Characterizing the Spatiotemporal Neural Representation of Concrete Nouns Across Paradigms

Gustavo Sudre

December 2012

Center for the Neural Basis of Cognition
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Tom Mitchell, Chair
Dean Pomerleau
Michael Tarr
Riitta Salmelin, Aalto University, Helsinki, FI

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Gustavo Sudre is supported by a graduate fellowship from the multi-modal neural training program (MNTP) and a presidential fellowship in the Life Sciences from the Richard King Mellon Foundation.

Keywords: conceptual knowledge representation, MEG, machine learning, zero-shot, semantic memory, neural code

To Kimmie, even though it's not really about computer worms and brain-machine interfaces.

Abstract

Most of the work investigating the representation of concrete nouns in the brain has focused on the locations that code the information. We present a model to study the contributions of perceptual and semantic features to the neural code representing concepts over time and space. The model is evaluated using magnetoencephalography data from different paradigms and not only corroborates previous findings regarding a distributed code, but provides further details about how the encoding of different subcomponents varies in the space-time spectrum. The model also successfully generalizes to novel concepts that it has never seen during training, which argues for the combination of specific properties in forming the meaning of concrete nouns in the brain.

The results across paradigms are in agreement when the main differences among the experiments (namely, the number of repetitions of the stimulus, the task the subjects performed, and the type of stimulus provided) were taken into consideration. More specifically, these results suggest that features specific to the physical properties of the stimuli, such as *word length* and *right-diagonalness*, are encoded in posterior regions of the brain in the first hundreds of milliseconds after stimulus onset. Then, properties inherent to the nouns, such as *is it alive?* and *can you pick it up?*, are represented in the signal starting at about 250 ms, focusing on more anterior parts of the cortex. The code for these different features was found to be distributed over time and space, and it was common for several regions to simultaneously code for a particular property. Moreover, most anterior regions were found to code for multiple features, and a complex temporal profile could be observed for the majority of properties. For example, some features inherent to the nouns were encoded earlier than others, and the extent of time in which these properties could be decoded varied greatly among them. These findings complement much of the work previously described in the literature, and offer new insights about the temporal aspects of the neural encoding of concrete nouns.

This model provides a spatiotemporal signature of the representation of objects in the brain. Paired with data from carefully-designed paradigms, the model is an important tool with which to analyze the commonalities of the neural code across stimulus modalities and tasks performed by the subjects.

Acknowledgments

Working in the intersection of two challenging fields such as machine learning and neuroscience is certainly not easy. On the plus side, it is great for doing science. We never run out of interesting questions to ask, and the ways to go about answering them are countless. On the other hand, it is often hard to convince people from both fields that what you are working on is indeed an advancement to their fields or at least as cool as you think they are, and there is always a lingering doubt about whether you managed to become an expert in either of the areas.

I, for one, am quite happy with the balance achieved in my work between machine learning and neuroscience. The neuroscience aspect of it is geared towards knowledge representation, and more specifically, to how concrete nouns are represented in the brain. This is an area in which I have always been interested, and I cannot express how lucky I was to find a group of people that was equally interested in it. Even better, they wanted to use computers and algorithms to study it, which is another area I have always favored. So, the first person I want to thank is the leader of that group and my advisor, Tom Mitchell, who has been an inspiration and career model for me from the start. His positive attitude towards every single issue makes it not only easy to be his student, but it is also an example on how to respond to any challenges with which one is faced in life. His way to take new approaches to tackle old problems, and to extract meaning from any poor result that I had been struggling to come up with for weeks, are also lessons I will take with me for the rest of my career. Finally, his enthusiasm to take a new student with questionable experience in a new imaging modality to study a mutually interesting problem (knowledge representation) is also something for which I'll always be grateful.

I also want to thank past and current members of the group. Dean Pomerleau, a creative force that has the power of shedding a special light into problems and make them look simpler than ever. Dean is a guy that has the skill of being successful at everything he tries, and I hope some day I can be half as gifted as him at that! He is a person with whom you can talk about a broad range of topics, and from whom I learned so much, so I can't even begin to express how important his lessons and motivation have been through every step of my research. Mark Palatucci, for his patience and skills in explaining the basics of machine learning and statistics, but also for being a great friend to discuss life topics in general. Alona Fyshe, for her friendship, honesty, and for always gracefully serving as a soundboard to many of my initially stupid ideas. Brian Murphy and Leila Wehbe, for their support in dealing with many of the issues every graduate student faces during their career.

The work shown here has also been marked by a critical collaboration with Riitta Salmelin's group, from Aalto University in Finland, and I hope that this collaboration continues to be successful in the future. The time I spent there has been of utmost importance to my knowledge of MEG and the study of language. Specifically, I'd like to thank Riitta and all her group for the friendship and patience while I was there. I sincerely appreciate the learning opportunity Riitta has given me, her straight-forward, timely, and thorough comments to any work that needed her feedback, the reminders about what people already know so we didn't re-invent the wheel, and her kind attempts to ground our crazy ideas in something close to reality! Additionally, special thanks to Mia Liljeström, Johanna Vartiainen, Annika Hultén, Hanna Renvall, Hannu Laaksonen, Claire Stevenson, and Jan Kujala, who served not only as colleagues and teachers in the lab, but also as friends and family during my stay in Finland.

My journey in MEG would have never started if it hadn't been for Doug Weber, with whom I did my MS degree and at some point realized that the lost Brazilian kid in his lab could be of use exploring this new technique to study sensorimotor processing. Along the same lines, I'd like to thank Wei Wang, who helped me through the growing pains with Matlab and the basics of signal processing. But I wouldn't be anywhere near where I am today in my studies on MEG if it wasn't for my friend Dr. Anto Bagic. I cannot thank him enough for entrusting that shy kid that was just beginning with a variety of projects so I could further develop my data analysis skills. Anto was also one of the first believers that we could use MEG to study knowledge representation the way we wanted, and it's because of him that we had the freedom to try many paradigms people would consider (at least) non-standard in the MEG scanner. While in this topic, I'd also like to thank Anna Haridis for her patience with us while we played around with these crazy paradigms, for helping to prep our subjects, and, most importantly, for making sure the MEG machine was well-behaved. Dr. Bagic is also behind introducing me to many of the key players in the MEG field, and I'll be forever grateful for that. Of all these key players, I have to single-out Lauri Parkkonen, the MEG oracle, who has been a good friend and is always there to share his knowledge no matter how basic my question is. From all these collaborations I've been involved with, I'd also like to thank Rich Randall, James Becker, and Melissa Fabrizio, for their friendship and patience with my delays in getting back to them, and for graciously pretending to understand the convoluted explanations I give about things people think I know.

Many of these collaborations would also not have been possible if it wasn't for the Center for the Neural Basis of Cognition, and all their amazing work in promoting this open and fluid environment between Carnegie Mellon University and University of Pittsburgh. Mike Tarr, the director of the CNBC in the CMU side, has been of unique importance for his vision and enthusiasm, as well as for his support for initiatives such as the open MEG course I taught. There, I would like to thank all the students for indulging my experiment at being a teacher, and some of the other people involved with MEG in Pittsburgh for their help (Avniel Ghuman, Erika Lang, Yang Xu, and Stephen Foldes). I'd also like to thank Intel Labs for allowing access to their Open Cirrus computing cluster (most of the computations of the results you see in this dissertation were performed there), their guidance in using it, and for providing me with an office space!

More formally, I'd like to thank my dissertation committee for their suggestions through all the steps of this work, and also the Program in Neural Computation, for providing an environment where students have the freedom to choose to study what they are interested in, and the independence to work with whomever they want to work.

Last, but not least, I'd like to thank the part of my family in Brazil for their unconditional support and for believing in me all these years. And my family in the US, especially Kimberlee Eberle-Sudre, my wife, for everything she's had to put up with in this journey and still be able to maintain the power to make me smile even in the toughest of times.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Knowledge representation has been the subject of several studies in neuroscience [29, 56, 71]. More specifically, fMRI has been used extensively to study how the human brain represents concrete objects in terms of neural activity [7, 38, 79]. fMRI offers relatively high spatial resolution (1 - 3 mm), but because it measures a slow signal (the BOLD response) it has a low temporal resolution on the order of seconds. Still, these studies have successfully identified locations in the brain that are associated with thinking about properties of the given objects. For example, Just et al. [38] show that areas of the sensorimotor cortex become more active when subjects think of different tools, while regions of the parahippocampal and fusiform gyri display increased activation when subjects focus on properties of different buildings. By looking at the results of many of these studies on semantic knowledge, one can attempt to describe "where we know what we know" in the brain [74].

Magnetoencephalography (MEG) measures the magnetic fields associated with neuronal activities in the brain. While its spatial accuracy is limited to the centimeter range in cognitive tasks, it has high temporal resolution on the order of milliseconds [24], which can be used to shed light on the temporal characteristics of the processes associated with knowledge representation. In other words, MEG can help to describe "when we know what we know". Previous MEG studies [77] have shown the evolution of neuronal activity through the cortex while subjects performed tasks such as word reading, picture naming, and listening to speech. These studies have also revealed regions in time and space that are affected by changes in the perceptual aspects and the semantic content of the stimuli. Such findings motivate the quest for a clearer picture of the type of information that these cortical regions encode over time during language comprehension.

Representation of knowledge in the brain has been the focus of decades of research [8, 29, 56]. Although the future goal is to be able to understand what happens in the brain while processing any length of text, sounds, or movie, this research question becomes more tractable when the task is divided into smaller pieces. Thus, it has been common to study the brain activity related to the processing of single nouns or verbs. In particular, our group has employed a combination of machine learning and neuroscience techniques to study what information is encoded in the brain during the half second it takes us to comprehend the meaning of a concrete noun.

One way to study this information is to train classifiers to predict features of the nouns from brain data. From that, it is possible to acquire insights about when and where information related to concrete noun comprehension is encoded in the MEG signal (as a proxy for where/when

it is encoded in the brain). This dissertation presents a model that can track the information related to knowledge representation that is encoded in the spatiotemporal evolution of the neural activity. The model is used to investigate distributional and compositional claims about the neural code, and also to show the specificity of the MEG signal in predicting different nouns. Also, by leveraging the data from different subjects and paradigms, the work described in this dissertation investigates how concrete nouns are represented across different stimulus modalities and experimental tasks.

More specifically, this dissertation investigates the following questions:

- Is there enough information in the MEG signal to decode different properties of concrete nouns?
- Can we leverage the compositional aspect of the code to predict nouns our models have never seen during training?
- What is the flow of information in different regions of the brain at specific time points?
- How do the results compare to previous findings in the literature?
- What is the spatiotemporal distribution of the neural code for these different properties?
- Is there a sequence of activation among properties inherent to the nouns (e.g. *is it alive?*, *is it usually inside?*)?
- How do the differences among experimental paradigms affect the results?

In the remaining of this chapter, the reader will be presented with a review of the most relevant literature to the work presented in the dissertation, and with a description of the datasets that will be analyzed in the next chapters. Then, chapter 2 presents the first model and use it to analyze data from the *answer-questions* paradigm. Chapter 3 argues for a different model and shows how the previous results can be reproduced, and then presents complementary questions that can be asked using the new model. Chapter 4 uses this model to study the *freely-think* paradigm, and chapter 5 uses the same model to study two new paradigms: *1back-text* and *1back-speech*. Finally, chapter 6 investigates the importance of the number of repetitions in the results of the different datasets, and tests a couple hypotheses about the neural code for representing concrete nouns.

1.1 Related Work

1.1.1 Different processes in noun comprehension

The spatiotemporal characteristics of the different processes involved in concrete noun comprehension have been extensively studied in MEG/EEG literature. This section summarizes some of the key findings related to the three stimulus modalities studied in this dissertation: text, speech, and pictures. Figure 1.1 shows the main processes that are thought to occur in the first 800 ms after stimulus onset in each modality. Please see [77] for a more in-depth review of these processes.

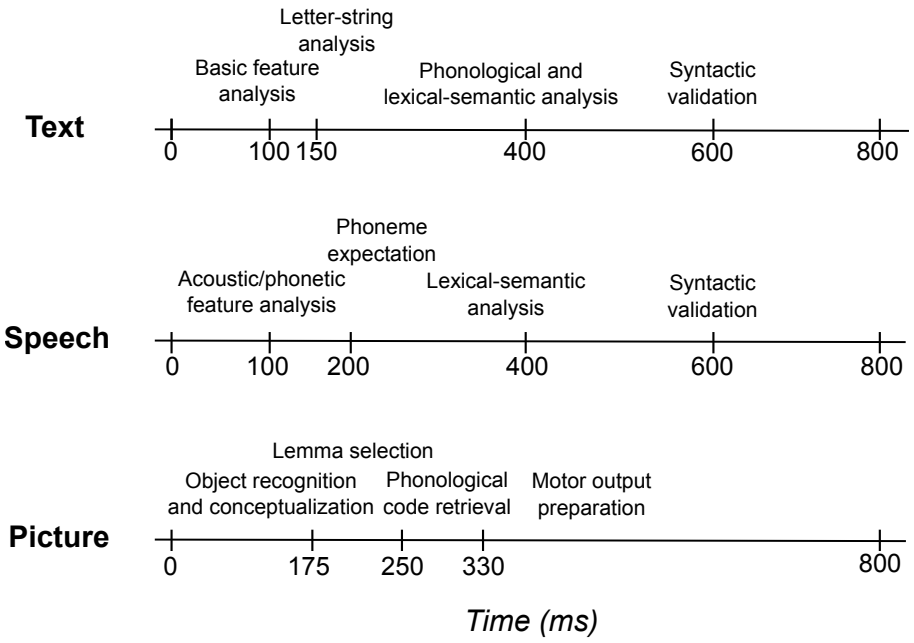


Figure 1.1: Schematic of the main processes that are thought to occur in the first 800 ms following stimulus presentation, according to MEG/EEG literature.

Text

When reading single words, the first peak of activation occurs bilaterally around the midline of the occipital cortex at about 100 ms, and it reflects basic feature analysis non-specific to the stimulus. It is directly modulated by the amount of noise through the character string, as well as by the number of characters in the string. Then, the inferior occipitotemporal junctions are activated around 150 ms, but only the left hemisphere is functionally specific to letter-string stimuli, regardless of whether they are words or nonwords, and so it does not respond as strongly to textual stimuli non-specific to language (e.g. symbols) [86]. In comparison to the initial response at 100 ms, this left occipitotemporal activation is negatively modulated by the amount of noise over the letter string (i.e. magnetic response of higher amplitude when less noise is present). Later, phonological and lexical-semantic analysis are reflected on the activation of the left-superior temporal cortex at 200-600 ms, with a peak at 400 ms (N400m) [32]. This activation is also modulated by words vs. nonwords (lexicality) [100], and semantic congruity, where the amplitude of the N400m response is weaker and shorter-lasting the more expected a word is in a given context. The location of the origin of the N400m has also been suggested to spread more towards anterior regions of the temporal cortex and the inferior frontal cortex [22, 55], when source localization was performed using distributed source models. This left-hemisphere activation is also known to be affected by phonological priming between 200-400 ms, and by semantic priming after 200 ms [95].

Speech

MEG/EEG activity evoked by listening to single words starts bilaterally at the superior temporal cortex at about 100 ms (N100). In the left hemisphere, the N100 is stronger and it builds up faster for speech sounds (either vowel or consonants) than for pure tones, but such distinctions are not found in the right hemisphere [65]. The N100 is followed by the N400, which starts at about 200 ms after stimulus onset and ends around 600-800 ms, with a peak at 400 ms [33, 55]. Although the activation up to 150 ms is modulated by acoustic/phonetic features, after 150 ms phonological categories start affecting the neural activity, and after 200 ms lexical-semantic analysis effects can be noticed in the superior temporal activation. N400 modulation can be better appreciated when the spoken word is put into a context, such as a sentence or a series of priming words. Specifically, the more expected the word is in the context, the weaker the amplitude of the N400 (an effect more visible in the left hemisphere). More frequent words also tend to evoke smaller N400 [21]. Semantic priming affects left hemisphere activity after 250 ms, but both phonological and semantic priming affect right hemisphere response after 400 ms (while phonological priming predominantly affects left hemispheric response at the N100) [95].

When a word is presented as part of a context, such as a sequence of words or a phrase, two other processes are salient: the N200 and P600/SPS (please refer to [21] for a more in-depth overview of speech comprehension). The N200 tends to overlap in time with the N400, making it harder to distinguish the neural generators of both effects. Still, the N200 seems to be present only for speech stimuli, and it emerges when the expected phoneme of a given word in context deviates from what was expected. The P600/SPS occurs when a syntactic constraint is violated in the sentence, and is also observed during reading [78]. In sum, the literature in speech comprehension has shown that there is no strict point in time when lexical selection ends and semantic integration begins, because the spoken word is evaluated with respect to its context in parallel to being identified based on the acoustic information alone.

It is important to note that in both textual and sound modalities there is a N400 activation during single noun processing. While the activity before 200 ms has different spatiotemporal characteristics based on the modality being studied, mostly starting at the primary sensory areas and spreading along the ventral streams, they converge in this peak around 400 ms in the anterior and superior temporal cortices and inferior prefrontal regions, primarily on the left hemisphere. This activity has been suggested to reflect a semantic network that does not depend on the stimulus modality [95], and it is likely to correspond to a representation of concepts that is independent of stimulus modality.

Pictures

Picture naming is a task that is worth studying in depth because it possibly includes all stages of word production, from conceptualization to selection of a lexical item and phonological encoding. However, the neuroscience behind picture naming is thought to be far from simple. Although the sequence of activation is somewhat consistent across different paradigms, it is hard to assign functional roles to specific regions and time periods. Several studies have consistently pointed out the involvement of the left temporal cortex in naming pictures of objects [50]. MEG data has been used to characterize the evolution of activity in this task starting at bilateral occip-

ital cortex before 200 ms after stimulus onset, followed by parietal and temporal regions after 200 ms, and later recruiting frontal regions (especially if overt naming is required) after 300 ms [35]. Finally, a meta-analysis has proposed that the evolution of activity in picture naming starts in occipital and ventro-temporal regions (before 175 ms following stimulus onset), which is related to visual object recognition and conceptualization. Then, the mid-section of the left middle temporal gyrus is involved at selecting the lemma (i.e. the canonical form of a given word) from the mental lexicon at 175-250 ms, followed by the activation of posterior parts of left middle and superior temporal gyri (i.e. Wernicke's area) from 250-330 ms reflecting phonological code retrieval. Finally, Broca's area in the left inferior frontal gyrus and bilateral sensorimotor areas is recruited for preparation for oral output after 330 ms [36, 77]. These final stages of syllabification and self-monitoring, roughly from 300 ms onward, are thought to occur somewhat in parallel.

1.1.2 Conceptual knowledge representation

In all three stimulus modalities studied in this dissertation, there is a part of the activation that is thought to reflect semantics, or the representation of the concepts for which the stimuli stand. In fact, a few studies [1, 9, 80, 81] have shown that it is possible to distinguish between categories of objects across stimulus modalities, which further corroborates the existence of a stable and identifiable semantic representation that is common for the different modalities.

Before we review the main findings in the literature about this representation, it is important to be clear about how we define semantics in this dissertation. In one side of the spectrum, semantics can be seen as a similarity structure that helps to define the role of different concepts, regardless of their perceptual properties. For example, according to that view objects such as paper, pencil, and typewriter would be very closely represented. This common internal representation mediates the mapping between multiple input and output modalities, such as text, pictures, tactile, or sounds. The more modalities it needs to mediate, the less similar to any modality it is [68]. This view also distinguishes between what is being encoded and what gets activated in the brain. For example, just because areas involved with visual imagery get activated when we think about an object, regardless of the modality with which the object is presented, this activation is not necessarily part of semantics. In this view, we use semantics to get to visual imagery, but the semantics part happens elsewhere. This similarity structure also needs to code for idiosyncrasies (e.g. penguins are birds, but they do not fly), and be able to support predictions. For example, we know that living things move differently than nonliving, so categorizing something as living will affect our predictions of how that object will move.

In this dissertation we adopt a more flexible definition of semantic representation, more along the lines of semantic memory and conceptual knowledge representation as extensively described in the review by Martin [56] and also in Mahon and Caramazza [51]. According to this definition, the semantic representation is encoded in the neural activity through perceptual and motor properties, along the lines of the embodied cognition model, so that these properties reflect how one interacts with the objects in the environment [3]. These properties are largely encoded in the same areas involved in processing them. For example, touch and motor properties would be encoded in sensory and motor cortices, respectively. Visual imagery and word concreteness, normally attributed to happen in the fusiform area, would also be considered part of the semantic

representation. The ventral temporal cortex is also related to processing and storing color information. In addition to these areas, some regions such as the ventro-lateral prefrontal cortex control and modulate the information stored elsewhere in the brain.

We also agree that this property-based model is combined with a domain-specific organization, such that concepts belonging to the same category might be encoded closer together. The representation of visual categories, such as faces, cats, and manmade objects has been found to be distributed and overlapping in regions as early as the ventral temporal cortex [31]. But it is conceivable that the domain-specificity seen in some regions of the brain arise directly from the innate connectivity with a network of regions responsible for processing the characteristic properties of that domain [52]. There has also been some debate about how different levels of specificity of objects presented to subjects can lead to the recruitment of different regions in the brain (see Tarr and Cheng [87] for a review). Regardless of these discussions, the objects used in the paradigms analyzed in this dissertation belong to the same basic level of specificity, so that the same neural circuitry should be involved in their distinction across modalities.

It is also safe to assume that different tasks performed by the subject will evoke different properties related to the nouns. For example, while the knowledge of how to use a screwdriver might be part of the representation for the tool, it is possible that the activity that encodes such property does not get evoked at every instance a screwdriver is presented. Instead, one might need to actively think of how to use the objects (e.g. in a semantic similarity task) in order to evoke such activity. To further highlight the importance of the task in evoking conceptual representation, it has also been shown that humans can make a distinction about which of two images contains an animal in less than 150 ms [40], which suggests that features related to animacy can be decoded very fast in the visual system. It is unlikely that all the properties related to an animal are evoked in parallel to make such fast judgements.

The involvement of multiple brain areas specialized in different types of information is a clear indication of a distributed code in the brain to store this semantic representation. Moreover, it is a sign that these representations can be decomposed into components, such that if one were to learn what components these are, then it would be possible to predict the code for new words. The model we use in this dissertation tests these distributed and composition principles by using the activity in multiple regions in the brain, at different time windows, to predict new words based on the mapping to a pre-defined set of semantic components. Our models also make no assumptions regarding where or when the semantic information is encoded in the neural activity. It uses whole brain data to make its predictions, which allows us to further explore the regions that were most important to the encoding of different properties.

1.1.3 Feature set

For the remainder of this dissertation, we define *perceptual features* as anything that is particular to a given stimulus modality (e.g. "how many letters are in the word?", or "how many lines are in the drawing?"). Conversely, any information that is inherent to a noun, regardless of the stimulus modality that was used to probe it, is called a *semantic feature* (e.g. "is it alive?"). We seek to identify parts of the observed MEG activity that encode perceptual as well as semantic features.

We acquired a semantic knowledge base for 1000 concrete nouns using the Mechanical Turk human computation service from Amazon.com (`https://www.mturk.com/mturk/`

welcome). Humans were asked 218 questions about the semantic properties of the 1000 nouns, and the questions were inspired by the game *20 Questions*. For example, some questions were related to size, shape, surface properties, context, and typical usage. These were selected based on conjectures that neural activity for concrete objects is correlated to the sensory and motor properties of the objects [57]. Example questions include *is it shiny?* and *can you hold it?* (the complete list of semantic and perceptual features used can be seen in Appendix C). Users of the Mechanical Turk service answered these questions for each noun on a scale of 1 to 5 (definitely not to definitely yes). At least three humans scored each question and the median score was used in the final dataset (collected by Dean Pomerleau). For some of the analyses, the set of 218 semantic features (human-218) was complemented by 11 perceptual features, such as *word length* and *number of white pixels*.

1.1.4 Selecting features for subcomponents

Although for this work we choose to use the human-218 semantic feature set to represent semantics, there is no consensus about the subcomponents that are put together by the brain when forming the meaning of concrete nouns. Several datasets have been used in the literature as an attempt to form an intermediate basis that correlates with the actual inner-working of the brain. These feature sets serve as intermediate levels for predicting brain activity. Ideally, if one finds an optimal set of features to predict brain activity, then a link can be established between these features and the subcomponents used in the brain. Still, it is important to keep in mind that there is a difference between predicting the MEG signals we measure in our experiments, and understanding the inner-workings of the brain. In this dissertation, we explore where and when we find correlates of concrete noun representation across stimulus modalities, but how the representation is implemented in a single neuron level is beyond the scope of this work.

For example, Mitchell et al. [58] used the word co-occurrence with 25 pre-selected sensori-motor verbs as an intermediate semantic feature set to predict fMRI images for concrete nouns with well-above chance accuracy. A comparison between two different sets of features to be predicted from fMRI data was performed by Palatucci et al. [64]. The success at predicting the word co-occurrence with the 5000 most frequent words from the Google Trillion-Word-Corpus was not as high as predicting the answers to the questions in the human-218 dataset. Murphy et al. [61] also compared different semantic feature sets derived from text corpora on how well they can predict fMRI images for different concrete nouns. They found that specific types of corpus-based semantic features can achieve accuracies similar to the human-218 dataset, but not better, at predicting the fMRI activity for that specific set of concrete nouns.

Using factor analysis, Just et al. [38] found that shelter, manipulation, eating, and word length were the four most differentiable factors underlying the representation of concrete nouns when subjects thought about the properties of different objects in the fMRI scanner. These factors were also shown to have similar locations in bilingual participants, regardless of the language in which the stimuli were presented [6], which corroborates the claims that many aspects of conceptual knowledge are stored in the brain in a manner unrelated to language [34]. The semantic factors found by Just et al. [38] also show similarities to the list of the 20 most predictable features from MEG data [83]. Using a different analysis approach (Canonical Correlation Analysis), Rustandi et al. [76] found semantic components related to manipulability, shelter, body parts, and

word length among the latent variables discovered using fMRI data from two different experiments, one using only the word of each concrete noun, and the other using words and pictures simultaneously [75].

Some work decoding nouns and verbs from MEG data has been published [19, 72, 84], but only a few studies have looked at the different semantic features that describe a noun. For example, Chan et al. [9] have successfully decoded whether a subject is considering a living vs. nonliving stimulus based on MEG activity, but this leaves open the question of what other semantic features may be encoded by the MEG signal at different times and cortical locations.

Considerable work has also been done related to the processing of perceptual features. For example, Kay et al. [39] modeled the receptive fields of voxels in early visual areas to describe properties such as spatial, orientation, and spatial frequency tuning, and used the fMRI activity of those regions to predict the image a subject was seeing. Employing a different approach, Formisano et al. [15] used a combination of fMRI activation and data mining techniques to distinguish among different vowels to which subjects were listening, as well as which speakers had recorded the sounds.

1.2 MEG datasets used in this dissertation

All subjects gave their written informed consent approved by the University of Pittsburgh (protocol PRO09030355) and Carnegie Mellon (protocol HS09-343) Institutional Review Boards. MEG data were recorded using an Elekta Neuromag device (Elekta Oy), which has a total of 306 channels. These channels are distributed in 102 sensor triplets, each containing one magnetometer and two gradiometers that measure the differential magnetic field in orthogonal directions. The data were acquired at 1 kHz, high-pass filtered at 0.1 Hz and low-pass filtered at 330 Hz. Eye movements (EOG) were also monitored by recording differential activity of muscles above, below, and lateral to the eyes. These signals captured vertical and horizontal eye movements, as well as eye blinks. Additionally, four head position indicator (HPI) coils were placed on the subject's scalp to record the position of the head with relation to the MEG helmet at the beginning of each session. These coils, along with three cardinal points (nasion, left and right pre-auricular), were digitized into the system and were later used for source localization. Finally, structural MRIs were also obtained for each subject to create surface models of their brains and facilitate source localization.

1.2.1 Experimental paradigms

This section describes each of the paradigms of which data was used in this dissertation. The paradigms *answer-questions* and *freely-think* were scanned back-to-back (after a small break), so the subjects for the two paradigms are the same. Similarly, the paradigms *1back-text* and *1back-speech* were also scanned sequentially. Not all subjects in *answer-questions* were scanned in *1back-text*, and vice-versa. However, the subject numbers are kept constant across paradigms.

answer-questions

Nine right-handed human participants were scanned in this study. They answered 20 questions (e.g. *Was it ever alive?*, *Can you pick it up?*) about 60 different concrete objects equally divided into 12 categories (tools, foods, animals, etc...). Each object was represented by a line drawing and corresponding written noun below it (the complete set of the 60 line drawings can be seen in Appendix A, and the set of 20 questions is shown in Appendix B). Picture and word were positioned as close to each other as possible in order to minimize saccades between the two. A question was presented first, then all 60 nouns were presented in a random order (see Figure 1.2). The subjects used a response pad to answer yes or no after each noun presentation. Each stimulus was displayed until one of the buttons was pressed. After answering the question for all 60 nouns, a new question would come up, and the 60 nouns were randomly presented again. This cycle continued for a total of 20 questions. The questions were divided into blocks of 3 questions each (i.e. a question followed by the 60 nouns randomly displayed, then another question, etc), and the subjects had as much time as needed to rest in between blocks (no more than 3 minutes considering all subjects). Each block lasted approximately 8 minutes, depending on the subject's reaction time.

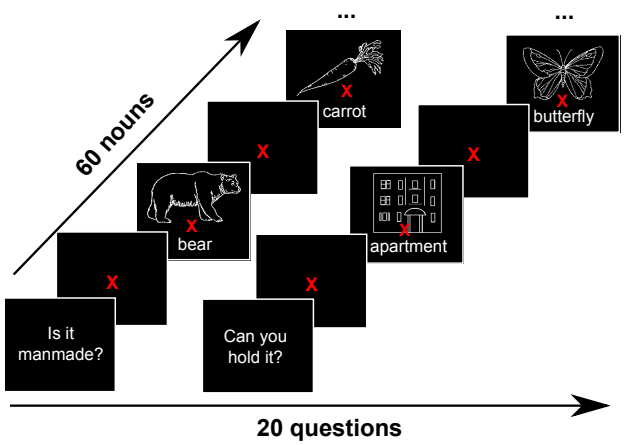


Figure 1.2: Experimental paradigm for the *answer-questions* task. Subjects are first presented with a question, followed by the 60 nouns (combination of picture and word) in a random order. Each stimulus is displayed until the subject presses a button to answer yes or no to the initially presented question. A new question is shown after all 60 nouns have been presented. A fixation point is displayed for 1 s in between nouns and questions. The set of stimuli is the same as in *freely-think*.

freely-think

The same nine right-handed subjects scanned in *answer-questions* also participated in this study. The exact same stimuli used to represent the 60 concrete objects in *answer-questions* were also used here. The main difference between the paradigms is that, instead of answering a particular question for each of the nouns, the subjects were instructed to think about the properties of each object as consistently as possible across repetitions. For example, for the noun *screwdriver*, one could think of how it feels to hold it, to use it, its shape, and weight. Despite the similar names, this paradigm is slightly different than the free-recall task used by Manning et al. [54] and Polyn et al. [69] among others. Although the subjects in our experiment have the freedom to think of

whatever object properties they choose, we still provide stimuli in the form of picture and word to probe reflection upon specific concrete nouns. The stimuli were presented between 6-10 times each, depending on the subject's level of comfort and willingness to continue participation in the experiment. The presentations were split in blocks of 60 words (each word was presented once per block), and the order of the words within a block was randomized. Each trial was preceded by a cross-hair placed in the center of the screen for 500 ms, on which the subject was instructed to focus until the stimuli appeared. The stimuli disappeared from the screen after 2 sec, and there was 1 sec of interval (blank screen) between repetitions.

1back-text

Six subjects that participated in the *answer-questions* and *freely-think* paradigms were also scanned on a different day while performing the *1back-text* task. A sequence of 1000 different words was presented a single time each to the subjects, and the participants had to press a button (positioned on the left hand) whenever the same word was immediately repeated (10% probability, 1-back identity task). To avoid motor contamination, the word following repeated words did not belong to the set of 1000 words and therefore was not used in the analysis. The 1000 words were a super-set of the 60 nouns previously presented in *answer-questions* and *freely-think*. Words were presented for 300 ms every 1.2 s, in blocks of 200 randomized words, after which the subjects were allowed a small break.

1back-speech

The same subjects that participated in *1back-text* were also scanned under *1back-speech*, on the same day, after a short break. The task performed by the subjects was the same as in *1back-text*: press a button on the left hand whenever two consecutive stimuli are the same. However, in *1back-speech* the stimuli presented to the subjects was the spoken word, instead of the written version shown before. Also, instead of 1000 different nouns, only 500 were used, and these were a subset of the 1000 nouns previously employed. These 500 nouns were chosen based on the length of the words and the intelligibility of the recordings. Intelligibility was measured by how well two different researchers could correctly transcribe the spoken words.

The paradigms analyzed in this dissertation vary along three major dimensions: the stimulus modality, the task performed by the subjects, and the number of repetitions of each noun. By analyzing data of multiple experimental paradigms we can get detailed insights about the commonalities in concrete noun representation, and investigate what are the specific contributions of the quality of the MEG signal, the task performed, and the stimulus modality to the results.

Overall, *answer-questions* and *freely-think* experiments are complemented by *1back-text* and *1back-speech* in many ways. One downside of the former two is the repetition of the stimuli, which has been shown to affect language processing [55, 94]. The MEG response in word reading tasks such as in *1back-text* is well-described in the literature [11, 86, 96], and so is the MEG response to spoken words [33, 65, 93]. Moreover, these latter tasks do not involve an explicit semantic decision task. Finally, *1back-text* and *1back-speech* do not require a button

press for every trial, and probe automatic language-related responses by only presenting words to the subjects, while in *answer-questions* and *freely-think* a pair of word and picture were shown in each trial. Any additional shortcomings will be presented during the analysis of each paradigm, and their impact in the results will also be considered in the appropriate sections.

1.2.2 MEG Data Preprocessing

The data were preprocessed using the Signal Space Separation method (SSS) [88, 89]. SSS divides the measured MEG data into components originating inside the sensor array vs. outside or very close to it, using the properties of electromagnetic fields and harmonic function expansions. The temporal extension of SSS (tSSS) further enables suppressing components that are highly correlated between the inner and close-by space. Finally, tSSS realigned the head position measured at the beginning of each block to a common location. Because they were scanned in sequence, the head position of blocks in *freely-think* were aligned to the first block of *answer-questions*. Similarly, the head position of blocks in *1back-speech* were aligned to the first block of *1back-text*.

The MEG signal was then low-pass filtered to 50 Hz to remove the contributions of line noise and down-sampled to 200 Hz. The Signal Space Projection method (SSP) [92] was subsequently applied to remove signal contamination by eye blinks or movements, as well as to remove MEG sensor malfunctions or other artifacts [95]. Freesurfer software (`http://surfer.nmr.mgh.harvard.edu/`) was used to construct the 3D model of the brain from the structural MRIs, and to automatically segment, based on each subject's anatomical data, the 67 regions of interest (ROIs) analyzed in this dissertation (Freesurfer 'aparc' annotation, Desikan-Killiany Atlas). The Minimum Norm Estimates method [23], which finds the distribution of currents over the cortical mantle that has the minimum overall power, was employed to generate source localized estimates of brain activity from MEG data (MNE Suite software, `http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php`). MNE test sources were evenly distributed in each subject's cortical sheet, and source localization was performed separately for each stimulus noun, using the average of the repetitions when available in the paradigm. In several analyses in this dissertation, data between -0.1 s and 0.8 s were used, where 0 was when the stimulus was presented. Note that while we do not expect that neural activity before stimulus onset contributes to the results shown, utilizing these time points works as a good sanity check for our results (e.g. decoding accuracies in that period should not be better than chance).

Chapter 2

Tracking Neural Coding of Perceptual and Semantic Features of Concrete Nouns

This chapter presents a methodological approach to explore the types of information that are encoded in the MEG activity over time and space. We show that using MEG data it is possible to not only classify which of two nouns a subject was thinking about, similarly to what has been done with fMRI [64], but also to investigate which features of the stimulus noun are encoded in the MEG signal at different times and cortical locations. We can also test whether data from regions previously shown in fMRI to be involved in semantic processing [70] yield satisfactory decoding results, and make a comparison to the regions that show highest decoding results using MEG. Finally, we infer from the results when and where different groups of features are encoded in the MEG signal.

The data used in this section comes from the *answer-questions* paradigm, described in section 1.2. Parts of this chapter have also been described elsewhere [83].

2.1 Methods

2.1.1 Training and Testing Decoders of Stimulus Features

To study the question of when and where MEG activity encodes various features of the noun stimulus, we used a machine learning approach.

Standard machine learning methods such as support vector machines, logistic regression, and linear regression learn a function $f : X \rightarrow Y$, that maps a set of predictive features X to some predicted value Y (Figure 2.1, left) [5, 62, 67]. For example, the predictive features X might be the observed activity in MEG sensors at some location, averaged over a particular time window, and Y might be a variable indicating whether the subject is reading the stimulus word "house" or "horse". Alternatively, Y could also be some feature of the stimulus word, such as the number of characters in the word (a perceptual feature) or a variable indicating whether or not the stimulus word describes a living thing (a semantic feature). Whatever the case, the machine learning algorithm creates its estimate $\hat{f}$ of the function f from a set of training examples consisting of given $\langle x, y \rangle$ pairs. The learned $\hat{f}$ can then be tested by giving it a new example x, and testing

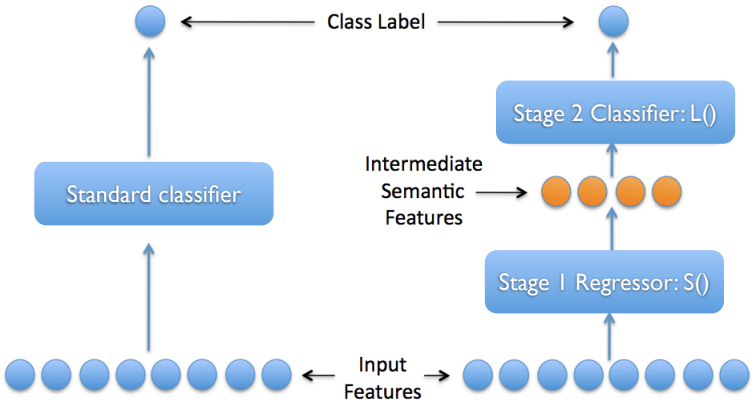


Figure 2.1: A typical single stage classifier (shown on left) compared to the semantic output code regressor (SOCR, shown on right). The SOCR is a two stage classifier that uses a layer of intermediate semantic features between the input features and the class label. These semantic features represent attributes of the class labels. In our experiments, the input features are the MEG data, the class labels are different nouns (e.g. bear, carrot), and the intermediate semantic features are data collected using Mechanical Turk about the different nouns (e.g. *is it alive? can you hold it?*). Modified from Palatucci et al. [64].

whether its prediction $\hat{f}(x)$ gives the correct value y.

Here, we use the success or failure of the learned function $\hat{f} : X \rightarrow Y$ in predicting Y over the test data to explore what information is encoded in the MEG activity X. In particular, if $\hat{f}$ can accurately predict the value y of variable Y for MEG images x, over pairs $\langle x, y \rangle$ that were not involved in training $\hat{f}$, then we conclude that the MEG activity X in fact encodes information about Y. In the analyses reported here, we trained many different functions $\hat{f}$ using different inputs X and outputs Y. We varied X to cover different spatial regions of source-localized MEG data, and different windows in time, to explore which of these spatial-temporal subsets of MEG activity in fact encode Y. We varied Y to cover hundreds of different semantic and perceptual features. We then tested the accuracy of these trained functions to determine which make accurate predictions, in order to study which of the different spatial-temporal segments of MEG activity X encode which of the different stimulus features Y. The stimulus features used here are described in section 1.1.3.

In addition to training individual functions to determine which features are encoded where and when in the MEG signal, we also considered the question of whether the features that were decodable by our functions were, in fact, decoded accurately enough to distinguish individual words from one another based on their decoded feature values. This test gives a crude measure of whether our approach captures the majority of the features represented by neural activity (sufficient to distinguish arbitrary pairs of words) or just a fraction of these features. To accomplish this, we trained all decoders for the 218 semantic features, then applied them to decode the features of a novel stimulus word (not included in the training set). To test the accuracy of this collection of decoded features, we asked the predictor which of two novel stimulus words the

subject was viewing when the test MEG image was captured, based on its collection of predicted features. We call this two-stage classifier the *semantic output code regressor* (SOCR, Figure 2.1, right), which was first introduced by Palatucci et al. [64].

A common question about this approach is why we leave two words out for testing, instead of only one. Regardless of the number of nouns left out of the training set, the important point about this model is that it can extrapolate to novel words it never saw during training. Leaving two words out is a straightforward way to evaluate these predictions using two estimates with comparable noise levels, and also a simple approach to measure the chance level (i.e. 50%, the discrimination between the word pair is either correct or not). Still, we have shown elsewhere [16] that the model performs significantly better than chance at choosing a single left-out word out of a set of 1000 possible words.

