Despite the inter-subject differences in the signal-to-noise ratio and the inter-item variance the within-subject prediction accuracies were markedly above the significance level for all the subjects, with the average accuracy 86.7% (sd. 4.6%). This suggest that the signal-to-noise ratio was acceptable because the decoding model was able to distinguish between the two target words in most of the leave-two-out cases. The best prediction accuracies were achieved during 150-500 ms which indicates that during this time window the representations of items are most different and the brain activity correlates best with the feature set. The peak decoding time starts approximately, when there is a peak in inter-item variance.

The amount of variance in the amplitude of the brain response to the different items did not directly correlate with the decoding accuracy. For example, the subjects s08 and s02 had the greatest variance but the within-subject prediction accuracies for these subjects were not especially precise or poor. Also the variances were relatively low for subject s13 but the prediction accuracy was still precise. It was found that the subjects (s03, s14 and s16) who had the best signal-to-noise ratio performed well in the within-subject

decoding, whereas the subject (s10) who had the lowest signal-to-noise ratio had also the lowest prediction accuracy. This finding is supported by the fact that typically a classifier performs better when it is tested on cleaner data (Shinkareva et al., 2011).

Successful within-subject decoding results demonstrate that the brain activity varies for different meanings and confirms the quality of the data and algorithms used.

6.2 Similar semantic representations in different individuals

The meanings of words were successfully decoded across subjects, which suggests there are commonalities between individuals in the time course of how words are represented in the brain. The vast majority of the testing pairs were well above chance level. The prediction accuracies (66.5%) were still notably worse than the within-subject decoding results (86.7%), which are in line with earlier studies that even though there are shared components of semantic representations, all aspects are not shared (Shinkareva et al., 2008; Shinkareva et al., 2011; Shinkareva et al., 2012; Just et al., 2010; Huth et al., 2012).

There were differences in how each testing pair performed. For example, the decoding model trained on subject s01 was able to predict the data of subjects s12 and s07 but not s08 and s09. Interestingly, the model trained on subject s20 was able to successfully decode the items from the data of all but one other subject. The signal-to-noise ratio was better for subjects s03, s14 and s16 than for the other subjects, but this did not affect the cross-subjects decoding performance.

These asymmetrical differences in decoding accuracies could be due to a number of factors. In this study, there was no link detected between the signal-to-noise ratio and cross-subject decoding performance, though the signal-to-noise ratio affected on within-subject decoding performance. It may

be that the methods used to determine the signal-to-noise ratio were not able to measure the ratio correct. An another possible factor affecting to cross-subject decoding could be that the semantic space of certain individuals are closer to the average space than others.

Hierarchical cluster analysis of the decoding results demonstrated that the within-cluster decoding accuracies were higher than the average decoding accuracy. The best within-cluster decoding accuracy was 75% which was achieved when the data of s01, s07, s12 and s20 were used for training and testing the algorithm. Nevertheless, these subjects did not perform particularly well compared to other subjects in the within-subject decoding task, so these results cannot be explained merely by a good model fit to the semantic feature model. Instead, the finding suggests that some subjects have more similar semantic spaces than others.

During the time window from 100 to 650 ms, more than 50% of the cross-subject decoding results were significant. This is also the time window where the within-subject decoding accuracies were highest. Thus, most of the inter-subject commonalities as well as within-subject properties in semantic processing occurred in that time window. The peak performance in cross-subject decoding was between 150 and 250 ms, and another smaller peak occurred between 350 and 550ms. The first peak agreed with the time window, when the inter-item variance was highest. The latter peak occurred in a time window, that is typically associated with phonological code retrieval and preparation of the oral output (Indefrey, 2011; Levelt et al., 1998; Salmelin, 2007). This is interesting, as it points to a possibility that the information in the Aalto-norms might correlate with the phonological properties of the words, even though these are not directly semantic in nature. During these time windows, the within-subject decoding results were also significant.

There were significant correlations between the inter-item distance matrices computed from the individual subjects' MEG data, which implies that there are commonalities in the internal structure of semantic representations. It was found that well-performing testing pairs in decoding had also a more

similar internal structure than other subjects. Nevertheless, a hierarchical clustering analysis found differences in the composition of subjects in different clusters, when analysis was done on the decoding results compared to the analysis performed on correlations between distance matrices. For example, there was a high correlation between distance matrices of subjects s16 and s20, but these subjects appeared in a completely different clusters when the decoding results were used. This suggests that even though there are commonalities in the internal structure of representation in terms of distances between item activations, the items are not necessarily encoded similarly in the brain of different individuals. This finding was also highlighted in the research conducted by Shinkareva et al. (2012).

6.3 Limitations of the methods

In this study, a common feature set was used for all the subjects, which resulted from the average responses to a behavioural feature survey. This means that the present study also incorporated the question of how similarly the different individuals interpreted the stimulus words; in other words, the question was not only about the similarity of brain responses. To obtain more accurate results concerning the similarity of brain representations between subjects, individual feature spaces could be used. That could also improve the within-subject decoding results, if certain features are indeed the building blocks of meaning for certain subjects.

When a common feature space is used for every subject, the feature set is a chosen group of attributes, which may not include all of the important properties. This could have been tested by using a survey where subjects would have been asked to determine, based on a certain set of features, what target word the features describe. If the subjects had been unable to choose a target word, there might have been too few properties. Due to the fairly large number of features (85), this kind of method could lead to a complicated and long survey, which would have affected to the quality of

the answers. Nonetheless, that kind of survey could have confirmed that the features represented the same item for everyone.

There are also some limitations related to decompositional theories of semantics. For example, when a meaning is cut into small features, the features themselves are already small concepts, and it has not been demonstrated how these verbal features are actually represented in the brain (Binder et al., 2016). Therefore, only the meaning is modelled by these features, but there are no particular neurons which would, for example, represent some primitive feature (Binder et al., 2016).

In this study, the semantic representations between individuals were compared by performing cross-subject decoding. An assumption was that, after coordinate transformation, the sensors picked up signals from the same general locations in the brain, across subjects. Nevertheless, the structure of the head is individual and the similarity of the sensor-locations is only an approximation.

The signal-to-noise ratio of the MEG data varied greatly between individuals. For further analysis, it is recommended to examine more the possible source of high standard deviation of average signal strength during the baseline to improve the data quality of subjects s08 and s10. One obvious way to improve signal-to-noise ratio in the future would be to increase the number of presentation times of each item during the measurements. In this study, it was limited to 18 times, because of the already long measurement times.

Long measurement times (1 hour) are likely to cause fatigue, which affects the quality of the measured brain signal (Grill-Spector et al., 2006). There were three measurement days, and stimuli were randomised, which minimises the effect of getting tired. The measurements were long because they also included stimuli that were not analysed in this particular study. By focusing on picture stimuli in future measurements the number of presentation times per item could be increased.

6.4 Future directions: Is 'beauty' in the brain of the beholder?

The present results imply that there are commonalities in how concrete nouns are represented in the brains of different individuals. These findings raise new questions: Are there more individual differences for some words than others? For example, are the semantic representations of abstract or emotional words more, or less, similar between different people than those of common concrete nouns? The concrete nouns used here were unambitious in their meanings, but what about words with more nuanced meanings that vary depending on the context? It would be interesting to investigate those differences in meanings and how they are mirrored in brain activity. If we say 'the beauty is in the eye of the beholder', are perceived differences in the meaning of 'beauty' actually grounded in brain activity? More generally, could this phenomenon be part of the basis for the many misunderstandings in communication?

In the beginning of the thesis, the importance of experiences in relation to semantic processing was discussed. It might be possible to experimentally create shared or individual experiences, for example, by using stories or movies as stimuli. It would be also interesting to see if within-subject decoding accuracies could be improved by using individual semantic feature sets which would correlate with individual experiences. If individual feature sets were available, it would be possible to compare, first, the similarity of the feature spaces themselves and, subsequently, investigate whether those similarities might be found in the brain responses.

Here, it was demonstrated that healthy subjects share common components of semantic representations. It has been suggested that emotionally-valenced words evoke different kinds of activation in patients with schizophrenia compared to healthy individuals (Klumpp et al., 2010). Such differences were not detected when depressed patients and healthy control groups were compared. It would be useful to further investigate the differences to find hidden indicators of illness and factors that could be taken into account when

planning treatment.

Findings of common components of semantic processing could be used in the design of new brain-computer interface applications. For example, there could be a communication tool for individuals who cannot move or communicate verbally. Using a similar machine learning approach as here, it might be able to facilitate communication for these patients by translating brain activation into words. For this implementation to work, more method development is needed to improve the signal-to-noise ratio of single trials, which is currently too poor to allow successful decoding.

Chapter 7

Conclusions

This study found commonalities between individuals in how items are represented in the brain. It was possible to decode meanings across subjects using MEG data and Aalto-norms as predictor features. High inter-item variance in the MEG data did not directly influence decoding performance. However, a good signal-to-noise ratio resulted in better within-subject decoding accuracy. Most commonalities between subjects occurred at 150 to 250 ms and 350 to 550 ms with respect to the stimulus onset. Testing pairs (two subjects with training performed on data of one subject and testing on data of the other subject) that resulted in high decoding accuracies seemed to also have more similar internal structures of semantic representations than others, but these phenomena did not correlate directly. As the cross-subject decoding results were worse than within-subject decoding results, it is suggested that there are commonalities in semantic representations in the brain but some of the components of semantic processing are individual. However, more research is needed to answer which factors cause the differences in semantic representations.

A better understanding of how the brain processes semantics will give us more knowledge about how information is stored and used in the brain and here, how it occurs in individuals. By using machine learning, it has become possible to decode relatively abstract aspects of brain function including

meanings of words. In this research, these new approaches made possible to examine shared and individual content of thought.

Bibliography

Baillet, S., Mosher, J. C. & Leahy, R. M. (2001). "Electromagnetic brain mapping". In: *IEEE Signal Processing Magazine* 18.6, pp. 14–30. DOI: `10.1109/79.962275`.

Binder, J. R. & Desai, R. H. (2011). "The neurobiology of semantic memory". In: *Trends in Cognitive Sciences* 15.11, pp. 527–536. DOI: `10.1016/j.tics.2011.10.001`.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M. & Desai, R. H. (2016). "Toward a brain-based componential semantic representation". In: *Cognitive Neuropsychology* 33.3-4, pp. 130–174. DOI: `10.1080/02643294.2016.1147426`.

Birbaumer, N. (2006). "Breaking the silence: brain–computer interfaces (BCI) for communication and motor control". In: *Psychophysiology* 43.6, pp. 517–532. DOI: `10.1111/j.1469-8986.2006.00456.x`.

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E. & Flor, H. (1999). "A spelling device for the paralysed". In: *Nature* 398.6725, p. 297. DOI: `10.1038/18581`.

Birbaumer, N., Strehl, U. & Hinterberger, T. (2004). "Brain-computer interfaces for verbal communication". In: *Neuroprosthetics: Theory and Practice*. World Scientific, pp. 1146–1157. DOI: `10.1142/9789812561763_0039`.

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E. & Piazza, M. (2016). "Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding". In: *NeuroImage* 143, pp. 128–140.