2.1.2 Regression Model and SOCR

We used a *multiple output linear regression* to estimate $\hat{f}$ between the MEG data X and the set of perceptual and semantic features Y (Figure 2.1, right). Each semantic or perceptual feature was normalized over the different nouns, such that the final vector of values for each feature used in the regression had a mean of 0 and variance of 1. A similar normalization process was applied to the MEG data, namely, the activity of a source at a given time point (i.e. a feature in the regression) was normalized such that its mean over the different observations of the data was 0, and the variance was 1. Let $\mathbf{X} \in \Re^{N*d}$ be a training set of examples from MEG data where N is the number of distinct noun stimuli and d is the number of dimensions of the neural activity pattern. Each row of $\mathbf{X}$ is the average of several repetitions of a particular noun (i.e. for this experiment, the average of several sources over time over the 20 repetitions of a given noun). Let $\mathbf{F} \in \Re^{N*p}$ be a *matrix* of p semantic features for those N nouns. We learn a matrix of weights $\widehat{\mathbf{W}} \in \Re^{d*p}$ which maps from the d-dimensional neural activity to the p semantic features. In this model, each output is treated independently, so we can solve all of them quickly in one matrix operation:

$$\widehat{\mathbf{W}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_d)^{-1}\mathbf{X}^T\mathbf{F} \tag{2.1}$$

where $\mathbf{I}_d$ is the identity matrix with dimension $d * d$ and λ is a scalar regularization parameter chosen automatically using the cross-validation scoring function [27, page 216][1]. A different λ is chosen for each output feature in Y. One disadvantage of Equation (2.1) is that it requires an inversion of a d by d matrix, which is computationally slow (or even intractable) for any moderate number of input features. With this form, it would be impossible to compute the model for several thousands of features without first reducing their number using some method of feature selection.

However, a simple computational technique can overcome this problem by rewriting Equation (2.1) in its *dual* form, also known as its *kernel form*. Following the procedure described in [37] we obtain:

[1]We compute the cross-validation score for each output (i.e. prediction of a particular semantic feature), and choose the parameter that minimizes the average loss across all outputs.

$$\widehat{\mathbf{W}} \;=\; \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}_N)^{-1}\mathbf{F} \tag{2.2}$$

This equation is known as *kernel ridge regression* and only requires inversion of an $N * N$ matrix. This is highly useful for neural imaging tasks where N is the number of examples which is typically small, while the number of features d can be very large. Another computational technique from [20] further shows that with a little pre-computation, it is possible to obtain the inverse for any regularization parameter λ in time $O(N)$, much faster than the time required for a full inverse $O(N^3)$. Combined with the cross-validation scoring function from Hastie et al. [27, page 216], the end result is an extremely fast method for solving the resulting regression even with thousands of input and semantic features, while automatically selecting the best regularization parameter λ from a large grid of possible parameter choices.[2]

Using this form, it is possible to quickly obtain the weight matrix $\widehat{\mathbf{W}}$. Then, given a novel neural image $\mathbf{x} \in \Re^{1*d}$, we can obtain a prediction $\widehat{\mathbf{f}} \in \Re^{1*p}$ of the semantic features for this image by multiplying the image by the weights:

$$\widehat{\mathbf{f}} = \mathbf{x} \cdot \widehat{\mathbf{W}} \tag{2.3}$$

We performed a *leave-two-out-cross-validation* and trained the model in Equation 2.2 to learn the mapping between 58 MEG images and the set of features for their respective nouns. For the second stage of the semantic output code classifier, we applied the learned weight matrix $\widehat{\mathbf{W}}$ to obtain a prediction of the 218 semantic features, and then we used a Euclidean distance metric to compare the vector of predictions to the true feature encodings of the *two held-out nouns* [64]. The labels were chosen by computing the combined distance of the two labeling configurations (i.e. the nouns with their true labels or the reverse labeling) and choosing the labeling that results in the smallest total distance [58]. For example, if *dist()* is the Euclidean distance, and p_1 and p_2 are the two predictions for the held out nouns, and s_1 and s_2 are the true feature encodings, then the labeling was correct if:

$$dist(p_1, s_1) + dist(p_2, s_2) < dist(p_1, s_2) + dist(p_2, s_1)$$

This process was repeated for all $\binom{60}{2} = 1,770$ possible leave-two-out combinations.

Because the experimental task in *answer-questions* involved a button press in every trial, we were also careful not to extend the decoding period past 0.75 s to avoid the contributions of cortical activity associated with the button press (group-level mean reaction time 1.1 s) [10]. As a further confirmation that the information contained in the button presses was not contributing to our decoding results, we ran our classifier for each subject with a single feature as the input, representing the button press value (yes or no), replacing the brain data previously used. Similarly to what was done to the brain data, the feature representing the button press was averaged over all 20 repetitions of a noun. The accuracy of the classifier was not better than chance (50%). We

[2]Computational speed was a large factor in choosing the kernel ridge regression model for the first stage of the classifier. A common question we receive is why not use a more modern method like Support Vector Machines or Support Vector Regression. Besides being significantly computationally slower, our tests found no performance advantage of these more complicated algorithms over the simpler kernel ridge regression model.

also performed a similar test by using only the EOG signal of each subject as the input to the SOCR. The decoding results were again not better than chance for any of the subjects, suggesting that any remaining eye movement artifacts possibly captured by the MEG channels did not contribute to the decoding results shown in this dissertation.

2.1.3 Feature scoring

In order to quantify how well each of the semantic features was predicted in the first stage of the SOCR, equation 2.4 was used:

$$score_{feature} = 1 - \frac{\sum_i (f_i - \hat{f}_i)^2}{\sum_i (f_i - \bar{f})^2} \tag{2.4}$$

where f_i is the true value of the semantic feature for noun i held-out from the training set in the ith cross validation fold, $\bar{f}$ is the mean true feature value over all nouns, $\hat{f}_i$ is the predicted value of the semantic feature for the held-out noun i, and the summation is over all cross-validation iterations. Equation 2.4 is a measure of the percent of variance in the feature that is explained by our learned function. So, the closer the semantic feature score is to 1, the better it is predicted by our classifier using MEG data. Equation 2.4 is also known in the literature as the coefficient of determination (R^2) [82]. However, because we use only the test data to calculate the feature score, some of the properties of R^2 do not hold (e.g. our feature score can actually become negative).

2.1.4 Statistical significance

Statistical significance was established by running the computational analysis several times with permuted data. More specifically, in each permutation set one subject was chosen at random according to a uniform distribution, and the trial labels for that subject were shuffled prior to averaging over the 20 repetitions. The computational analysis, including source localized estimates, was conducted with the shuffled data set. These analyses were performed over three hundred times, and the results were combined to form a null distribution. Finally, the p-values of the reported results were obtained for each individual subject by using a normal kernel density function to estimate the cumulative distribution associated with the empirical null distribution. P-values across subjects were combined using Fisher's method [14], and correction for multiple comparisons (over time, features, and/or regions; indicated in the pertinent parts of the results section) was done using False Discovery Rate [4] with no dependency assumptions, at an alpha level of 0.01.

2.2 Results

Prior to using the methods described above, an initial analysis of the data revealed that the cortical dynamics generated in this paradigm matched what has been shown in the literature to occur

while subjects view pictures or words [77]. More specifically, we noticed a peak in activation around 100 ms in posterior regions of the brain such as the lateral occipital cortex, followed by a sustained activation, starting at about 200 ms and peaking at 400 ms, in more anterior regions such as the banks of the left superior temporal sulci (Figure 2.2). These results made us comfortable to carry out an analysis to address the following questions.

2.2.1 Can we discriminate between two novel nouns only using semantic features predicted from MEG activity?

The input of the classifier consisted of all estimated sources and their activity over time, and only the 218 semantic features were predicted using the data in the first stage of the SOCR. Based on the results shown in Table 2.1, we see that, for each of the nine participants, it was possible to discriminate with better than 85% accuracy (mean 91%) accuracy between two held-out nouns based on semantic features predicted from observed MEG activity, even though neither noun appeared in the training set.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
Leave-two-out accuracy	88.25	95.82	86.16	95.20	93.62	92.49	92.54	85.25	91.36	**91.19**

Table 2.1: Accuracies for the *leave-two-out* experiment using simultaneously all time points from stimulus onset to .75 s and all sources in the cortex as input features for the classifier. The classifiers are able to distinguish between two different concrete nouns that the MEG-based classifier has never seen before with 91% mean accuracy over the nine participants S1 through S9. Chance accuracy was 50.0%. For a single-subject model, 62.5% corresponds to $p < 10^{-2}$. The result for the subject with the lowest accuracy (85.25%) corresponds to $p < 10^{-6}$. The p value associated with observing that all nine independently trained participant models exhibit accuracies greater than 62.5% is $p < 10^{-11}$.

We also evaluated the decoding accuracy over time by training distinct functions to decode each of the 218 semantic features using consecutive 50-ms time windows of MEG data. The black solid lines in the bottom plots of Figure 2.2 show that, when distinguishing between two left-out nouns using all regions of the cortex, the accuracies started to exceed the chance level at 50-100 ms. The peak accuracy was reached at 350-400 ms.

2.2.2 Can we discriminate between two concrete nouns in the same category?

It is also informative to investigate which pairs of nouns displayed best classification results. Figure 2.3 displays the success rate for distinguishing each pair of nouns. Although most noun pairs were correctly classified across subjects, the matrix in Figure 2.3 has some noticeable spots where the classifier failed. Along the diagonal of the matrix lie the regions where the nouns belong to the same category, and therefore are closer in the semantic space. Because our classifier relies on these semantic differences, it had more difficulty distinguishing between nouns with similar than markedly dissimilar semantic representations. This behavior is also observed in other regions of the matrix, such as when we classify man-made objects against tools. If we look at a similarity matrix that summarizes how close each pair of nouns exist in the semantic space (Figure 2.4), we can observe that the pairs with which the classifier had difficulties are exactly the ones that are closer in semantic space. Still, there are a few word pairs that the classifier can

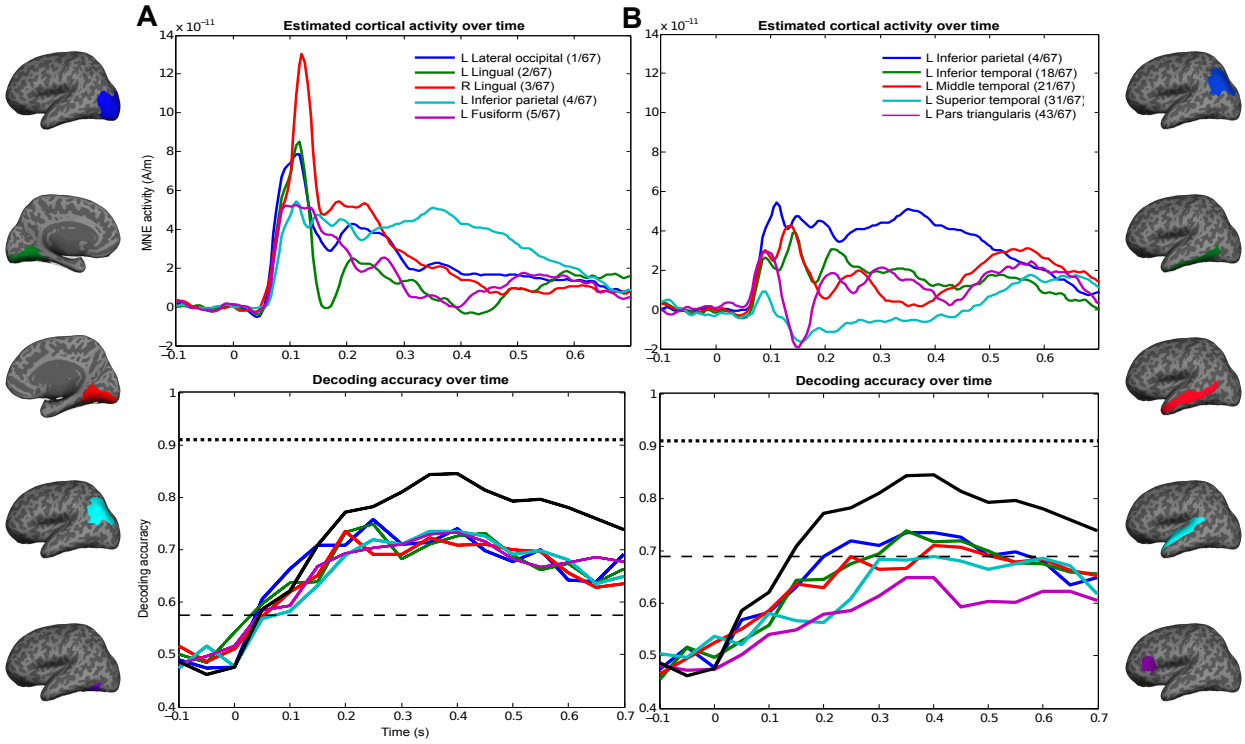


Figure 2.2: Time courses of activity (top) and decoding accuracy (bottom) in different brain areas. (A) Five ROIs with best decoding results, (B) ROIs pre-selected based on fMRI literature [70]. The ROIs used in the analysis are displayed on an inflated brain. In each column, top plots denote time courses of activation in the ROIs and bottom plots corresponding time courses of decoding accuracy. Estimated MEG activity for all sources in the different ROIs (different traces in the plot), was averaged over all 60 nouns and across subjects. Decoding accuracy over time was averaged over all subjects. There are clear differences between the time course of MEG activity and that of decodable semantic information. Each time window is 50-ms wide, taken at steps of 50 ms, starting at -0.1 s. For example, the time point at 0.3 s in the bottom graphs shows the decoding accuracy for the window 300 - 350 ms. Time zero indicates stimulus onset. Chance accuracy is 50%, and light dashed line shows the accuracy above which all values are significant for the different ROIs and time points ($p < .01$, FDR corrected for multiple comparisons over time and all ROIs). Darker dashed line denotes the mean accuracy over all subjects when the classifier is allowed to observe all time points and sources (i.e. no averages within time windows, same as Table 2.1). Black solid line indicates decoding accuracy over time when all sources on the cortex were used for each 50-ms time window. Legends in the top row also indicate the rank order of the ROI out of the 67 possible ROIs.

accurately distinguish for most subjects, despite their closeness in semantic space, suggesting that the information for such subtle distinctions can also be encoded in the MEG signal. And it is important to note that we can still perform better than chance even when classifying within some categories, which shows the specificity of this compositional method (Table 2.2).

Even in the cases where the stimulus figures shared similar perceptual features (e.g. carrot and spoon), or words shared similar lexical form (e.g. cat and car), the classifier was still able to perform significantly above chance level (the complete set of the 60 line drawings can be seen in Appendix B). These results suggest that the classifier relies on the semantic information of the different nouns, and not on the perceptual aspects of the figures and words.

2.2.3 How do the prediction results vary over time and space?

The relatively high temporal and spatial resolution of MEG allows us to address the question of what regions of the brain and what time intervals are responsible for these results. For this computational experiment, only the Freesurfer-based pre-specified ROIs were used, and all non-

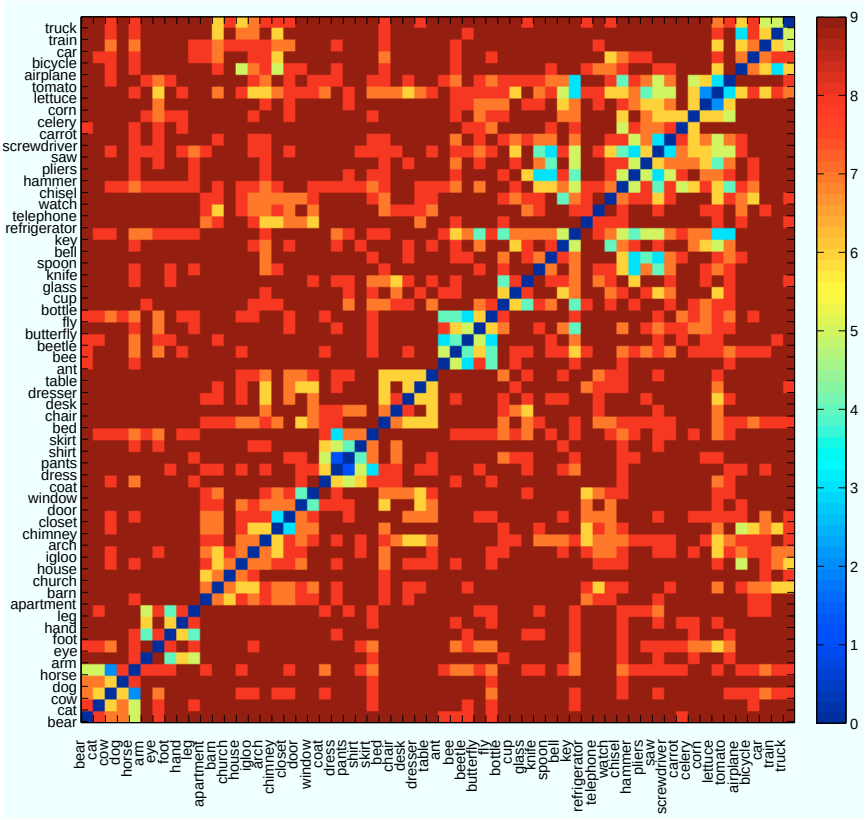


Figure 2.3: Confusion matrix for results using all sources and time points. Pairs of nouns that are closer in the semantic space (i.e. closer to the diagonal) are harder to be distinguished using MEG data. X and Y labels indicate the noun pair being held-out during training, and color bar shows for how many subjects the algorithm correctly classified the given pair. Nouns are sorted based on their category (i.e. every 5 nouns belong to the same category). Total of 9 subjects.

overlapping 50-ms time windows between -0.1 s and 0.75 s were considered.

The activity and decoding accuracy curves of the 5 regions that showed best decoding accuracy over all subjects are displayed in Figure 2.2A. The different regions were ranked based on how many significant time points they showed. Then, in the situations when there was a tie in the number of significant accuracies over time, the ROIs in the tie were sorted by maximal decoding accuracy. Most ROIs (63 out of 67) had at least one time point with significant decoding accuracy, and for the top 2 ROIs all windows starting at 50 ms after stimulus onset displayed significant decoding accuracies. In the next best 14 ROIs, all windows starting at 100 ms showed significant decoding accuracies. The complete rank of ROIs can be found in the supplemental website (`http://www.cs.cmu.edu/afs/cs/project/theo-73/www/neuroimage2012.html`).

Using all sources and time points together resulted in even better decoding accuracies (dashed horizontal black line in Figure 2.2, bottom). Using all sources over 50 msec time windows also produced higher decoding accuracy than just using individual ROIs, especially after 200 ms (see difference between black solid curve and other curves). Another way to look at the evolution of the activity and decodability over time is to plot such curves for regions described in fMRI literature to participate in semantic processing [70], as seen in Figure 2.2B. We selected

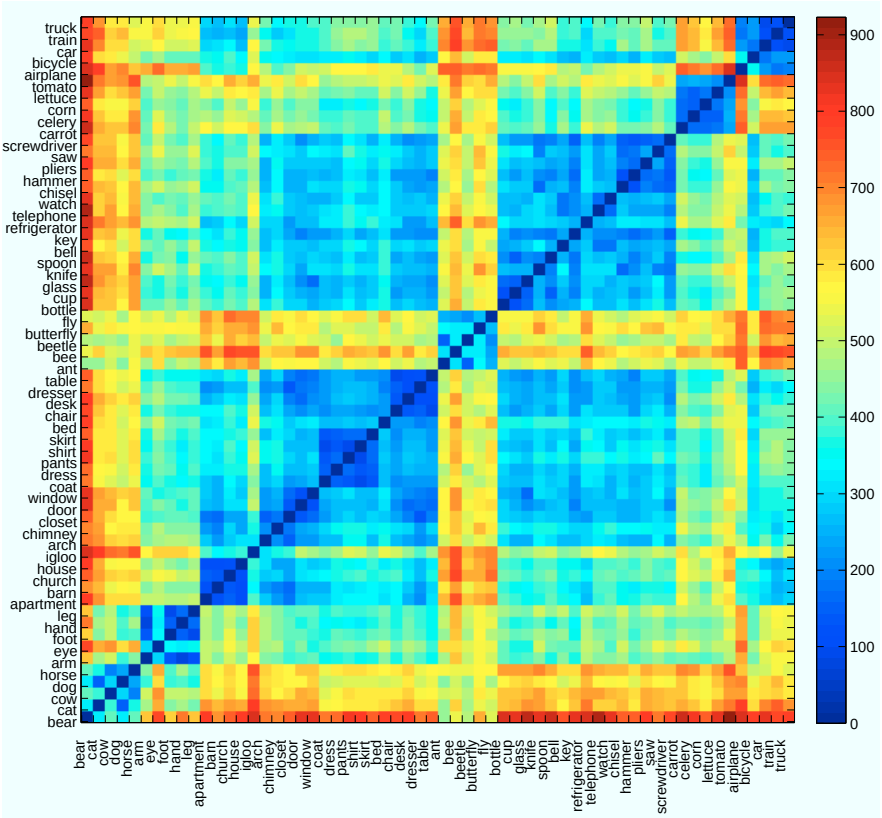


Figure 2.4: Confusion matrix for 218 semantic features. Pairs of nouns in the same category are closer in semantic space, and therefore have smaller differences between their semantic vectors. X and Y labels indicate the noun pair being evaluated, and color bar shows the squared distance between the two semantic vectors. Nouns are sorted based on their category (i.e. every 5 nouns belong to the same category).

seven ROIs: left pars opercularis, pars orbitalis, pars triangularis, inferior, middle, and superior temporal cortices, and inferior parietal cortex. One of these regions (left inferior-parietal cortex) was among the top ROIs for decoding (Figure 2.2, left), with significant accuracies in all windows from 50 to 700ms, but the other regions did not rank as well (see legend of the top plots). It is not necessarily surprising that these pre-selected ROIs would not perform as well as the regions shown in Figure 2.2A, and also not rank high among all regions that were analyzed, since regions active in fMRI may not fully correspond to the MEG activations [63].

Note that the time course of activity (top plots in Figure 2.2) does not coincide with the time course of decoding accuracy for the majority of the regions plotted. For example, although activity in left lateral occipital cortex peaks at around 115 ms, its peak decoding happens in the window 200 - 250 ms. This result shows that higher activation does not necessarily imply more information for decoding between nouns. It is also interesting to note the gap between decoding over time that includes all sources and decoding within each individual ROI (i.e. solid black line versus colored lines in the bottom plots of Figure 2.2). As none of the single regions reached the decoding accuracy indicated by the solid black line, one can infer that the combination of the activity of several regions markedly contributes to processing of semantic information.

Category	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
animals	70.00	70.00	40.00	60.00	80.00	40.00	90.00*	90.00*	70.00	**67.78**
body parts	70.00	80.00	60.00	90.00*	90.00*	90.00*	50.00	70.00	70.00	**74.44**
buildings	90.00*	90.00*	40.00	60.00	100.00*	90.00*	60.00	60.00	100.00*	**76.67**
building parts	60.00	70.00	70.00	90.00*	90.00*	80.00	60.00	70.00	60.00	**72.22**
clothing	40.00	40.00	40.00	90.00*	60.00	50.00	70.00	70.00	60.00	**57.78**
furniture	70.00	70.00	30.00	60.00	80.00	80.00	80.00	90.00*	100.00*	**73.33**
insects	40.00	60.00	80.00	60.00	80.00	40.00	50.00	30.00	40.00	**53.33**
kitchen items	90.00*	90.00*	80.00	70.00	60.00	100.00*	100.00*	80.00	80.00	**83.33**
tools	70.00	90.00*	90.00*	80.00	80.00	70.00	100.00*	90.00*	90.00*	**84.44**
vegetables	40.00	100.00*	50.00	80.00	90.00*	50.00	50.00	50.00	30.00	**60.00**
vehicles	40.00	70.00	100.00*	70.00	80.00	60.00	40.00	90.00*	60.00	**67.78**
man-made items	50.00	70.00	60.00	70.00	90.00*	70.00	50.00	90.00*	60.00	**67.78**
Mean	60.83	75.00	61.67	73.33	81.67	68.33	66.67	73.33	68.33	**69.91**

Table 2.2: Accuracies for the *leave-two-out* experiment for the intra-category items. The classifiers used simultaneously all time points from stimulus onset to .75 s and all sources in the cortex as input features for the classifier. Results show that for some categories our classifiers are able to distinguish between two different concrete nouns that the MEG-based classifier has never seen before with accuracies significantly better than chance despite their similar semantics. Chance accuracy was 50.0%. Accuracies that are significant at $p < 0.05$ are marked with a star.

2.2.4 What features are best predicted?

Mean score ($\pm$ SD)	Semantic Feature
0.59 ($\pm$ 0.07)	CAN YOU PICK IT UP?
0.57 ($\pm$ 0.08)	IS IT TALLER THAN A PERSON?
0.57 ($\pm$ 0.04)	IS IT ALIVE?
0.57 ($\pm$ 0.08)	IS IT BIGGER THAN A CAR?
0.57 ($\pm$ 0.08)	CAN YOU HOLD IT?
0.56 ($\pm$ 0.04)	IS IT MANMADE?
0.56 ($\pm$ 0.08)	CAN YOU HOLD IT IN ONE HAND?
0.56 ($\pm$ 0.09)	IS IT BIGGER THAN A LOAF OF BREAD?
0.55 ($\pm$ 0.06)	IS IT BIGGER THAN A MICROWAVE OVEN?
0.55 ($\pm$ 0.06)	IS IT MANUFACTURED?
0.55 ($\pm$ 0.06)	IS IT BIGGER THAN A BED?
0.54 ($\pm$ 0.05)	DOES IT GROW?
0.54 ($\pm$ 0.06)	IS IT AN ANIMAL?
0.54 ($\pm$ 0.05)	WAS IT EVER ALIVE?
0.53 ($\pm$ 0.08)	DOES IT HAVE FEELINGS?
0.53 ($\pm$ 0.04)	CAN IT BEND?
0.53 ($\pm$ 0.08)	CAN IT BE EASILY MOVED?
0.53 ($\pm$ 0.06)	IS IT HAIRY?
0.51 ($\pm$ 0.06)	WAS IT INVENTED?
0.51 ($\pm$ 0.04)	DOES IT HAVE CORNERS?

Table 2.3: Top 20 semantic features sorted by mean feature score when using data from all time points and sources in the cortex ($p < 0.01$, FDR corrected for multiple comparisons over features). Features related to *size*, *manipulability*, and *animacy* are among the top semantic features predicted from MEG data.

It is also interesting to take a step back and look at the semantic features that were best predicted by the first stage of the classifier (Table 2.3). When the activity in all sources and time points was used simultaneously as the input to the SOCR (e.g. results in Table 2.1), a single feature score could be calculated for each semantic feature (Equation 2.4). Out of the 218 semantic features, 184 were predicted with statistically significant accuracy by our method ($p < 0.01$, FDR corrected for multiple comparisons across features). The larger the number of semantic features that were significantly predicted for a subject, the better was the accuracy for

that subject in distinguishing between the two left-out nouns (r = 0.93, $p < 10^{-3}$). Table 2.3 shows the top 20 features based on their mean feature score across subjects. It is possible to see a pattern across the well-predicted features. They group around three general categories: size (*is it bigger than a car? is it bigger than a loaf of bread?*), manipulability (*can you hold it? can you pick it up?*), and animacy (*is it man-made? is it alive?*). The complete list of decodable features can be found in the accompanying website (`http://www.cs.cmu.edu/afs/cs/project/theo-73/www/neuroimage2012.html`).

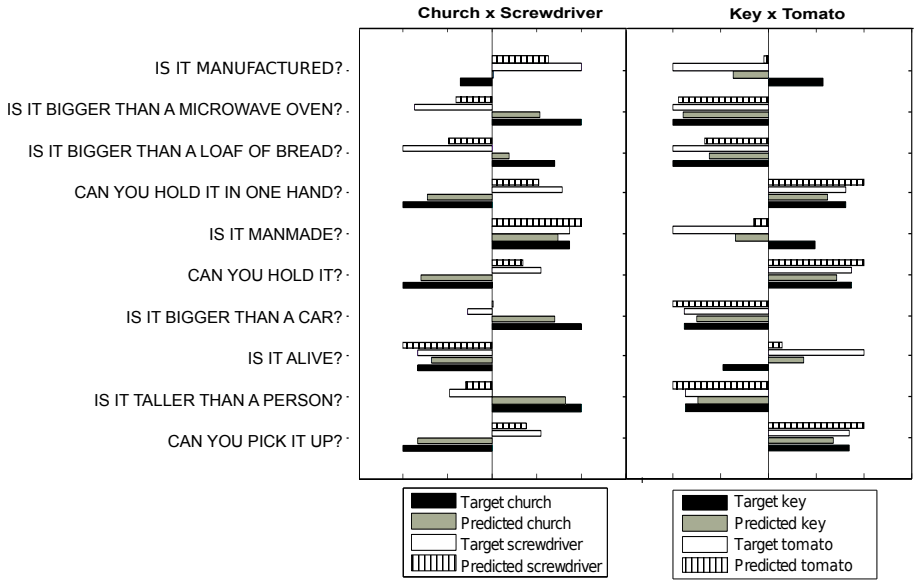


Figure 2.5: Illustrative example showing predictions for the 10 most predictable semantic features (Table 2.3) for subject S5, for two representative pairs of nouns (i.e. the two nouns that were left out of the training set). The SOCR performs well even when all features are not predicted correctly. Target denotes the actual semantic feature value for the noun, and the predicted one is the result of the classifier. For this figure, the feature values were normalized to the maximum absolute value per feature (i.e. the longest bar in each feature shares the same absolute value, and the other bars were scaled relative to that).

To better illustrate the inner workings of the algorithm to distinguish between two nouns, we may analyze the prediction results for a few of the best predicted features and pairs of representative nouns (Figure 2.5). Although the predictions are not perfect, they provide a good estimate of the semantic features for the selected nouns. Even when the predicted value does not agree with the target value, that information is still useful to the second phase of the classifier because of the comparative score that was used. For example, although a small positive value was predicted for a screwdriver being bigger than a car, which is obviously not true, that value was nevertheless closer to the true feature value of screwdriver than church, and thus useful for classification. However, the pair key and tomato is not properly classified over the subjects. From Figure 2.5, we can see that the target values for those semantic features for the two nouns are very similar, and although they are reasonably well-predicted, it is not enough to differentiate between the two nouns.

2.2.5 When and where are perceptual and semantic features best predicted?

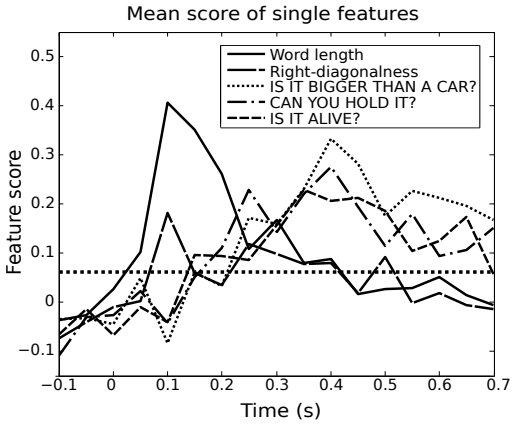


Figure 2.6: Evolution of the mean feature score for five representative features. Perceptual features such as *word length* are decoded from MEG data earlier than semantic features. Mean score was taken over feature scores for all subjects using all cortical data. Perceptual features were the two best predicted among subjects. Semantic features were taken from Table 2.3 and represent 3 distinct groups of features (size, manipulability, and animacy). Black dotted line shows the score above which all scores in the plot become significant ($p < 10^{-3}$, FDR corrected for comparisons over features and time).

Analyzing how decodable each type of feature is over time also yields interesting insights (Figure 2.6). For the following computational experiments, 11 perceptual features were added to the set of 218 semantic features, and we probed when during the response each feature could be reliably decoded. We see that perceptual features such as *word length* and *right-diagonalness* have high scores across subjects during the early processing of the stimuli (starting at 50 ms). Later during the course of the response, the score for semantic features starts to rise, while the score for perceptual features drops.

The feature *word length* was the best decoded perceptual feature. This might be because the majority of subjects reported upon completion of the experiment that the word was the first part of the stimuli to which they actively attended, and the picture was attended to later (if at all). Still, we can see from the plot the second-best decoded perceptual feature, *right diagonalness*, also showed an early rise around 50 - 100 ms, earlier than the semantic features.

It is also interesting to check which regions of the cortex are the source of MEG signal that encodes different features. Figure 2.7 shows the top three features decoded from MEG data in different regions of the cortex. Only the sources within each ROI were used to predict both semantic and perceptual features over time (50-ms non-overlapping windows) in the first stage of the SOCR. The ROIs displayed in Figure 2.7 are composed of the ROIs that yielded best decoding accuracy (see Figure 2.2) and pre-selected regions based on literature [70].

The different ROIs were also ranked based on how many features they significantly encoded. Then, in situations when there was a tie in the number of significant features, the ROIs in the tie were sorted by the number of semantic features they encoded. All 67 ROIs encoded at least one feature. Only two regions encoded more than 60 of the 229 features: left lateral occipital and left inferior parietal regions. The histogram in the inset of Figure 2.7 summarizes the number of decoded features per ROI. The supplemental website contains a list of the most accurately decoded features for each ROI investigated (`http://www.cs.cmu.edu/afs/cs/project/theo-73/www/neuroimage2012.html`).

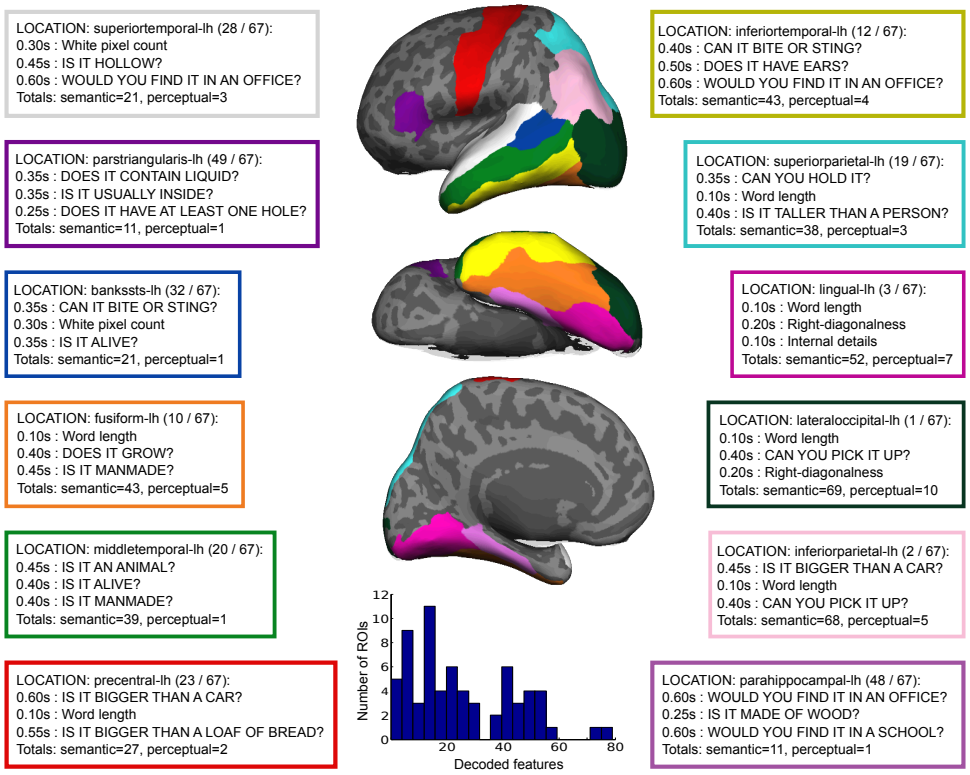


Figure 2.7: Spatio-temporal characterization of feature decodability in the brain. Perceptual features were better decoded earlier in time using MEG data from posterior regions in the cortex, and semantic features later in time from data of more anterior and lateral regions. Each table shows the three most accurately decoded features for a given ROI, where the color of the table border matches the region marked in the brain plots. Features within a ROI were ranked based on the median peak score over 9 subjects. A row in the table also shows the median time point when that feature reached its peak score (medians taken over 9 subjects). All features shown were significantly predicted with $p < 10^{-5}$ (corrected for multiple comparisons over features, regions, and time). Each table also shows the ROI rank in predicting features (based on how many features the ROI significantly decoded), and the total number of semantic and perceptual features that were significantly decoded using data from that region. Inflated brain plots display the different ROIs in three different views of the left brain hemisphere: lateral (top), ventral (center, where right is occipital cortex and left is frontal cortex), and medial (bottom). Inset shows a histogram of the total number of features encoded in the 67 ROIs.

As one might expect, low-level visual properties of the stimuli, such as *word length*, were best predicted from the neural data in the lateral-occipital cortex around 150 ms. In fact, across regions all perceptual features were usually decoded before 250 ms, with the majority of them decoded at 100 - 200 ms. Another region that showed preference for decoding perceptual features is the left lingual gyrus, for which all top 3 features were perceptual. Here again, *word length* was the best predicted feature, with the other perceptual features following close behind.

In most regions the semantic features were predictable from neural data beginning around 250 ms, with a peak in the window at 400 - 450 ms when considering all regions. There was a clear evolution of the peak time window when the top semantic features were decoded. More specifically, the peak decoding window for semantic features occurred earlier (around 300 - 400 ms) in posterior regions such as the banks of the superior temporal sulci or the inferior parietal gyrus than in the more anterior regions, such as the precentral cortex and left pars opercularis. Finally, we were interested in whether specific cortical regions were associated with specific se-

mantic feature types. For example, the work by Just et al. [38] shows that manipulability is one of the 4 most important factors decoded from fMRI activation while subjects thought about concrete nouns, and one of the main regions representing this factor is the precentral cortex. Hence, can we predict features related to manipulability better in the precentral region? Features such as *can you hold it?* and *can you hold it in one hand?* were among the top 6 decoded features in that region (out of 27 significantly decoded semantic features, $p < 10^{-5}$ FDR corrected for multiple comparisons over features, time, and space). Another region showing such specificity was the parahippocampal gyrus in the left hemisphere. Also in agreement to what was shown in Just et al. [38] regarding the parahippocampal region involved in representing a shelter-like semantic factor, features related to a specific location, such as *would you find it in an office?*, *would you find it in a house?*, and *would you find it in a school?* were ranked among the top 6 decoded features using data from the parahippocampal region (out of 11 significantly predicted semantic features, $p < 10^{-5}$ FDR corrected for multiple comparisons over features, time, and space). At this point, it is important to remind the reader that regions such as the parahippocampal cortex are fairly distant to the MEG sensors, so activity assigned to such deep regions is susceptible to issues of spatial resolution in MEG, as well as any errors associated with the parcellation process based on the subjects' anatomies and the common atlas.

While Figure 2.7 gives an overview of the top features predicted in different brain regions, it is worthwhile to explore the evolution of feature encoding in the MEG signal in more detail. For example, Figure 2.8 shows the feature score for all perceptual and semantic features when using the data from the whole brain as the input to the classifiers.

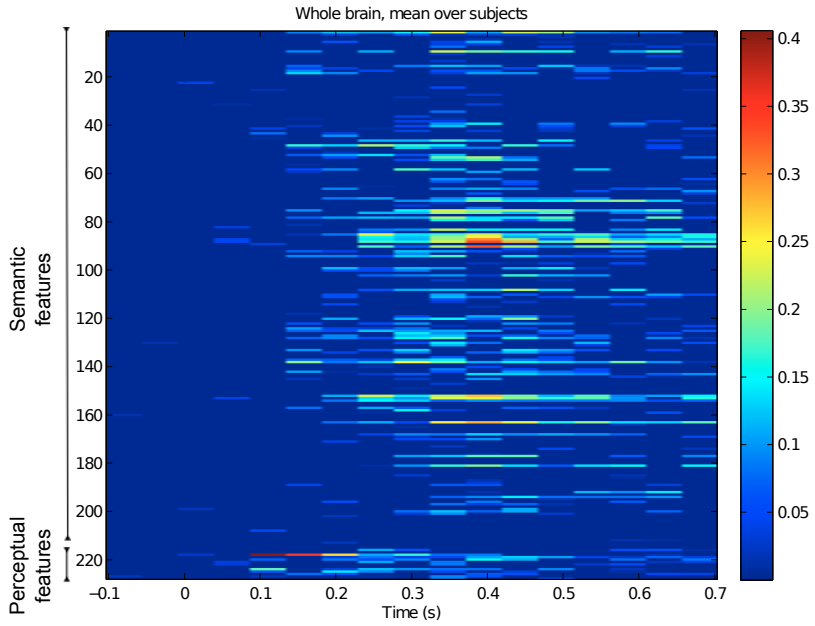


Figure 2.8: Spatio-temporal characterization of feature decodability using data from the whole brain. Perceptual features were better decoded earlier and semantic features later in time. The color in the plots shows the feature score of individual features averaged over subjects. X-axis shows the beginning of each time window (e.g. first column shows results for time window -100 to -50 ms, then -50 to 0 ms, etc). Time 0 is when the stimulus is presented to subjects. Only values significant at $p < 10^{-5}$ are shown in the plot (i.e. all others were set to 0 and shown as dark blue).