Broster, P. (2010). "File: Bay Horse (6014766934).jpg". In: *Wikimedia Commons*. URL: `https://commons.wikimedia.org/wiki/File:Bay_Horse_(6014766934).jpg#metadata`. accessed:12.2.2019.

Chan, A. M., Halgren, E., Marinkovic, K. & Cash, S. S. (2011). "Decoding word and category-specific spatiotemporal representations from MEG and EEG". In: *NeuroImage* 54.4, pp. 3028–3039. DOI: `10.1016/j.neuroimage.2010.10.073`.

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. (2014). "Unique semantic space in the brain of each beholder predicts perceived similarity". In: *Proceedings of the National Academy of Sciences* 111.40, pp. 14565–14570. DOI: `10.1073/pnas.1402594111`.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A. & Hasson, U. (2017). "Shared memories reveal shared structure in neural activity across individuals". In: *Nature Neuroscience* 20.1, p. 115. DOI: `10.1038/nn.4450`.

Clark, E. V. (1973). "What's in a word? On the child's acquisition of semantics in his first language". In: *Cognitive Development and Acquisition of Language*. Elsevier, pp. 65–110. DOI: `10.1016/B978-0-12-505850-6.50009-8`.

Clarke, A., Devereux, B. J., Randall, B. & Tyler, L. K. (2014). "Predicting the time course of individual objects with MEG". In: *Cerebral Cortex* 25.10, pp. 3602–3612. DOI: `10.1093/cercor/bhu203`.

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J., Velliste, M., Boninger, M. L. & Schwartz, A. B. (2013). "High-performance neuroprosthetic control by an individual with tetraplegia". In: *The Lancet* 381.9866, pp. 557–564. DOI: `10.1016/S0140-6736(12)61816-9`.

Contini, E. W., Wardle, S. G. & Carlson, T. A. (2017). "Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions". In: *Neuropsychologia* 105, pp. 165–176. DOI: `10.1016/j.neuropsychologia.2017.02.013`.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B. & Bonte, M. (2014). "Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe". In: *Journal of Neuroscience* 34.1, pp. 332–338. DOI: 10.1523/JNEUROSCI.1302-13.2014.

Cree, G. S. & McRae, K. (2003). "Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns)." In: *Journal of Experimental Psychology: General* 132.2, p. 163. DOI: 10.1037/0096-3445.132.2.163.

Friederici, A. D. & Singer, W. (2015). "Grounding language processing on basic neurophysiological principles". In: *Trends in Cognitive Sciences* 19.6, pp. 329–338. DOI: 10.1016/j.tics.2015.03.012.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L. & Hämäläinen, M. S. (2014). "MNE software for processing MEG and EEG data". In: *NeuroImage* 86, pp. 446–460. DOI: 10.1016/j.neuroimage.2013.10.027.

Grill-Spector, K., Henson, R. & Martin, A. (2006). "Repetition and the brain: neural models of stimulus-specific effects". In: *Trends in Cognitive Sciences* 10.1, pp. 14–23. DOI: 10.1016/j.tics.2005.11.006.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. (1993). "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain". In: *Reviews of Modern Physics* 65.2, p. 413. DOI: 10.1103/RevModPhys.65.413.

Hari, R. & Salmelin, R. (2012). "Magnetoencephalography: from SQUIDs to neuroscience: neuroimage 20th anniversary special edition". In: *NeuroImage* 61.2, pp. 386–396. DOI: 10.1016/j.neuroimage.2011.11.074.

Harley, T. A. (2008). *The Psychology of Language: From Data to Theory*. 3rd ed. New York: Hove: Psychology Press. Chap. 11, pp. 321–360. ISBN: 1841693820.

Haumann, N. T., Parkkonen, L., Kliuchko, M., Vuust, P. & Brattico, E. (2016). "Comparing the performance of popular MEG/EEG artifact correction methods in an evoked-response study". In: *Computational Intelligence and Neuroscience* 2016, p. 3. DOI: `10.1155/2016/7489108`.

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". In: *Neuron* 76.6, pp. 1210–1224. DOI: `10.1016/j.neuron.2012.10.014`.

Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T. & Gallant, J. L. (2016a). "Decoding the semantic content of natural movies from human brain activity". In: *Frontiers in Systems Neuroscience* 10, p. 81. DOI: `10.3389/fnsys.2016.00081`.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. (2016b). "Natural speech reveals the semantic maps that tile human cerebral cortex". In: *Nature* 532.7600, pp. 453–458. DOI: `10.1038/nature17637`.

Hyvärinen, A. (1999). "Fast and robust fixed-point algorithms for independent component analysis". In: *IEEE Transactions on Neural Networks* 10.3, pp. 626–634. DOI: `10.1109/72.761722`.

Indefrey, P. (2011). "The spatial and temporal signatures of word production components: a critical update". In: *Frontiers in Psychology* 2, p. 255. DOI: `10.3389/fpsyg.2011.00255`.

Just, M. A., Cherkassky, V. L., Aryal, S. & Mitchell, T. M. (2010). "A neurosemantic theory of concrete noun representation based on the underlying brain codes". In: *PloS One* 5.1, e8622. DOI: `10.1371/journal.pone.0008622`.

Kaufman, L. & Rousseeuw, P. J. (2005). *Finding groups in data: an introduction to cluster analysis*. 1st ed. New Jersey: John Wiley & Sons. Chap. 1, pp. 1–68. ISBN: 0471735787.

Kiefer, M. & Pulvermüller, F. (2012). "Conceptual representations in mind and brain: theoretical developments, current evidence and future direc-

tions". In: *Cortex* 48.7, pp. 805–825. DOI: `10.1016/j.cortex.2011.04.006`.

Klumpp, H., Keller, J., Miller, G. A., Casas, B. R., Best, J. L. & Deldin, P. J. (2010). "Semantic processing of emotional words in depression and schizophrenia". In: *International Journal of Psychophysiology* 75.2, pp. 211–215. DOI: `10.1016/j.ijpsycho.2009.12.004`.

Kotsiantis, S. B., Zaharakis, I & Pintelas, P (2007). "Supervised machine learning: A review of classification techniques". In: *Emerging Artificial Intelligence Applications in Computer Engineering* 160, pp. 3–24.

Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P. & Salmelin, R. (1998). "An MEG study of picture naming". In: *Journal of Cognitive Neuroscience* 10.5, pp. 553–567. DOI: `10.1162/089892998562960`.

Lopes da Silva, F. (2010). "Electrophysiological Basis of MEG Signals". In: *MEG: An Introduction to Methods*. Ed. by P. Hansen, M. Kringelbach & R. Salmelin. 1st ed. Oxford University Press. Chap. 1, pp. 1–23. ISBN: 0195307232.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. Taylor & Francis.

McRae, K., Cree, G. S., Seidenberg, M. S. & McNorgan, C. (2005). "Semantic feature production norms for a large set of living and nonliving things". In: *Behavior Research Methods* 37.4, pp. 547–559.

Mellin, I. (2010). *Todennäköisyyslaskenta ja Tilastotiede: Kaavat*. Unigrafia Oy.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A. & Just, M. A. (2008). "Predicting human brain activity associated with the meanings of nouns". In: *Science* 320.5880, pp. 1191–1195. DOI: `10.1126/science.1152876`.

Müllner, D. (2011). "Modern hierarchical, agglomerative clustering algorithms". In: *Preprint*. DOI: `arXiv:1109.2378`.

Murphy, B., Wehbe, L. & Fyshe, A. (2018). "Decoding Language from the Brain". In: *Language, Cognition, and Computational Models*. Ed. by T.

Poibeau & A. Villavicencio. Studies in Natural Language Processing. Cambridge University Press, 53–80. DOI: `10.1017/9781316676974.003`.

Oldfield, R. C. (1971). "The assessment and analysis of handedness: the Edinburgh inventory". In: *Neuropsychologia* 9.1, pp. 97–113. DOI: `10.1016/0028-3932(71)90067-4`.

Palatucci, M., Pomerleau, D., Hinton, G. E. & Mitchell, T. M. (2009). "Zero-shot learning with semantic output codes". In: *Advances in Neural Information Processing Systems*, pp. 1410–1418.

Parkkonen, L. (2010). "Instrumentation and Data Preprocessing in MEG: An Introduction to Methods". In: *MEG: An Introduction to Methods.* Ed. by P. Hansen, M. Kringelbach & R. Salmelin. 1st ed. Oxford University Press. Chap. 2, pp. 24–64. ISBN: 0195307232.

Raizada, R. D. & Connolly, A. C. (2012). "What makes different people's representations alike: neural similarity space solves the problem of across-subject fMRI decoding". In: *Journal of Cognitive Neuroscience* 24.4, pp. 868–877. DOI: `10.1162/jocn_a_00189`.

Rupp, K., Roos, M., Milsap, G., Caceres, C., Ratto, C., Chevillet, M., Crone, N. E. & Wolmetz, M. (2017). "Semantic attributes are encoded in human electrocorticographic signals during visual object recognition". In: *NeuroImage* 148, pp. 318–329. DOI: `10.1016/j.neuroimage.2016.12.074`.

Salmelin, R. (2007). "Clinical neurophysiology of language: the MEG approach". In: *Clinical Neurophysiology* 118.2, pp. 237–254. DOI: `10.1016/j.clinph.2006.07.316`.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M. & Just, M. A. (2008). "Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings". In: *PLoS One* 3.1, e1394. DOI: `10.1371/journal.pone.0001394`.

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M. & Just, M. A. (2011). "Commonality of neural representations of words and pictures". In: *NeuroImage* 54.3, pp. 2418–2425. DOI: `10.1016/j.neuroimage.2010.10.042`.

Shinkareva, S. V., Malave, V. L., Just, M. A. & Mitchell, T. M. (2012). "Exploring commonalities across participants in the neural representation of objects". In: *Human Brain Mapping* 33.6, pp. 1375–1383. DOI: 10.1002/hbm.21296.

Snodgrass, J. G. & Vanderwart, M. (1980). "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity". In: *Journal of Experimental Psychology: Human Learning and Memory* 6.2, p. 174. DOI: 10.1037/0278-7393.6.2.174.

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R. & Mitchell, T. (2012). "Tracking neural coding of perceptual and semantic features of concrete nouns". In: *NeuroImage* 62.1, pp. 451–463. DOI: 10.1016/j.neuroimage.2012.04.048.

Taulu, S. & Simola, J. (2006). "Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements". In: *Physics in Medicine & Biology* 51.7, pp. 1759–1768. DOI: 10.1088/0031-9155/51/7/008.

Vigário, R., Sarela, J., Jousmäki, V., Hämäläinen, M. & Oja, E. (2000). "Independent component approach to the analysis of EEG and MEG recordings". In: *IEEE Transactions on Biomedical Engineering* 47.5, pp. 589–593. DOI: 10.1109/10.841330.

Vigliocco, G. & Vinson, D. P. (2007). "Semantic representation". In: *The Oxford Handbook of Psycholinguistics*, pp. 195–215.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A. & Mitchell, T. (2014). "Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses". In: *PloS One* 9.11, e112575. DOI: 10.1371/journal.pone.0112575.