It is clear that perceptual features, such as *word length* (219), peak around 100 ms when using

whole brain data, and semantic features peak later in time (from 250 to 500 ms). The semantic features with highest feature score in Figure 2.8 are related to size (*is it bigger than a car?* (88), *is it bigger than a person?* (90)), which peak at 400 ms, and manipulability (*can you hold it?* (152) and *can you pick it up* (163)). Following these results, two complementary questions can be asked. First, where and when are some of the perceptual and semantic features best decoded? Then, can we provide a more comprehensive map of what features are best encoded in different parts of the brain?

We answer the first question by providing a comparison of the feature scores for different perceptual and semantic features, over time and space (Figure 2.9). The order of ROIs in the Y-axis was chosen based on their peak feature score for the feature *word length*, and this order was kept consistent across plots (as seen in Figure 2.10).

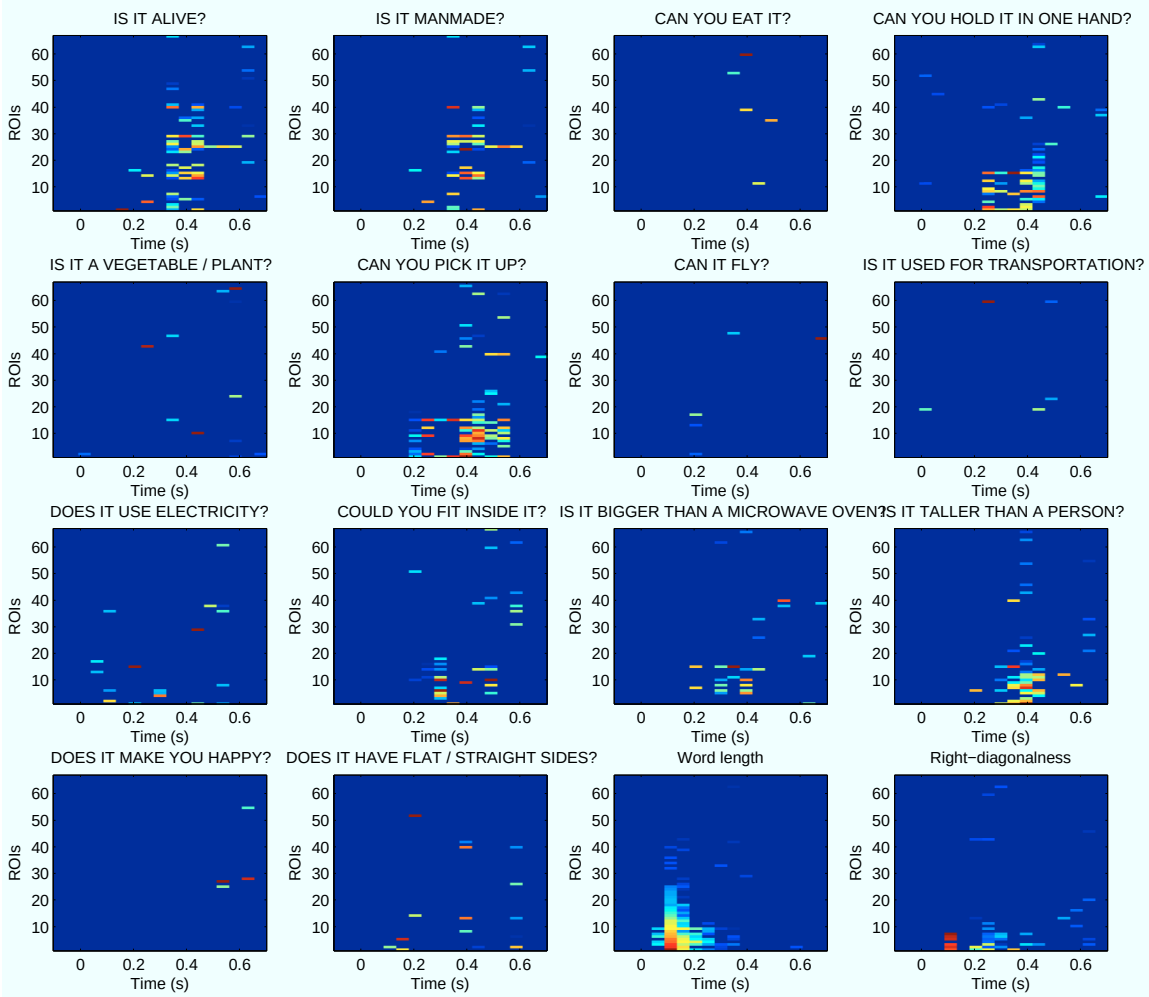


Figure 2.9: Spatio-temporal characterization of feature decodability using data from the different ROIs. The color in the plots shows the feature score of individual features averaged over subjects, with the order of ROIs kept constant in the Y axis. Time 0 is when the stimulus is presented to subjects. Only values significant at $p < 10^{-5}$ are shown in the plot (i.e. all others were set to 0 and shown as dark blue). Each plot has its individual color scale.

The most immediate conclusion one draws from Figure 2.9 is a confirmation of what has been shown so far: while perceptual features are encoded early in time, and mostly focusing on

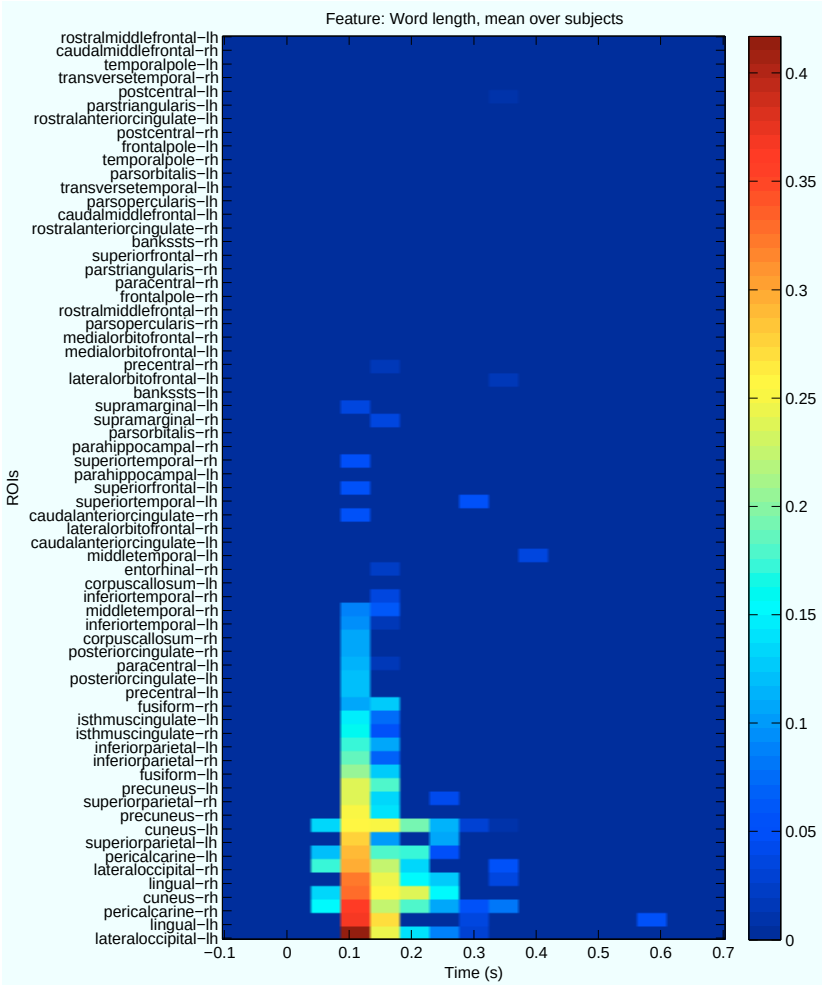


Figure 2.10: Spatio-temporal characterization of *word length* decodability using data from the different ROIs. The color in the plots shows the feature score averaged over subjects, where the different ROIs are listed in the Y axis. The order of ROIs was chosen based on their peak feature score for the feature. Time 0 is when the stimulus is presented to subjects. Only values significant at $p < 10^{-5}$ are shown in the plot (i.e. all others were set to 0 and shown as dark blue).

posterior regions, the semantic features are better decoded from data from later time windows and more anterior ROIs. This point is made clear when comparing two stereotypical features for each group: *word length* and *is it alive?*. Although feature scores for *word length* are more time-locked, it is clear that the peak decodability for *is it alive?* happens much later in time, around 400 ms. It is also worth mentioning that highly correlated features, such as *is it alive?* and *is it manmade?* show similar spatiotemporal signatures, as one would expect.

From observing the evolution of feature scores for these prototypical features of different categories, it is clear that there is a gradual process of recruiting different ROIs in time, which encode different features of the stimuli (i.e. instead of a single peak of activation that would represent a word being ”figured out” all at once). Some semantic features, such as *is it bigger than a microwave?, does it have flat / straight sides?*, and *does it use electricity?* show a profile that is more spread over time and space, which is different than features like *is it alive?* and *is it taller than a person*, which have a narrower encoding profile. We can also observe some

semantic features being encoded earlier than others. For example, features such as *can you eat it?* and *does it make you happy?* only show significant feature scores after 400 ms, while features such as *can you hold it in one hand?* and *is it alive?* show significant feature scores as early as 250 ms. Finally, another important result shown in Figure 2.9 relates to the spatial distribution of feature scores. It shows that a particular semantic feature can be encoded in the MEG signal from several different regions. For example, the size-related features showed feature score peaks in left inferior parietal cortex, bilateral occipital cortex, and bilateral cuneus starting at 250 to 300 ms. In addition to those, manipulability-related features showed such early peaks in the inferior and superior parietal cortices. However, animacy-related features showed higher feature scores in more anterior regions, such as the middle temporal gyrus and supra marginal region, happening around 350 to 400 ms, although inferior and middle temporal gyri were also represented.

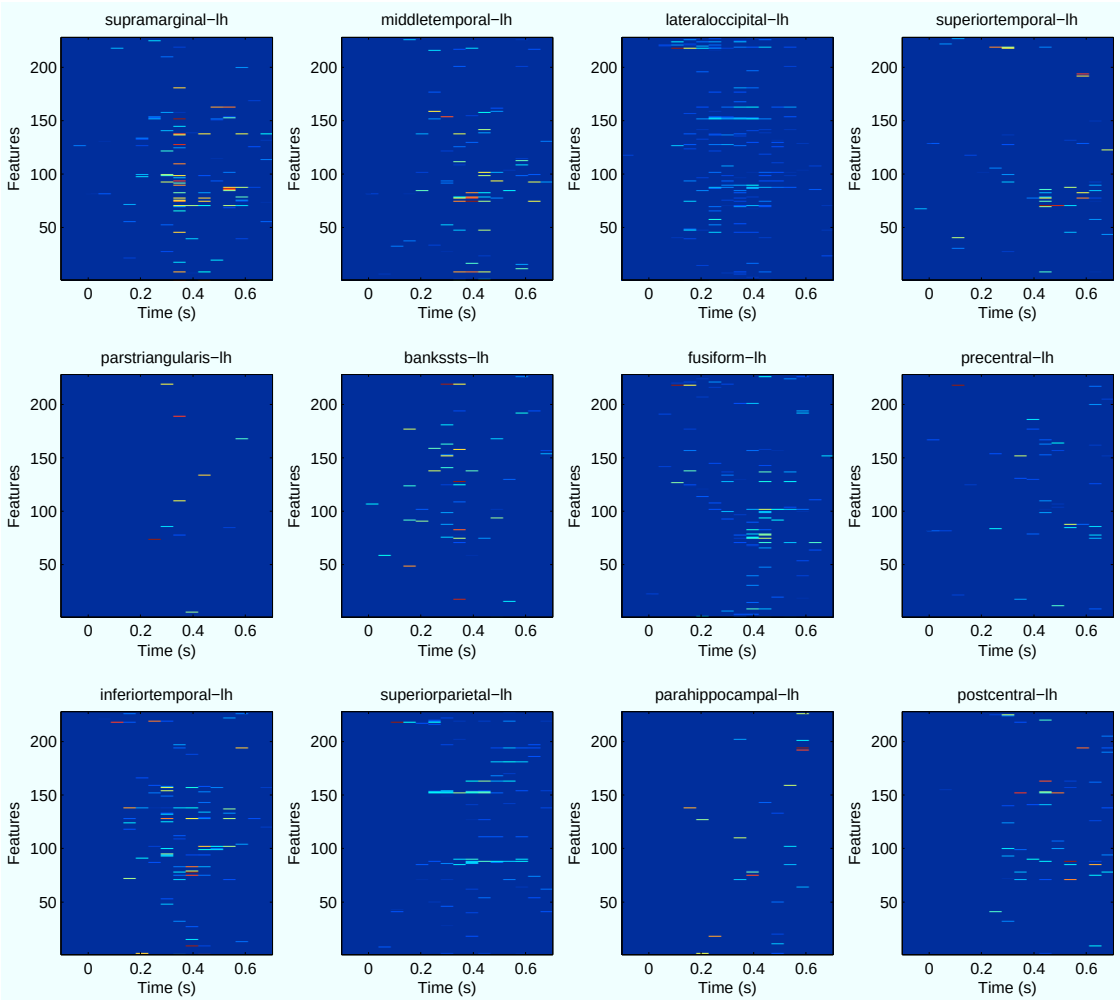


Figure 2.11: Different view of feature decodability characterization using data from the different ROIs. The color in the plots shows the feature score of individual features averaged over subjects, where features 1 to 218 are semantic, and the rest are perceptual. Time 0 is when the stimulus is presented to subjects. Only values significant at $p < 10^{-5}$ are shown in the plot (i.e. all others were set to 0 and shown as dark blue). Each plot has its individual color scale.

While the results in Figure 2.9 corroborate the views of distributed coding for semantic features, it is still important to ask whether certain regions of the brain show preference for different

semantic features. Figure 2.11 shows a different view of the results first displayed in Figure 2.9, but now focusing on what features were best decoded using only data from the different ROIs as inputs to the regression functions. The ROIs shown were selected based on their high accuracies in distinguishing between two novel nouns (see section 2.2.3). We can clearly see different profiles of feature encoding for the different ROIs. While regions such as the left lateral occipital cortex show stronger encoding for *word length* at 100 ms, regions such as the left superior temporal gyrus equally significantly decoded semantic features after 350 ms. Regions such as the left inferior and middle temporal gyri, and the supra marginal region seem to significantly encode several different semantic features, which does not happen in the left pars-triangularis. In corroboration to what has been shown before with fMRI data [38], features related to manipulability were well encoded in the MEG signals from the left post central gyrus, supra marginal gyrus, left pre-central, and left inferior temporal gyrus, all between 350 to 450 ms. Similarly, a feature related to shelter (*can you fit inside it?*) showed high feature scores in ventral regions such as the right fusiform gyrus and bilateral precuneus at 300 ms. However, none of the regions associated with eating-like features (e.g. pars triangularis) showed up in our results.

2.3 Discussion

This chapter presented a methodological approach based on machine learning methods that can be used to identify where and when the MEG signal collected during stimulus comprehension encodes hundreds of semantic and perceptual features. Applying this analysis we found that semantic features related to *animacy*, *manipulability*, and *size* were consistently decoded across subjects. These semantic features were encoded in later time windows, after 250 ms, than perceptual features related to the visual stimuli, which were best decoded before 200 ms. Finally, we have shown that MEG activity localized to certain regions of the brain, such as the lateral occipital cortex and the lingual region, were preferentially related to encoding perceptual features, whereas activity localized to other brain areas, such as the inferior parietal and inferior temporal regions, showed preference for encoding semantic features.

We have also shown that our algorithms can decode a diverse set of semantic features with sufficient accuracy that these decoded semantic features can be used to successfully distinguish which of two words the subject is considering. We report 91% accuracy in distinguishing between pairs of stimulus words based on semantic features decoded from MEG data, even when neither word was present in the training data. When this decoding task was performed using different 50 msec windows in time, the best decoding time window turned out to be from 250 to 450 ms post stimulus onset. This time frame of decoding further supports the hypothesis that the classifier uses semantic information for classification, as several studies have demonstrated the importance of this time window for semantic processing [47, 77, 95, 99]; purely perceptual characteristics of the stimuli influence cortical activity earlier. We also found that the time course of MEG activity and the time course of decoding accuracy based on semantic features did not coincide, suggesting that only a fraction of the total MEG activity is involved in encoding stimulus semantics. The present results also support the notion that different decoding accuracies at different time windows is not simply a result of higher signal-to-noise ratio in the MEG data, the peak of which should presumably occur at the peak of MEG activity.

2.3.1 Comparing the method with existing approaches in the literature

This dissociation between the time at which the MEG activity peaks and the time at which the information about semantic features peaks highlights the importance of using a novel data analysis method like ours, that decodes specific information from MEG signals, in contrast to more standard approaches that merely focus on differences in the magnitude of this activity. The current method is also novel because it uses an intermediate feature set to decode attributes of nouns on which the classifier was never trained. Combined with MEG, this approach allows experimenters to test the relationship of different sets of semantic and perceptual features to brain data, and also to analyze when and where different features are best encoded in the MEG-derived brain activity. Additionally, this method uses cross-validation to test several hypotheses about how perceptual and semantic features are encoded in brain activity. Examples of such hypotheses are the degree of distribution of information encoding, the type of features encoded in the neural signal, the time points that contain most semantic information, the order in which perceptual and semantic features are encoded, etc. In this dissertation, we showed that such hypotheses can be tested with highly significant results.

It is important to compare the proposed method with other techniques previously proposed in the literature. Many of them resort to multi-variate pattern analysis (MVPA) to classify data from a particular scan into different categories, such as animals and objects (see [9, 31] for examples using fMRI and MEG, respectively). While these approaches offer many interesting insights to neuroscience, one of the main differences between MVPA-approaches and what we presented is the possibility of testing different perceptual and semantic feature sets as hypotheses of how knowledge is represented in the brain. Here, we have used a set of 218 semantic features augmented by a few perceptual features. However, we could have used any other set of features that one hypothesizes to better represent knowledge storage in the brain. Finally, throughout this chapter we used the concept of encoding/decoding features with the idea that the MEG activity *encodes* a feature (at least relative to our stimuli) if the algorithms can *decode* that feature from observed MEG data of a subject processing a novel stimulus. It does not necessarily follow that the brain represents the object in question using the same code, because the information we are decoding could represent some correlate of the processing the brain performs in such tasks.

A different method that deserves further exploration is representational dissimilarity matrices (RDMs) [42]. In short, Kriegeskorte et al. use this method to compare representations in two different domains. For example, [43] uses RDMs to compare the categorical representation of objects of man and monkey in inferior temporal cortex. Similarly, one could use RDMs to compare between a given feature set and brain data. In RDMs, one matrix of correlations is created per domain being compared. In our case, one would create a 60 x 60 matrix where each item in the (symmetric) matrix represents the correlation of two nouns in the specific domain. The correlation is calculated using the two vectors of length 218 containing the semantic features for the two words, or the two longer vectors containing brain activity for the words. Finally, the correlation between the two 60 x 60 matrices is computed, for a final single number representing the dissimilarity (1 - similarity) between the two domains. Moreover, Kriegeskorte et al. use Spearman correlation as the metric for similarity, because it is invariant to differences in the mean and variability of the values being compared (unlike Euclidean distance, for example), and also to avoid assuming a linear match between domains.

There lies one of the main differences between RDMs and the method described here. Our approach assumes a linear model, so that the more distant two words exist in semantic space, the more distant they are expected to exist in brain space. We also perform our analysis in a cross-validation framework, and assure that our model can generalize to nouns it has never seen during training. In general, one uses a correlation metric (in RDMs' case, Spearman correlation) when there are two measured variables and the quantity of interest is how these variables are related. Regression, on the other hand, is better suited when one of the variables is to be experimentally manipulated (e.g. a semantic feature set to be tested), and in the end it is possible to write an equation with which one can predict one variable from the other. Such prediction is not possible using Spearman correlation.

2.3.2 Limitations of the paradigm

While this chapter describes an innovative method for tracking the flow of decodable information during language comprehension, the cognitive implications of the results are necessarily limited by the paradigm used while collecting the MEG data. Here, the choice for double stimulation (word and line drawing) was influenced by the group's previous successful results with a similar paradigm in fMRI [58, 79]. This set of stimuli was exported to MEG without major modifications, with the primary intent to evaluate the methodological approach described in this chapter. One advantage of the present stimulus set is that the line drawings help to disambiguate the word meanings. However, the processing of the word-drawing pair by the subject is likely to include neural subprocesses that would not be present using either picture or word stimuli alone. For example, some of the neural activity we observe might reflect post-perceptual processes of matching the word and the line drawing. It can also be the case that the order in which subjects viewed the word and the picture affect the differences in latencies between the decoded perceptual and semantic features. Additionally, there was considerable correlation between perceptual and semantic features, and also within the semantic feature set. This correlation is a plausible explanation for the early rise of accuracies even when only semantic features were decoded from the MEG data (Figure 2.2). In chapter 5 we apply a similar analysis approach to experimental paradigms using stimuli that avoid the confounds of the double stimulus and are better optimized for the high temporal sensitivity of MEG [95]. Such paradigms (*1back-text* and *1back-speech*, as described in section 1.2) allow us to study these timing issues directly, better disambiguate between the different perceptual and semantic features, and also align better with existing studies of the neural correlates of language processing [30].

There is an on-going debate in the language literature about serial versus parallel models of language comprehension (see Pulvermüller et al. [73] for a review). Proponents of the parallel model cite evidence for early ($<$ 250 ms after stimulus onset) manifestations of processing of semantic information [28, 59]. The results in this chapter show that most of the semantic features were encoded later in the trial. While this observation may seem to disagree with the parallel model hypothesis, our two-stimulus protocol was probably not ideal for asking this question, nor was it designed for this specific purpose. These issues will need to be addressed in more detail by future studies that use more optimally designed stimuli and experimental protocols.

2.3.3 Perceptual and semantic feature sets

The analyses here were based on a set of 218 semantic features, chosen to span a broad space of semantics for concrete noun stimuli. One eventual goal of our research is to identify the key semantic subcomponents, or factors, that are combined by the brain to form neural encodings of concrete word meanings. Our present experimental results show that the 218 semantic features did indeed span a sufficiently diverse semantic space to capture subtle distinctions among our concrete noun stimuli: our approach was able to distinguish which of two novel words was being considered with highly significant accuracy when using only the decoded estimates of these 218 features. The high performance of the model suggests that the stimuli studied here may share a fairly similar grouping in brain signal space and the semantic features space spanned by the 218 features. Furthermore, our results suggest that only a fraction of the 218 features need to be accurately decoded in order to reliably distinguish between the two left-out nouns. The most decodable of the 218 semantic features could be grouped into three main classes: features about *animacy*, *manipulability*, and *size*. These groups intersect the 3 factors (*shelter*, *manipulability*, *eating*) previously shown to be relevant to the representation of concrete nouns in the brain [38], even though that work was done using fMRI, and employed a completely different type of analysis and task performed by the subjects. The other factor singled out by Just et al. [38], *word length*, was also consistently decoded in this chapter. Moreover, several other studies have shown the importance of *animacy* in the neural representation of concrete nouns [48, 51, 60]. In the future, we plan to narrow down the number and type of semantic features being decoded in order to fully characterize a minimal semantic space that is sufficient to obtain comparable results. This could be done by applying dimensionality reduction algorithms to our set of 218 semantic features, as well as to other sets of features such as word co-occurrences, and then interpreting the minimum number of uncorrelated features that are needed to obtain significant accuracies. Also, experiments that use a larger and specially-chosen set of stimuli will help to decrease the between-feature correlation that affects the current results.

Another issue worth discussing regards the blurry line between perceptual and semantic features. For example, consider the question of whether to define the feature *does it have 4 legs?* as perceptual or semantic. Of course the dividing line between the two is a matter of definition. We find it useful to adopt the following operational definition: any feature of a concrete noun that a person considers regardless of stimulus modality (e.g. independent of whether the stimulus item is presented as a written noun, spoken noun, or picture) we define to be a semantic feature of that noun. Clearly, this definition allows some sensory features such as the shape of a pine tree or the sound of a bell to potentially be semantic features, to the degree that we think of those sensory features automatically when we read the word. Given previous work suggesting that neural representations of concrete noun meanings are largely grounded in sensory-motor cortical regions [38], this seems appropriate. To the degree that we think of a bell's sound when reading the word *bell*, then it seems appropriate to consider it part of the word's semantics – a part that is also activated directly when we hear the bell instead of read about it. It is also hard to prove, with the current paradigm, that some of the semantic features we decoded here are not simply a correlate of perceptual features that we did not include in our original set. While it would not be possible to list all possible perceptual features for decoding, increasing the variety of features in the set can help with this issue, as well as employing a broader and better-designed set of stim-

uli that would control for most of these correlations between semantic and perceptual features, especially data with both auditory and visual stimuli (see chapter 5).

2.3.4 Comparison to results using fMRI

Previous reports that used fMRI data to decode cognitive states showed regions of the brain contributing to decoding results that were not found in this study. For example, parts of the brain that are commonly associated with semantic processing in the fMRI literature [70], such as the left pars opercularis and pars orbitalis, did not show high decoding accuracies over time in our MEG study. Regardless of the complications of capturing MEG signatures from different parts of the brain (e.g. from subcortical structures [24]), there are two important points to be made. First, it is common for subjects in fMRI experiments to have a significant amount of time to think about the different properties of an object. Although this time is necessary in order to capture the slowly-rising BOLD response, that also gives the subject the opportunity to think of specific properties of a noun that will involve those regions of the brain. For example, while it is debatable that the distributed representation of a screwdriver involves motor cortex every time the subject thinks of it, if the subject is asked and given enough time to imagine certain properties of a screwdriver, and to do this as consistently as possible across repetitions of screwdriver, it is very likely that holding a screwdriver will come to mind in every repetition, and this way motor cortex will be active [79]. In our MEG experiments, subjects spent only about 1 s on average considering the noun stimulus. Thus, the MEG signals on which our analysis was based reflect mostly the automatic brain responses when the subjects think briefly of one of the nouns. Another aspect of the fMRI signal that can affect the analysis regards the inherent averaging of signals over time. For example, if the subject thinks of picking an apple and then eating it in one trial, and thinks of the same properties in the opposite sequence in another trial, the resulting image will likely be similar in fMRI, but not MEG. Moreover, because our analysis was conducted on data averaged across trials, the chances of activity that is not time-locked to the stimulus to be washed away in the averaging process is high. It has also been conjectured that MEG gives stronger emphasis to bottom-up processes, while fMRI tends to emphasize more top-down processes [96], based on the observation that certain regions of the brain show task effects in fMRI but not in MEG. If a region is more affected by top-down modulation than others, it is possible that non-synchronous activation will dominate that region, and hence be more likely captured by fMRI (because of the summation of metabolic demands of the region over time) than by MEG, which measures the synchronous post-synaptic activation of thousands of pyramidal neurons. This type of activation would likely be elicited in the longer trials of fMRI experiments and contribute to activation in areas not seen in MEG.

To summarize, we showed that several semantic features can be decoded from the evoked responses in MEG. Nevertheless, it is likely that some of the responses resembling what is usually seen in fMRI (assuming an unlikely 1-1 correspondence between MEG and fMRI activity), were either not activated by the paradigms, or were offset in time and were not represented in the averages over trials. Regions pre-selected from the fMRI literature did not perform as well as other regions in the cortex in distinguishing which of two novel nouns was being presented. It is clear that the tasks used in fMRI and MEG, as well as the nature of the signals being measured, influence the regions activated in the brain and the results we obtained in the different

experiments. The context in which the nouns are thought of can also help to justify some of the differences between MEG and fMRI results noted in this dissertation. On the other hand, we saw some intriguing results regarding certain regions of the cortex displaying a bias towards decoding specific types of features. For example, some of the top semantic features decoded from motor cortex were related to manipulability (e.g. *can you hold it?*), and the top features decoded from the parahippocampal region were associated with a location (e.g. *would you find it in an office?*). It is common to see motor cortex involved in the representation of tools, and the parahippocampal region associated with shelter-like features [38, 58]. We hope that future results using this method to analyze the data of better-designed paradigms will help elucidate the role of such regions in decoding these features.

2.4 Future work

The work presented here still leaves several topics to be investigated, of which some have already been mentioned in the Discussion section. This section lists additional topics to be explored in the future:

- We have shown that we can distinguish between two nouns within certain categories (i.e. very close in semantic space) with statistically significant accuracies. It is unclear whether adding semantic features that better distinguish among the words in the same category would benefit the model. In theory, yes, but that highly depends on whether these new semantic features are encoded in the recorded MEG signal, and if the signal-to-noise ratio obtained by averaging a few trials is enough to capture this information.
- Similar to the item above, exploring different semantic feature sets will be crucial to advance this research. Some work has been conducted in our group [61] to compare different semantic feature sets when decoding fMRI data. Many of the feature sets compared were derived from text corpus data, which is certainly an interesting approach to collect the representation of words based on every-day language. It is also, in theory, easier to collect representation of new words, compared to using Mechanical Turk to collect representations for the 218 features collected using crowd-sourcing. Moreover, one could arbitrarily choose the number of dimensions to be used when employing corpus data sets, but then one runs into one of the limitations of the method: interpreting what each dimension actually means. To that extent, if one's goal is to only compare the similarity of a given feature set to brain data, then RDMs as employed by Kriegeskorte et al. [42] can also be a valid method, especially if one considers that the relationship could be highly non-linear.
- While we presented a method that can estimate where and when different perceptual and semantic features are encoded in the MEG activity, it remains unclear how the information is transformed by the brain from perceptual to semantic features. More work needs to be done to understand how the transformation takes place between the different levels of representation, and also to investigate how the encoding is actually performed in the brain.
- The spatial resolution of MEG (about 1-3 cm in our case - based on the source localization technique and the highly cognitive paradigm used) might also become an issue when comparing these results to what is found in fMRI. The source localization method used here is

susceptible to blurring, an effect that smears the activation coming from a region to other nearby locations. That is more likely to happen in regions with lower SNR, such as deeper parts of the brain that stay farther from the helmet, or that are obscured by other parts of the cortex. That certainly contributes to regions analyzed here showing similar feature evolution profile as others. For example, some results for the superior parietal cortex show similar profiles as the ones obtained for the lateral occipital cortex. It is the case that the two ROIs share boundaries, and when activation is located nearby this boundary, it is likely that it will be attributed to both regions. Different methods of source localization might be employed in the future to hopefully overcome such issues, or paradigms that provide more trials to increase SNR might also be a viable option.

- The way minimum-norm source estimation was used sums the amount of activity within a cortical area to a single unsigned value per time point that represents the total amount of activity for the source at that instant. However, it carries no information about the direction of current flow that can be a functionally highly relevant parameter and might well be helpful in decoding the different perceptual and semantic features used in the analysis.
- The results shown in this chapter rely on the assumption that perceptual and semantic features are coded by the magnitude of the MEG signal, but there may also be relevant information in other attributes of the signal. Several other types of attributes, such as the power and phase of different frequency bands, or different functional connectivity measures [45], might work as well, if not better, in decoding features [16], and it is possible that some combination of these different attributes would be the best approach. Using smaller time windows might also benefit the analysis of feature evolution over time and space, hence providing better time resolution of onset of decoding of different perceptual and semantic features. Another pitfall of using the amplitude of the MEG signals, and averaging them over different trials is the temporal alignment of the features being decoded. One could argue that the problem is ameliorated because we average different repetitions of the same noun, but it is still unlikely that the subjects would think about a given noun with strictly consistent timing across the 20 repetitions. The signal used for feature decoding might thus be washed out by averaging across the repetitions. In fact, that is true for any cognitive processing that is not time-locked to the stimuli, and hence exploring representations that are not as restricted by this time-lock assumption could greatly improve these results in the future. An interesting future direction could be to perform correlation analysis between the time series of activation in the various areas identified in this study, as it could provide better understanding of how these regions interact to communicate and encode information.

Chapter 3

Predicting MEG activity associated with the meanings of nouns

We have presented in chapter 2 an effective method for studying the temporal sequence and cortical locations of perceptual and semantic features encoded by observed MEG neural activity, while subjects were presented with concrete objects stimuli. Here, we invert the direction of the model in order to predict MEG activity from a set of perceptual and semantic features (similarly to what was done in [58] with fMRI data). This model corroborates the results obtained in the previous chapter, and also allows us to ask complementary questions to what was asked before.

3.1 Motivations for inverting the direction of the model

The equations used here are the same as the ones used in chapter 2. Namely,

$$\widehat{\mathbf{f}} = \mathbf{x} \cdot \widehat{\mathbf{W}} \tag{3.1}$$

where:

$$\widehat{\mathbf{W}} \;=\; \mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}_N)^{-1}\mathbf{F} \tag{3.2}$$

However, while in the previous chapter X stood for the MEG data and f for the feature set, now $\mathbf{X} \in \Re^{N*p}$ contains the feature set (for example, p = 218 when using the set of 218 semantic features), and $\mathbf{F} \in \Re^{N*d}$ becomes the MEG data. The model accuracy is still evaluated by leaving two words out, training on the remaining set, and predicting the MEG activity for the 2 left-out words ($\mathbf{x}$). The distance between predicted and measured values of MEG activity is calculated using the cosine distance between the two vectors of activations. As usual, the important fact here is that the model can extrapolate to predicting activity for words it never saw during training. The 2-vs-2 method is simply a way to evaluate it, without biasing the comparison results towards items the model has used for learning.

The direction of the model (i.e. whether the end result is predicted MEG activity or feature values) is mostly dependent on the questions we want to answer (the two directions are illustrated

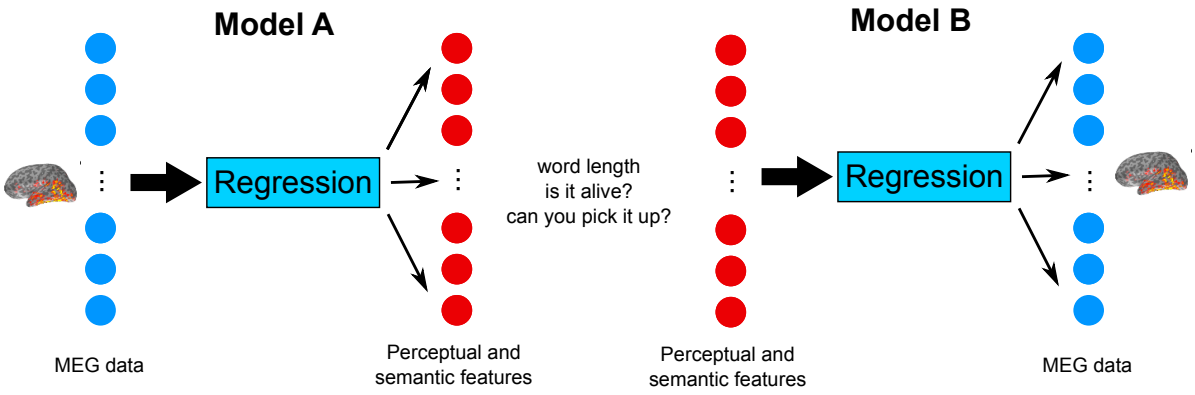


Figure 3.1: Schematic of the two model directions. While model A uses MEG data as the input to predict individual perceptual and semantic (A)ttributes, model B uses the perceptual and semantic feature set as its input to make independent predictions for each (B)rain data point in space and time.

in Figure 3.1). When brain activity is being predicted (model B), we can ask questions such as "what brain regions and time windows do our semantic feature model best account for?" and "how much does each feature contribute to predicting neural activity in that brain region / time window, compared to the other features' contribution?". Additionally, predicting the activity allows for a more direct comparison to the group's previous results with fMRI data [58].

Predicting MEG activity also allows for more versatility in testing the predictions of different brain regions. While in chapter 2 a different instance of model A had to be trained for each ROI or time window being investigated, here all we need to do is separate the MEG activity that belongs to the ROI or time window of interest and only use those when evaluating the model. This versatility is only true because each brain attribute is predicted independently of the others, similarly to how each feature was independently predicted of the others in chapter 2.

One drawback of predicting the MEG activity is that the prediction target is noisier than the semantic and perceptual features, so we need better feature selection in the output side. More specifically, there are two sides to this issue. Firstly, many brain regions exhibit activity that is insensitive to the stimulus. Secondly, the MEG signal for a region that is sensitive to the stimulus might be corrupted by a significant amount of noise. In other words, if one uses all brain predicted data for scoring the results, the noisier MEG activity will not be predicted as well, and thus drag down the accuracies. Conversely, when using brain activity as the input for the regression model as in chapter 2 (model A), the regularization parameter can effectively penalize the weights of noisy features.

3.2 Considerations of using sensor and source space data

Another methodological consideration that is worth discussing here is whether to use sensor or source domain data in the models. In short, does using one over the other improve the results? The direct answer is no. Because we are using linear regressions and a linear source localization method, the results should be very similar. But let's provide some details to convince the reader of this conclusion.

Equations 3.1 and 3.2 summarize the regressor we use to predict MEG data from an inter-

mediate feature set. There is a vast literature on source localization methods [25], and here we used the Minimum Norm Estimates method [23] (MNE) to estimate the sources of the measured MEG activity. Most source localization methods can be summarized in equation 3.3:

$$\widehat{\mathbf{s}} = \mathbf{y} \cdot \widehat{\mathbf{M}} \tag{3.3}$$

where $\widehat{\mathbf{M}}$ is called the linear inverse operator, $\mathbf{y}$ is the measured MEG sensor data and $\widehat{\mathbf{s}}$ is the estimated source data. In MNE, $\widehat{\mathbf{M}}$ is estimated by:

$$\widehat{\mathbf{M}} \;=\; \mathbf{R}\mathbf{G}^T(\mathbf{G}\mathbf{R}\mathbf{G}^T + \lambda^2\mathbf{C})^{-1} \tag{3.4}$$

where $\mathbf{R}$ is the source covariance matrix, $\mathbf{G}$ is the forward solution, which maps the influence of a point source in different locations of the cortex to all sensors, $\mathbf{C}$ is the sensor noise covariance matrix, and λ is a regularization term.

When constructing a forward solution, it is common to use 3 orthogonal vectors per source location to represent the orientations the current can take at that location [24]. However, many times only the orientation normal to the cortical mantle is used, because it is believed that most of the captured MEG activity is generated by large cortical pyramidal neurons with dendrites normally oriented to the local cortical surface [2]. Whether there are 3 orientations per source location, or only one, MNE is still a linear operation. So, if our regression function originally would learn the weights $\mathbf{W_1}$ to map from feature set to sensor data, and $\mathbf{W_2}$ to map from feature set to source data, then it should have no problems to learn $\mathbf{W_3}$, which is a linear combination of $\mathbf{M}$ and $\mathbf{W_2}$ to map directly from feature set to sensor data, such that $\mathbf{W_3} = (\mathbf{W_2}\mathbf{M^{-1}})$.

It is important to note that when visualizing source space results it is common to combine the results of the 3 vectors per source location into one single value per location. For example, a traditional way to combine the 3 orthogonal vectors in a source location is to take the Euclidean norm among them, which is a non-linear operation. Evidently, if similar non-linear methods are used to combine the different orientations of a source, then the results for sensor and source data will not be the same.

Because of the properties mentioned above, as long as the linear relationship is retained in the chosen data transformations we can train our regressors to predict the sensor data, and then whenever necessary just localize that predicted activation, which reduces the computational time by several factors.

3.3 Replication of results from *answer-questions* dataset

For sanity check, it is important to obtain comparable results with model B using the *answer-questions* data, to what was done before in Chapter 2, using the model that predicted perceptual and semantic features from the MEG data (model A).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
Leave-two-out accuracy	86.44	86.38	73.79	89.60	92.54	88.40	86.10	82.20	88.70	**86.02**

Table 3.1: Accuracies for the *leave-two-out* experiment using perceptual and semantic features to predict all time points from -.1 to .8 s and 306 MEG sensors in *answer-questions* (model B). The classifiers are able to distinguish between two different concrete nouns that the MEG-based classifier has never seen before with 86% mean accuracy over the nine participants S1 through S9. Chance accuracy was 50.0%. For a single-subject model, 62.5% correspond to $p < 10^{-2}$. The p value associated with observing that all nine independently trained participant models exhibit accuracies greater than 62.5% is $p < 10^{-11}$.

3.3.1 Can we discriminate between two novel nouns using predicted MEG activity?

The input to the regressor consisted of all perceptual and semantic features, and the MEG data in all 306 sensors between -100 and 800 ms were predicted in the first stage of the SOCR (as described in section 2.1.2). The results for each of the nine participants are shown in Table 3.1. Although the results are still well above chance level, we notice an overall decrease in accuracies across all subjects compared to the results using model A (Table 2.1). This decrease was already expected given the noisier attributes in the output side of the regressor, but it is encouraging that, even without any type of feature selection, we can still discriminate with a mean accuracy of 86.0% between two novel nouns the model has never seen during training using predicted MEG activity.

3.3.2 How does feature selection improve the discrimination results?

For the remainder of this subsection, we call a feature the MEG datum at a specific sensor and time point, which is the output of the model in this chapter (not to be confused with semantic and perceptual features).