Appendix A

Measurement documents

Tietoa tutkimuksesta:
Yksittäisten käsitteiden hermostolliset merkitysedustumat

Tämä tutkimus on osa Neurotieteen ja lääketieteellisen tekniikan laitoksen (Aalto-yliopisto, Otaniemi) työtä ja sen tarkoituksena on selvittää, miten, missä ja milloin yksittäiset kielen merkitysyksiköt ovat edustettuina ihmisaivojen sähköfysiologisessa toiminnassa. Jotta tämän kaltaisesta aivotoiminnasta saadaan luotettava käsitys, **tutkimukseen kuuluu 3 MEG mittauskertaa** (yksi mittauskerta kestää noin 2 tuntia, josta tunti varsinaista mittausta). MEG-mittaus tapahtuu osoitteessa Puumiehenkuja 2A, 02150 Espoo (Nano-talo).

Koehenkilöksi soveltuvat 18-40 vuotiaat, oikeakätiset, äidinkielenä suomea puhuvat terveet henkilöt. Mikäli käytätte hermostoon vaikuttavia lääkkeitä (esim. masennuslääkkeet), teillä on todettu luki- tai muu kielellinen häiriö tai kehossanne on metallia (esim. hammasraudat), ette voi osallistua koehenkilöksi. Mikäli epäilette soveltavuuttanne, ottakaa yhteys tutkijaan. Tutkimukseen ei voi osallistua päihtyneenä.

Tutkimuksessa mitataan heikkoja aivosolutoiminnan tuottamia sähkömagneettisia signaaleja magnetoenkefalografia (MEG) -menetelmää käyttäen. Mittauksessa tutkittava istuu mukavassa tuolissa MEG-kypärän alla samalla, kun mittalaite kerää pään ulkopuolelta tietoa aivoissa syntyvistä magneettikentistä. Magneettiset kentät syntyvät, kun aivosolun sisällä tieto siirtyy sähköisenä aaltona. MEG-tutkimus on täysin turvallinen ja kivuton, eikä siihen liity esimerkiksi säteilyä tai lääkeaineita. Tutkimuksessa kiinnitetään iholle pään asennon paikantamiseen käytettävät pienet mittakelat sekä silmän liikkeitä seuraavat elektrodit (elektro-okulografia, EOG). Nämäkin ovat turvallisia ja kivuttomia.

Koska aivojen signaalit ovat hyvin pieniä, mittauslaite on sijoitettu magneettisesti suojattuun huoneeseen, johon on puhe- ja videokamerayhteys. Jos koette suljetun paikan kammoa, se saattaa estää osallistumisenne tutkimukseen. **Tutkimuksen aikana keskitytte Teille esitettäviin sana-, kuva- ja ääniärsykkeisiin**. Toisinaan mittauksessa tehtävänä on myös yhdistää esitetyn ärsykkeen merkitys lauseeseen. Tarkempia ohjeita annamme mittauspäivänä ja tutkija vastaa mielellään, jos mikään tuntuu epäselvältä. Mittauksen aikana tutkijat ovat jatkuvassa videokuva ja ääniyhteydessä teihin, ja te voitte ilmaista puheella tai liikkeellä, jos teillä on jokin hätä.

Teitä pyydetään välttämään liikkumista ja lihasjännitystä (esim. niskajännitys) mittauksen aikana, koska se aiheuttaa häiriöitä MEG-mittaukseen. Istukaa siis vain liikkumatta paikoillaan mittauksen aikana. Turhia silmänräpäytyksiä ja -liikkeitä tulisi myös välttää, koska nekin aiheuttavat häiriöitä. Katse pyydetään keskittämään projektoriruudun keskiosaan. Mittauksen aikana on säännöllisesti taukoja joiden aikana voitte lepuuttaa silmiänne, mutta muuten pyydämme Teitä pysymään paikallanne näidenkin aikana.

MEG-mittauksen lisäksi keräämme anatomiset aivokuvat magneettikuvauksella (MRI-kuvat). Tätä tietoa tarvitaan MEG-signaalien lähteiden paikantamiseen. Jos Teistä on jo anatominen kuva aiemmasta kokeesta Aalto-yliopistolla, niin kirjallisella suostumuksellanne

tätä voidaan hyödyntää myös tässä kokeessa. MRI-kuvat mitataan Otaniemen kampuksella sijaitsevassa AMI (Advanced Magnetic Imaging) -keskuksessa sovittuna aikana. Kuvaus on turvallinen ja kivuton terveelle tutkittavalle, mutta vaatii, ettei kehossa ole metallia (esim. sydämentahdistin, metallinen proteesi). Ennen MRI-kuvausta Teille selitetään huolellisesti mittauksen kulku sekä tehdään turvallisuuskysely. Mittauksen aikana Te makaatte kuvauslaitteessa. Kuvauksen kesto on noin 10 minuuttia. AMI-keskus sijaitsee osoitteessa Otakaari 5 (I-siipi, Magneettitalo).

Osallistuminen tähän kokeeseen on täysin vapaaehtoista. Teillä on oikeus keskeyttää tutkimus missä vaiheessa tai mistä syystä tahansa. On erittäin epätodennäköistä, että aivokuvissa havaittaisiin poikkeavuutta, joka saattaisi vaikuttaa terveyteenne. Mikäli poikkeavuuksia kuitenkin havaitaan, tutkijat konsultoivat ryhmässä työskentelevää lääkäriä ja ilmoitamme löydöksistä teille. Koehenkilöille, jotka eivät ole töissä Neurotieteen ja lääketieteellisen tekniikan laitoksella tai AMI-keskuksessa, korvataan matkakustannukset ja menetetystä työajasta maksetaan pieni korvaus (36 euroa/kerta).

Kiitämme mahdollisesta osallistumisesta tutkimukseemme, jolla pyrimme lisäämään kielen sanojen merkitysten aivoperustan ymmärtämistä! Jos Teillä on mitään kysyttävää, ottakaa yhteyttä alla olevaan vastuulliseen tutkijaan:

Annika Hultén, PhD

Lotta Lammi, tutkimusavustaja

MEG-ohjeistus koehenkilölle

MEG-kokeessa esitetään yksitellen kirjoitettuja ja puhuttuja sanoja sekä kuvia. Kokeen aikana tehtävänäsi on lukea sanoja, katsoa kuvia ja nimetä ne mielessä sekä kuunnella puhuttuja sanoja. Suorittaessasi tehtävää sinun tulee ajatella esitettävien ärsykkeiden merkityksiä. Toisinaan ärsykettä seuraa lause, jonka ensimmäinen sana on korvattu kolmella pisteellä (...).. Jos juuri esitetty sana/kuva/ääni sopii merkitykseltään lauseen ensimmäiseksi sanaksi, vastaa **"kyllä" nostamalla oikeaa etusormea**. Jos juuri esitetty sana ei sovi lauseen ensimmäiseksi sanaksi, vastaa **"ei" nostamalla vasenta etusormea.**

Esimerkki 1

"KYLLÄ"

Esimerkki 2

"EI"
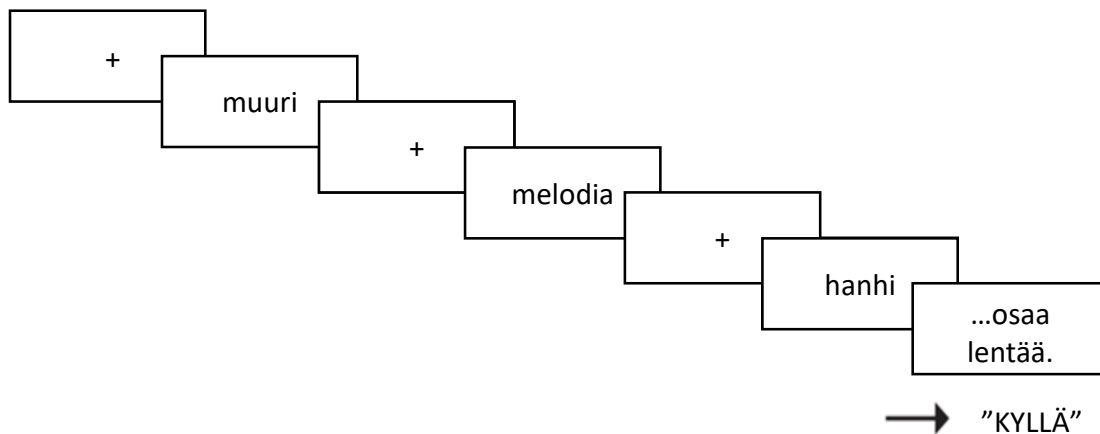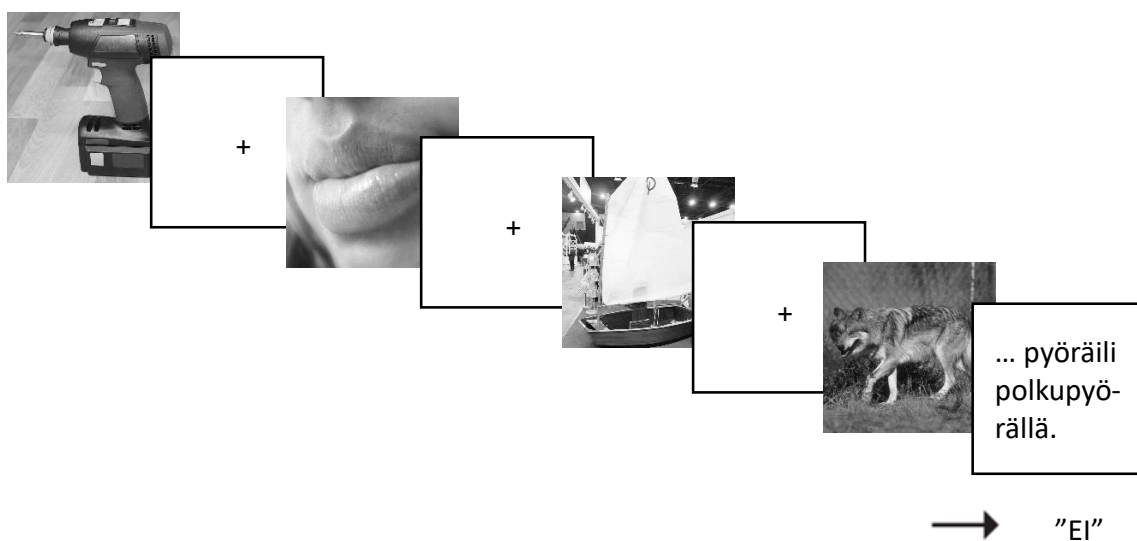
Koe jaksotetaan noin 7 minuutin osiin. Pyri olemaan mittausten aikana niin liikkumatta kuin mahdollista. Pyri räpyttämään silmiäsi mahdollisimman vähän tai räpäyttele sanojen välissä esitettävän ristin aikana. Yritä myös pitää katse koko ajan ruudun keskellä, jolloin vältät tehokkaammin silmänliikkeitä.

Kokeen ensimmäisessä osassa on yhteensä 6 jaksoa, joiden välissä pidetään 2 hieman pidempää taukoa. Näiden taukojen aikana voit vapaasti räpyttää silmiä ja liikkua maltillisesti.

Näemme sinut ja kuulemme sinua koko mittauksen ajan, joten voit missä tahansa vaiheessa ilmoittaa, jos kokeen aikana ilmenee ongelmia (esim. et ymmärtänyt ohjeita) tai haluat keskeyttää kokeen.