Feature stability

There are several methods one can use for feature selection (see [67] for a review), but here we used feature stability [58]. In short, a feature is deemed stable when it exhibits consistent variation in activity across the repeated presentations of 58 training stimuli (we compute independent stability scores for all features in each cross-validation loop). This variation is measured across the different repetitions of the noun. Consider the synthetic example shown in Figure 3.2. On the left, we see the values for a feature (e.g. MEG sensor amplitudes) across different repetitions of 58 different words. The stability score for that feature is calculated by averaging the pairwise correlation over all pairs of rows in the table. For example, if a feature were to exhibit the same 58 responses during each presentation, it would have an average pairwise correlation of 1.0. On the other hand, consider the table on the right, which shows the values for an unstable feature. Because this is a synthetic example, the values in the table were chosen at random in order to illustrate the point that the stability score for such a feature would be close to 0.

More specifically, feature stability for feature f was obtained according to equation 3.5:

$$stability(f) = \frac{\sum_{i=1}^{R} \sum_{j=i+1}^{R} corr(x_{f,i}, x_{f,j})}{C(R,2)} \tag{3.5}$$

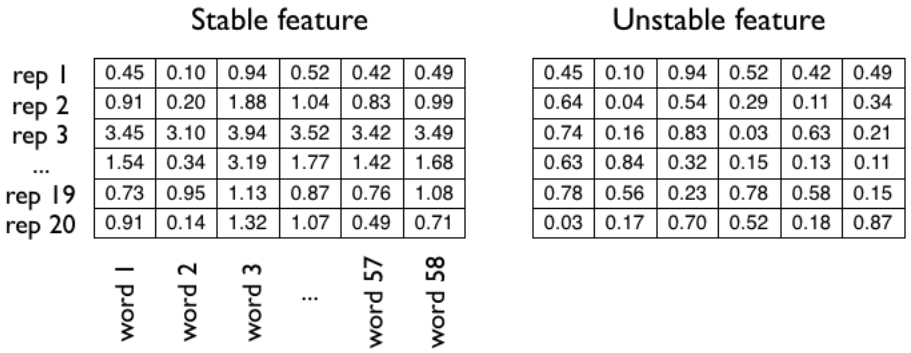


Figure 3.2: Example of a brain feature (e.g. MEG sensor amplitude) with high (left) and low (right) stability score. This measure ranks higher features that behave most similarly across repetitions of the same stimuli.

where R is the number of repetitions in the paradigm (20 in the case of the *answer-questions* data), $C(a, b)$ is the number of combinations of a choose b, $corr(a, b)$ is the correlation between vectors a and b, and $x_{f,i}$ is the vector of length 58 representing the measured values for feature f in the repetition i of each word in the training set. Each feature (i.e. MEG sensor amplitude at a particular time point) has a stability score for each cross-validation fold. More specifically, because we used all possible combinations of pairs out of the 60 possible words (1770) as the test set during cross validation, each feature has 1770 stability values. In each cross validation fold, the stability score is computed using only the 58 words for selected for training, and later applied to the 2 left out words.

Analysis of results after feature selection

The initial question to be asked is how much improvement in the 2-vs-2 discrimination task one can get by selecting a subset of features based on their feature stability. Figure 3.3 shows how the overall accuracy is affected by using only the top-most stable features of the predicted MEG signal in *answer-questions*. It is clear that there is an intermediate number of most stable features where the best accuracy is achieved, and this number varies among subjects. More specifically, when analyzing the *answer-questions* experiment data from -.1 to .8 s in sensor space that are about 55,000 spatiotemporal MEG activity features that are independently predicted by the regression functions. On average, using the best 10,000 features (out of 55,000) for each subject seems to yield the best results, and the accuracies start to drop the more features are used for scoring, which corroborates the statement about many noisy output features hurting the results.

Hence, we can show a different table for 2-vs-2 accuracy after feature selection (Table 3.2). It is clear that some subjects, like S3 and S8, improve greatly by doing feature selection. There are a few other factors that can account for the differences between the results from the previous chapter (first row in Table 3.2 and the results shown here). For example, the data in the previous chapter had been source localized and the 3 orientations in each source location were combined using Euclidean norm, which is a non-linear operation. Additionally, we used data only from 0 to 750 ms before, while now we score using predicted MEG data from -100 to 800 ms (although this should not be a problem when using feature stability, assuming that the non-informative or

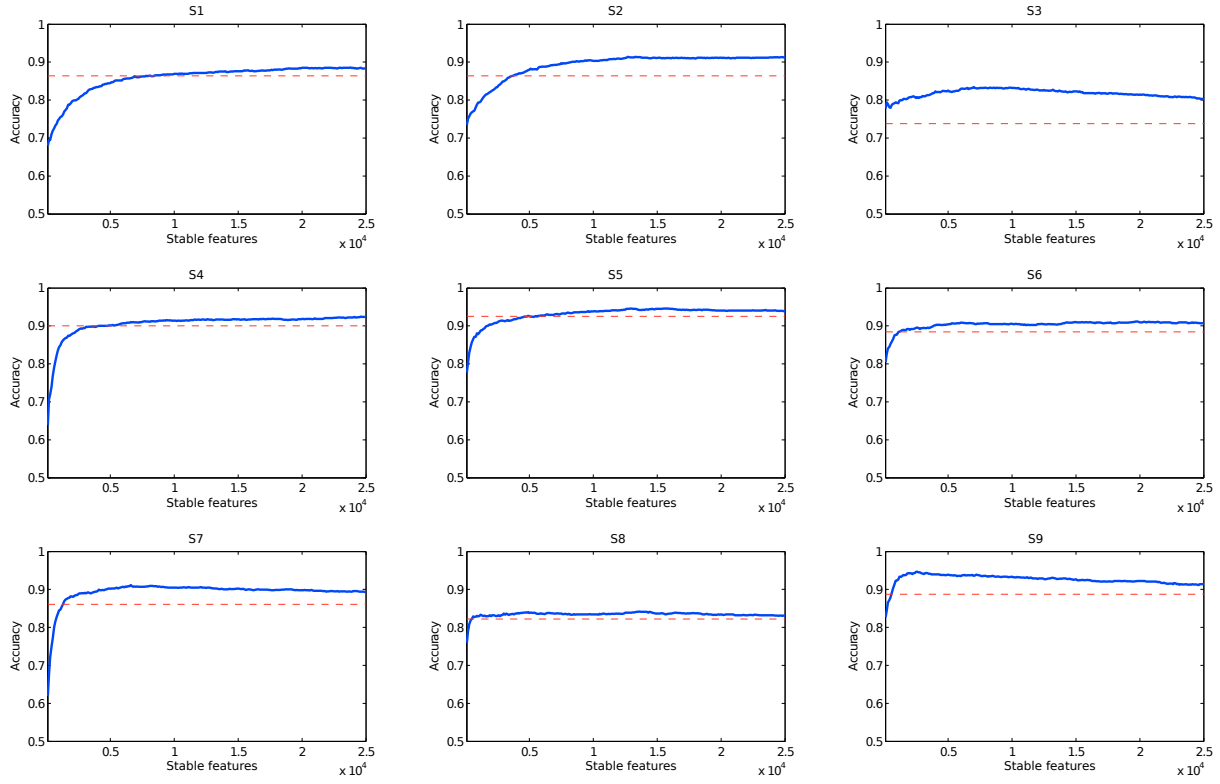


Figure 3.3: Effects of number of stable features used for scoring on the 2-vs-2 accuracy. Model predicts MEG activity in *answer-questions* using a set of perceptual and semantic features. One plot per subject. In each plot, all 55,000 features are sorted in descending order based on their feature stability score, and the accuracy is computed using only a few of the top-most features. Plots also show the accuracy obtained using all features (red dashed line). If the X-axis were extended to 55,000, blue and red traces would meet.

noisier features - before 0 or after 750 ms - would have low stability scores).

To finalize the investigation of how feature stability affects the accuracies, it is important to show how the accuracies evolve over time. Figure 3.4 shows how the 2-vs-2 decoding accuracy is affected by adding more unstable features (sensors in the helmet) at each separate time point. For example, the horizontal line at $Y = 1$ represents the evolution of the 2-vs-2 accuracy over time when using only the most stable sensor at different time points for scoring. Similarly, the horizontal line at $Y = 50$ shows the time evolution of the 2-vs-2 accuracy using only the top 50 most stable MEG sensors for scoring. The first point to observe is that the accuracies start rising at around 200 ms for most subjects. It is also interesting to note that between 200 and 600 ms, only a small subset of the sensors (i.e. only the few most stable ones) are needed to achieve similar accuracy as using all 306 sensors at that time point, as indicated by the long red vertical lines at those time points.

This observation invites a comparison to the results using all sensors and time windows simultaneously (Table 3.3 and Figure 3.3). More specifically, why does the inclusion of less stable features hurt the 2-vs-2 decoding accuracy when predicting all time windows and sensors simultaneously, but the results do not change so drastically when predicting the value of all 306 sensors in a specific time point, compared to predicting only more stable subsets of sensors? Wouldn't one expect that, within a time point (i.e. a vertical line in the subplots of Figure 3.4), the accura-

Scoring features	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
Semantic	88.25	95.82	86.16	95.20	93.62	92.49	92.54	85.25	91.36	**91.19**
MEG data	86.44	86.38	73.79	89.60	92.54	88.40	86.10	82.20	88.70	**86.02**
Best stable MEG	88.59	91.36	83.39	92.43	94.68	91.13	91.13	84.18	94.63	**90.17**
10K stable MEG	86.89	90.34	83.16	91.41	93.84	90.56	90.51	83.54	93.33	**89.29**

Table 3.2: Comparison of accuracies for the *leave-two-out* experiment using *answer-questions*. First row shows the results reported in chapter 2, and the second row repeats the results in table 3.1 for comparison. The last row in the table shows the accuracy when using only the top 10,000 most stable voxels for scoring. Row 3 (Best stable) is displayed for comparison, and the accuracies there show the top accuracy obtained for each subject (i.e. peak of blue traces in Figure 3.3).

cies would go down as we add more unstable brain features? We answer this question by noting that the accuracies shown in Figure 3.4 are computed independently for each time point. So, it is clear that the predictions of several time points do not results in good accuracies (e.g. most of the signal before 200 ms). These are likely the features regarded as unstable when predicting all the data simultaneously, which is also seen in Figure 3.5. So, by avoiding the use of predicted data in those time points for scoring, the accuracy of the classifier that uses predicted data for all sensors and time points would indeed improve. It is also worth noting that there is high spatial correlation among MEG sensors, so using the data of nearby sensors within a time point likely does not add that much information. Additionally, we can still observe instances where adding less stable sensors in the same time point do hurt the accuracies, as in subject S4 around 300 ms, S8 also around 300 ms, and S9 around 250 ms.

Figures 3.4 and 3.5 are examples of another advantage of using model B (i.e. predicting MEG activity): while the predictions were performed for individual time points, we can score them that way, or after averaging the features over time to study the size of different time windows in the results.

In conclusion, it is clear that we can still distinguish between novel words using predicted MEG activity. Switching from a model that predicts perceptual and semantic features using brain data, to a model that predicts brain data using the feature set, costs an average of 5% increase in error. However, most of this increased error can be removed with feature selection, and this way achieve similar accuracies to what we obtained before when using the MEG activity as the input to the SOCR.

3.3.3 Where and when are the most stable features in the MEG helmet?

The feature stability measure also allows us to investigate what spatiotemporal aspects of the MEG signal are most stable across repetitions of the nouns (Figure 3.5). In the plots we show a ranked stability score, which is calculated by averaging the rank of a given feature (i.e. MEG sensor at a particular time point) over all cross validation folds. For example, if a feature has the highest feature stability (equation 3.5) in every cross validation fold, it will have a ranked stability score of 1. Conversely, if a feature is the least stable feature in every cross validation fold, its ranked stability score will be 55,000 (the total number of features).

It is noteworthy how consistent the results are across subjects. We can see that the most stable features start at about 100 ms in occipito-parietal sensors, and also posterior temporal regions, staying stable until about 500 ms. The spatial location of the stable sensors also does not vary greatly over time.

Overall, these results show that, across subjects, the sensors in the bottom half of the helmet

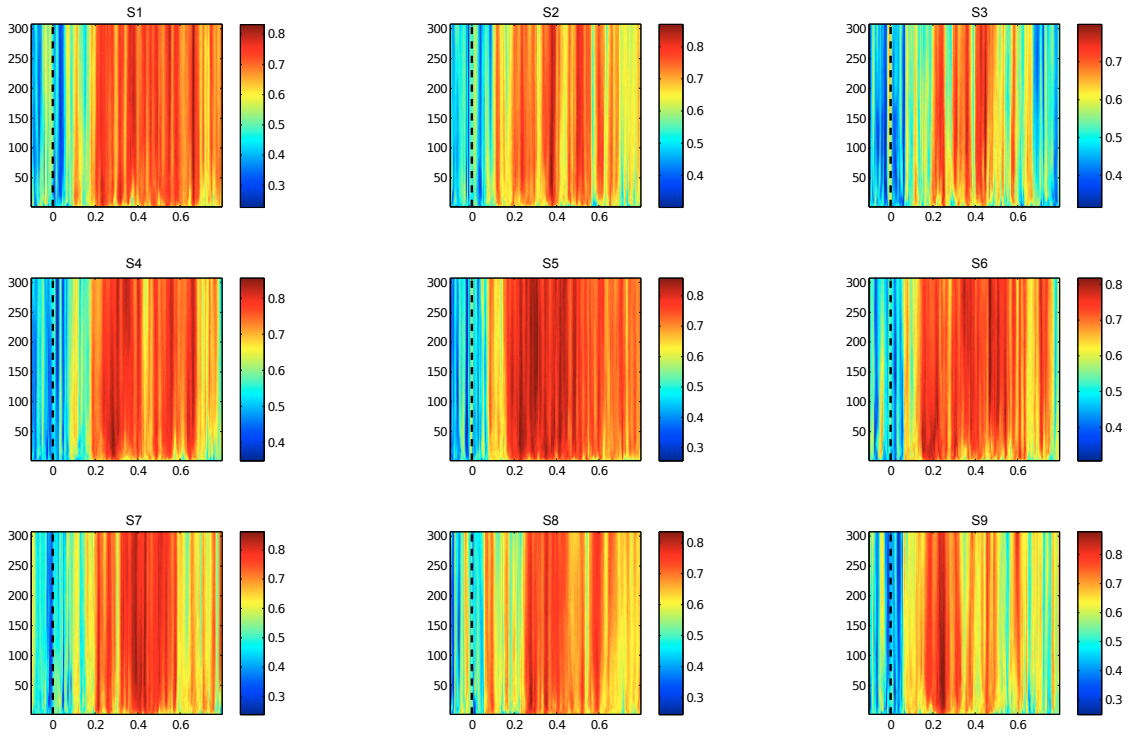


Figure 3.4: Effects of number of stable features used for scoring on the 2-vs-2 accuracy, evaluated for each time point. One plot per subject. In each plot, the 306 features (one per MEG sensor) in one time point (X-axis) are sorted in descending order based on their feature score, and the accuracy is computed using only a subset of the top-most features (Y-axis). The colors in the plot show the accuracy. For example, the top most line of each plot (Y = 306) shows the accuracies over time using all MEG sensor predictions.

respond consistently to the presentation of different words across repetitions, especially in the time window between 100 and and 500 ms after stimulus onset. These sensors cover temporal, parietal, and occipital regions of the brain, which we have seen in the results from chapter 2 to be involved in the processing of several perceptual and semantic features.

3.4 What are the brain regions and time points best accounted for by our model?

This model allows us to investigate what parts of the MEG activity are best predicted by model B using our perceptual and semantic feature set. The measure we will use here is the percent of variance explained (POVE, Equation 3.6).

$$POVE_b = 1 - \frac{\sum \left(f_{b,i} - \hat{f_{b,i}}\right)^2}{\sum \left(f_{b,i} - \bar{f}_b\right)^2} \tag{3.6}$$

where $f_{b,i}$ is the true value of the MEG activity at a particular region and time point b for noun i, $\bar{f}_b$ is the mean activity value over all nouns, $\hat{f_{b,i}}$ is the predicted spatiotemporal MEG activation for the noun i, and the summation is over all possible nouns. The reader will notice that this

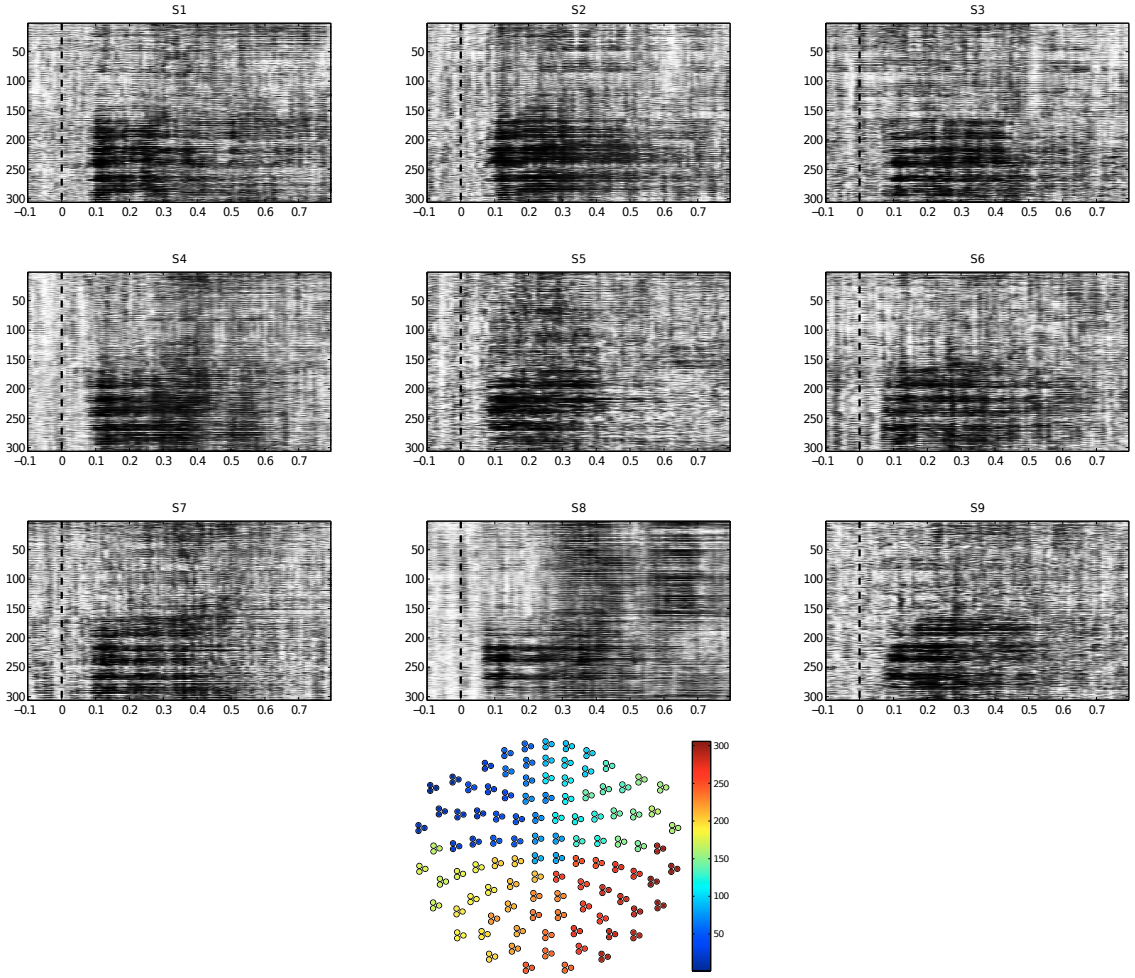


Figure 3.5: Ranked stability scores (see text for description) for all spatiotemporal features in the MEG helmet. The more stable the feature, the darker the shade of gray in each plot. One plot per subject. Time is in the X axis, and sensor position in Y axis. Guide to sensor positions in the MEG helmet (bottom plot) has the following orientation: nose is up, left ear on the left, right ear on the right. The color of each sensor indicates the sensor number, with blue starting at sensor 1.

is the same method we used before (equation 2.4) to quantify feature score. However, now we are using the whole dataset to calculate POVE (i.e. the data for all 60 nouns), and therefore the POVE is bounded between 0 and 1 (where 1 means that our model can explain all the variance in the data). We feel comfortable in using POVE this way because we have already shown that the model can generalize well to new items, and we can assess the reliability of our POVE results by applying permutation tests. More specifically, we calculate the POVE for thousands of iterations after permuting the labels of all trials (before averaging) in each iteration. That generates a distribution under the null hypothesis that there is no relationship between MEG activity and the perceptual and semantic feature set, such that the POVE values generated with permuted data are similar to the ones generated with data that are correctly labeled. The p-value of a POVE result is computed by adding up the iterations when the POVE obtained with the permuted data was bigger than the one obtained with the correct labels (divided by the total number of iterations). Note that POVE is invariant to the amount of activity in a region.

To deepen our neuroscience insights, we first source localized the data before computing POVE (but after averaging over repetitions of the same noun). We used only the normal orientation to the cortex per location, thus avoiding the need to combine different orientations, which makes the entire process linear (as seen in section 3.2).

Figure 3.6 shows how well we can predict the activation in each ROI in the brain over time. The POVE of individual subjects was combined by first setting to zero the POVE of any source in the brain with a p-value higher than 10^{-1}. Then, all the sources in a given ROI were averaged within subjects, and then that result was averaged among subjects.

We can observe in the plot a very early peak in POVE before 100 ms, but the most time-locked peak happens at about 250 ms, starting to rise around 150 ms. This peak in POVE is better captured in posterior regions such as the bilateral lingual, pericalcarine, and lateral occipital cortices. We later find spatiotemporal regions that are also well-explained by the model, but they are not as time-locked to the stimulus as the first one. These peaks are best observed in the inferior parietal region (325 to 475 ms) and other more anterior regions, such as the banks of the superior temporal sulcus and the supra-marginal gyrus. We finally see more scattered high POVE values, which happen even later in time, such as in the right temporal pole and entorhinal region, peaking between 525 and 650 ms, and even later at subregions of the frontal cortex (starting at 600 ms).

While the results from model B offer a bit more detail compared to what we had seen before, most of the activity we can explain with this model happens between 200 and 500 ms after stimulus onset, as expected based on the results shown in Chapter 2 using model A (specifically, Figures 2.8 and 2.9). Also, the location of these sources evolve from posterior to more anterior positions over time, and it seems like different regions over time are being best predicted across both hemispheres.

Because of the versatility of model B, we actually have predictions for each source location in the brain over time. So, we can conduct this analysis by averaging over ROIs and then over subjects (as in Figure 3.6), or take a different look at the results by first morphing each subject's brain to a canonical brain, and then averaging the results over subjects (Figure 3.7). The figure shows the three time points (275, 425, and 630 ms) where the POVE was highest when combined across subjects. Each subject-specific spatiotemporal POVE map was first multiplied by the subject's 2-vs-2 decoding accuracy, and then morphed to the common brain. Then, the set of morphed POVE values was averaged over subjects.

We can see that several sources in both hemispheres are best explained by our model in the first time point, specifically along the left inferior parietal and posterior middle temporal cortices, as well as in the right lateral occipital cortex. However, at 425 ms most of the sources that are best predicted by the model are in the left hemisphere, more specifically in the posterior part of the superior and middle temporal cortices, the banks of the superior temporal sulci, and the inferior parietal cortex. The few sources predicted at 425 ms in the right hemisphere are focused along the ventral part of the posterior temporal cortex. Finally, at 630 ms the majority of predicted sources are again in the left hemisphere, following the shift towards more anterior regions that began at 425 ms, and now focusing on sources in the left anterior temporal lobe, along with sources in the ventral temporal region.

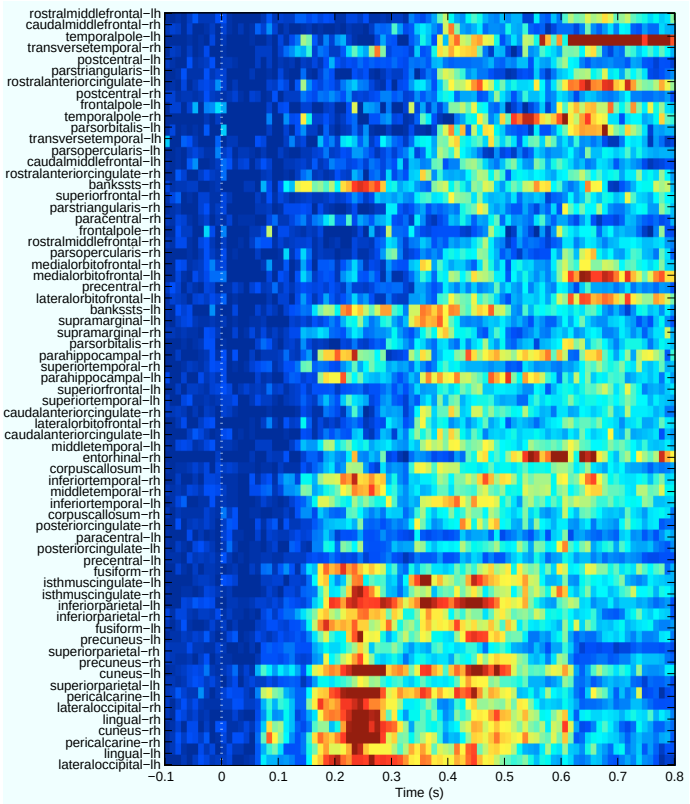


Figure 3.6: Percent of variance explained (POVE) in different regions of the brain over time in the *answer-questions* paradigm. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is the same as in Figure 2.10. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

3.5 How much does each feature contribute to predicting the spatiotemporal MEG activity?

While it is important to check what spatiotemporal MEG activity we can best predict with our model, it is also interesting to understand how much each stimulus feature contributes to predicting a particular space-time point of MEG activity, in comparison to the other features in the model. The main issue with this analysis relates to the question of how to compare the weights of different brain activations. Suppose we are predicting the MEG activity in only two brain regions, as described in equation 3.7:

$$\begin{bmatrix} M \\ P \end{bmatrix} = \begin{bmatrix} w_{hM} & w_{sM} \\ w_{hP} & w_{sP} \end{bmatrix} \begin{bmatrix} h \\ s \end{bmatrix} \tag{3.7}$$

To illustrate, say M represents a source of MEG activity in motor cortex at some time window, and P a source in the parahippocampal region. Because there is no normalization of the MEG data, and MEG has better sensitivity to M than the deeper source P, the SNR in M is bigger. Then, we employ a model similar to model B, but using only 2 semantic features *can you hold it?* and *is it a shelter?* (h and s in equation 3.7) to predict the activity in the two areas.

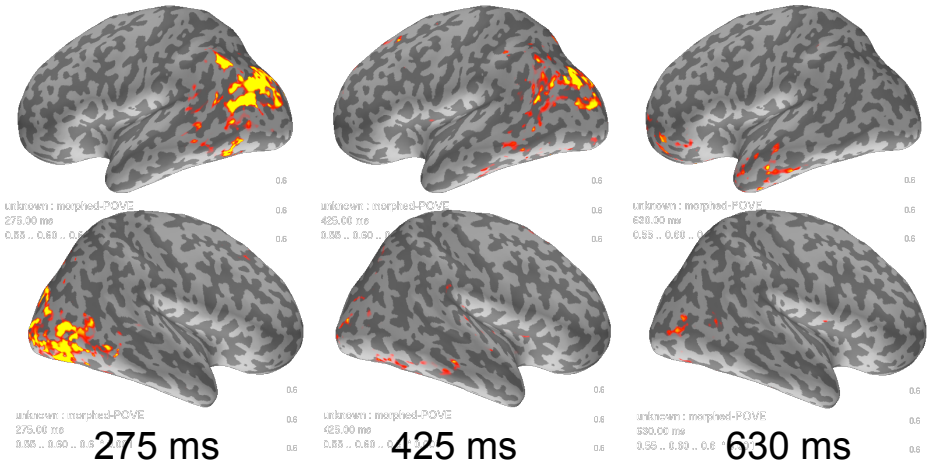


Figure 3.7: Average POVE over subjects of individual source locations after morphing to a canonical brain, at pre-selected time points. Top and bottom figures correspond to lateral views of the left and right hemisphere, respectively.

Suppose that h is twice more important than s to predict M, so w_{hM} (the weight from h to M) is 100, and w_{sM} is 50. Conversely, to predict P, w_{sP} is 1 and w_{hP} is .5, because s has twice more weight than h at predicting P, but the signal amplitude in P is smaller than in M, and therefore the weights to P are in general smaller. If we plot the weights for the semantic feature s, we will see a weight 50 times bigger to M than to P, which blurs some of the conclusions one can make about the influence of such features on brain activity.

Here, we overcome this limitation by normalizing the weights as follows. First, we divide the absolute value of all the weights to a given brain attribute (i.e. MEG activity at space-time) by the maximum over all weights to it. In the example above, w_{hM} becomes 1 and w_{sM} turns into .5, while w_{hP} stays .5 and w_{sP} also keeps its value of 1. This assures that the relationship we initially intended to portrait in the example is clearly observable (namely, h is twice more important to predict M than s is, and vice-versa for predicting P). Finally, we multiply these weights by the POVE at the specific spatiotemporal location to mark down any weights to sources we cannot accurately predict. This method produces scaled weights for each source and time point. Because of the scaling applied, these weights can be compared over time and space, even though they were predicted independently, and also compared among different features, since the scaling applied to individual features (POVE) is the same.

The plots in Figure 3.8 show how much each feature contributes to predicting the MEG activity in a ROI (again, sources were averaged within ROIs to better compare to previous results). The first observation when looking at those results is that the same peaks seen in the POVE results are also present in these pictures (namely, early peak at 250 ms and later, less time-locked, at 450 ms). However, the two perceptual features show a weaker later peak than most semantic features, especially when compared to the first peak of each individual plot. For example, the later peak in the weight distribution of *word length* is almost negligible compared to its earlier peak, and this later peak is also smaller than the second peak of semantic features such as *is it alive?* and *is it bigger than a microwave oven?*. Similarly to what was seen in Figure 2.9, se-

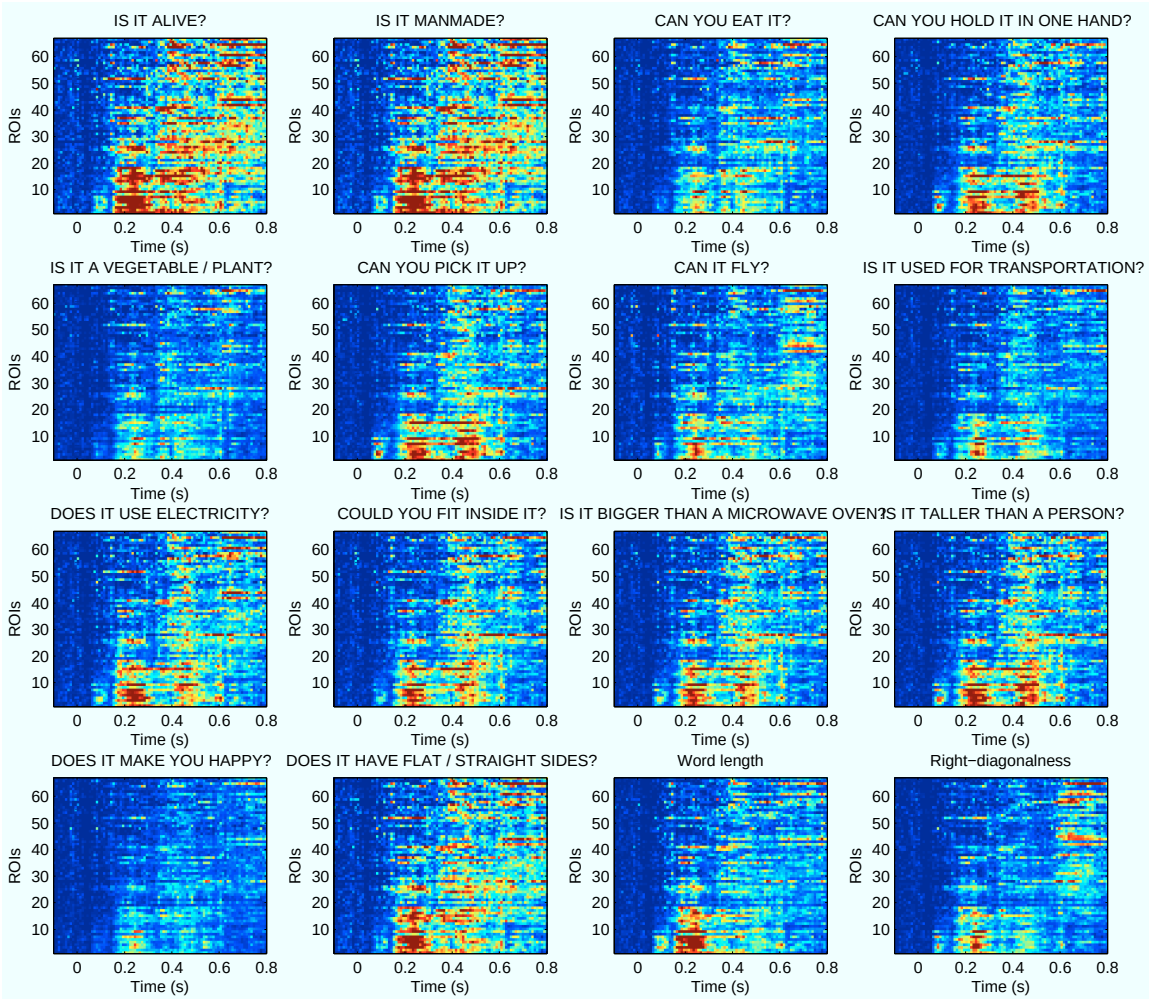


Figure 3.8: Weights for different perceptual and semantic features averaged over ROIs and subjects. Weights were first scaled to allow for comparing across regions and time, and then multiplied by the POVE in each spatiotemporal attribute. All plots share the same color scale, and the Y axis is sorted to match Figure 3.6. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

mantic features show higher weights in later time points than perceptual features, recruiting more anterior ROIs in the process. Also, highly correlated features still show similar spatiotemporal signatures (e.g. *can you hold it in one hand?* and *can you pick it up?*). We can also observe the gradual process of increasing weights over time for semantic features, which we conjecture to represent the recruitment of different ROIs that encode different features of the stimuli. Finally, while the lines among time windows are a bit blurrier here (which is, in a way, a strength of the current method - not being tied down to a pre-defined window size), there are still differences among the spatiotemporal profiles of semantic features. For example, the second peak of the animacy-related features starts earlier than for the other ones. So early, in fact, that it starts to blend with the earlier peak, making it harder to distinguish whether there are one or two peaks for *is it alive?*. The early decodability of animacy-related features has also been reported in Chapter 2 (specifically, Figure 2.9), which leads to speculation about a sequence of processing among semantic features, possibly guided by evolutionary constraints.

3.6 Discussion

This chapter presented a model (model B) that predicts MEG activity using a set of perceptual and semantic features. The results shown are in line with what has been previously observed in chapter 2, especially when feature selection is used. This model provides several advantages over the previous one when it comes to studying the MEG signal. It is more versatile on studying different brain regions, and also on analyzing different time windows. More specifically, while we had to train a new model every time we wanted to investigate a specific brain region or time window with model A (chapter 2), model B allows for predicting the MEG activity in all regions at once, and then only evaluate the space and time regions in which we are interested, combining them in any way that we find most informative for the analysis (e.g. combine sources in a ROI, morph the whole brain, etc). Finally, it allows for asking questions not previously asked with model A, such as what parts of the MEG signal are best explained by the model, and what is the contribution of each feature used in the model in explaining the MEG activity.

On the other hand, model A is more robust to the noise in MEG data, especially in situations when feature selection is not effective in discarding the less informative features in the output of model B. Additionally, model A might be better-suited for studying what information is encoded in specific parts of the MEG signal, because we can train the model with different ROIs and time windows, and predict individual perceptual and semantic features independently of each other. In sum, the two models provide different ways to look at the same data, and the selection of the model to be used is highly dependent on the questions the researcher wants to answer.

Combining the results seen here and also in the previous chapter, we see that the feature representation is coded in the MEG signal in a distributed fashion. This code is distributed not only over time, where perceptual features are encoded earlier than semantic features, but also in space, where more anterior regions are involved in encoding semantic features and posterior regions focusing on perceptual and also semantic features. Based on those results, we conjecture that the same region can encode several features, especially when combined with the activity of other regions. The results also suggest that the same feature can be encoded by several different regions over time, but we do not have enough evidence to make claims about whether the information n encoded in these different regions is complementary or redundant. Finally, our model supports a compositional code in the sense that we can learn where and when the properties of different nouns are encoded in the MEG signal, and then predict the activity for nouns we have never trained on by using the perceptual and semantic properties of these novel nouns. In other words, because we can learn the MEG activity related to properties of concrete nouns and use that mapping to predict the signal for novel nouns, one can infer that the individual representation of concrete nouns is not encoded independently in the MEG signal (e.g. not an extrapolation of the grandmother cell idea), but we speculate that the meaning of a noun is formed by putting together several basic properties that characterize the concept. Considering that we are predicting a signal that is correlated to what the brain is doing, our findings support the ideas of a distributed and compositional code for knowledge representation in the brain.

Chapter 4

Predicting MEG activity associated with the meanings of nouns when freely thinking about their properties

We have demonstrated in chapters 2 and 3 that it is possible to track the flow of perceptual and semantic features in the MEG signal, and observed that perceptual attributes were decoded early in the MEG response, while semantic attributes were decoded later. While the results were encouraging, the task required subjects to explicitly answer questions about the nouns, which lead us to investigate whether we could observe similar results when the subjects performed different tasks. One step towards that direction is taken in the current chapter.

Here, we first run model B (chapter 3) using a different paradigm (*freely-think*), and then compare those results with the ones obtained with *answer-questions*.

4.1 Results

Both *answer-questions* and *freely-think* paradigms are fully described in section 1.2, but we remind the reader that the main difference between the paradigms is that while in *answer-questions* the subjects performed a semantic task in each trial (answering one of the 20 questions with respect to the noun), in *freely-think* the subjects openly thought about the properties of the noun. The same subjects were scanned in both paradigms. Additionally, *answer-questions* has more repetitions of a single noun than *freely-think* (20 vs. 6-10, respectively).

4.1.1 Can we discriminate between two novel nouns using predicted MEG activity from *freely-think*?

Although we have shown that is possible to discriminate between two novel nouns using *answer-questions* data, it is interesting to check whether we can perform the same analysis using data from a paradigm with less repetitions and a more open-ended task performed by subjects. So, let's analyze the 2-vs-2 decoding results for *freely-think* and how they are affected by only using stable MEG features for scoring (Table 4.1).

Scoring features	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
MEG data	80.5	88.1	77.8	81.4	86.2	75.4	78.9	53.4	81.5	**76.4**
Best stable	85.0	87.9	78.6	89.9	90.8	86.3	83.6	69.9	87.2	**83.5**
5K stable	82.5	85.0	75.8	87.9	90.7	86.2	82.7	66.2	86.2	**81.6**

Table 4.1: Comparison of accuracies for the *leave-two-out* experiment using *freely-think*. First row shows the results using all MEG data features for scoring, and the last row in the table shows the accuracy when using only the top 5,000 most stable voxels for scoring. Row 2 (Best stable) is displayed for comparison, and the accuracies there show the top accuracy obtained for each subject (i.e. peak of blue traces in Figure 4.1).

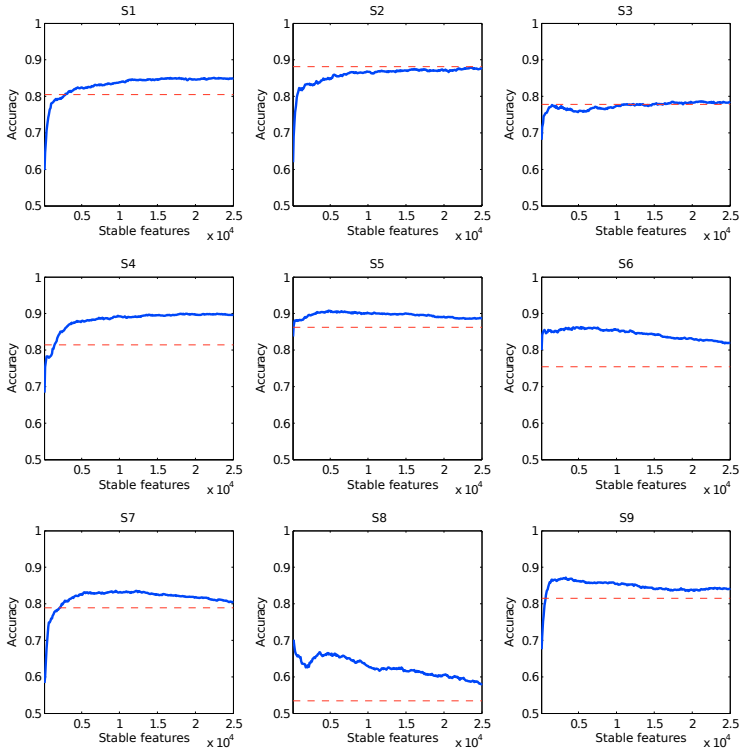


Figure 4.1: Effects of number of stable features used for scoring on the 2-vs-2 accuracy in the *freely-think* paradigm. One plot per subject. In each plot, all 55,000 features are sorted in descending order based on their feature score, and the accuracy is computed using only a few of the top-most features. Plots also show the accuracy obtained using all features (red dashed line). If the X-axis were extended to 55,000, blue and red traces would meet.