Älä epäröi kysyä, jos joku asia jää epäselväksi!

Lue alla olevat sanat ja ilmoita kokeenjohtajalle, jos joku niistä on sinulle vieras tai merkitykseltään epäselvä.

jalka	pilvi
sotilas	tie
lammas	puisto
lääkäri	kuningas
leijona	kotka
sakset	linna
nenä	saari
lapio	silta
laiva	museo
haarukka	sormus
kampa	poliisi
lapsi	rekka
juna	lusikka
pappi	opettaja
bussi	selkä
kirkko	tuomari
vene	pesä
ankka	kissa
karhu	joki
silmä	kirjasto
kirjasto	suu
kallio	koira
hevonen	
vankila	
aalto	
vanki	
tehdas	
hiiri	
meri	
varvas	
korva	
auto	
torni	
käsi	
vuori	
pallo	
sormi	
saha	

Taustatietolomake

Nimi: _____

Sukupuoli_____

Ikä:_____

Puhelin numero:_____

Sähköposti: _____

Koulutustaso, tutkinto:_____

Ammatti:_____

Kätisyys:_____

Äidinkieli:_____

Onko koulussa ollut vaikeuksia lukemaan tai kirjoittamaan oppimisessa (luki-häiriö)?:

_____

Onko todettu neurologisia sairauksia (onko tutkittu aivoja, epilepsia, kallovamnmoja, pidempi tajuttomuus tms.)?:

_____

Onko kehossasi metallia (esim. hammasraudat, kirurgiset klipsit/naulat, sydämmentahdistin tms.)

_____

Muuta:

_____

_____

_____

Tutkittavan nimi: _____

Päivämäärä: _____

Edinburghin kätisyyskartoitus (Edinburgh Handednes Inventory)

Laita yksi rasti oikeaan sarakkeeseen sen mukaan, kumpaako kättä käytät seuraavissa toiminnoissa tai seuraavia esineitä käyttäessäsi. Jos käytät toiminnossa ainoastaan toista kättä, etkä käyttäisi toista missään tilanteessa, laita kaksi rastia oikeaan sarakkeeseen. Jos taas voisit käyttää kumpaa kättä tahansa, laita rasti molempiin sarakkeisiin. Täytä kaikki kohdat, ja jätä tyhjää vain, jos sinulla ei ole lainkaan aiempaa kokemusta toiminnosta.

Jotkut toiminnot listassa vaativat molempien käsien käyttöä. Näissä tapauksissa se osa toiminnosta, johon käytettävästä kädestä ollaan kiinnostuneita, lukee suluissa.

Toiminto	Oikea	Vasen
Kirjoittaminen		
Piirtäminen		
Heittäminen		
Sakset		
Hammasharja		
Veitsi (ilman haarukkaa)		
Lusikka		
Lattiaharja (ylempi käsi)		
Tulitikun sytytys (tikku)		
Laatikon avaus (kansi)		
YHTEENSÄ (rastien lukumäärä sarakkeissa)		

Aalto-yliopisto

Tietoon perustuva suostumus magnetonenkefalografialla (MEG) suoritettavaan tutkimukseen Aalto-yliopiston Neurotieteen ja lääketieteellisen tekniikan laitoksessa.

Tutkittava:

Nimi: _____

Henkilötunnus: _____

Osoite: _____

Sähköposti: _____

Puhelinnumero: _____

Olen lukenut ja ymmärtänyt saamani tutkimustiedotteen ja saanut riittävästi tietoa tutkimuksen kulusta. Olen ymmärtänyt, että osallistumiseni tutkimukseen on täysin vapaaehtoista ja että voin missä tutkimuksen vaiheessa tahansa keskeyttää tutkimuksen sen vaikuttamatta kohteluuni jatkossa. Minulle on selvitetty, että halutessani saan alla nimetyltä tutkijalta lisätietoja tutkimuksen yleisistä periaatteista ja edistymisestä tai itseäni koskevista tuloksista.

Tutkimustulokseni ovat vain tutkijaryhmän tutkijoiden käytettävissä eikä niitä anneta ulkopuoliselle taholle ilman erikseen antamaani kirjallista suostumusta. Tutkimuksesta vastaava tutkija voi kuitenkin antaa muiden yhteistyökumppaniensa analysoida tutkimustuloksiani tieteellisiä tarkoituksia varten ilman erillistä suostumusta edellyttäen että tulosten anonyymisyys on varmistettu. Tulosten kaikenlainen kaupallinen hyödyntäminen on kielletty. Ymmärrän, että aineistoa ja tutkimustietoa kerätään pelkästään tieteellistä tutkimusta varten eikä sitä luovuteta osittainkaan koehenkilölle itselleen. Allekirjoituksellani vahvistan osallistumiseni tähän tutkimukseen ja suostun vapaaehtoisesi tutkimushenkilöksi.

_____
Paikka ja aika

_____
Tutkittavan henkilön allekirjoitus

_____
Vastuullisen tutkijan allekirjoitus (nimen selvennys)
Aalto University, / The Department of Neuroscience and Biomedical Engineering
Puumiehenkuja 2 B, 00076 Aalto

Appendix B

Stimulus words

koira/dog	varvas/toe	vanki/prisoner	kirja/book
hevonen/horse	sormi/finger	pappi/priest	pallo/ball
ankka/duck	nenä/nose	opettaja/teacher	saha/saw
kotka/eagle	kirkko/church	lääkäri/physician	sormus/ring
kissa/cat	tie/road	tuomari/judge	sakset/scissors
leijona/lion	tehdas/factory	lapsi/child	lusikka/spoon
hiiri/mouse	linna/castle	joki/river	lapio/shovel
karhu/bear	silta/bridge	saari/island	kampa/comb
lammas/sheep	vankila/prison	meri/sea	auto/car
selkä/back	torni/tower	puisto/park	laiva/ship
käsi/hand	kirjasto/library	vuori/mountain	juna/train
silmä/eye	museo/museum	kallio/rock	vene/boat
jalka/foot	kuningas/king	aalto/wave	bussi/bus
korva/ear	sotilas/soldier	pilvi/cloud	rekka/truck
suu/mouth	poliisi/police	pesä/nest	

Table B.1: Stimulus words in Finnish/English

Appendix C

Feature questions

Liittyykö se taiteeseen?/Is it related to art?