Although the accuracies are not as high as the ones obtained for *answer-questions* (Table 3.2), they are mostly significant at $p < .01$, except for subject S8. This is also the subject who most benefitted from feature selection. Curiously, subject S2 is the only one who gets better accuracy by using all features at the same time. Although almost all subjects had their results improved when the score was calculated only using the top stable features, this number of stable features that worked best for most subjects was smaller than what was seen for *answer-questions* (Figure 3.3). There is actually a good amount of variation among the subjects regarding the number of top stable features that yields best decoding results (Figure 4.1). Still, we use the heuristics that the peak number of top stable features across subjects happens at about 5,000 features.

Another point worth mentioning is that, compared to what was seen for *answer-questions*, the best accuracies over time are sustained for shorter time periods (Figure 4.2). In other words, while long red horizontal bands are seen in Figure 3.4, indicating that high 2-vs-2 accuracies were obtained in several consecutive (but independently scored) time points, that happened less

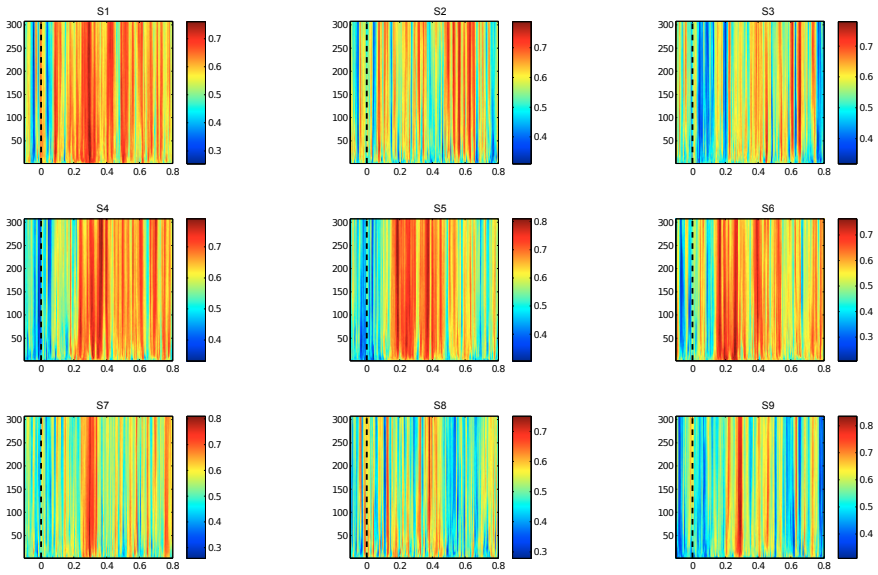


Figure 4.2: Effects of number of stable features used for scoring on the 2-vs-2 accuracy in the *freely-think* paradigm, evaluated for each time point. One plot per subject. In each plot, the 306 features (one per MEG sensor) in one time point (X-axis) are sorted in descending order based on their feature score, and the accuracy is computed using only a few of the top-most features (Y-axis). The colors in the plot show the accuracy. For example, the top most line of each plot (Y = 306) shows the accuracies over time using all MEG sensor predictions.

frequently in *freely-think*, where the high accuracy bursts were shorter and more intermittent. Those plots also show that less stable sensors within a time window were required in *answer-questions* to reach the maximal accuracy, while in *freely-think* more stable features were required to reach the highest decoding accuracy (i.e. vertical bands reach their hottest possible color at bigger Y values). The time window between 250 and 450 ms still required less sensors in the helmet to reach maximal accuracy than other time windows, similarly to what was seen for *answer-questions*.

Finally, we note that the spatiotemporal signature of the most stable features is not as well-defined as it was for *answer-questions* (Figure 4.3). For the subjects that showed a more well-defined pattern of stable features (e.g. subjects S1 and S4), the spatiotemporal distribution of these features clearly resembled the ones in *answer-questions* in that they were located in occipito-parietal sensors and evolved from 150 to about 400 ms. However, not all subjects showed a clear pattern of stable features. These subjects were also the ones whose 2-vs-2 accuracy results did not show any improvement by scoring using only stable features.

4.2 What are the brain regions and time points in *freely-think* best accounted for by our model?

Following the same sequence of questions analyzed with data from *answer-questions*, we then need to check what parts of the MEG signal from the *freely-think* experiment were best predicted by model B (Figure 4.4). Similarly to what was seen with *answer-questions*, the signal in posterior regions such as the bilateral cuneus and lingual gyri, around 250 ms, were among the spatiotemporal brain features best explained by the model. However, more anterior regions

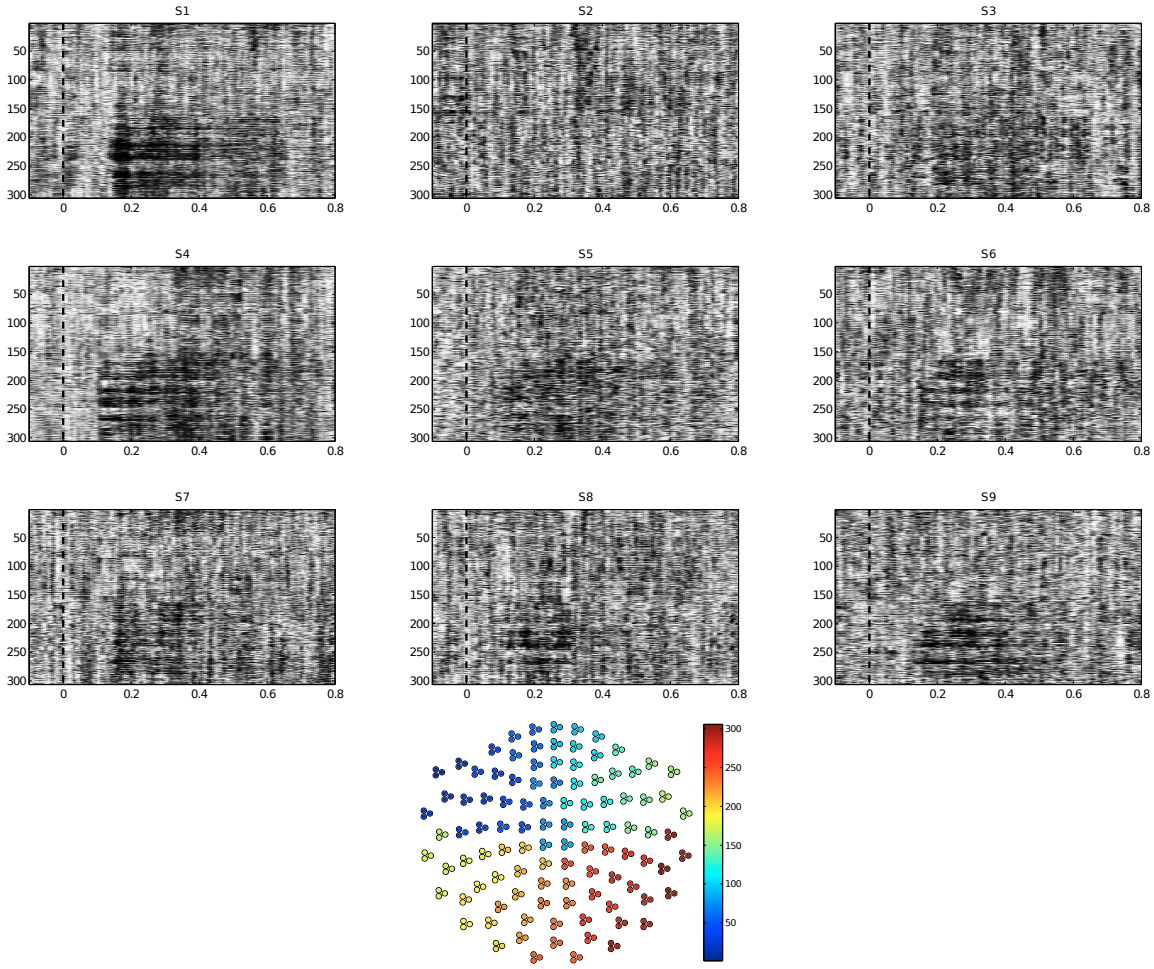


Figure 4.3: Ranked stability scores (see Chapter 3 for description) for all spatiotemporal features in the MEG helmet. The more stable the feature, the darker the shade of gray in each plot. One plot per subject. Time is in the X axis, and sensor position in Y axis. Guide to sensor positions in the MEG helmet (bottom plot) has the following orientation: nose is up, left ear on the left, right ear on the right. The color of each sensor indicates the sensor number, with blue starting at sensor 1.

such as the left superior temporal sulcus and the right caudal middle frontal gyrus also showed heightened POVE values. The results after that time period were very much scattered over the brain, not repeating the more well-defined pattern seen with *answer-questions* (Figure 3.6).

4.3 How much does each feature contribute to predicting the spatiotemporal MEG activity in *freely-think*?

The last results of this section identify what semantic and perceptual features were most important at predicting different parts of the MEG signal in the *freely-think* paradigm (Figure 4.5). The double peak pattern previously seen in *answer-questions* was not as clear, if at all present, in the *freely-think* dataset. Most of the semantic features do show higher weights than the perceptual features in time windows after 400 ms, but the difference is certainly not as clear as before. There

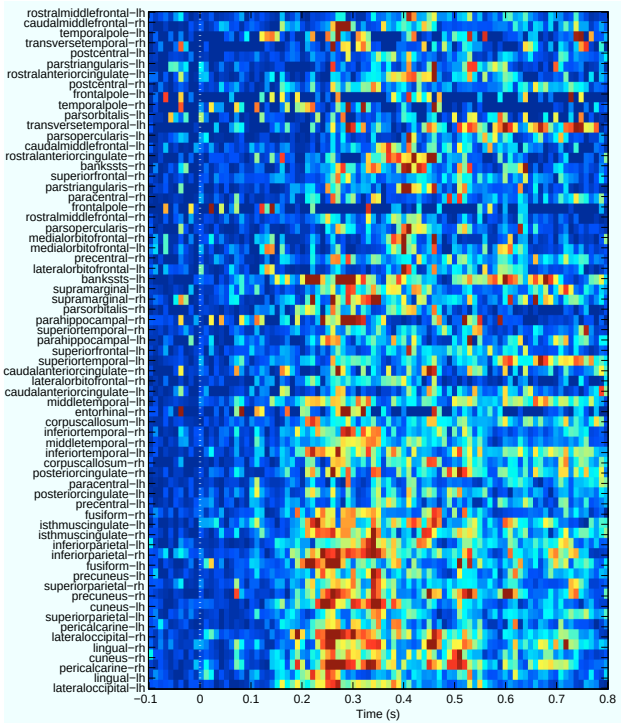


Figure 4.4: POVE in different regions of the brain over time in the *freely-think* paradigm. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is the same as in Figure 2.10. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

are also higher weights in anterior regions for semantic features around 400 ms, but it is harder to infer any type of sequence within semantic features.

4.4 Discussion

In this chapter model B was employed to predict MEG data for the *freely-think* paradigm using the perceptual and semantics feature set as its input. Together with the results from Chapter 3, several comparisons can be made between the two paradigms analyzed so far.

4.4.1 Task and repetitions effects

Subjects that had best results in *answer-questions* were not necessarily the same ones with best results in *freely-think*. Subject fatigue might have played a role there, since *freely-think* was scanned on the same session, but after *answer-questions*. It is also possible that these subjects had more difficulty keeping the same representation for the different nouns across repetitions, which is an implicit assumption when we average over repetitions of the same word.

Overall, 2-vs-2 decoding results were better with *answer-questions* than with *freely-think*. We conjecture that while the more well-defined task in *answer-questions* might have played an important role in the results, it is more likely that the number of repetitions in the two paradigms is the biggest factor behind the differences (this conjecture is further analyzed in chapter 6). More

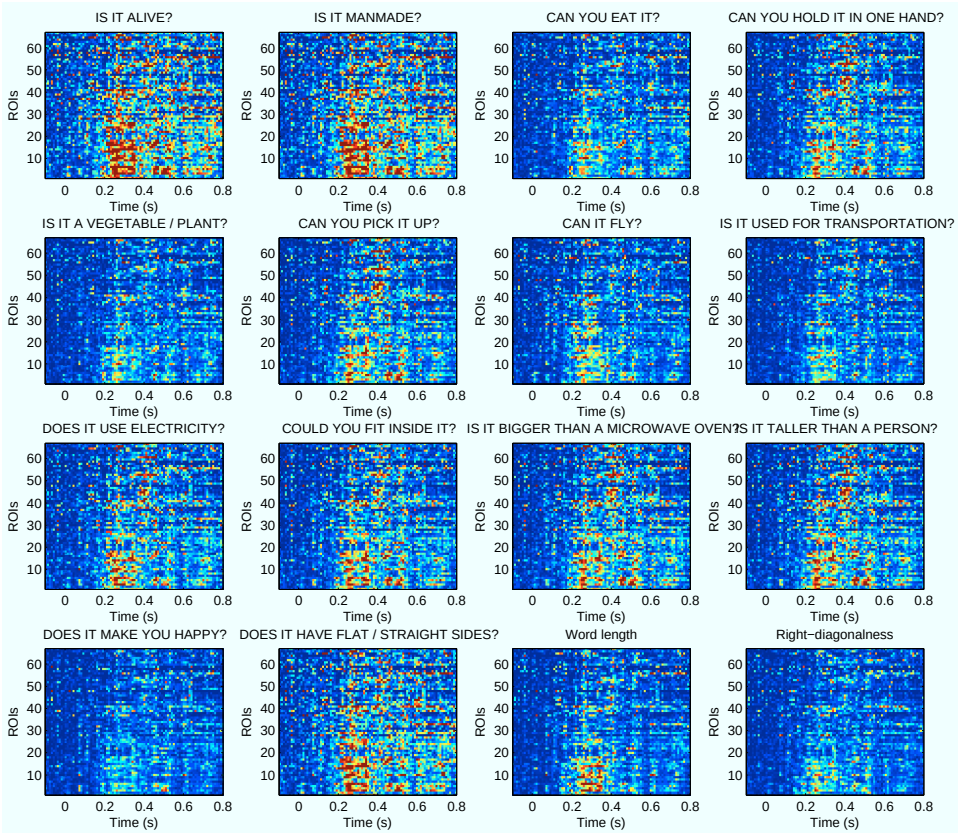


Figure 4.5: Weights for different intermediate features averaged over ROIs and subjects in the *freely-think* paradigm. Weights were first scaled to allow for comparing across regions and time, and then multiplied by the POVE in each spatiotemporal attribute. All plots share the same color scale, and Y axis remains constant across plots. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

repetitions increase SNR, and there were at least twice as many repetitions in *answer-questions*. Furthermore, the *freely-think* subjects with smaller number of repetitions also had worse results than the other subjects.

The percentage of variance in the MEG activity explained by our model, which was higher in *answer-questions* than in *freely-think*, might also be affected by the difference in the number of repetitions between the two paradigms. An analysis of the effects of the number of repetitions in these results will be conducted in Chapter 6.

When comparing the weights of different semantic and perceptual features across the two tasks, a few differences are clear. First, the two peaks observed for semantic features in the *answer-questions* results are not as obvious in *freely-think*. The overall results in *freely-think* also appear to occur later in time. From the onset of the most stable features (Figures 3.5 and 4.3), to the early peak of the percent of signal explained by the model (Figures 3.6 and 4.4), and the contribution of each feature in explaining the signal (Figures 3.8 and 4.5), we can observe a 50 to 100 ms offset in time. We suspect that this might align with the subject's level of engagement in the task, since the *answer-questions* paradigm is reportedly more engaging than *freely-think*.

In sum, we can use the predicted MEG activity for both tasks to distinguish between two novel nouns with highly significant accuracies. That, combined with the POVE results, show that

our perceptual and semantic feature set can capture crucial information related to the meaning of concrete nouns in the MEG signal. These results only use MEG data up to 800 ms after stimulus onset, although the subjects had up to 2 s to think about the properties of objects in *freely-think*. The choice of limiting the timeline at 800 ms comes from the goal of better comparing the results to the ones obtained in *answer-questions*. In that paradigm, analyzing data after 800 ms might obscure some of the results because we would be approaching the average time in which subjects performed a button press. We also assume that such smaller time window ensures that we are using only the more automatic signals related to the recognition of the nouns, instead of the more variable (both in time and across repetitions) process of actively thinking about properties of the objects.

4.4.2 Remarks on the neural encoding of perceptual and semantic features

While some of these results corroborate what was found before with fMRI data, they are not all in agreement. For example, there has been much discussion in the literature about the dorsal and ventral visual pathways [91]. In short, the dorsal pathway corresponds to the left middle temporal gyrus, posterior parietal cortex, and inferior parietal lobe, and reflects object processing, spatial analysis ("where") and object-directed actions, while the ventral pathway encompasses regions such as medial temporal cortex, the fusiform and the parahippocampal gyrus, and is related to object recognition ("what" the object is) [53]. The results presented here are well in line with this framework. The regions considered to be part of both streams were among the locations best predicted by our model over time, and most of the semantic features in both paradigms evoked larger weights in regions considered to be in the ventral stream. In particular, features related to shelter showed high weights in the parahippocampal region (which also agrees with [38]), and animacy features showed high weights in the fusiform gyrus. However, these were not the regions with highest weights for these features. Also in agreement with Just et al. [38], post- and pre-central gyri showed high weights for manipulable features. Nevertheless, unlike the fMRI results reported by them, we did not find evidence for edibility features being encoded in pars-triangularis or other reported regions. To that extent, the results with fMRI and MEG data are quite complementary. We conjecture that the methodological choices in the analyses, combined with the evident difference between the signal both modalities measure, are acceptable explanations for the differences in the results.

Both in *answer-questions* and *freely-think*, the second wave of high weights (to the extent that it can be observed in *freely-think*) occurred earlier for the animacy-related features than for other semantic features. This again suggests a recruitment order among semantic features. Another commonality across paradigms was that the perceptual features did not show a second peak of weight magnitude, or at least not as high as when compared to the semantic features. This result indicates that, regardless of the task being performed, perceptual features are not as important in predicting MEG activity as semantic features after about 350 ms.

4.4.3 Future directions

The work shown in this chapter will benefit from future analysis to more quantitatively compare the results of the two paradigms. For example, feature selection (using feature stability or any

other measure) could be done across paradigms, in order to use the top features of one paradigm to score the other, or even a mix of the two. A similar strategy could be done using regions previously seen in fMRI to be related to the semantic features we are studying here, and only use the brain signal from those regions for scoring.

A few peculiarities of the *freely-think* paradigm can also be further explored. For example, what happens to the 2-vs-2 accuracies after 800 ms? How are the semantic features represented then? It will also be interesting to check if the less clear divide between perceptual and semantic features, as well as the lower percentage of variance explained, are mostly an effect of the small number of repetitions in *freely-think*, or if the more open-ended nature of the paradigm is partly responsible for that.

Chapter 5

Predicting MEG activity without stimulus repetition

We have presented so far an effective method for studying the temporal sequence and cortical locations of perceptual and semantic features encoded by observed MEG neural activity, while subjects perceive concrete objects stimuli. The experiments showed intriguing results (albeit limited by the choice of experimental paradigms), and thus encourage further, more carefully controlled and cognitively relevant studies. In this chapter we again use model B to study the representation of concrete nouns in the neural code, but now we predict data from two paradigms better-suited to be studied using MEG: *1back-text* and *1back-speech*.

We refer the reader to section 1.2 for an in-depth description of the paradigms. The main characteristic of the two paradigms is that the words are presented only once to the subjects. While that allows for a richer set of stimuli to be used, it presents new signal processing challenges related to low signal-to-noise ration (SNR) of a single trial in MEG.

5.1 Challenges of single trial analysis

The idea of averaging multiple repetitions of the same condition in order to increase SNR has been crucial in most neuroimaging modalities. This approach results in no further concerns when the processes being investigated are mostly automatic in nature (e.g. somatosensory stimulation [26], basic auditory processing [41]), and are mostly time-locked to the stimuli, because then the assumption that the same activation occurs during all repetitions is hardly violated. However, the more cognitive-demanding the paradigm becomes, the less likely it is that the same brain processes will be repeated over time in every trial. Finally, there is also a concern about repetition effects, which have been shown to dampen the neural signal [18].

Another way to evaluate the importance of the number of repetitions of a stimulus is to plot how the 2-vs-2 decoding results shown in chapter 2 vary with the number of repetitions averaged (Figure 5.1). It is clear that the more repetitions averaged, the better the decoding accuracy. The number of repetitions needed to obtain the best result varies by subject (i.e. the traces asymptote at different points), which can be a function of several variables, including how noisy the machine and the environment were during the scan, the quality of the pre-processing performed with each

subject, and how motivated a subject actually was during the scan. But another important point to make is that statistically significant results can be obtained at least with some of the subjects, which encouraged us to go on with analyzing *1back-text* and *1back-speech*.

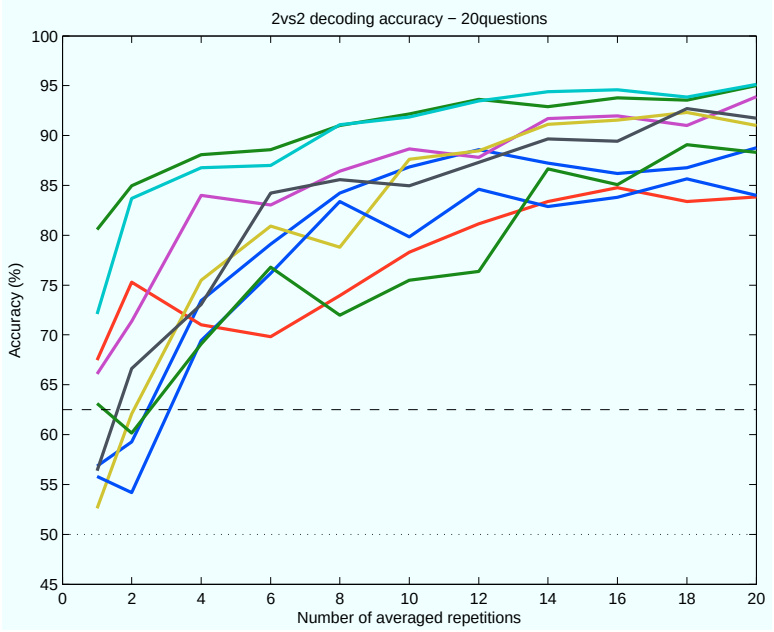


Figure 5.1: Effect of number of repetitions averaged (X axis) in distinguishing between two novel nouns only using predicted semantic features from MEG data (*answer-questions* paradigm). One trace per subject. Dotted line shows chance accuracy, and dashed line shows the statistically significant results at $p < 0.01$. The end result of each trace matches what was shown in chapter 2.

In summary, the two paradigms analyzed in this chapter play with an important trade-off: while it is advantageous to present a richer set of stimuli (500 to 1000 different words), new methods need to be employed to extract as much information as possible from single trials.

5.2 Concatenating trials

In principle, it would be possible to use model B without modifications to analyze data from *1back-text* and *1back-speech*. After all, the method predicts one brain image (i.e. MEG signal over space and time) per word, so in theory it should be possible to predict the words used in the two paradigms analyzed in this chapter. The difficulty comes from the realization that the brain image predicted for *answer-questions* and *freely-think* comes from an average of several repetitions of the noun, which increases the SNR of each brain image. This alternative is not available in *1back-text* and *1back-speech*, in which case the model would have to predict very noisy images.

The main technique we will use to classify single trials using the same 2-vs-2 framework from before is concatenation. The rationale behind the method comes from the realization that we do not suffer from data scarcity during training, but during testing instead. In other words, even though we might learn a very good model because of the diversity of stimuli used for training, testing the predictions against the two noisy single trial brain images will likely be troublesome.

By concatenating the single trials we still compare two true brain signals against two predicted brain signals during testing, but now the brain signals come from the presentation of multiple nouns to the subjects. Still, none of the words used for testing are used for training.

An analogy might be the best way to understand how the concatenation tool helps here. Suppose your task is to tell between two very blurry pictures, which one is the cat and which one is the dog. In this analogy, the blurriness of the pictures corresponds to the MEG signal noise and, in the single trial case, it is quite high. For the example, say the blur is so high that it is very hard to say which picture is the dog and which one is the cat. Now, we add another object to each picture: a carrot and a screwdriver. They are also equally blurry, but now your task is to decide which picture has a cat and a carrot, and which one has the dog and the screwdriver. It is easy to see how adding more objects to each picture, despite of how blurry they are, makes it easier to tell which picture is which. Of course, if the amount of noise is so high that one cannot infer any information in the picture of any of the objects, then no matter how many objects get concatenated, distinguishing between the two pictures will not be possible.

Another way we can concatenate words is across subjects. Namely, instead of concatenating different words presented to one subject to form the two groups of words to be tested, we concatenate the brain activity associated with the presentation of the same word across subjects. In the analogy above, this would be the same as adding several pictures of dogs together, and then several pictures of cats together. However, the cameras that took the pictures had different amounts of distortions (i.e. different amounts of noise in the MEG data), and there were different types of cats and dogs as well (i.e. representing the innate variability of the representation of dogs and cats across subjects).

It is important to note that while the 2-vs-2 task from before asked the question of whether we can distinguish between two novel words, now the question becomes whether we can distinguish between two novel groups of words. The compositional assumptions, as well as the validity of how we test the generalization capabilities of the model, remain the same.

5.3 Can we discriminate between novel groups of words never used for training using data without repetitions?

We have seen in Chapters 3 and 4 that the perceptual and semantics features employed in model B provide a good basis set that we can use to predict the MEG activity associated with the processing of word and picture stimuli representing concrete nouns. By analyzing the contributions of the different features to the predicted signals, we can also have further insights on how the neural activity can be decomposed into patterns associated with the different perceptual and semantic features.

We start the results in this section by investigating whether model B can predict the MEG activity of single words accurately enough that the predicted signal can be used to distinguish between groups of words. Because now we compare groups of words instead of single words in the second phase of the SOCR, a few modifications need to be made to the cross-validation framework.

When using data of paradigms with only 60 words (*freely-think* and *answer-questions*), we

tested the model with every possible pair of two left-out words (1770 cross-validation folds), while training the model with the remaining 58 in each fold. Now, even if we were predicting only two words (i.e. two groups of words, with only one word in each group), the problem becomes more computationally challenging, since we would be conceivably calculating every possible pair of two left-out nouns out of 500 (or even 1000!). The situation becomes even tougher when we compare groups of words with more than 1 word in each, because the order in which the words in the group are chosen matters. For example, a group with words *cat - dog - cow* is different than a group with the words *dog - cow - cat*, even though these two groups would never be compared against each other because a word can only be part of one group. So, an even bigger number of possible two groups can exist out of the 500 or 1000 words.

In order to overcome this problem, for each subject we decide on a smaller number of cross-validation folds, choosing the words in each group at random. From that, we estimate an accuracy, and then repeat the process many times in order to estimate a distribution of accuracies. For instance, in *1back-text* there were 100 folds in each cross-validation set, and in each fold 200 words were held out (i.e. 100 words in each group). The number of cross validation folds within a set was the same for *1back-speech*, but only 100 words were held-out (i.e. 50 words in each group) in each fold. The numbers on table 5.1 show the mean $\pm$ the standard deviation of a distribution of 150 cross-validation sets (i.e. distributions with 150 accuracy data points). The pseudocode for this procedure is described in Algorithm 1.

Algorithm 1 Calculating 2-vs-2 accuracy for a single subject using groups of words

for set = 1 to 150 **do**
  for fold = 1 to 100 **do**
    $accuracy_{fold} \leftarrow 0$
    $testGroup_1 \leftarrow$ set of K randomly held out words
    $testGroup_2 \leftarrow$ set of K different randomly held out words
    $trainWords \leftarrow (allWords - group_1 - group_2)$
    $trainData \leftarrow$ perceptual and semantic features for $trainWords$
    $trainTargets \leftarrow$ MEG images for $trainWords$
    $testDataGroup_1 \leftarrow$ perceptual and semantic features for $group_1$
    $testDataGroup_2 \leftarrow$ perceptual and semantic features for $group_2$
    $testTargetsGroup_1 \leftarrow$ concatenated MEG image for $group_1$
    $testTargetsGroup_2 \leftarrow$ concatenated MEG image for $group_2$
    $W \leftarrow$ regularizedLinearRegression(trainData,trainTargets)
    $predictedTargetsGroup_1 \leftarrow W * testDataGroup_1$
    $predictedTargetsGroup_2 \leftarrow W * testDataGroup_2$
    $distance_{1,1} \leftarrow$ cosineDistance($predictedTargetsGroup_1$,$testTargetsGroup_1$)
    $distance_{2,2} \leftarrow$ cosineDistance($predictedTargetsGroup_2$,$testTargetsGroup_2$)
    $distance_{1,2} \leftarrow$ cosineDistance($predictedTargetsGroup_1$,$testTargetsGroup_2$)
    $distance_{2,1} \leftarrow$ cosineDistance($predictedTargetsGroup_2$,$testTargetsGroup_1$)
    if $(distance_{1,1} + distance_{2,2}) < (distance_{1,2} + distance_{2,1})$ **then**
      $correct \leftarrow 1$
    else
      $correct \leftarrow 0$
    end if
    $accuracy_{fold} \leftarrow accuracy_{fold} + correct$
  end for
  $accuracies(set) \leftarrow \frac{accuracy_{fold}}{100}$
end for
$accuracy_{subject} \leftarrow mean(accuracies) \pm std(accuracies)$

Note that the number of words in each group is another variable we could potentially tweak. As discussed in the previous section, if there is even a small amount of signal we can predict in

each single word, the more words we concatenate the better the accuracy we will get (given that there is still enough data left for training the model). However, if there is no signal that we can predict, the number of concatenated words does not affect the results that much.

Because of the complexities of this testing framework, it is hard to find a parametric approach to adopt in order to draw confidence intervals for the results (chance accuracy is still 50%). To that end, we ran hundreds of permutations of the framework described in Algorithm 1 for each subject, where in each permutation the labels of the trials were shuffled. We then calculated how often the accuracies with the shuffled labels were higher than the ones with the original data, and that ratio is reported as the empirical p-value for the accuracies.

	S1	S2	S3	S4	S6	S9	SX
1back-text	69.81±4.76	**97.13±1.74**	78.97±3.80	64.47±4.63	**91.70±2.59**	76.39±4.04	64.16±4.41
1back-speech	39.54±5.05	43.10±5.55	55.47±4.79	63.99±4.82	71.20±4.51	57.76±4.78	77.53±3.56

Table 5.1: Accuracies for the 2-vs-2 decoding task using perceptual and semantic features to predict all time points from -.1 to .9 s and 306 MEG sensors in the two single trial experiments. The classifiers are able to significantly distinguish between two different groups of concrete nouns that the MEG-based classifier has never seen before for a few of the subjects. SX is a subject not scanned for the two previous paradigms. Chance accuracy was 50.0%. Accuracies significant at $p < 0.01$ are marked in bold.

It is clear from Table 5.1 that distinguishing between two groups of words is not possible for many of the subjects in either paradigm. Still, the results for subjects such as S2 and S6 in *1back-text* are good enough to encourage us to later explore what parts of the MEG signal our model best explains.

But before we ask that question, we should investigate whether concatenating trials across subjects helps the decoding results. For this set of results we do not distinguish between groups of different words anymore. Instead, each group is formed by the same word across all subjects, for a total of 7 MEG trials (1 word) in each group. This could be seen as an extension of the 2-vs-2 decoding task used in chapters 3 and 4, but now we include the brain image of a word for all subjects, instead of a single subject. Therefore, we could potentially calculate all possible combinations of 2 left-out words, but we again would face some computational difficulties. The approach we take in the *1back-text* paradigm is to select a set of 500 pairs of words, and a different pair is to be held-out from training in each cross-validation fold. Here, a given word is only present once in all pairs, such that if one analyzes all possible held-out sets across folds, that word only appears once. Hence, because we never test on the same word twice, we can use the binomial distribution to assess the significance of our results.

More specifically, let the random variable X represent the number of folds for which a correct response was made (out of 500). Under the null hypothesis that in each fold we guess among the two predicted brain images for the sets of words, X has a binomial distribution with parameters n=500 and p=1/2. For example, say we get a 56% accuracy, which corresponds to 280 folds out of 500. How likely would we be to get that result under the null hypothesis? That is similar to looking for the p-value of $P(X \geq 280)$ assuming p = 1/2, which can be found by summing the binomial probabilities for 280, 281, etc, up to 500 successes, which is the same as $1\text{-}P(X < 279)$, or 1 minus the cumulative distribution function with parameters 500 and 1/2, at the value 279. For *1back-speech*, the same testing framework was employed, but there were only 250 pairs of words.

Table 5.2 shows the results of concatenating the data for all subjects in both paradigms, as well as their individual results (i.e. performing a single-trial 2-vs-2 decoding task similar to

	S1	S2	S3	S4	S6	S9	SX	Combined
set 1	49.2	49.4	49.4	54.8	56.8	56.0	50.4	59.8
set 2	48.8	52.0	52.8	52.0	55.2	53.4	53.2	56.6
set 3	49.6	52.8	52.8	49.4	51.4	58.8	53.4	55.8
set 4	52.8	51.6	51.4	54.6	58.8	56.4	52.6	58.6
set 5	51.4	53.8	46.6	52.2	54.2	54.2	49.8	55.4
set 6	51.2	53.8	49.0	52.6	53.8	56.4	53.2	56.6
set 1	48.0	51.2	52.4	46.0	54.4	50.8	52.8	54.0
set 2	52.0	52.0	47.2	48.8	56.0	49.2	49.2	53.6
set 3	51.2	52.8	44.4	46.8	55.6	50.8	52.0	56.4
set 4	51.2	51.6	50.8	54.0	51.2	49.6	48.0	52.0
set 5	51.2	52.0	49.6	47.2	54.0	55.6	49.6	55.6
set 6	44.8	47.2	48.8	49.6	52.0	46.4	52.4	45.6

Table 5.2: Accuracies for the 2-vs-2 decoding task using perceptual and semantic features to predict all time points from .05 to .9 s and 306 MEG sensors in *1back-text* paradigm. Table shows results for individual subjects and then for the concatenation of words across subjects. SX is a subject not scanned for the two previous paradigms. There were 6 sets of 500 random pairs of combined words in each set (250 pairs for *1back-speech*), where the top 6 sets represent results with *1back-text* and the bottom 6 with *1back-speech*. Using a binomial distribution, accuracies above 55.2% are significant at $p < 0.01$ for *1back-text* and at $p < 0.05$ in *1back-speech*. Distinguishing between two different groups of the same word was almost always better with the combined data across subjects than for any individual subjects.

the previous chapters). First considering the sets for *1back-text*, we see that classifying with the combined data was almost always better than classifying with any individual subject's data. Also, despite the non-impressive results (especially when compared to the ones in previous chapters), in *1back-text* we could always distinguish significantly better than chance ($p < .01$) with the data combined across subjects, but that was certainly not true for most of the subjects individually. Unfortunately that same finding was not true for *1back-speech*. While the results for the data combined across subjects were generally better than for single subjects, they were mostly not significantly better than chance.

We conclude that for a couple subjects it was possible to use the perceptual and semantic feature set to predict groups of words. This result does not discourage the use of these features to predict MEG activity, because we conjecture that the main reason behind the poor results was the amount of noise in the signal (which will be further studied in Chapter 6). Furthermore, combining data across subjects does seem to improve the signal and, consequently, the 2-vs-2 decoding results. For these two *1back-text* subjects, interesting patterns of weight distributions were also obtained (see later in section 5.6) which corroborate the results obtained with previous paradigms.

5.4 Negative results

It is attributed to Thomas Edison, the American inventor and businessman behind the phonograph, the motion picture camera, and the light bulb (`http://en.wikipedia.org/wiki/Thomas_Edison`), the saying: "Negative results are just what I want. They are just as valuable to me as positive results. I can never find the thing that does the job best until I find the ones that do not." (`http://www.goodreads.com/quotes/9788-negative-results-are-\just-what-i-want-they-re-just-as`). This thought is very applicable to this chapter of the dissertation because much time was spent trying to get better results with the *1back-text* and *1back-speech* datasets. While the results reported here were the best we are able to produce, it is important to discuss some of the failed attempts to hopefully learn from them, and that way figure out new approaches that will work better in the future.

First, a few data pre-processing attempts were made in order to clean the single trial data. The current results employed data after tSSS (see section 1.2.2 for more details), but we also tried using the original Signal Space Separation method [89] (without the temporal extension). Additionally, we have also done preliminary experiments using an extension to SSS that reduces sensor noise [90], but the results using that method were also not satisfactory. One of the possible explanations would be that this method works best in higher frequencies (e.g. above 30 Hz), and it could be that the information we are trying to decode is not encoded in such bands.

It could also be that the attributes of the data we used in our classifiers (namely, the MEG signal amplitude) are especially deteriorated by the single trial scenario. We have shown in previous chapters that the MEG signal amplitude does carry decodable perceptual and semantic information about concrete nouns, but it could be that other attributes would have better results in single trials. To that end, we also carried out a preliminary exploration of wavelet coefficients (using Haar wavelets) to represent the *1back-text* data. Unfortunately the 2-vs-2 decoding results for individual subjects were not much different than the ones presented here. However, because of the potential utility of this data transformation as indicated by other works in our group and elsewhere [16, 101], this is an alternative that should be further explored in the future.

Other attempts to pre-process the data were also performed following what has been done in the literature. Following Wang et al. [97], we tried smoothing the single trial signals (i.e. low pass filter at 8Hz and downsampling at 20 Hz) and then using a lower dimensional representation after Principal Component Analysis [27]. We also tried choosing specific time intervals of the trial [9]. Neither approaches resulted in better 2-vs-2 decoding results.

Along the same lines of experiments, we tried a few methods to select the data attributes being used (i.e. feature selection). The most evident method to try involved selecting the features based on their POVE, before using the features for scoring the predicted data. Although minor improvement was observed for a few subjects (about 2-5 % points increase in the 2-vs-2 decoding accuracy), the results were not as consistent across subjects nor as drastically increased as what was shown in chapter 3. Two additional methods were attempted for feature selection: mutual information and feature stability. In the first, we calculated the mutual information between the semantic features and the brain data for individual data attributes, and then ranked the features for scoring based on their mutual information score. The idea behind the method is that the best data attributes (i.e. the ones that need to be picked first to score the predictions) are the ones that have the most amount of mutual information with the semantic feature set. Feature stability, on the other hand, cannot be calculated in its original form from single trial data. If the reader recalls from chapter 3, the stability of a given feature is calculated based on that feature's behavior across repetitions. Because there are no repetitions in *1back-text* and *1back-speech*, we cannot compute these measures. However, we did try using the most stable features as ranked in the *answer-questions* and *60words* paradigms, both for individual subjects or averaging over subjects. Unfortunately, neither mutual information nor (borrowed) feature stability helped to improve the results. It is possible that in the future a method akin to feature stability can be developed for single trials, for example, drawing on the different trials across subjects, and this way be more readily applicable to such paradigms.

Finally, another approach we pursued dealt with testing different representations of the semantic feature space. It is entirely possible that the set of 218 semantic features is a good representation of the 60 words presented in *freely-think* and *answer-questions*, but it might not

generalize well to the 500 or 1000 words presented in the paradigms analyzed in the current chapter. We tried different transformations of semantic feature sets obtained from text corpus data (similar to [61]), but none of the results were better than the ones presented here. This is no proof that the 218 semantic feature sets is still the best one representing the larger set of words, and it can certainly happen that a different representation of the data, partnered with a different semantic feature set, will yield better results with the single trial data.

5.5 What are the brain regions and time points best accounted for by our model?

Differently than chapters 2 and 3, where we were able to look at the results for all subjects, here we will restrict ourselves to the results of subjects that showed statistically significant performance in the 2-vs-2 decoding task. Namely, we will focus on the results for subjects S2 and S6 in *1back-text*.

In order to better compare with the results shown in previous chapters, we again source localize the MEG activity prior to computing POVE and statistical significance. Then, we find the spatiotemporal features that are significant at $p < 0.01$ in a per-subject basis, and average these significant features over the pre-defined ROIs.

The first salient result from Figure 5.2 is a cluster of significantly predicted features from 475 to 600 ms. These features were located around the bilateral frontal cortices, such as the left superior frontal and the lateral orbitofrontal cortices, and the right frontal pole and pars triangularis regions. But again, it is important to keep in mind that the noise level increases in such areas distant to the MEG helmet, which potentially makes spatial resolution less clear. Still, this later wave of activation is part of a noticeable two-wave pattern, similar to what has been seen before in *freely-think* and *answer-questions*. While this later wave starts at about 375 ms, mostly focused on anterior regions of the brain, there is an earlier wave of regions well-explained by model B that peaks at around 100 ms, mostly focused on posterior regions of the brain. However, in *1back-text* there is a less clear boundary between features in the inter-trial interval and this earlier activation wave that is explained by the model. The boundary between the two waves of explained signal is also not as clear as it was for *answer-questions*.

5.6 How much does each feature contribute to predicting the spatiotemporal MEG activity in *1back-text*?

The contribution of perceptual and semantic features to these predictions is shown in Figure 5.3. As we have seen before, *word length* has most of its high weights in more posterior regions such as the lateral occipital and lingual gyri, and they contribute to the model in the early part of the signal (from 100 to 200 ms). Interestingly, these weights remain high later in the activation, up until 450 ms. The behavior of animacy-related features also corroborates what has been seen in previous paradigms. Namely, their weights are high in later parts of the trial (rising at about 400 ms) in more anterior regions in the brain, such as the left superior temporal sulcus.

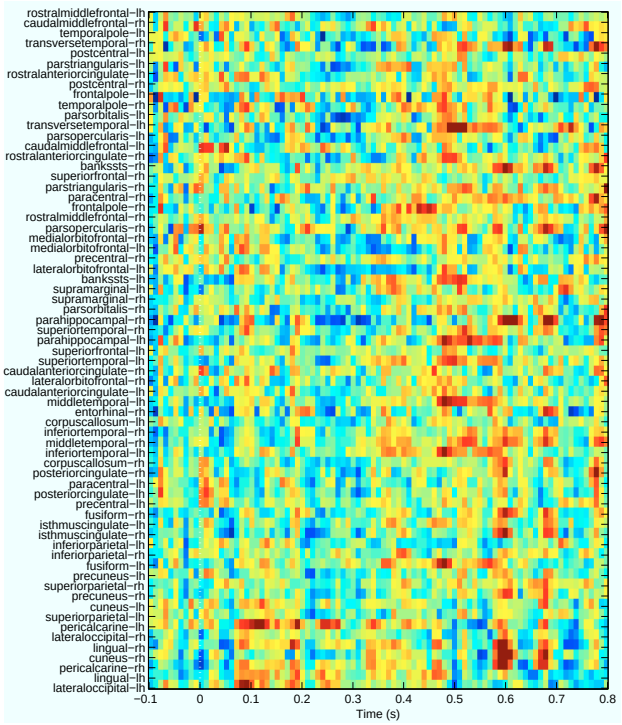


Figure 5.2: POVE in different regions of the brain over time in the *1back-text* paradigm. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is the same as in previous similar figures. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

An interesting result not seen before is the disambiguation between *word length* and the semantic features. In other words, the brain regions and time points with high weights for *word length* did not stay high for semantic features. For example, compare the plot for *word length* and *is it alive?* in Figures 5.3 and 3.8. While in *answer-questions* the same posterior regions of the brain show high weights for *word length* and *is it alive?* in the time window between 150 and 250 ms (a pattern also seen in *freely-think*), in *1back-text* those spatiotemporal regions only show high weights for *word length*.

Among the semantic features, we notice a fair amount of sparsity in the spatiotemporal distribution of weights. Semantic features that were correlated within the stimulus set of 60 nouns used in *answer-questions* and *freely-think*, given more diverse set of stimuli, show a few more disparities, such as the higher weights in anterior regions for *is it manmade?* compared to *is it alive?*. However, given the amount of sparsity of the weight distributions, it is difficult to infer any sequence of activation among the semantic features.

Overall, Figure 5.2 showed a pattern of POVE in the MEG signal quite different than what was seen for *answer-questions* and *freely-think* (Figures 3.6 and 4.4, respectively). We conjecture that this could be due to the less reliable (noisier) data summary in this sparse data situation, but we cannot discard the hypothesis that the nature of the task performed by the subjects also contributed to the different patterns seen in the plots. Regarding the contributions of different perceptual and semantic features to the signal, Figure 5.3 looks quite similar to Figures 3.8 and 4.5. So, in order to better compare the weight distributions across paradigms, in Figure 5.4 we

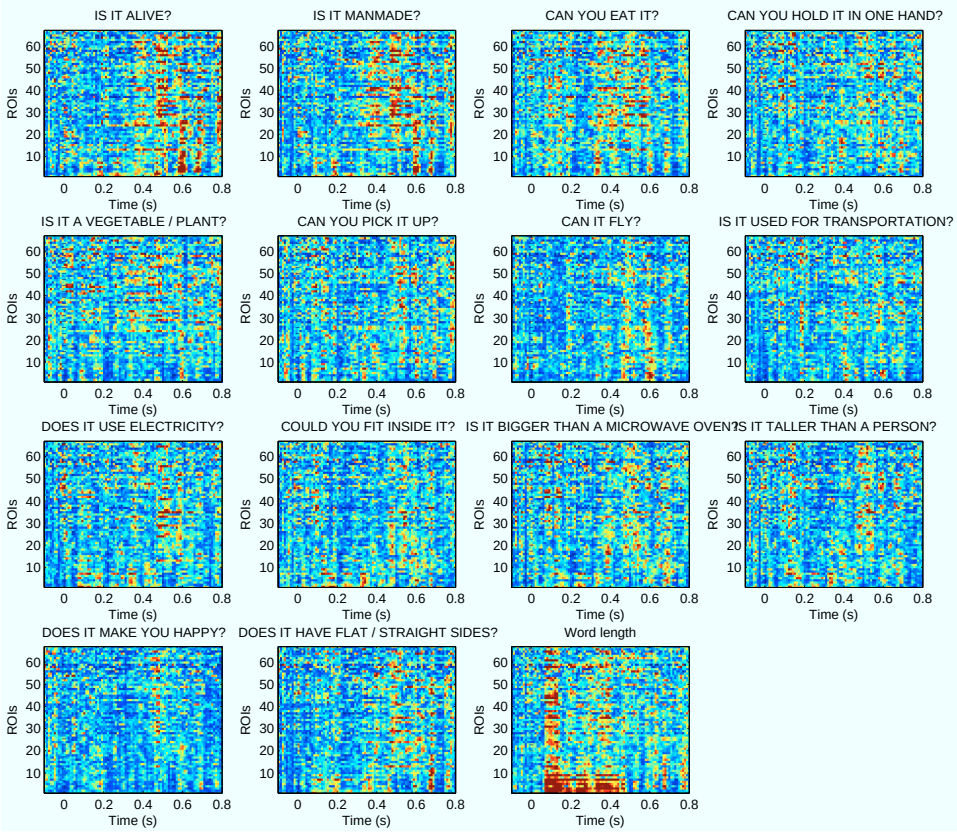


Figure 5.3: Weights for different intermediate features averaged over ROIs and subjects in the *1back-text* paradigm. Weights were first scaled to allow for comparing across regions and time, and then multiplied by the POVE in each spatiotemporal attribute. All plots share the same color scale. Note that *word length* was the only perceptual feature used in the model. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

combined the neural signature of the representative semantic and perceptual features we have been analyzing so far. Before the weight distributions were averaged across paradigms, they were first normalized so that the highest weight over time and space across all features was 1 in a given paradigm.

We can see in Figure 5.4 that the contributions of the features to explaining the MEG signal show several patterns of overlap across paradigms. For example, *word length* seems to heavily contribute to predicting the early MEG response (starting at 90 ms) in posterior regions such as the bilateral cuneus, lateral occipital cortex, and pericalcarine region. The contribution of the semantic features to the signal starting at about 400 ms is also clear, especially in more anterior ROIs such as the left medial orbito-frontal and the left superior temporal gyrus. Among semantic features, animacy-related properties show a remarkably different neural signature, considerably stronger in the later part of the signal than the other semantic features, and the two waves of higher weights are more fluid across paradigms than for other features. Semantic features such as *can you pick it up?* evoked a more well-defined second peak of activity, although not all semantic features (e.g. *can it fly?*) were able to elicit this second wave of high weights.

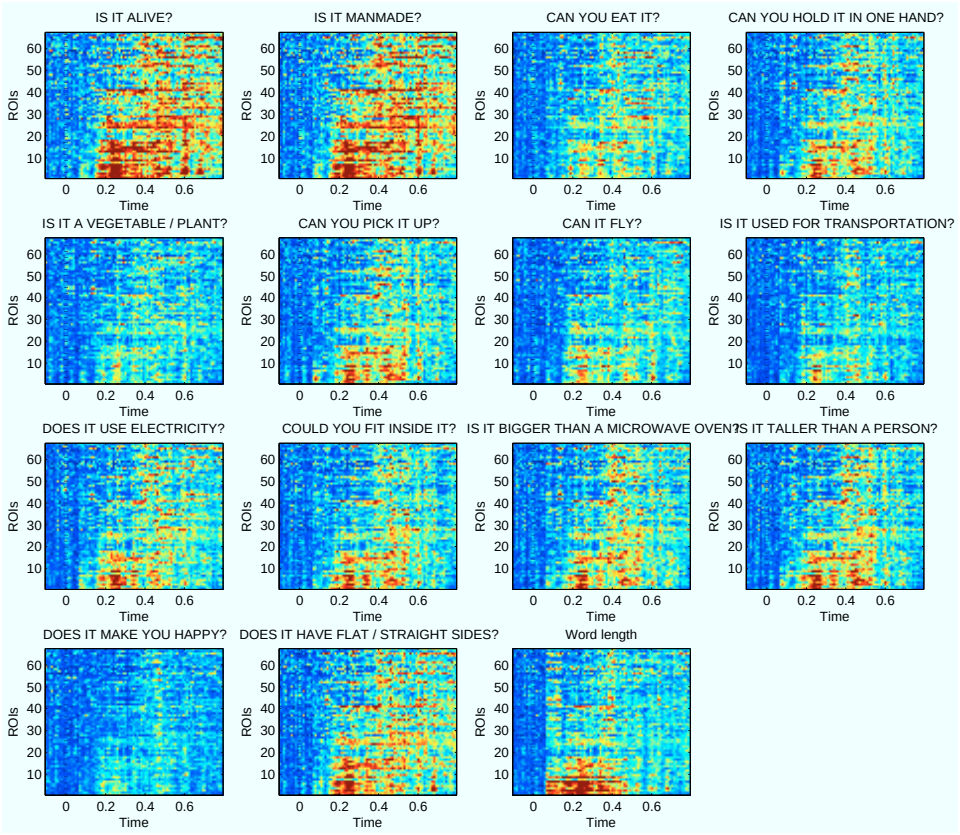


Figure 5.4: Weights for different intermediate features averaged over ROIs and subjects, combined over all 3 paradigms. Weights were first scaled within paradigms prior to the averages. All subplots share the same color scale, and the order of ROIs in the Y axis remains the same. Color scale goes from 0 to mean plus 3 standard deviations of the data pulled together over all subplots.

5.7 Discussion

The results we have seen in these three chapters provide insights about the flow of information encoded in the MEG signal associated with the processing of concrete nouns. Whereas classifiers trained using fMRI can establish where in the brain neural activity distinguishes between different nouns, our results show when and where MEG data localized to different regions of the cortex encode perceptual or semantic attributes of the nouns. We also demonstrated that it is possible to use model B to predict the MEG activity associated with concrete nouns even when no explicit semantic task was being performed by the subjects, and observed that perceptual attributes contributed mostly to the early MEG response, while semantic attributes contributed to the later part.

5.7.1 Comparison of results among paradigms

The paradigms analyzed vary along three major dimensions:

- the modality of the stimulus presented to the subjects
- the task these subjects were performing

- the trade-off between the number of stimuli presented (i.e. number of concrete nouns) and how many repetitions of each stimulus were acquired

We conjecture that many of the differences seen across paradigms, such as decoding accuracies and weight distribution sparsity, can be attributed to the number of trials in each task. More specifically, the fewer the trials averaged in the experiment, the lower the accuracies and the more sparse the distribution of the weights to the different ROIs. One could also argue that the task being performed by the subjects goes from semantically more demanding to less demanding (i.e. think of specific property, think freely, no explicit thinking about semantics), so that *answer-questions* would not only have a clearer signal, but also evoke more semantic activity.

Besides the number of repetitions and the task being performed, the type of stimulus also plays a role in the results. While *answer-questions* and *freely-think* showed a combination of picture and text to the subject, *1back-text* showed only the text. This choice of stimulus modality not only helps to disentangle the effects of double-stimuli of the previous two paradigms, but it also avoids recruiting areas that were previously involved mostly with image processing. This can be a possible explanation for why areas with high weights for perceptual features in *answer-questions* and *freely-think*, which also showed high weights for semantic features in those paradigms, did not show a similar pattern of weight distribution in *1back-text*, not displaying high weights for the semantic features. Another plausible explanation is that the more diverse set of stimuli helped to de-correlate the perceptual and semantic variables, which was one of the goals of the *1back-text* paradigm from the beginning.

A number of similarities is observed across paradigms. Firstly, it is possible to significantly predict MEG data using a set of perceptual and semantic features regardless of the task being performed by the subjects, and from that distinguish which of two groups of words subjects were thinking of. Another constant across paradigms was perceptual features being predicted earlier in time, usually associated with posterior regions of the brain, while semantic features showing weights related to predicting the signal in more anterior regions, in later time windows. In contrast, a number of disparities among the results of the different paradigms is also evident. For example, the results in *1back-text* show more spatiotemporal features being significantly explained by the model before stimulus onset than the other paradigms (Figure 5.3). While this could be one of the effects of noise in single trial analysis, it could also be hypothesized that the timing of the task was conducive to subjects to store the meaning of the previous word in memory, while processing the current word. More specifically, while there was not only more time between stimuli in *freely-think* and *answer-questions*, there was also no relationship between consecutive trials. In *1back-text*, it was advantageous for the subject to record the word meaning in memory in case the word was repeated (1-back task).

Another aspect in which the paradigms differ is the spatiotemporal location of the activity best explained by model B. While the model did a better job at explaining posterior activity early in the trial, and later activity in anterior regions for *answer-questions* and *freely-think*, only the latter was more evident for *1back-text* (Figure 5.3). The location of these activations was not in good agreement either. While the difference in the modality of the stimuli likely contributes to these differences, the noise in the single trials also makes source localization more challenging, increasing the chances of errors in location estimates.

Finally, there were some particularities in the weight distributions among the different paradigms. For example, weights for *word length* started earlier in *1back-text* than for the other 2 paradigms, which we speculate to be related to the double stimulation in the latter.

5.7.2 Considerations about paradigms without stimulus repetitions

Because only single words were presented in *1back-text*, there was no issue regarding double-stimulation. Hence, it is possible to make more clear statements about the serial vs parallel processing discussion, initiated in chapter 2. In short, the serial/cascade processing proponents argue that there are sequential stages of activation when interpreting the meaning of a word: first there is an analysis of the physical properties of the word stimulus, then lexical category information is retrieved, which is followed by semantic access and context integration [73]. On the other hand, the parallel processing advocates suggest that all these processes occur in parallel, near-simultaneously within 100 to 200 ms after stimulus onset, and the later activations (e.g. N400 and P600) reflect secondary lexical, semantic and syntactic information processing. By analyzing the weights for different features over time and space, there is no clear boundary between perceptual and semantic features, especially for the features that seem to be processed earlier (e.g. animacy-related features). To be clear, it is evident that the bulk of high weights for perceptual features happens earlier than for semantic features, but there is a blurry line around 250 ms where the brain seems to encode both types. Moreover, weights for *word length* seemed to extend well over 300 ms, while some of the semantic features did have high weights before 200 ms. Because we were only able to analyze data for 2 subjects in *1back-text*, we are prevented from strongly taking any sides on the debate.

It is also important to comment on the poor results obtained with the *1back-speech* paradigm. The single trial aspect of the data might not have been the only factor contributing to the results obtained. The words chosen to be presented as speech to the subjects were controlled with respect to length, and were also normalized based on the onset of the sound. However, there was no control for the energy envelope of the speech. After the scans, we noticed that the prototypical pattern of a peak at about 100 ms in temporal-parietal sensors, followed by a more sustained activation at about 400 ms was not found in the data (for an example of the sought pattern, see [77]). Still, we hoped to find some information in the data because the parts of the signal encoding the semantics of each word should still be present, despite their onset not being properly aligned across trials. We also tried centering the trials based on the peak of the energy of individual spoken words (instead of the spoken word onset), but that measure did not help the decoding results. In the end, neither approach yielded 2-vs-2 decoding accuracies that were significantly above chance ($p < 0.01$), which seems to speak to the importance of alignment of the processes involved in spoken word understanding.

5.7.3 Future directions

The possibility of providing a richer set of stimuli is very attractive to many aspects of research. However, that usually happens at the cost of repetitions of the same stimulus, which reduces SNR. In those cases, signal processing becomes crucial to remove artifacts that could have been previously ignored, because the averaging process over repeated trials could potentially get rid

of those sources of noise. We have already covered in section 5.4 some of the signal processing techniques that were tested with the data, but there are always different parameters that can be tweaked within those techniques to obtain better results. In addition, techniques such as Time-shift Denoising Source Separation [12] have been shown in the literature to properly clean up weak brain signals from electrophysiological recordings, so there is hope that they might help with results of single trial paradigms.

Another future direction would be to design new paradigms specifically structured to tease apart the differences along the major axis in which the paradigms studied here vary (i.e. task, modality, repetitions). For example, a task similar to *answer-questions* only using words, and then only using figures would be very positive. Along the same lines, performing the *1back-text* task with pictures, and also with (normalized) spoken words, would provide new insights to the differences in processing stimuli from different modalities.

Finally, this chapter took initial steps towards combining data of multiple subjects by concatenating their brain images for the same word. More robust techniques for combining subjects need to be tested not only to improve decoding accuracy results (as it was done here), but most importantly to investigate what is common across subjects in solving a particular task. Another aspect of these datasets that cannot be left unnoticed is that most of the subjects were scanned under all 3 paradigms. Although it is not true that the best subjects in *answer-questions* and *freely-think* are also the best ones in *1back-text*, this is another dimension in which the data can be combined (i.e. same subject across experiments). Models such as Canonical Correlation Analysis have been applied to neuroimaging data by Rustandi et al. [76] with great success, and exploring the latent space discovered by such models should provide further insights into a common representation of concrete objects in the brain.

Chapter 6

Towards a spatiotemporal description of the neural code for concrete objects

In this chapter we elaborate on the answers to questions that have risen while comparing the results for the different paradigms. Then, we explore a few new hypotheses about the neural code for concrete objects.

6.1 The effect of the number of trials in the results

As discussed in section 5.7.1, one of the major differences among the paradigms analyzed in this dissertation is the number of repetitions of the stimuli. While providing less repetitions allows for a richer set of stimuli, and also reduces repetitions effects, the improvement in signal-to-noise ratio that is usually gained by averaging multiple repetitions of the same stimulus is not available anymore. In this section, we explore the changes in the results obtained with *answer-questions*, *freely-think*, and *1back-text* after reducing the number of repetitions available.

6.1.1 Reducing the repetitions in *answer-questions*

The *answer-questions* paradigm is the experiment with the highest amount of repetitions among the ones analyzed in this dissertation. We want to investigate the effects that reducing the amount of data being averaged (and, consequently decreasing the SNR) has on the findings discussed so far. Moreover, by approximating the SNR in *answer-questions* to the one in *freely-think*, we will have a chance to better compare the results of these two paradigms that share the same set of stimuli, but vary on the task performed by the subjects.

We have already seen a preview of what happens to the 2-vs-2 decoding results in *answer-questions* after reducing the number of repetitions available (Figure 5.1). Here, we reduce the number of repetitions in *answer-questions* to match the number of repetitions in *freely-think* (6 - 10 repetitions depending on the subject) by selecting only the first presentations of a stimulus for each subject. This approach was chosen in order to reduce repetitions effects, but we acknowledge that choosing the left out repetitions at random or by some other heuristics (e.g. smallest reaction time in pressing the response button) would also have its advantages.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
answer-questions-redux	66.8	82.2	59.8	74.3	75.4	67.1	69.4	64.4	54.8	**68.2**
freely-think	80.5	88.1	77.8	81.4	86.2	75.4	78.9	53.4	81.5	**76.4**

Table 6.1: Accuracies for the *leave-two-out* experiment using perceptual and semantic features to predict MEG activity in all time points from -.1 to .8 s and 306 MEG sensors. Most subjects perform better under *freely-think* than under *answer-questions*-redux, where the number of repetitions in *answer-questions* was reduced to match the number in *freely-think*. Chance accuracy was 50.0%. For a single-subject model, 62.5% correspond to $p < 10^{-2}$.

Table 6.1 compares the 2-vs-2 decoding results of *answer-questions* with fewer repetitions and the original *freely-think* results. Compared to the results in Table 3.2, it is clear that the *answer-questions* 2-vs-2 results deteriorate after reducing the number of repetitions. In fact, as it is seen in Table 6.1, the results with *freely-think* are actually better than the ones with *answer-questions*-redux for the majority of subjects. This result is quite surprising, since the *freely-think* task is more open-ended compared to the one in *answer-questions*, and the SNR increase obtained by averaging repetitions relies heavily on the subject's ability to consistently think of the same properties of an object across repetitions. Another factor that would have played against better results in *freely-think* is subject fatigue, since *freely-think* was scanned right after *answer-questions*, but apparently that was not the case. Note that neither row of results in Table 6.1 uses feature selection, which was shown in chapter 3 to improve the 2-vs-2 decoding results.

We conjecture that the *freely-think* experiment data performed better for a couple of reasons. First, effects of the button press (present for every trial in *answer-questions*) might be more problematic when averaging fewer repetitions (i.e. not enough repetitions averaged to downplay its effects). Secondly, after the initial recognition phase present in both paradigms, the *freely-think* task should recruit regions of the brain involved in representing different semantic aspects of the concrete noun, instead of only one semantic feature prompted by the *answer-questions* paradigm. For example, one of these properties might be visual imagery of the concrete noun, which is not necessarily required to perform the task in *answer-questions*. Hence, it could be that not only the process of recruiting more semantic features benefits the decoding with the *freely-think* data, but also that activity in the regions involved in processing these extra features is well-captured in MEG.

It is important to recall that the task of "thinking about the properties of objects" has been borrowed from fMRI [58] but up to this point there have been concerns about how adaptable it is to MEG settings. The open-endedness nature of the task seemed more suitable to the low temporal resolution and consequently integration over time of fMRI, compared to the focus on a semantic-task offered in *answer-questions*. While the double-stimulation with word and picture should be re-considered in the future, especially to allow for a better control of perceptual features to be used in predicting MEG activity, it is clear that the task performed by the subjects in *freely-think* is still sufficient to promote semantic processing activity to be captured in MEG. Furthermore, using similar paradigms in fMRI and MEG will facilitate the combination of data across modalities in the future.

We end this comparison of results by analyzing Figures 6.1 and 6.2, which respectively show the percent of variance explained (POVE) and the weights of different perceptual and semantic features, using data from *answer-questions*-redux. Compare these figures with Figures 3.6 and 3.8. First looking at the POVE figures, the two distinguishable peaks of signal explained by model B, clearly visible in Figure 3.6, are blended in *answer-questions*-redux. The early part of

the explained activity is still visible, but the later peak in the posterior regions (bottom part of the figures) is not as clear. The signal in more anterior parts of the brain (for example, the left and right medial orbito-frontal regions) is still well-explained by the model between 350 and 400 ms, as well as the activity in later time windows (after 500 ms). The reduction in the number of repetitions also makes Figure 6.1 resemble more Figure 4.4, which showed the POVE for *freely-think*, especially in the blending of early and later peaks. The spatiotemporal regions in both figures that are best explained by the model are also more sparse than what is depicted in the original figure for *answer-questions*, which suggests that the fewer repetitions one uses, the more sparse these plots become.

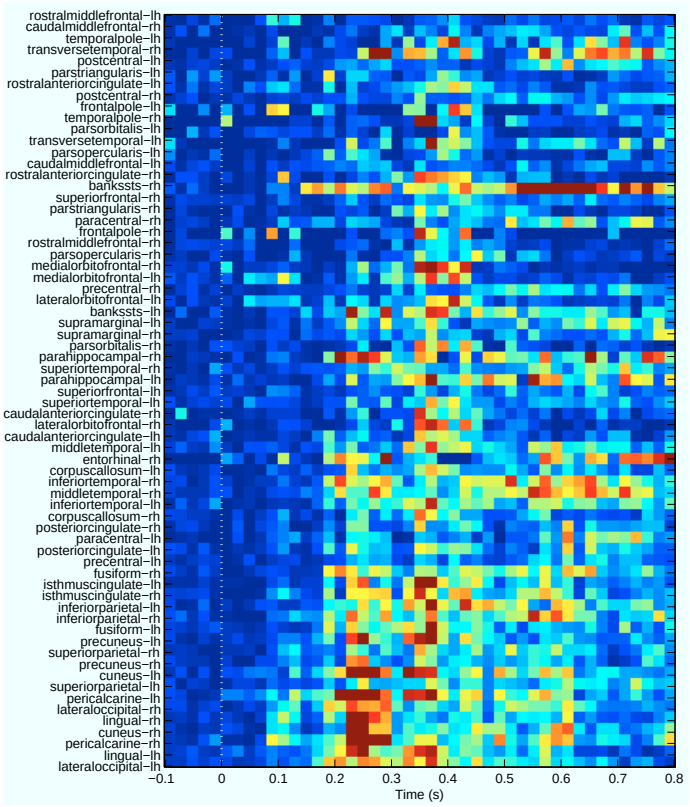


Figure 6.1: POVE in different regions of the brain over time for *answer-questions*-redux. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is repeated from Figure 2.10. POVE-values were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

The most salient observation that comes from comparing the weight distribution in the original *answer-questions* data and the version with reduced repetitions (Figures 3.8 and 6.2, respectively) is their similarities. Again, the fewer number of repetitions in the redux version seems to affect the sparsity of the weight plots. The early, perceptual peak is still clear in both figures, but the semantic peak (after 300 ms) seems to be more affected by the decrease in the number of repetitions. The semantic feature *is it taller than a person?* is a good example of that: while the peak in posterior regions around 200 ms is still visible, the peak around 450 ms is less clear. As it was in the case of POVE plots, the reduction in the number of repetitions makes the plots for *answer-questions*-redux resemble the weight plots for *freely-think* (Figure 4.5).

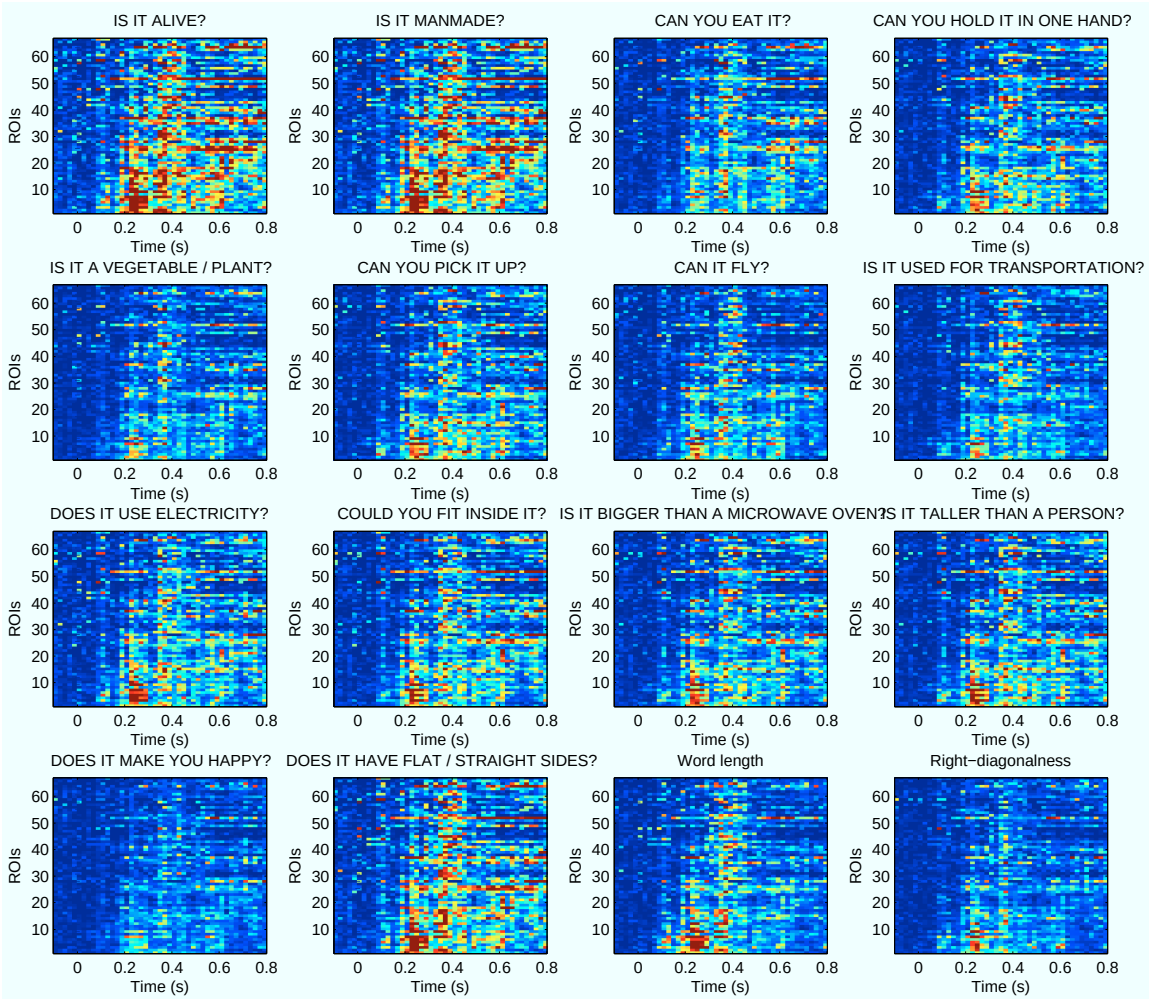


Figure 6.2: Weights for different intermediate features averaged over ROIs and subjects for *answer-questions*-redux. All plots share the same color scale, and the order of ROIs in the Y axis is constant, as seen in Figure 2.10. Weights were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

In summary, by reducing the number of repetitions in *answer-questions* to exactly the number of repetitions available in the *freely-think* data we have better grounds to compare the two paradigms. There was a more pronounced encoding of semantic features around 400 ms in the *answer-questions* dataset than in the *freely-think* dataset, evident in the different weight distributions in Figures 6.2 and 4.5. We conjecture this is due to the presence of the question-answering task in the *answer-questions* experiment, and its demands for neural processing to support answering questions about word semantics, in contrast to the less targeted "think about the word" task in the *freely-think* experiment. The neural signature for each perceptual and semantic feature looked less distributed over time in *answer-questions* as well (again, comparing Figures 6.2 and 4.5), suggesting that the presence of the question might be an effective way to better align in time processes related to semantic representation. Finally, the signal explained by model B was more spread in time and space for *freely-think* than *answer-questions* (see Figures 6.1 and 4.4), which we again conjecture reflects the consequences of the goal-directed task (i.e. focus on this

particular property) to a more narrow signature of semantic-related activations.

6.1.2 Reducing paradigms to single trials

In order to better compare the effects of the number of repetitions among *answer-questions*, *freely-think*, and *1back-text*, they all need to be reduced to a single repetition. In the case of *1back-text*, although it is already single trials, it would not be fair to keep all 1000 words in the comparison, so this paradigm is reduced to only contain the 60 nouns also used in *answer-questions* and *freely-think*. For these latter two, only the first trial was kept for each noun, again with the intent of getting rid of repetitions effects.

We start the comparison by showing the results for 2-vs-2 decoding for the three reduced paradigms (Table 6.2). Here we choose to go back to model A, in which we predicted semantic and perceptual features from MEG data. The regularization of the model seemed to best deal with the noise in MEG activity features, and because feature selection by stability would not be an option in the current case, there was a higher chance of obtaining better results with that model.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean
answer-questions	50.68	**74.75**	66.33	**70.40**	67.85	53.79	56.05	59.10	54.18	**61.46**
freely-think	57.46	49.04	45.65	68.31	51.47	53.05	60.00	55.59	39.60	**53.35**
1back-text	47.32	39.77	50.40	43.62	n/a	52.03	n/a	n/a	51.07	**47.37**

Table 6.2: Accuracies for the *leave-two-out* experiment using MEG activity in all time points from -.1 to .8 s and 306 MEG sensors to predict the perceptual and semantic features in single trials. Subjects marked with n/a were not scanned for the *1back-text* paradigm. Chance accuracy was 50.0%. Subjects with accuracies significant at $< 10^{-2}$ are shown in bold.

The 2-vs-2 decoding for which results are shown in Table 6.2 was performed in much the same way as it was done in chapter 2. Namely, every possible pair of 2 words were left out for testing, while training on the remaining 58 words in each cross-validation fold. Significance was assessed by performing the same decoding hundreds of times with permuted data, and then adding up how often the accuracy obtained with permuted labels was higher than the accuracy of the data with correct labels. More specifically, in each iteration with permuted labels (per paradigm), a subject was chosen at random and the label of the data were shuffled prior to decoding.

As could be expected based on the results in chapter 5, reducing each paradigm to a single trial per concrete noun drastically affects the results. Only a few subjects still retain significant decoding accuracy. It also appears as though the *answer-questions* paradigm has a slight advantage over the other paradigms, and that the paradigms that showed pictures and had a more semantic-related task to the subjects also evoked somewhat better results. At this point, it is unclear whether the nature of the stimulus or the task performed by the subjects causes this difference, but given the poor results with single trials, such claims would be better made using paradigms that have at least a few repetitions to be averaged to improve SNR.

Despite the already expected poor results of 2-vs-2 decoding with single trials, it is still informative to study how much of the MEG activity model B can explain, as well as the distribution of weights for individual features, and compare to how these maps look in their full-data counterpart paradigms. We start by looking at the POVE values for *answer-questions* single trials, and then at the weight distribution maps.

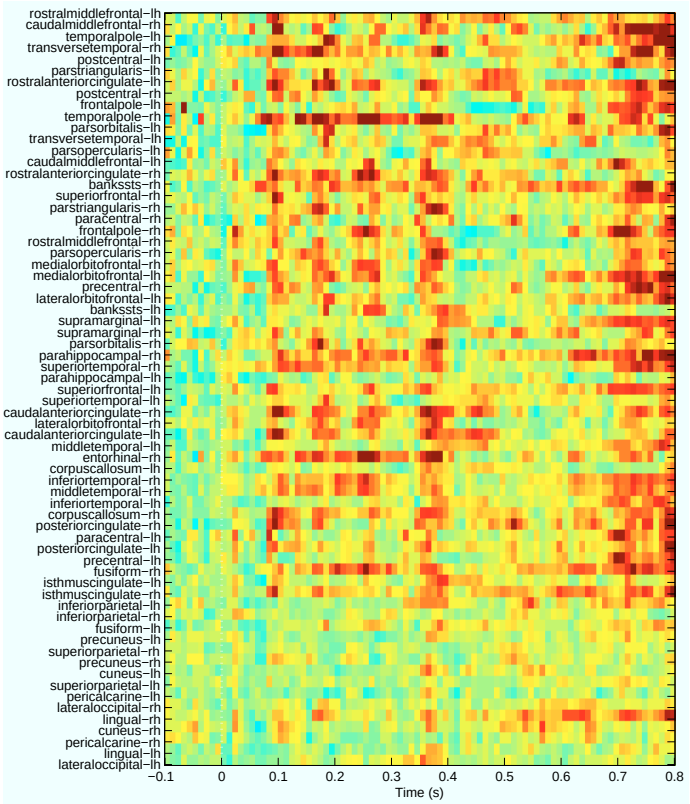


Figure 6.3: POVE in different regions of the brain over time for *answer-questions* single trials. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is repeated from Figure 2.10. POVE-values were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

Figure 6.3 shows an interesting pattern: we do not see the early perceptual peak around 150-200ms anymore (compared to Figure 3.6), which is normally located in posterior regions such as the bilateral cuneus and pericalcarine regions (bottom part of the plots). Instead, the plot is dominated by higher POVE in anterior regions, especially after 700ms. Analyzing Figure 6.4 helps us understand this pattern. While the participation of posterior regions in the early part of the activity is still unclear, we can see that the animacy-related features have a decisive participation in predicting the signal after 700 ms, especially when compared to other, less correlated semantic features. This behavior is explained by going back to a few limitations of the paradigm design: there was a button-press response for each trial, and the questions were asked in a block-design. More specifically, the first question asked to all subjects was "Is it manmade?", and the average reaction time to answer this specific question was 952 ± 254 ms, which satisfactorily explains the high weights for the animacy-related features shown in Figure 6.4. In other words, there is a high correlation between the button press and the animacy-related features in every first repetition of the *answer-questions* paradigm. Activation related to pressing a button will be spread over anterior regions in the brain, but focusing on the sensorimotor cortices and starting about 250 ms before movement onset. This timeline matches perfectly to the ramp up of weights in the weight signature plots for animacy-related features. Any semantic feature that is correlated to the animacy-related features will see a similar ramp up in the weight signatures, which is directly

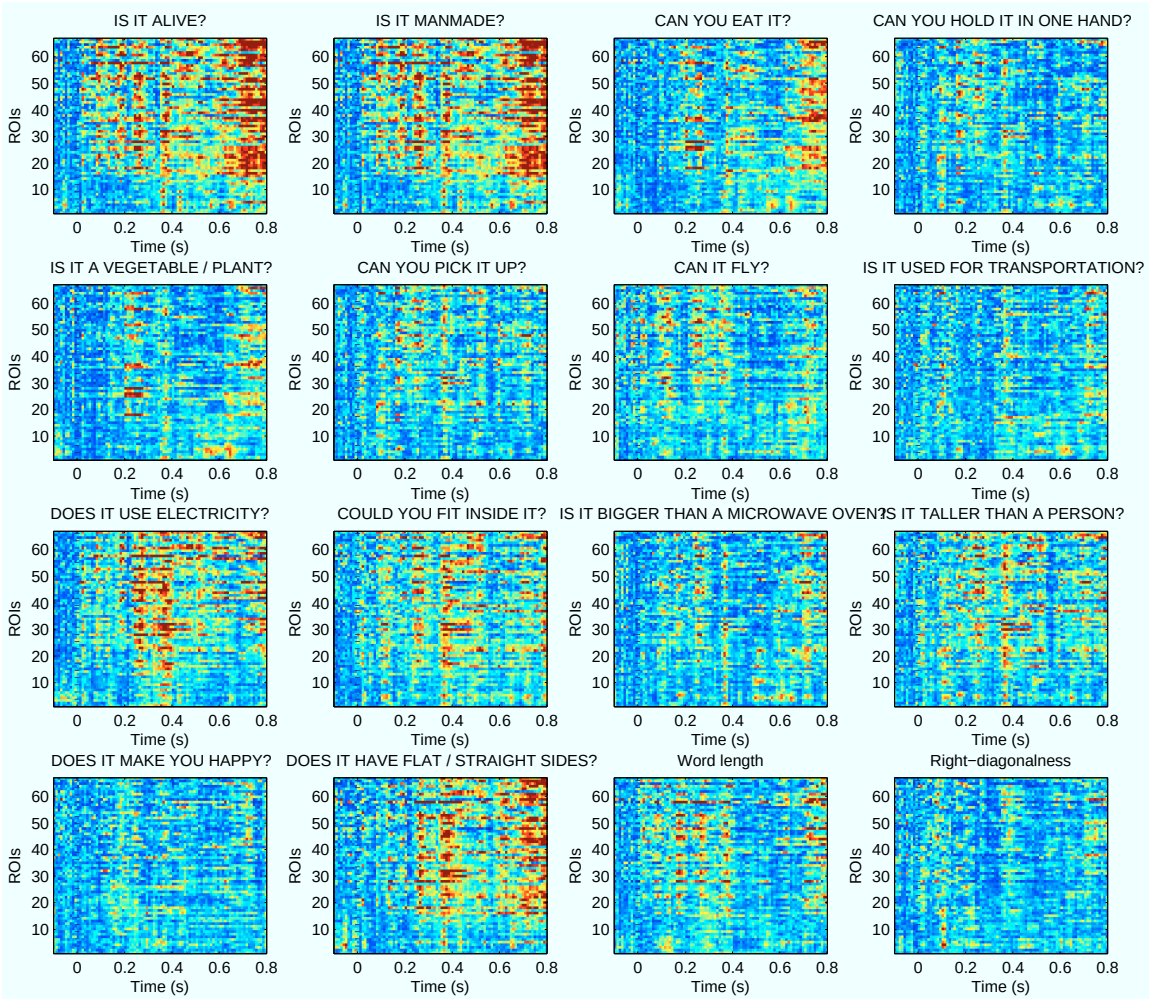


Figure 6.4: Weights for different intermediate features averaged over ROIs and subjects for *answer-questions* single trials. All plots share the same color scale, and the order of ROIs in the Y axis is constant, as seen in Figure 2.10. Weights were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

related to the button press (which is indirectly related to the feature itself, because of the question the subjects were answering in that trial). This could also explain why *answer-questions* was the only paradigm to show significant 2-vs-2 decoding accuracy results using single trials.

It is important then to re-ask the question of how the button press affects the *answer-questions* results. When the analysis is run in single trials, and the trials are all taken from the same block, then it is important to take into consideration the average reaction time in answering the question in the block because the activation related to motor preparation and sensorimotor feedback might corrupt the results. However, when the analysis is performed by averaging over different blocks, the high variation in reaction times (both across subjects but also across blocks) washes out button-press effects. Moreover, as noted in chapter 2, we performed an analysis that determined that the button press values (i.e. the yes/no response, which correlates to the brain pattern associated with pressing the respective buttons), after averaging over trials, are not by themselves enough to obtain an accuracy above chance levels in performing the 2-vs-2 task.