Onko se nisäkäs?/Is it a mammal?

Onko sillä tai liittyykö siihen jalat?/Does it have or is it related to legs?

Voiko sen omistaa?/Can it be owned?

Onko se esine?/Is it an object?

Liittyykö se valtioihin?/Is it related to states?

Onko se ihminen?/Is it a human?

Voiko sitä oppia tai opiskella?/Can you learn or study it?

Onko sitä erikokoisina?/Is it different in size?

Liittyykö se lapsuuteen?/Is it related to childhood?

Liittyykö siihen vihreä väri?/Is it related to green?

Onko se rakennettu tai valmistettu?/Is it built or manufactured?

Liittyykö se rikoksiin?/Is it related to crimes?

Voiko siinä olla tekstiä?/Can it contain text?

Voiko sen sisälle mennä?/Can somebody go inside of it?

Onko sillä tai liittyykö siihen kädet?/Does it have or is it related to hands?

Pidetäänkö sitä hyvänä asiana?/Is it a good thing?

Voiko se olla vaarallinen?/Can it be dangerous?

Liittyykö se aikaan?/Is it related to time?

Onko se kulkuneuvo?/Is it a vehicle?

Koostuuko se sanoista?/Does it consist of words?

Liittyykö se kuninkaallisuuteen?/Is it related to royalism?

Liittyykö se vesistöihin?/Does it relate to waterways?

Tarvitaanko sen käyttämiseen sormia?/Is it needed to use fingers when using it?

Koostuuko se monesta osasta?/Does it consist of many parts?

Liittyykö se kasvoihin?/Does it relate to the face?

Onko se fysikaalinen suure?/Is it a physical quantity?

Voiko sitä kehittää tai parantaa?/Can it be developed or improved?

Käytetäänkö sitä johonkin?/Is it used for something?

Voiko sen havaita merellä?/Can it be detected at sea?

Liittyykö siihen raha?/Does it relates to money?

Onko se osa jotakin?/Is it part of something?

Sisältääkö se tarinan?/Does it contain a story?

Onko se pitkä tai korkea?/Is it long or high?

Liittyykö se oikeuslaitokseen?/Does it relate to the judiciary?

Pidetäänkö sitä arvokkaana?/Is it considered valuable?

Onko se eläin?/Is it an animal?

Liittyykö se asumiseen?/Is it related to housing?

Voiko sitä kontrolloida?/Can it be controlled?

Voiko sen havaita metsässä?/Can it be detected in the woods?

Onko siinä nestettä?/Does it contain liquid?

Liittyykö se koulunkäyntiin?/Is it related to education?

Voiko se olla huono?/Can it be bad?

Liittyykö siihen numero neljä?/Does the number four relates to it?

Voiko se mennä rikki tai vaurioitua?/Can it be broken or damaged?

Liittyykö siihen paperi?/Is it related to paper?

Voiko sen nähdä?/Can it be seen?

Liittyykö se ajatteluun?/Is it related to thinking?

Liittyykö se menneisyyteen?/Is it related to the past?

Onko se konkreettinen?/Is it concrete?

Onko siinä varsi?/Is there a stem?

Pidetäänkö sitä viisaana?/Is it wise?

Liittyykö se sairauksiin?/Is it related to diseases?

Onko se pieni?/Is it small?

Onko se tärkeä?/Is it important?

Voiko sitä pitää kädessä?/Can it be held in hand?

Onko siinä jotain terävää?/Is it something sharp?

Liittyykö se luontoon?/Is it related to nature?

Liittyykö siihen virkapuku?/Does it relate to a uniform?

Voiko se olla aikuinen?/Can it be an adult?

Onko se ihmisen keksimä?/Is it invented by man?

Pidetäänkö sitä lemmikkinä?/Is it considered a pet?

Herättääkö se paljon tunteita?/Does it awaken a lot of emotions?

Liittyykö se kehonosiin?/Is it related to body parts?

Osaako se lentää?/Can it fly?

Voiko sen ostaa?/Can it be bought?

Liikkuuko se laumoissa?/Does it move in hordes?

Voiko niiden määrän laskea?/Is it countable?

Voiko siinä olla metallia?/Can it be metal?

Lisääkö sen olemassaolo turvallisuuden tunnetta?/Does its existence add security?

Liittyykö siihen sininen väri?/IS it related to blue?

Käytetäänkö sitä vahingontekoon?/Is it used for damage?

Onko se eloton?/Is it inanimal?

Liittyykö se pelaamiseen?/Is it related to gaming?

Voiko sen kuulla?/Can it be heard?

Liittyykö se työntekoon?/Is it related to work?

Onko siinä jotakin valkoista?/Is there something white?

Liittyykö se kauneuteen?/Is it related to beauty?

Onko se karvainen?/Is it hairy?

Sisältääkö se puuta?/Does it contain wood?

Liittyykö se johonkin tapahtumaan?/Is it related to an event?

Liittyykö siihen moottori?/Is tit related to an engine?

Onko useimmissa kodeissa sellainen?/Is it typically located at home?

Voiko se olla henkilökohtaista?/Can it be personal?

Liikkuuko se?/Does it move?