But Figure 6.4 also shows other interesting patterns. For example, there is a slight increase in weight values for the perceptual features in posterior regions around 100 ms, despite the low POVE values in Figure 6.3. Also, many of the semantic features display another increase in weights in anterior regions between 250 and 400 ms. In fact, that is a pattern that is also seen in Figures 3.8 and 6.2.

This semantic-related pattern is also observed in Figure 6.6, which shows the weight distribution in the *freely-think* paradigm, single trials. Despite the sparsity of the plots, this anterior activity between 250 and 400 ms is clearly visible in several semantic features, and because of the nature of the paradigm, there are no confounds with button presses. Since this weight signature is present in the single-trial analysis for both *answer-questions* and *freely-think* experiments, we conjecture that it reflects the semantic processing relative to understanding the meaning of the concrete noun being displayed regardless of the task being performed. A similar comparison can be made about the POVE plot for *freely-think* single trials (Figure 6.5). While the early part of the signal in posterior regions is not properly explained by model B (unlike Figure 4.4), we see a similar vertical stripe between 300 and 400 ms, which we speculate to be attributable to the semantic processing for the nouns in both paradigms.

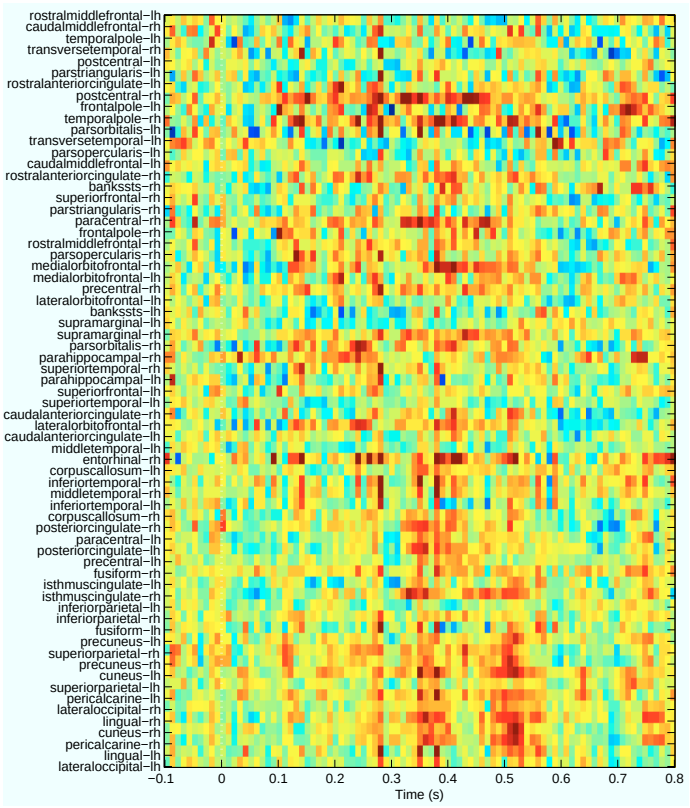


Figure 6.5: POVE in different regions of the brain over time for *freely-think* single trials. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is repeated from Figure 2.10. POVE-values were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

We end this subsection by comparing the results for *1back-text* single trials (Figures 6.7 and 6.8) with the single-trial versions of *answer-questions* and *freely-think*, and then with the full-

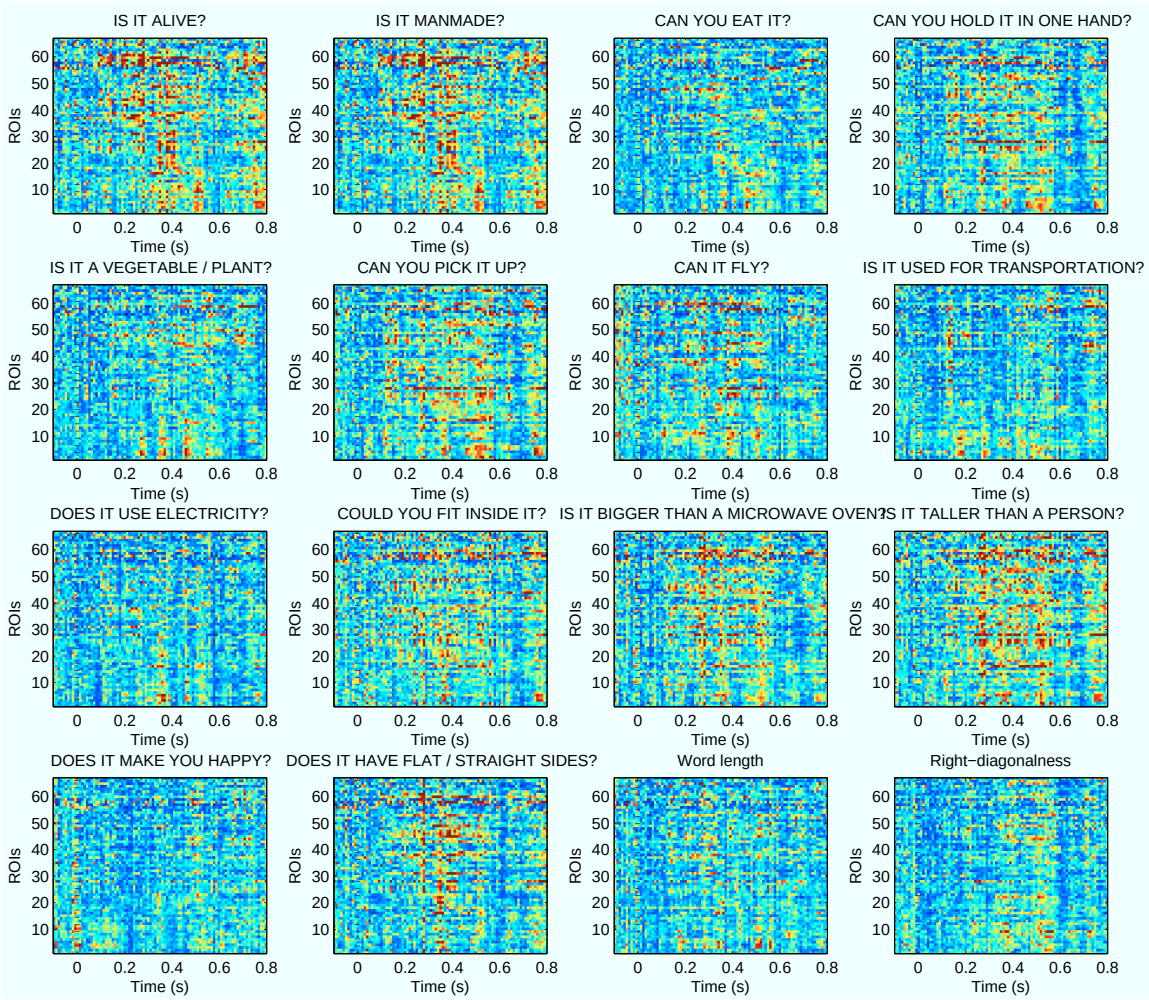


Figure 6.6: Weights for different intermediate features averaged over ROIs and subjects for *freely-think* single trials. All plots share the same color scale, and the order of ROIs in the Y axis is constant, as seen in Figure 2.10. Weights were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

fledged *1back-text* paradigm. Unfortunately, using only 60 nouns in the *1back-text* paradigm did not prove to be very informative. The plots were too sparse and did not reveal any meaningful patterns. This could be partly because only subjects S2 and S6 were used in combining the POVE and weight values (following chapter 5), but could also be a function of the task performed by the subject or the word-only stimuli. Still, there were interesting patterns to be studied in Figures 5.2 and 5.3, which suggests that if the type of stimulus and task were to be kept, then a wide and diverse collection of words are necessary in order to obtain meaningful results.

In summary, we have shown that even in single trials it is possible to observe interesting patterns in the data that can be predicted by a computational model using a set of perceptual and semantic features. In such cases, a rich set of words is important, and doing so helps to disentangle among the contributions of the many perceptual and semantic features in the model. However, special techniques such as feature selection and concatenation become crucial when decoding with single trials. Decoding results might also benefit from using different attributes

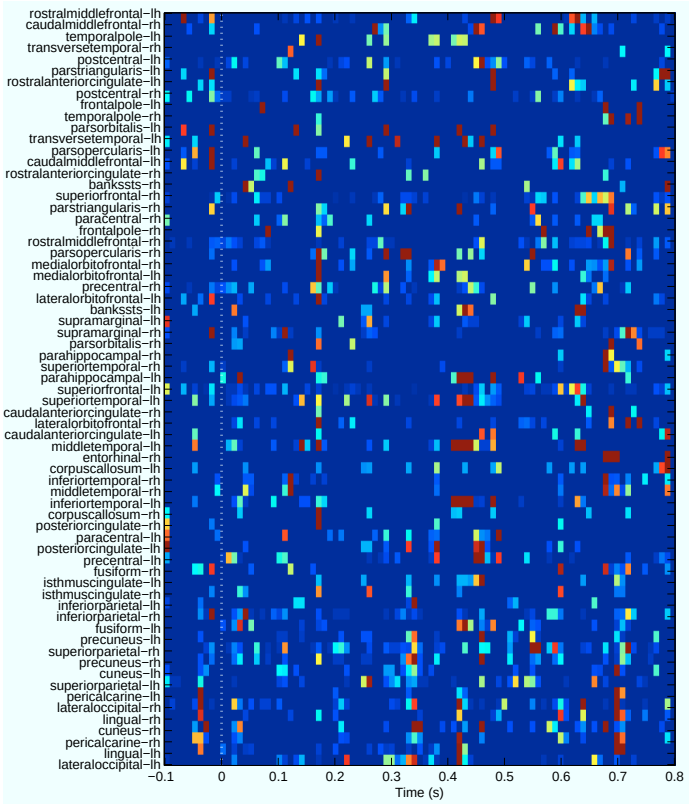


Figure 6.7: POVE in different regions of the brain over time for *1back-text*, using only the 60 nouns that were also part of *answer-questions* and *freely-think*. The hotter the color in the plot, the better we can predict the spatiotemporal activity using model B. The order of the ROIs in the Y axis is repeated from Figure 2.10. POVE-values were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data shown.

of the MEG data (e.g. wavelet coefficients), and from further signal processing techniques (as detailed in chapter 5). When repeating stimuli is not an issue, then we have shown that it is possible to obtain significant results with as few as 8 repetitions, but the more repetitions the better the decoding results.

Across paradigms, we see a consistent pattern of perceptual features contributing to explain the early parts of the signal, mostly in the posterior regions of the brain, while semantic features better contribute to explaining the MEG signal in more anterior regions, normally starting around 250 ms after stimulus onset. This pattern can be observed regardless of the task being performed or the nature of the stimuli. Following the differences in weight distribution for semantic features observed in the comparison between *answer-questions*-redux and *freely-think* (Figures 6.2 and 4.5), we conjecture that the effects of the question in the task make the neural signature for specific semantic features less distributed over time and space when compared to the more open-ended task in *freely-think* (and in *1back-text* to the extent that conclusion can be made given the sparsity of the plots). The weight signature plots also invite a speculation about early activation of semantic features in *answer-questions* when compared to the other paradigms, especially for the features correlated to the question being asked, but the confounds of the button press would need to be taken into consideration before investigating this conjecture any further.

The results in this section help to shed light into one of the variables of the paradigms ana-

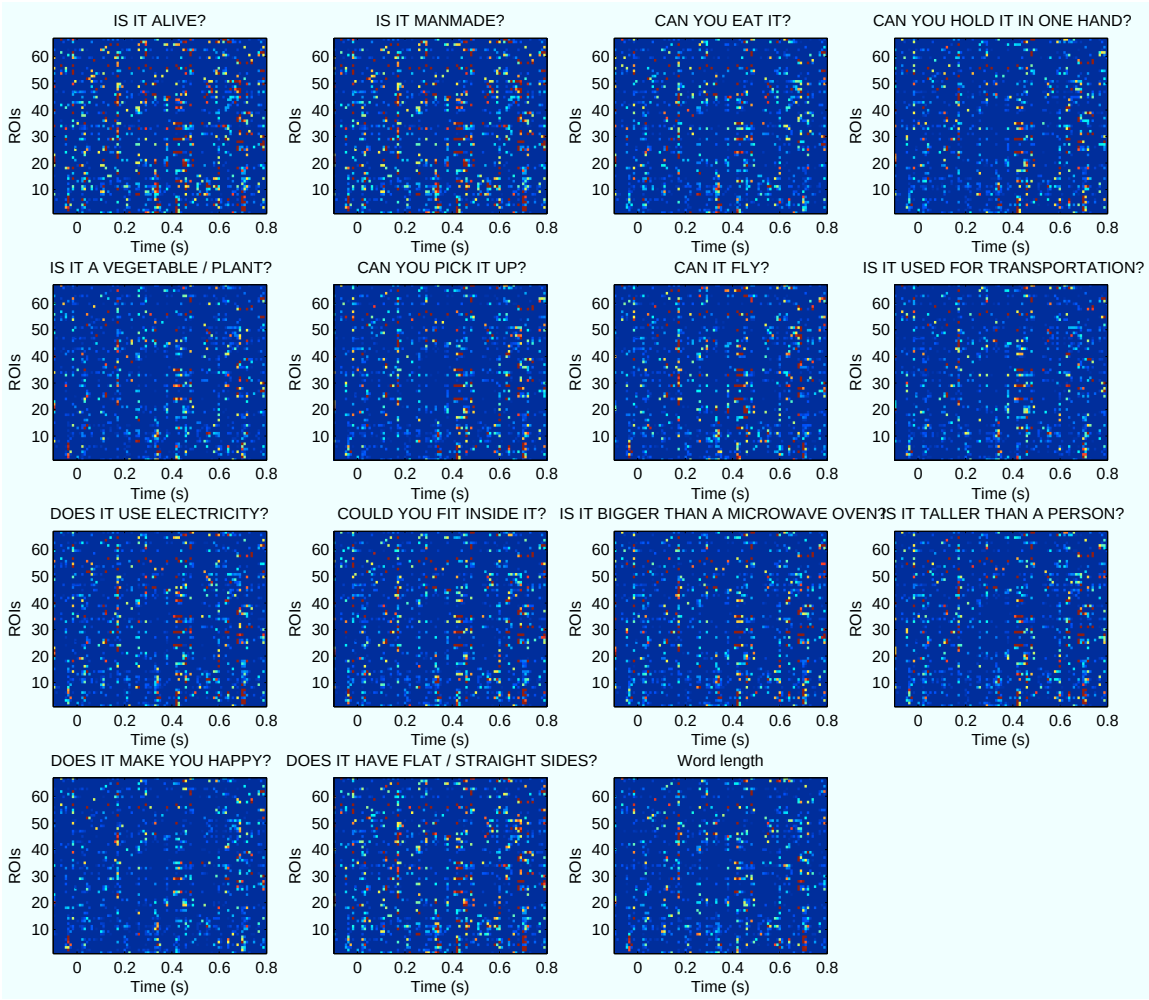


Figure 6.8: Weights for different intermediate features averaged over ROIs and subjects for *1back-text* using only the 60 nouns that were also part of *answer-questions* and *freely-think*. All plots share the same color scale, and the order of ROIs in the Y axis is constant, as seen in Figure 2.10. Weights were combined across subjects similarly to what was described in chapter 3. Color scale goes from 0 to mean plus 3 standard deviations of the data pooled together over all subplots.

lyzed here: the number of repetitions. We could only make suggestions about the other two major variables (type of stimulus and task performed), given the data at hand. For example, when the number of repetitions was balanced, the task in *freely-think* appeared to be as effective if not better than the task in *answer-questions* in 2-vs-2 decoding accuracies. However, more MEG data for different paradigms (e.g. a version of *freely-think* with word only) need to be collected to provide more insights about these other differences among the three paradigms.

6.2 Hypotheses testing

In this dissertation we have presented a couple models that provide further insights into the spatiotemporal characteristics of the neural code related to comprehending concrete nouns. So far, they have both been used to ask complementary questions about the nature of the neural

code. Here, we return to the model A to test a couple new hypothesis about the representation of concrete nouns in the brain.

We start by looking at Figure 6.9, which shows the POVE averaged over subjects for semantic feature *is it alive?* in the *answer-questions* paradigm. Recall that each result in the plot is obtained independently (i.e. a separate regressor is trained for each ROI-time window pair). This plot is used to illustrate a couple patterns that are true for many perceptual and semantic features. First, many ROIs encode the same feature at a particular time point. Additionally, we observe many ROIs encoding the same feature in different time windows.

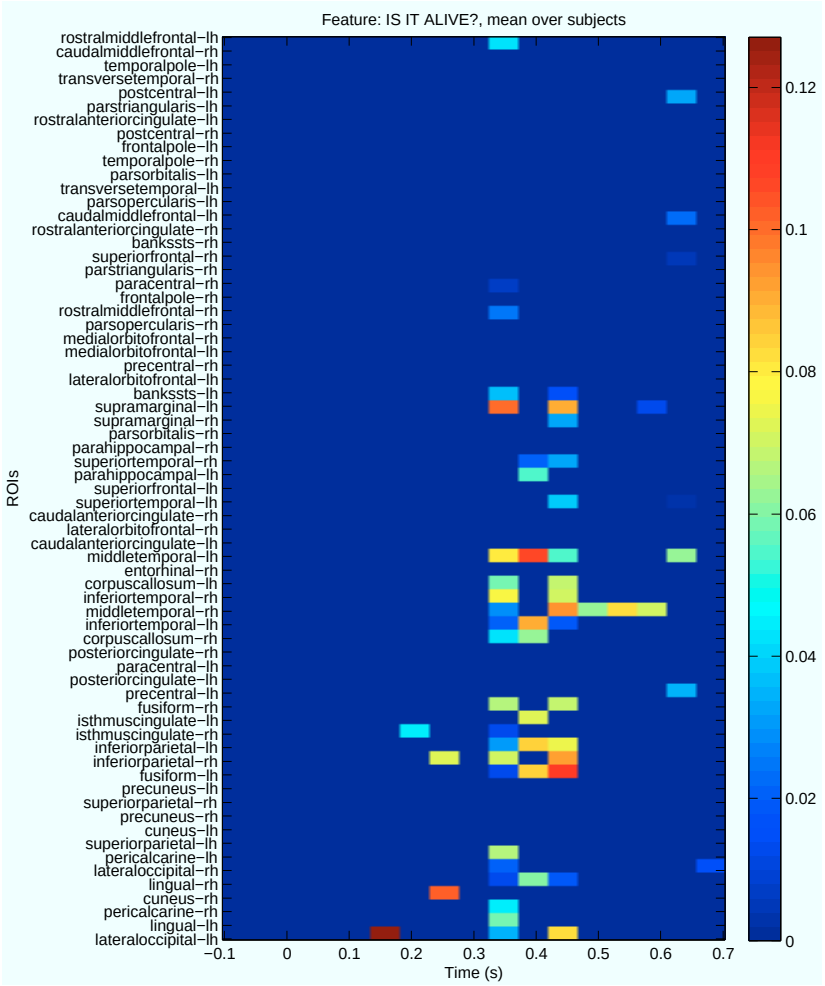


Figure 6.9: POVE averaged over all subjects for semantic feature *is it alive?* in *answer-questions* paradigm. Plot is reproduced from one of the subplots in Figure 2.9. Time 0 is when the stimulus is presented to subjects. Only values significant at $p < 10^{-5}$ are shown in the plot (i.e. all others were set to 0 and shown as dark blue).

6.2.1 H1: For regions that encode a feature in parallel, the information encoded in each region is complementary to what is encoded in the other regions.

This hypothesis is closely related to the argument for distributed coding in the brain. While the argument is most often described in terms of population of neurons, we can extrapolate the concept to the encoding of concrete nouns in several regions of the brain at a particular time point. The results in Figure 6.9 show that at a particular time window, many regions independently encode the particular semantic feature. We expect that, if the different regions are encoding the same information, then no gains in decoding accuracy will happen for that feature when combining their data. On the other hand, if combining the data of the different ROIs increases the feature decodability, then that is evidence towards complementary information being encoded in parallel among these regions.

To illustrate what these different types of coding for a semantic feature would be, consider the following synthetic example. Let the feature *is it alive?* be coded by a gradual score, such that 1 represents concepts that are not alive, and 5 represents living concepts. A distributed but redundant coding could be accomplished by each activated region encoding the rating based on its amount of activity, such that more activation in the region represents a higher score. On the other hand, one way to envision a distributed but complementary code is the summation of the activation over all regions representing the score. Under those scenarios, distinguishing between two nouns based on their animacy feature could be easily done by simply looking at one region in the redundant case, but the information of all involved regions would be necessary in the case of the complementary code. In the latter, partial distinction between the concepts could be accomplished by examining the activation of a subset of the regions, but a more accurate decoding of the semantic feature would only be obtained by using the information of all involved regions.

To test the hypothesis that complementary information is encoded simultaneously in different regions, we trained a regression function to predict representative semantic and perceptual features at the time window of their peak decodability for each subject, using as input to the function the MEG activity of all sources in different ROIs averaged over the time window. For example, Figure 6.9 indicates that, for the semantic feature *is it alive?*, the regressor would use data from the time window 450 to 500 ms. Once the time window is selected, the algorithm picks the top 10 regions encoding the feature and train regression functions with every possible combination of the 10 regions (i.e. first combining all 10 regions, then all possible combinations of 9 regions, then 8 regions, until eventually using only each region by itself). This is done in a leave-one-out cross validation framework, where one of the 60 possible nouns is left out in each fold, and POVE is calculated for the feature in a similar fashion as the feature score metric previously described in chapter 2.

The challenge here is that the larger the number of ROIs that are combined, the more terms are added to the regression. More specifically, the time-window averaged MEG activity in the sources within the ROIs is concatenated to the similar attributes of different ROIs. It is expected that cross-validating the results reduces data over fitting [67], but just to answer any critiques that using more data provides better SNR, and therefore better results, we also performed a similar

task without using the top 10 regions. In other words, for this comparison task we still use the top region encoding the feature in that time window, but the other 9 ROIs to be combined come from the bottom of the list. If an increase in results is just a function of adding more data, then we should also see an increase in results for this comparison case.

Figure 6.10 shows the results of the test described above for four representative features and Figure 6.11 shows the results of the comparison test. In the results shown, each plot represents the maximum POVE taken over all possible combinations of ROIs (indicated in the X axis) for a given subject. Missing values in the plots mark POVE results that were below a baseline, which was established by running Gaussian random data through the same testing procedure.

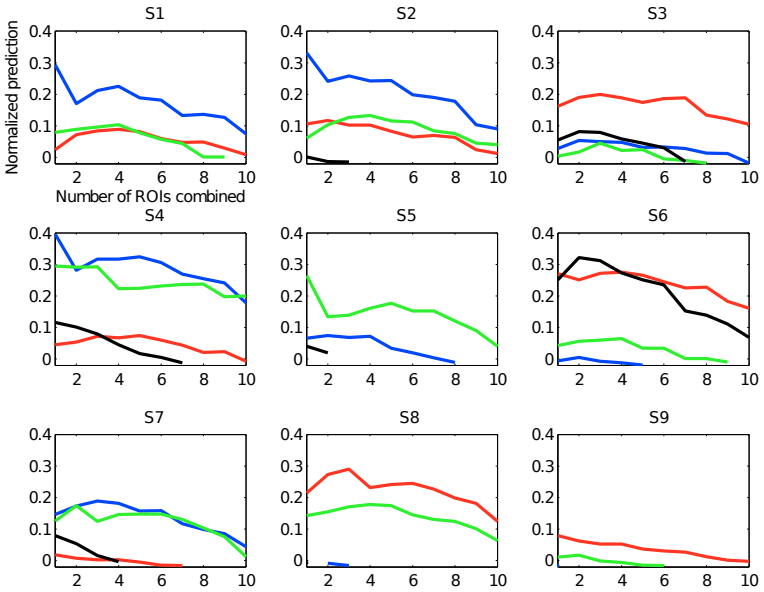


Figure 6.10: POVE for 2 perceptual and 2 semantic features in the *answer-questions* paradigm. One plot per subject, where the Y axis represents the maximum POVE obtained by combining the number of ROIs shown in the X axis. Blue is *word length*, red is *right diagonalness*, green is *is it alive?* and black is *can you pick it up?*. Adding data from different ROIs to the decoder often increased feature decodability up to a saturation level.

It is clear from comparing Figures 6.10 and 6.11 that just including more regions does not necessarily help the decoding results of the different features. In Figure 6.10, we can notice examples of an increase in feature decodability as we add more ROIs in the test, but that is certainly not the major trend in the plots. Overall, the more ROIs we add, the decoding results decrease. Because the level of noise also goes up (many of the sources being concatenated are noisy), this result is not surprising. The decodability of many features also stays relatively unchanged for many combinations of ROIs, which seems to indicate that the information in the ROIs is very similar. This again is not alarming, since the data from source localization comes from a linear combination of the original data, which by themselves are correlated in nearby sensors in the helmet.

The most interesting result then is the fact that some of the features do show better decodability by combining a small number of ROIs. For example, feature *can you pick it up?* is best decodable when combining the data for 2 ROIs (subject S6), and *is it alive?* by combining 4 to 5 ROIs for many subjects. Unfortunately we could not observe any clear distinctions among

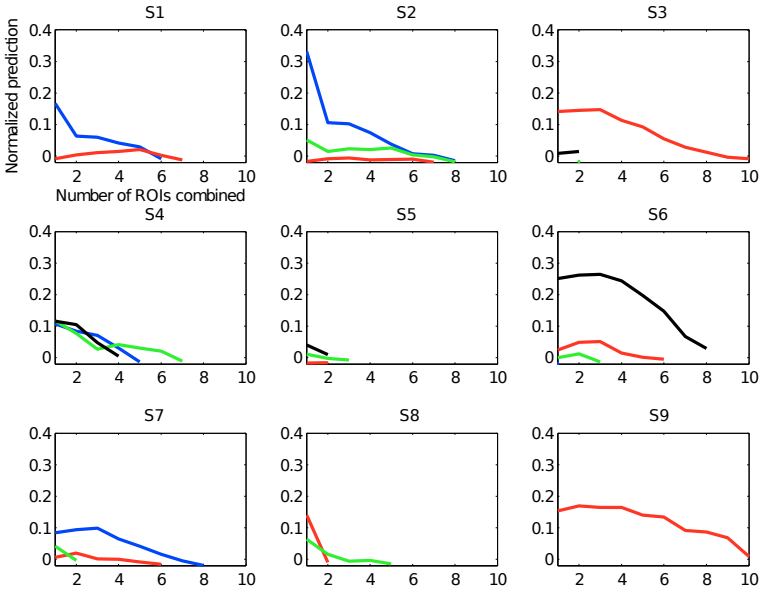


Figure 6.11: POVE for 2 perceptual and 2 semantic features in the *answer-questions* paradigm. One plot per subject, where the Y axis represents the maximum POVE obtained by combining the number of ROIs shown in the X axis. Data from the top ROI at decoding the features, plus 9 other ROIs are used in the test. Blue is *word length*, red is *right diagonalness*, green is *is it alive?* and black is *can you pick it up?*. Adding data from different ROIs to the decoder does not necessarily increase feature decodability.

perceptual and semantic features regarding which one benefits the most from combining ROIs.

Finally, it is possible that the challenges related to spatial localization in MEG, especially when it comes to the spread of activation over ROIs positioned close together, might have contributed to these results. Namely, the same activation might have been localized to sources that spill over different ROIs, resulting on different ROIs coding the same information. Regardless of this shortcoming, the current results suggest that ROIs showing parallel high decodability for a feature can encode complementary information, because in many cases the decodability of a feature increases when using data from different ROIs simultaneously. Therefore, these regions in the brain might encode complementary information about each feature. This also corroborates the observation previously made in chapter 2, when the 2-vs-2 decoding accuracies over time using the whole brain data were higher than the decoding with any individual ROI by itself.

6.2.2 H2: For regions that encode a feature in different time windows, the neural encoding is different at those times.

Before we test this hypothesis, it is important to be clear about what the neural encoding being tested means. When we state that a region encodes a feature using model A, what we are actually saying is that when predicting the value of a feature using all sources in that region (for a particular time window), we get a significant POVE value for that feature. Therefore, a direct way to test this hypothesis is to compare the weight distribution across the time windows when the region encoded the feature.

We focused this analysis on two representative semantic features: *is it alive?* and *can you pick it up?*. For each feature, we choose the ROIs with best decodability but that also show an

interesting temporal profile (i.e. show high feature decodability at different, non-consecutive time windows), and get the decoding weights for the time window with highest decoding (i.e. seed time window). Then, we compute the decoding weights for every time window from -100 to 600 ms in non-overlapping intervals of 50 ms, and calculate the cosine distance between the set of weights and the weights for the seed time window. This measure is then averaged over all subjects. Figure 6.12 shows this distance metric for the two pre-selected semantic features.

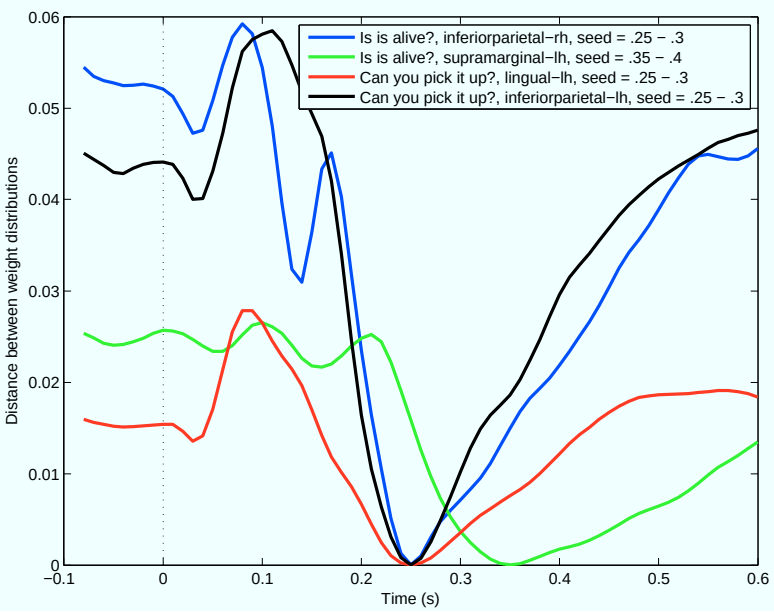


Figure 6.12: Differences in the neural code between time windows and a seed window in the *answer-questions* paradigm. Y axis represents the cosine distance between the weight distribution of the seed time window and the other time windows, averaged over subjects. Different time windows seem to have different neural code, regardless of how well they decode the feature. Time 0 is when the stimulus is presented to subjects.

By definition, the distance between the weight distribution in the seed time window and itself is zero for all feature-ROI combinations. If there were other time windows that shared a similar neural code to the seed window, one would expect to see the distance dipping towards zero again in the plot. However, that does not seem to be the case. In fact, weight distributions seem to be closer to the seed window's distribution in the time windows adjacent to the seed. It is true that we commonly see adjacent time windows with good feature decodability, and that seems to correlate with the distance to the seed (i.e. if adjacent time windows have good feature decoding, their weight distributions is also similar). But it is also true that often times a feature shows good decodability in quite distant time windows. For example, *is it alive?*, as depicted in Figure 6.9, appears to be significantly encoded in the right inferior parietal cortex at 250 and also at 450 ms, but the weight distributions in those time windows do not seem to be close (blue trace in Figure 6.12), at least not when compared to other time windows of the same feature-ROI pair.

Therefore, the current results suggest that good decodability for a feature using MEG data for a region at different time points does not necessarily imply that the neural encoding is similar at the different instants. This conclusion becomes stronger the farther apart the time windows are chosen.

Further analysis of Figure 6.9 prompts a few more questions that will be left for future work.

For example, when two regions encode a feature in parallel, are they more functionally connected at those times? Several measures of functional connectivity should be taken into consideration [17, 46, 66] when answering this question, because it is unclear whether the connectivity pattern would favor one technique over the other. Moreover, specifically regarding the *answer-questions* paradigm, it would be interesting to investigate the effects of the question being asked on the decodability of the related semantic feature. The paradigm would have to be re-designed to remove any artifacts created by block effects, but it would be interesting to find out whether the semantic feature related to the question being asked is among the top decoded semantic features using those data. Also, it is unclear whether the question would alter the timing of the feature decodability itself. More specifically, can we decode the semantic feature related to the question being asked earlier than other semantic features? What about with relation to the *freely-think* paradigm?

6.2.3 H3: Animacy-related features are encoded earlier in the neural code than other types of features.

Throughout this dissertation we have seen results that suggested an order of decobability among semantic features. More specifically, semantic features related to animacy seemed to be decoded earlier than other semantic features.

It is not surprising that animacy-related features might be preferentially encoded in the brain. Humans start distinguishing between animate and inanimate objects as young as 5-month-old [44]. Much of the work investigating animacy in the brain involves the analysis of different nouns within a context, such as in sentence processing. For example, it has been reported that animacy violations in a sentence affect both the N400 and the P600 evoked responses [85], and these changes can be seen as early as 200 ms after the stimulus onset [98]. There have also been reports of semantic properties reflected by neurophysiological indicators within the first 200 ms after word onset [73], and saccadic eye movements towards scenes containing an animal have been shown to occur in as little as 120 ms [40].

Although all these studies have shown very early effects of animacy in the brain signal, it is important to note that they mostly investigate the role of animacy by contrasting it to other categories, or by analyzing words or pictures within a context. Picture naming, on the other hand, can be performed outside of an underlying context for the different nouns, and living items are still named faster than nonliving ones [49]. Similarly, the name of living items is matched to their pictures faster than for nonliving concepts [13]. Such paradigms are more similar to the ones we analyzed in this dissertation because we decode different semantic features from MEG data while subjects were presented with concrete noun stimuli outside of a particular context.

Here, we use the methodological framework described in this dissertation as a tool to compare the timing of animacy-related features to other features that describe concrete nouns. More specifically, we will test the following two hypotheses:

1. The decoding peak of *is it alive?* over time is earlier than the decoding peak of other semantic features.
2. The earliest significantly-decoded time window for *is it alive?* is earlier than the one for other semantic features.

The semantic feature *is it alive?* is used as a prototypical exemplar of animacy-related features, and the other features to which *is it alive?* is compared are selected based on their correlation with *is it alive?*. More specifically, the Pearson's correlation coefficient was calculated between *is it alive?* and each of the remaining 217 semantic features. Of those, 87 features did not show enough evidence to reject the null hypothesis of no correlation (alpha at 0.05), and hence were selected as uncorrelated features for further analysis. This procedure corresponds to selecting all semantic features for which the absolute correlation to *is it alive?* was not greater than 0.25.

The decoding analysis was conducted in a per-subject basis, using model A to predict the semantic features from MEG data. A separate model was trained for each individual 50-ms time window (non-overlapping), starting at zero (stimulus unset) until 800 ms. Feature scores were calculated using equation 2.4, and the significance of these scores was established using the same model, but employing MEG data with permuted labels as its input (hundreds of different permutations). Data from *answer-questions* and *freely-think* experiments were used in the analysis.

Subject: paradigm	Is it alive?	Others (mean)	Others (std.)	Difference	Features
S1: answer-questions	150	313	184	-163	58
S2: answer-questions	100	269	160	-169	71
S3: answer-questions	200	286	176	-86	62
S4: answer-questions	200	270	165	-70	66
S5: answer-questions	150	224	174	-74	70
S6: answer-questions	150	238	178	-88	67
S7: answer-questions	150	249	152	-99	64
S8: answer-questions	250	237	184	13	58
S9: answer-questions	150	272	160	-122	61
S1: freely-think	250	362	176	-112	52
S2: freely-think	200	406	160	-206	63
S3: freely-think	400	402	176	-2	56
S4: freely-think	150	396	176	-246	62
S5: freely-think	200	319	172	-119	64
S6: freely-think	250	377	171	-127	56
S7: freely-think	250	418	186	-168	50
S8: freely-think	*	439	146	*	36
S9: freely-think	400	405	161	-5	52
Mean:	212	327	170	-108	59

Table 6.3: Start time of the first significantly-decoded window (in milliseconds) for *is it alive?* and for the other 87 uncorrelated semantic features. Each time window is 50-ms wide, taken at steps of 50 ms, and time zero indicates stimulus onset. Difference values are taken between columns 2 and 3. Features column shows how many of the 87 uncorrelated features were significantly decoded for each subject. Subject S8 did not have a significant decoding window ($p < .05$) for *is it alive?* in the *freely-think* paradigm.

Table 6.3 shows the timing of the first significantly-decoded time window ($p < .05$) of *is it alive?* and the uncorrelated semantic features for all subjects. Note that the data for subject S8 in *freely-think* did not predict *is it alive?* better than chance, so it was excluded from the analysis. The columns reporting the mean and standard deviation for the uncorrelated features were computed by first extracting the starting point of the earliest significant time window for each of the features that were significantly predicted for that subject (e.g. 58 time points for the first data row in the table), and then the mean or standard deviation was taken over all those values. The feature *is it alive?* was first significantly-decoded on average at around 212 ms, while the uncorrelated semantic features were first significantly-decoded at around 327 ms (mean over subjects).

Finally, a Wilcoxon signed rank test was used to test the null hypothesis that the difference between the earliest significantly-decoded time windows (columns 2 and 3 in Table 6.3), for any

random subject, comes from a distribution with zero median. The one-sided alternative hypothesis states that the time windows in which *is it alive?* is first significantly decoded occur earlier than the ones for the other semantic features. The results of this analysis show that **the earliest significantly-decoded time window for *is it alive?* happens earlier than for the uncorrelated semantic features ($p < .001$, right-tail)**.

Subject: paradigm	Is it alive?	Others (mean)	Others (std.)	Difference	Features
S1: answer-questions	500	472	173	28	58
S2: answer-questions	300	409	166	-109	71
S3: answer-questions	500	365	157	135	62
S4: answer-questions	650	420	198	230	66
S5: answer-questions	400	359	190	41	70
S6: answer-questions	700	415	165	285	67
S7: answer-questions	450	384	175	66	64
S8: answer-questions	350	377	200	-27	58
S9: answer-questions	350	344	175	6	61
S1: freely-think	500	444	192	56	52
S2: freely-think	200	499	169	-299	63
S3: freely-think	400	424	178	-24	56
S4: freely-think	150	482	158	-332	62
S5: freely-think	200	373	179	-173	64
S6: freely-think	400	421	173	-21	56
S7: freely-think	250	504	174	-254	50
S8: freely-think	*	464	138	*	36
S9: freely-think	650	462	156	188	52
Mean:	409	423	173	-12	59

Table 6.4: Start time of the decoding peak window (in milliseconds) for *is it alive?* and for the other 87 uncorrelated semantic features. Each time window is 50-ms wide, taken at steps of 50 ms, and time zero indicates stimulus onset. Difference values are taken between columns 2 and 3. Features column shows how many of the 87 uncorrelated features were significantly decoded for each subject. Subject S8 did not have a significant decoding window ($p < .05$) for *is it alive?* in the *freely-think* paradigm.

A similar analysis was performed using the timing of the peak decoding window for *is it alive?* and the other semantic features (Table 6.4). The feature *is it alive?* reached its peak decoding time at around 409 ms, while the uncorrelated semantic features reached their peak decoding time at around 423 ms. As it can be observed in the table, the peak decoding for *is it alive?* does not occur significantly earlier than for the uncorrelated semantic features. Figure 6.13 provides a graphical interpretation of the same data, illustrating the differences between analyzing the sequence of activation of semantic features using the earliest significantly-decoded window and the time of the decoding peak.

These results provide compelling evidence that animacy is encoded substantially earlier than dozens of other features. Furthermore, they show that **the timing difference is between the first encodings of the features, and not in the timing of their peak accuracies**. One could conjecture that the "understanding" of a word requires the peak of several different semantic features, which happen generally at around the same time, but some features start being recruited earlier than others, which is the case for animacy. We suggest that this ordering among semantic features might be motivated by evolutionary constraints, since it would have been advantageous to quickly discriminate between living and non-living objects, because quickly recognizing and reacting to dangerous living things (e.g. a tiger in the bushes) would have helped our ancestors survive.

The novelties of these results are two-fold. First, we assess the recruitment order of several different features of the stimuli for a wide range of nouns, without an implicit comparison between classes built into the task performed by the subjects. That is complementary to most of

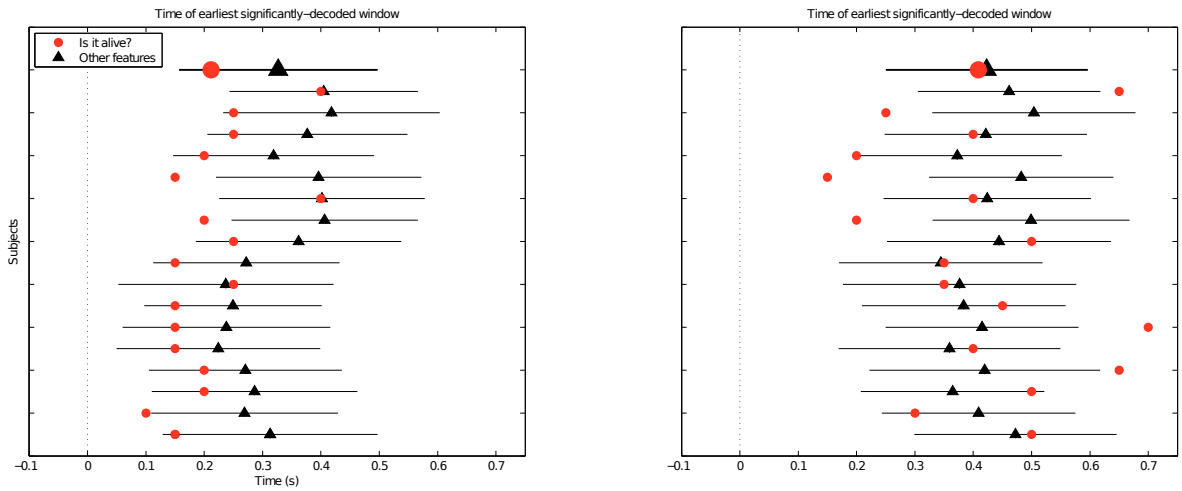


Figure 6.13: Means and standard deviations for comparing the sequence of activation between *is it alive?* and other uncorrelated semantic features. Animacy-related features are encoded significantly earlier than other types of features when examining the earliest significantly-decoded time window, but not when the peak decodability window is analyzed. Each time window is 50-ms wide, taken at steps of 50 ms, starting at -0.1 s. For example, the time point at 0.3 s in the graphs shows results for the window 300 - 350 ms. Y-axis represents each individual subject in the *answer-questions* and *freely-think* experiments, except for subject S8 in the latter (see text for more details). Top-most result in y-axis (thicker lines) show the mean over all subjects. Mean and standard deviation (black markers and bars) are taken over significantly-decoded features uncorrelated to *is it alive?* (last columns in Tables 6.3 and 6.4). Time zero indicates stimulus onset.

the previous work that has been done in the field, which predominantly measured the reaction time at different tasks for specific classes of objects (e.g. a picture-text matching task using only living/non-living and manipulable/non-manipulable items [13]). Secondly, to the best of our knowledge, this is the first time a distinction between the timing of the peak of decodability and the earliest significantly-decoded time window is made. Looking at these two characteristics of the feature decodability curve seems to be valuable in understanding the data, and also helpful for conjecturing a model of word comprehension in which a specific set of features, such as animacy, might first be inferred separately with low confidence. Later, during peak accuracy, these features might have their confidence increased by integrating evidence of other types of features into a coherent set that represents the concrete noun.

Chapter 7

Conclusions

In this concluding chapter we review the main methodological and scientific contributions of this dissertation, and list some future work that will help develop the ideas explored here even further. Finally, we draw on the collection of results shown in this dissertation and provide our conjecture of how the brain encodes the representation of concrete objects.

7.1 Methodological conclusions

The models described here help the study of the representation of concrete objects in the brain by providing:

- a prediction framework to test how well a set of perceptual and semantic features explains the observed MEG activity at individual space and time points, along with how much each of these features contributes to the predictions (i.e. their neural signatures). These measures can be used to test different models of knowledge representation, under the assumption that the better the model, the more accurate the predictions of the neural activity will be. Also, by studying the neural signature for each feature, one can compare what regions in the brain and at what time points that feature influences the neural signal, in the context of the other features present in the model.
- a decoding task that, combined with the prediction framework, tests composition hypothesis about the neural code by predicting the neural activity of nouns that the model has never seen during training. This neural activity is predicted over time and space, and makes no assumptions about what regions in space or time are best at encoding the information, so that the experimenter can drawn her own conclusions about distributive properties in the neural signal.
- the ability to generalize to words within a category and still distinguish between these two novel words with results above chance accuracy.
- a model that tracks what features are encoded in each specific region of the brain at different time points. This model can be used to directly test previous claims in the literature about which regions encode specific types of information, and helps to answer deeper neuroscientific questions, such as the recruitment order of different features when processing

the meaning of a concrete noun.

Additionally, there were a few practical advantages to using model B (i.e. predicting neural data using a perceptual and semantic feature set as input), especially regarding the interpretation of later results. We have also shown that it is possible to locate in space and time what parts of the MEG signals are particularly stable across repetitions of the stimuli, and we can improve the decoding results in model B by restricting the features predicted to only the most stable ones. These stable parts of the signal are very consistent across subjects, and to a certain extent also consistent across paradigms that show the same stimuli. Based on their temporal and spatial location, as well as on the fact that restricting the analysis to these consistent patterns improves results, we suggest that these stable patterns predominantly encode the perceptual and semantic attributes of the nouns being processed.

Another technique that showed promising results in increasing decoding results was concatenation, both over subjects and over nouns. In the future, it will be important to try different transformations of the MEG data, as well as different feature sets, which might have better results with the single trial paradigms. Furthermore, new signal processing techniques will be crucial to reduce the amount of noise in single trials.

7.2 Conclusions about the representation of objects in the neural signal

In chapter 1 we outlined a few questions to be investigated in this dissertation. Here we will summarize the answers to those questions and review the evidence on which we draw the conclusions.

- *Is there enough information in the MEG signal to decode different properties of concrete nouns?* Yes. Our predictions of semantic features using MEG data (model A) were accurate enough that we could distinguish between two novel nouns with an accuracy of 91% (average over subjects). When predicting MEG data using the set of perceptual and semantic features (model B), we achieved accuracies between 80 and 90%, depending on the experimental paradigm used.
- *Can we leverage the compositional aspect of the code to predict nouns our models have never seen during training?* Yes. The highly significant accuracy results of both models corroborate the ideas about a compositional code in the brain, instead of an unitary system (e.g. an extrapolation of the grandmother cell idea). Through a cross-validation framework in which examples of predicted nouns are never seen during training, the generalization to novel nouns depends on the compositionality assumption. Our results are strong across paradigms to infer that the neural code for the meaning of nouns is based on a collection of properties recruited at different degrees to represent the individual objects.
- *What is the flow of information in different regions of the brain at specific time points?* Across experimental paradigms, there was a gradual process of recruitment of different regions of the brain, which encoded different perceptual and semantic properties. Features related to the physical attributes of the stimulus (i.e. perceptual features, such as *word*

length and *number of white pixels*) are encoded in the early parts of the signal (before 250 ms have elapsed since stimulus onset), in regions related to low-level signal processing, mostly in the posterior regions of the brain, such as the bilateral occipital cortex and lingual gyri. On the other hand, features that are inherent to the nouns (i.e. semantic features, such as *is it alive?* and *can you pick it up?*) are better decoded in time points after 250 ms, using data from regions more anterior in the brain, such as the left middle temporal cortex and the left supra marginal region.

- *How do the results compare to previous findings in the literature?* While some of the regions shown to encode categories of features agreed with the fMRI literature, that was not true for all of them. For example, features related to a specific location, such as *would you find it in an office?* and *would you find it in a school?* were consistently decoded using data from the parahippocampal region, which has been shown in the literature to represent shelter-like semantic factors [38]. However, regions normally associated with eating-like features (e.g. left middle frontal gyrus) did not show up in our results as such. Similarly, many regions frequently reported in the fMRI literature to be important in semantic processing, such as the left pars-triangularis region, did not show good results decoding semantic features in general using MEG data. We take these results as complementary instead of opposing, because there are several plausible explanations to why different results should be expected (e.g. different signals being measured, different sensitivity to the areas being discussed, different paradigms, etc).
- *What is the spatiotemporal distribution of the neural code for these different properties?* Our results corroborate the claims of a distributed neural code for representing concrete objects. This distributed representation shows the contribution of different perceptual and semantic properties at various space and time points to represent a given noun. For example, the same feature can be decoded from different time points and regions in the MEG signal, and it is also true that particular regions in the brain participate in the encoding of multiple features. Furthermore, we not only investigated the start time for feature decodability, but also the temporal profile for decoding individual features. We noticed several features that are encoded for sustained amounts of time, while some seem to appear and disappear from focus. When the latter happens, our results suggest that the neural code changes among the different time windows. It was unclear from the results whether regions that encode a feature in parallel, at a particular time window, provide complementary or redundant information about that feature.
- *Is there a sequence of activation among properties inherent to the nouns?* Yes. Our results show that animacy-related features, such as *is it alive?*, are recruited earlier than other types of semantic features, and we conjecture that evolutionary constraints might be behind such ordering among different types of features. More specifically, the time window in which *is it alive?* was first significantly-decoded happened consistently earlier than the window for other semantic features, but this relationship was not observed when analyzing the time window in which the decodability of the semantic features peaked.
- *How do the differences among experimental paradigms affect the results?* The methods explored in this dissertation were used to analyze data from 4 different paradigms. Despite

the issues described with one of them (*1back-speech*), the results reported for the other 3 datasets were remarkably similar. Although the neuroscience conclusions drawn from *answer-questions* and *freely-think* are limited by the shortcomings of the two paradigms, it is important that many of those findings were also observed with degrees of variation in *1back-text*, especially considering that the paradigm did not involve an explicit semantic task. We should also consider that the predictions only worked above chance accuracy for 2 subjects in that paradigm, but we strongly believe that these results can be improved in the future by using the methods previously mentioned to better separate the weaker signals from the noise. The model also provided meaningful results without dependencies on the task being analyzed, but the fewer repetitions used, the sparser the neural signatures derived by the model B became. As expected, decreasing the number of repetitions used in each paradigm also affected decoding results, since the signal-to-noise ratio of each brain image decreases as fewer repetitions of the nouns are averaged together. When equalizing the number of repetitions between *answer-questions* and *freely-think*, data from the latter yielded higher decoding accuracies in discriminating between 2 novel nouns.

Our findings about the representation of concrete objects in the neural signal complement many of the current views in the literature by going beyond the description of where the information is encoded. It provides a richer description of the representation by distinguishing between perceptual and semantic parts of the information flow over time and space. Future studies should be designed to better untangle the order of activation of different semantic features, and well as to further investigate the effects of task and type of stimulus in the neural encoding of concrete nouns.

7.3 On the neural code for representing concrete nouns

We have presented several results based on the data of different paradigms to shed light into the spatiotemporal characteristics of the neural representation of concrete objects. The 2-vs-2 decoding results were used to show the generalization capabilities of the models to predict nouns with which they have never been trained before, as well as to test compositional assumptions about the neural code. Then, percent of variance explained and scaled regression weights were explored to analyze how much of the MEG signal (or the perceptual and semantic features) the models could explain, and to further understand what parts of the model contributed to the predictions. Through this combination of results we were able to track when and where perceptual and semantic features are encoded in the MEG data while subjects processed the meaning of concrete nouns across paradigms. This last section integrates these results to what has been shown in the literature to form a series of conjectures about how the brain processes the meaning of concrete nouns.

Regardless of the paradigm or the task being executed, when presented with a stimulus representing a concrete noun the low level sensory regions are the first to respond to the stimulus. The regions involved in this early processing vary based on the stimuli being presented. For word-only stimuli, regions such as the bilateral occipital cortex are activated at around 100 ms, and then the occipitotemporal junctions are activated at around 150 ms. For picture stimuli, occipital and ventro-temporal regions are the first to respond, before 175 ms following stimulus onset.

These early processing, posterior regions encode perceptual features of the stimuli, such as word length and features related to the shape of the object.

Between 200 and 250 ms a mixture of perceptual and semantic features are encoded in the neural activity. This could be partly because there is no real boundary between perceptual and semantic features, other than what is artificially pre-defined by researchers. However, after 250 ms, most of the spatiotemporal code reflects semantic properties of the different objects. Several regions are involved in encoding these semantic features. They are usually more anterior than the regions involved in perceptual processing, and a few of these regions consistently appear to be involved in encoding many semantic features, such as the left supra marginal gyrus and the left middle temporal cortex. It is also true that some semantic features, such as animacy, size, and manipulability-related features are more robustly encoded in the neural code, and there is also some evidence that these semantic features are encoded earlier than others. The robustness of the code for a given feature, as well as its timing, are likely dependent on our representation of the world as we interact with it, and on evolutionary clues (e.g. *is it alive* and *can you pick it up* seem like very important properties to quickly figure out about objects as they come into scene).

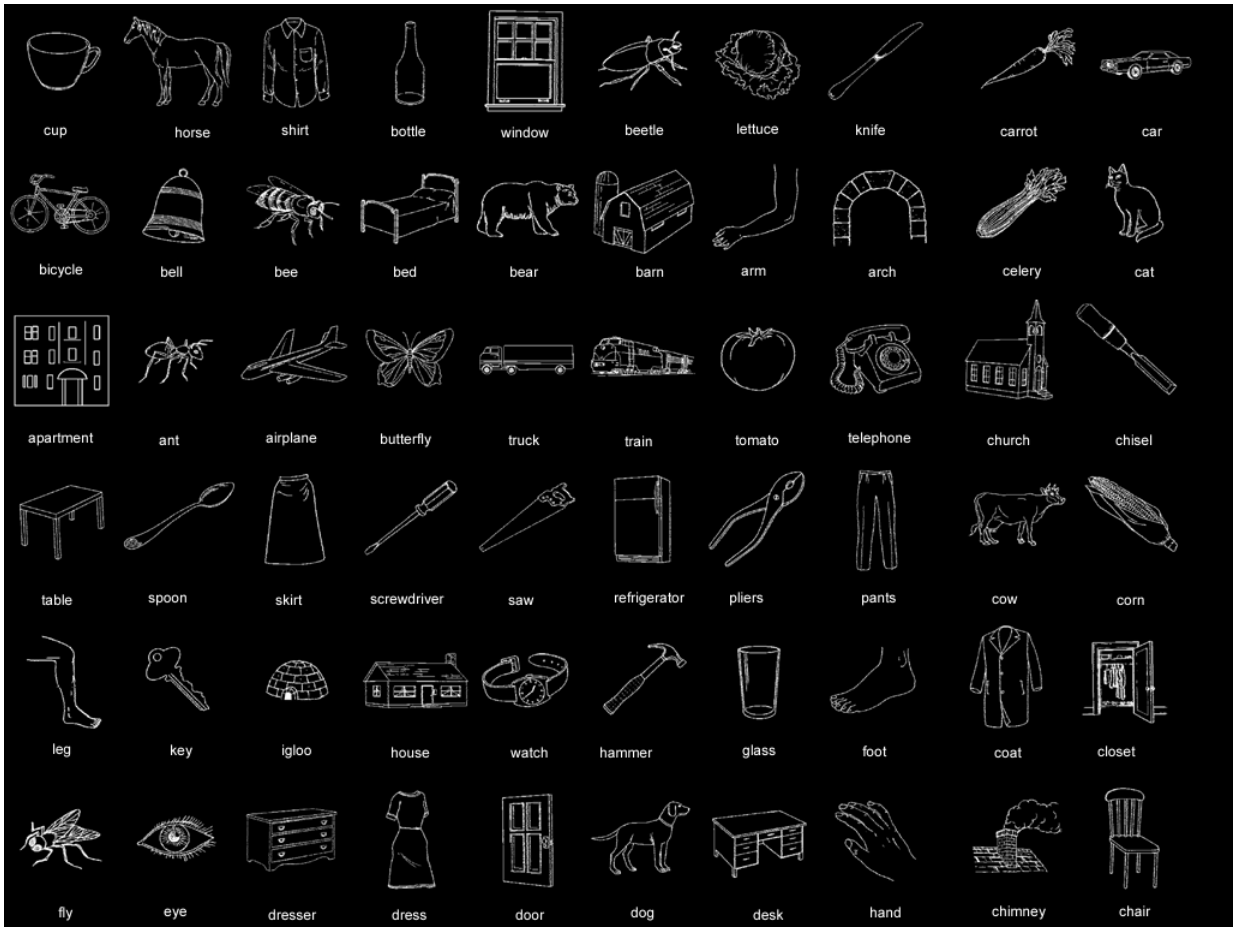
Moreover, the neural code is distributed in the sense that several regions can encode complementary parts of a feature, but also because a given region is often involved in the encoding of many features. Another feature of this code is compositionality. The brain encodes different nouns not by activating regions that are specific to the word, but that encode features of the object (e.g. *is it alive?*, *can you pick it up?*). Because of that, we are able to train classifiers on how the brain represents features of a set of words, and then predict how the activity of novel words would look like, since they are a combination of these learned features. Such extrapolation would not be possible if the representation of nouns were specific to each noun (e.g. some generalization of the grandmother cell argument).

The regions in the brain that are responsible for encoding these features are similar to the ones involved in processing them in the first place (e.g. motor cortex for manipulability features). That is likely related to a simple mechanism of repeating the brain patterns that occurred during learning when you think of the word. To illustrate, say you go to a zoo and see an animal for the first time. Without knowing its name, based on its features (e.g. four legs, furry) you can probably infer basic properties, such as how it moves or if you can pick it up. These predictions are automatic because patterns that encode these basic properties are also associated with other animals that share the same perceptual attributes. On the other hand, if you saw an animal whose perceptual features you had never encountered before, you would have difficulties making any predictions about it. This is another way of saying that semantic features follow perceptual features based on our past experiences, and they are encoded in regions that were also involved in the neural pattern during learning. It is unlikely that we evoke every property of a noun automatically, but there is a core activation that is present every time the object comes into scene, and the context in which the object is presented (e.g. experimental task, sentence, etc) selects what other properties should be evoked.

It is still unclear, however, how these features are put together by the brain to form the meaning of words, and also how the transition between perceptual and semantic features truly happens. New paradigms and methods will need to be employed in the future to specifically address and provide answers to these important questions.

Appendix A

Set of 60 line drawings



Appendix B

Set of 20 questions

- Is it manmade?
- Is it made of metal?
- Is it hollow?
- Is it hard to catch?
- Does it grow?
- Was it ever alive?
- Could you fit inside it?
- Does it have at least one hole?
- Can you hold it?
- Is it bigger than a loaf of bread?
- Does it live in groups?
- Can it keep you dry?
- Is part of it made of glass?
- Is it bigger than a car?
- Can you hold it in one hand?
- Is it manufactured?
- Is it bigger than a microwave oven?
- Is it alive?
- Does it have feelings?
- Can you pick it up?

Appendix C

List of semantic and perceptual features

Single features

Semantic features

1. IS IT AN ANIMAL?
2. IS IT A BODY PART?
3. IS IT A BUILDING?
4. IS IT A BUILDING PART?
5. IS IT CLOTHING?
6. IS IT FURNITURE?
7. IS IT AN INSECT?
8. IS IT A KITCHEN ITEM?
9. IS IT MAN-MADE?
10. IS IT A TOOL?
11. CAN YOU EAT IT?
12. IS IT A VEHICLE?
13. IS IT A PERSON?
14. IS IT A VEGETABLE / PLANT?
15. IS IT A FRUIT?
16. IS IT MADE OF METAL?
17. IS IT MADE OF PLASTIC?
18. IS PART OF IT MADE OF GLASS?
19. IS IT MADE OF WOOD?
20. IS IT SHINY?
21. CAN YOU SEE THROUGH IT?
22. IS IT COLORFUL?

23. DOES IT CHANGE COLOR?
24. IS ONE MORE THAN ONE COLORED?
25. IS IT ALWAYS THE SAME COLOR(S)?
26. IS IT WHITE?
27. IS IT RED?
28. IS IT ORANGE?
29. IS IT FLESH-COLORED?
30. IS IT YELLOW?
31. IS IT GREEN?
32. IS IT BLUE?
33. IS IT SILVER?
34. IS IT BROWN?
35. IS IT BLACK?
36. IS IT CURVED?
37. IS IT STRAIGHT?
38. IS IT FLAT?
39. DOES IT HAVE A FRONT AND A BACK?
40. DOES IT HAVE A FLAT / STRAIGHT TOP?
41. DOES IT HAVE FLAT / STRAIGHT SIDES?
42. IS TALLER THAN IT IS WIDE/LONG?
43. IS IT LONG?
44. IS IT POINTED / SHARP?
45. IS IT TAPERED?
46. IS IT ROUND?
47. DOES IT HAVE CORNERS?
48. IS IT SYMMETRICAL?
49. IS IT HAIRY?
50. IS IT FUZZY?
51. IS IT CLEAR?
52. IS IT SMOOTH?
53. IS IT SOFT?
54. IS IT HEAVY?
55. IS IT LIGHTWEIGHT?
56. IS IT DENSE?

57. IS IT SLIPPERY?
58. CAN IT CHANGE SHAPE?
59. CAN IT BEND?
60. CAN IT STRETCH?
61. CAN IT BREAK?
62. IS IT FRAGILE?
63. DOES IT HAVE PARTS?
64. DOES IT HAVE MOVING PARTS?
65. DOES IT COME IN PAIRS?
66. DOES IT COME IN A BUNCH/PACK?
67. DOES IT LIVE IN GROUPS?
68. IS IT PART OF SOMETHING LARGER?
69. DOES IT CONTAIN SOMETHING ELSE?
70. DOES IT HAVE INTERNAL STRUCTURE?
71. DOES IT OPEN?
72. IS IT HOLLOW?
73. DOES IT HAVE A HARD INSIDE?
74. DOES IT HAVE A HARD OUTER SHELL?
75. DOES IT HAVE AT LEAST ONE HOLE?
76. IS IT ALIVE?
77. WAS IT EVER ALIVE?
78. IS IT A SPECIFIC GENDER?
79. IS IT MANUFACTURED?
80. WAS IT INVENTED?
81. WAS IT AROUND 100 YEARS AGO?
82. ARE THERE MANY VARIETIES OF IT?
83. DOES IT COME IN DIFFERENT SIZES?
84. DOES IT GROW?
85. IS IT SMALLER THAN A GOLFBALL?
86. IS IT BIGGER THAN A LOAF OF BREAD?
87. IS IT BIGGER THAN A MICROWAVE OVEN?
88. IS IT BIGGER THAN A BED?
89. IS IT BIGGER THAN A CAR?
90. IS IT BIGGER THAN A HOUSE?

91. IS IT TALLER THAN A PERSON?
92. DOES IT HAVE A TAIL?
93. DOES IT HAVE LEGS?
94. DOES IT HAVE FOUR LEGS?
95. DOES IT HAVE FEET?
96. DOES IT HAVE PAWS?
97. DOES IT HAVE CLAWS?
98. DOES IT HAVE HORNS / THORNS / SPIKES?
99. DOES IT HAVE HOOVES?
100. DOES IT HAVE A FACE?
101. DOES IT HAVE A BACKBONE?
102. DOES IT HAVE WINGS?
103. DOES IT HAVE EARS?
104. DOES IT HAVE ROOTS?
105. DOES IT HAVE SEEDS?
106. DOES IT HAVE LEAVES?
107. DOES IT COME FROM A PLANT?
108. DOES IT HAVE FEATHERS?
109. DOES IT HAVE SOME SORT OF NOSE?
110. DOES IT HAVE A HARD NOSE/BEAK?
111. DOES IT CONTAIN LIQUID?
112. DOES IT HAVE WIRES OR A CORD?
113. DOES IT HAVE WRITING ON IT?
114. DOES IT HAVE WHEELS?
115. DOES IT MAKE A SOUND?
116. DOES IT MAKE A NICE SOUND?
117. DOES IT MAKE SOUND CONTINUOUSLY WHEN ACTIVE?
118. IS ITS JOB TO MAKE SOUNDS?
119. DOES IT ROLL?
120. CAN IT RUN?
121. IS IT FAST?
122. CAN IT FLY?
123. CAN IT JUMP?
124. CAN IT FLOAT?

125. CAN IT SWIM?
126. CAN IT DIG?
127. CAN IT CLIMB TREES?
128. CAN IT CAUSE YOU PAIN?
129. CAN IT BITE OR STING?
130. DOES IT STAND ON TWO LEGS?
131. IS IT WILD?
132. IS IT A HERBIVORE?
133. IS IT A PREDATOR?
134. IS IT WARM BLOODED?
135. IS IT A MAMMAL?
136. IS IT NOCTURNAL?
137. DOES IT LAY EGGS?
138. IS IT CONSCIOUS?
139. DOES IT HAVE FEELINGS?
140. IS IT SMART?
141. IS IT MECHANICAL?
142. IS IT ELECTRONIC?
143. DOES IT USE ELECTRICITY?
144. CAN IT KEEP YOU DRY?
145. DOES IT PROVIDE PROTECTION?
146. DOES IT PROVIDE SHADE?
147. DOES IT CAST A SHADOW?
148. DO YOU SEE IT DAILY?
149. IS IT HELPFUL?
150. DO YOU INTERACT WITH IT?
151. CAN YOU TOUCH IT?
152. WOULD YOU AVOID TOUCHING IT?
153. CAN YOU HOLD IT?
154. CAN YOU HOLD IT IN ONE HAND?
155. DO YOU HOLD IT TO USE IT?
156. CAN YOU PLAY IT?
157. CAN YOU PLAY WITH IT?
158. CAN YOU PET IT?

159. CAN YOU USE IT?
160. DO YOU USE IT DAILY?
161. CAN YOU USE IT UP?
162. DO YOU USE IT WHEN COOKING?
163. IS IT USED TO CARRY THINGS?
164. CAN YOU PICK IT UP?
165. CAN YOU CONTROL IT?
166. CAN YOU SIT ON IT?
167. CAN YOU RIDE ON/IN IT?
168. IS IT USED FOR TRANSPORTATION?
169. COULD YOU FIT INSIDE IT?
170. IS IT USED IN SPORTS?
171. DO YOU WEAR IT?
172. CAN IT BE WASHED?
173. IS IT COLD?
174. IS IT COOL?
175. IS IT WARM?
176. IS IT HOT?
177. IS IT UNHEALTHY?
178. IS IT HARD TO CATCH?
179. CAN YOU PEEL IT?
180. CAN YOU WALK ON IT?
181. CAN YOU SWITCH IT ON AND OFF?
182. CAN IT BE EASILY MOVED?
183. DO YOU DRINK FROM IT?
184. DOES IT GO IN YOUR MOUTH?
185. IS IT TASTY?
186. IS IT USED DURING MEALS?
187. DOES IT HAVE A STRONG SMELL?
188. DOES IT SMELL GOOD?
189. DOES IT SMELL BAD?
190. IS IT USUALLY INSIDE?
191. IS IT USUALLY OUTSIDE?
192. WOULD YOU FIND IT ON A FARM?

193. WOULD YOU FIND IT IN A SCHOOL?
194. WOULD YOU FIND IT IN A ZOO?
195. WOULD YOU FIND IT IN AN OFFICE?
196. WOULD YOU FIND IT IN A RESTAURANT?
197. WOULD YOU FIND IN THE BATHROOM?
198. WOULD YOU FIND IT IN A HOUSE?
199. WOULD YOU FIND IT NEAR A ROAD?
200. WOULD YOU FIND IT IN A DUMP/LANDFILL?
201. WOULD YOU FIND IT IN THE FOREST?
202. WOULD YOU FIND IT IN A GARDEN?
203. WOULD YOU FIND IT IN THE SKY?
204. DO YOU FIND IT IN SPACE?
205. DOES IT LIVE ABOVE GROUND?
206. DOES IT GET WET?
207. DOES IT LIVE IN WATER?
208. CAN IT LIVE OUT OF WATER?
209. DO YOU TAKE CARE OF IT?
210. DOES IT MAKE YOU HAPPY?
211. DO YOU LOVE IT?
212. WOULD YOU MISS IT IF IT WERE GONE?
213. IS IT SCARY?
214. IS IT DANGEROUS?
215. IS IT FRIENDLY?
216. IS IT RARE?
217. CAN YOU BUY IT?
218. IS IT VALUABLE?

Perceptual features

1. Word length
2. White pixel count
3. Internal details
4. Verticality
5. Horizontalness
6. Left-diagonalness
7. Right-diagonalness

8. Aspect-ratio: skinny->fat
9. Prickiliness
10. Line curviness
11. 3D curviness

Bibliography

[1] Hiroyuki Akama, Brian Murphy, Li Na, Yumiko Shimizu, and Massimo Poesio. Decoding semantics across fMRI sessions with different stimulus modalities: a practical MVPA study. *Frontiers in Neuroinformatics*, 6(August): 24, January 2012. ISSN 1662-5196. doi: 10.3389/fninf.2012.00024. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3426793&tool=pmcentrez&rendertype=abstract. 1.1.2

[2] Sylvain Baillet, J C Mosher, and RM Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30, 2001. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=962275. 3.2

[3] Lawrence W. Barsalou. Grounded cognition. *Annual Review of Psychology*, 59:617–45, 2008. ISSN 0066-4308. doi: 10.1146/annurev.psych.59.103006.093639. URL http://www.ncbi.nlm.nih.gov/pubmed/17705682. 1.1.2

[4] Yoav Benjamini and Daniel Yekutieli. The control of the False Discovery Rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001. URL http://www.jstor.org/stable/10.2307/2674075. 2.1.4

[5] Michel Besserve, Karim Jerbi, Francois Laurent, Sylvain Baillet, Jacques Martinerie, and Line Garnero. Classification methods for ongoing EEG and MEG signals. *Biological Research*, 40(4):415–37, January 2007. ISSN 0716-9760. doi: /S0716-97602007000500005. URL http://www.ncbi.nlm.nih.gov/pubmed/18575676. 2.1.1

[6] Augusto Buchweitz, SV Shinkareva, and RA Mason. Identifying bilingual semantic neural representations across languages. *Brain and Language*, 120(3), October 2012. ISSN 1090-2155. doi: 10.1016/j.bandl.2011.09.003. URL http://www.ncbi.nlm.nih.gov/pubmed/21978845http://www.sciencedirect.com/science/article/pii/S0093934X11001568. 1.1.4

[7] Roberto Cabeza and Lars Nyberg. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12(1):1–47, January 2000. ISSN 0898-929X. URL http://www.ncbi.nlm.nih.gov/pubmed/10769304. 1

[8] Alfonso Caramazza and Bradford Z Mahon. The organisation of conceptual knowledge in the brain: The futures past and some future directions. *Cognitive Neuropsychology*, 23 (1):13–38, 2006. 1

[9] Alexander M Chan, Eric Halgren, K Marinkovic, and Sydney S Cash. Decoding word

and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, 54(4):3028–39, October 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.10.073. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811910013819. 1.1.2, 1.1.4, 2.3.1, 5.4

[10] Douglas Cheyne, Leyla Bakhtazad, and William Gaetz. Spatiotemporal mapping of cortical activity accompanying voluntary movements using an event-related beamforming approach. *Human Brain Mapping*, 27(3):213–29, March 2006. ISSN 1065-9471. doi: 10.1002/hbm.20178. URL http://www.ncbi.nlm.nih.gov/pubmed/16037985. 2.1.2

[11] Piers Cornelissen, Antti Tarkiainen, Päivi Helenius, and Riitta Salmelin. Cortical effects of shifting letter position in letter strings of varying length. *Journal of Cognitive Neuroscience*, 15(5):731–46, July 2003. ISSN 0898-929X. doi: 10.1162/089892903322307447. URL http://www.ncbi.nlm.nih.gov/pubmed/12965046. 1.2.1

[12] Alain de Cheveigné. Time-shift denoising source separation. *Journal of Neuroscience Methods*, 189(1):113–20, May 2010. ISSN 1872-678X. doi: 10.1016/j.jneumeth.2010.03.002. URL http://www.ncbi.nlm.nih.gov/pubmed/20298717. 5.7.3

[13] Jillian H Filliter, Patricia a McMullen, and David Westwood. Manipulability and living/non-living category effects on object identification. *Brain and Cognition*, 57(1):61–5, February 2005. ISSN 0278-2626. doi: 10.1016/j.bandc.2004.08.022. URL http://www.ncbi.nlm.nih.gov/pubmed/15629216. 6.2.3, 6.2.3

[14] Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925. URL http://psychclassics.yorku.ca/Fisher/Methods. 2.1.4

[15] Elia Formisano, Federico De Martino, Milene Bonte, and Rainer Goebel. Who is saying what? Brain-based decoding of human voice and speech. *Science*, 322(November):970–973, 2008. 1.1.4

[16] Alona Fyshe. Decoding word semantics from magnetoencephalography time series transformations. Technical report, Carnegie Mellon University, 2012. URL http://www.ml.cmu.edu/research/dap-papers/dap_fyshe.pdf. 2.1.1, 2.4, 5.4

[17] Avniel Singh Ghuman, Jonathan R McDaniel, and Alex Martin. A wavelet-based method for measuring the oscillatory dynamics of resting-state functional connectivity in MEG. *NeuroImage*, 56(1):69–77, May 2011. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.01.046. URL http://www.ncbi.nlm.nih.gov/pubmed/21256967. 6.2.2

[18] Stephen J Gotts, CC Chow, and Alex Martin. Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*, pages 1–15, 2012. URL http://www.tandfonline.com/doi/abs/10.1080/17588928.2012.670617. 5.1

[19] Marcos Perreau Guimaraes, Dik Kin Wong, E Timothy Uy, P Suppes, and Logan Grosenick. Single-trial classification of MEG recordings. *IEEE Transactions on Bio-medical Engineering*, 54(3):436–43, March 2007. ISSN 0018-9294. doi: 10.1109/TBME.2006.888824. URL http://www.ncbi.nlm.nih.gov/pubmed/17355055. 1.1.4

[20] Isabelle Guyon. Kernel ridge regression. Technical Report 10, Notes on Kernel Ridge Regression. ClopiNet., August 2005. URL http://clopinet.com/isabelle/Projects/ETH/KernelRidge.pdf. 2.1.2

[21] Peter Hagoort. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 363(1493):1055–69, March 2008. ISSN 0962-8436. doi: 10.1098/rstb.2007.2159. URL http://www.ncbi.nlm.nih.gov/pubmed/17890190. 1.1.1

[22] Eric Halgren, Rupali P Dhond, Natalie Christensen, Cyma Van Petten, Ksenija Marinkovic, Jeffrey D Lewine, and Anders M Dale. N400-like magnetoencephalography responses modulated by semantic context, word Frequency, and lexical class in sentences. *NeuroImage*, 17(3):1101–1116, November 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1268. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811902912681. 1.1.1

[23] Matti S Hämäläinen and Risto J Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical and Biological Engineering and Computing*, 32(1):35–42, 1994. URL http://www.springerlink.com/index/4474345W652H8581.pdf. 1.2.2, 3.2

[24] Matti S Hämäläinen, Riitta Hari, Risto J Ilmoniemi, J Knuutila, and Ol Lounasmaa. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–507, 1993. 1, 2.3.4, 3.2

[25] Peter C Hansen, Morten L Kringelbach, and Riitta Salmelin, editors. *MEG: An Introduction to Methods*. Oxford Univ Press, 2010. 3.2

[26] Riitta Hari and Nina Forss. Magnetoencephalography in the study of human somatosensory cortical processing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 354:1145–1154, 1999. URL http://scholar.google.com/scholar?q=intitle:Magnetoencephalography+in+the+study+of+human+somatosensory+cortical+processing#0. 5.1

[27] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning*. Springer, 2011. 2.1.2, 2.1.2, 5.4

[28] Olaf Hauk. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *NeuroImage*, 21(4):1612–21, 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2003.12.018. URL http://www.ncbi.nlm.nih.gov/pubmed/15050585. 2.3.2

[29] Olaf Hauk, Matthew H Davis, Ferath Kherif, and Friedemann Pulvermüller. Imagery or meaning? Evidence for a semantic origin of category-specific brain activity in metabolic imaging. *European Journal of Neuroscience*, 27(7):1856–66, 2008. ISSN 1460-9568. doi: 10.1111/j.1460-9568.2008.06143.x. URL http://www.ncbi.nlm.nih.gov/pubmed/18380676. 1

[30] Olaf Hauk, Friedemann Pulvermüller, M Ford, W D Marslen-Wilson, and Matthew H Davis. Can I have a quick word? Early electrophysiological manifestations of psycholin-

guistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, 80(1):64–74, January 2009. ISSN 1873-6246. doi: 10.1016/j.biopsycho.2008.04.015. URL http://www.ncbi.nlm.nih.gov/pubmed/18565639. 2.3.2

[31] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–30, September 2001. ISSN 0036-8075. doi: 10.1126/science.1063736. URL http://www.ncbi.nlm.nih.gov/pubmed/11577229. 1.1.2, 2.3.1

[32] Päivi Helenius, Riitta Salmelin, Elisabet Service, and John F Connolly. Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, 121:1133–42, June 1998. ISSN 0006-8950. URL http://www.ncbi.nlm.nih.gov/pubmed/9648548. 1.1.1

[33] Päivi Helenius, Riitta Salmelin, Elisabet Service, John F Connolly, Seija Leinonen, and Heikki Lyytinen. Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *The Journal of Neuroscience*, 22(7):2936–44, April 2002. ISSN 1529-2401. doi: 20026244. URL http://www.ncbi.nlm.nih.gov/pubmed/11923458. 1.1.1, 1.2.1

[34] Christopher J Honey, Christopher R Thompson, Yulia Lerner, and Uri Hasson. Not Lost in Translation: Neural Responses Shared Across Languages. *The Journal of Neuroscience*, 32(44):15277–15283, October 2012. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1800-12.2012. URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1800-12.2012. 1.1.4

[35] Annika Hultén, Minna Vihla, Matti Laine, and Riitta Salmelin. Accessing newly learned names and meanings in the native language. *Human Brain Mapping*, 30(3):976–89, March 2009. ISSN 1097-0193. doi: 10.1002/hbm.20561. URL http://www.ncbi.nlm.nih.gov/pubmed/18412130. 1.1.1

[36] Peter Indefrey and Willem J M Levelt. The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–44, 2004. ISSN 0010-0277. doi: 10.1016/j.cognition.2002.06.001. URL http://www.ncbi.nlm.nih.gov/pubmed/15037128. 1.1.1

[37] Michael I Jordan and Dave Latham. Linear and ridge regression, and kernels. Technical report, Lecture Notes: Advanced Topics in Learning and Decision Making. UC Berkeley, 2004. URL http://www.cs.berkeley.edu/~jordan/courses/281B-spring04/lectures/lec4.ps. 2.1.2

[38] Marcel Adam Just, Vladimir L Cherkassky, Sandesh Aryal, and Tom M Mitchell. A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PloS One*, 5(1):e8622, January 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0008622. URL http://dx.plos.org/10.1371/journal.pone.0008622. 1, 1.1.4, 2.2.5, 2.2.5, 2.3.3, 2.3.4, 4.4.2, 7.2

[39] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452:352–355,

2008. URL http://www.nature.com/nature/journal/v452/n7185/abs/nature06713.html. 1.1.4

[40] Holle Kirchner and Simon J Thorpe. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, 46(11):1762–76, May 2006. ISSN 0042-6989. doi: 10.1016/j.visres.2005.10.002. URL http://www.ncbi.nlm.nih.gov/pubmed/16289663. 1.1.2, 6.2.3

[41] Erika Kirveskari, Riitta Salmelin, and Riitta Hari. Neuromagnetic responses to vowels vs. tones reveal hemispheric lateralization. *Clinical Neurophysiology*, 117(3):643–8, 2006. ISSN 1388-2457. doi: 10.1016/j.clinph.2005.11.001. URL http://www.ncbi.nlm.nih.gov/pubmed/16403672. 5.1

[42] Nikolaus Kriegeskorte, Marieke Mur, and Peter a Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2(November):4, January 2008. ISSN 1662-5137. doi: 10.3389/neuro.06.004.2008. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2605405&tool=pmcentrez&rendertype=abstract. 2.3.1, 2.4

[43] Nikolaus Kriegeskorte, Marieke Mur, Douglas a Ruff, Roozbeh Kiani, Jerzy Bodurka, Hossein Esteky, Keiji Tanaka, and Peter a Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–41, 2008. ISSN 1097-4199. doi: 10.1016/j.neuron.2008.10.043. URL http://www.ncbi.nlm.nih.gov/pubmed/19109916. 2.3.1

[44] Valerie a Kuhlmeier, Paul Bloom, and Karen Wynn. Do 5-month-old infants see humans as material objects? *Cognition*, 94(1):95–103, November 2004. ISSN 0010-0277. doi: 10.1016/j.cognition.2004.02.007. URL http://www.ncbi.nlm.nih.gov/pubmed/15302330. 6.2.3

[45] Jan Kujala, Joachim Gross, and Riitta Salmelin. Localization of correlated network activity at the cortical level with MEG. *NeuroImage*, 39(4):1706–20, 2008. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.10.042. URL http://www.ncbi.nlm.nih.gov/pubmed/18164214. 2.4

[46] Jan Kujala, Johanna Vartiainen, Hannu Laaksonen, and Riitta Salmelin. Neural Interactions at the Core of Phonological and Semantic Priming of Written Words. *Cerebral Cortex*, November 2011. ISSN 1460-2199. doi: 10.1093/cercor/bhr307. URL http://www.ncbi.nlm.nih.gov/pubmed/22056541. 6.2.2

[47] Marta Kutas and Kara D Federmeier. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–70, December 2000. ISSN 1879-307X. URL http://www.ncbi.nlm.nih.gov/pubmed/11115760. 2.3

[48] Matthew a Lambon Ralph, Christine Lowe, and Timothy T Rogers. Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain*, 130(Pt 4):1127–37, April 2007. ISSN 1460-2156. doi: 10.1093/brain/awm025. URL http://www.ncbi.nlm.nih.gov/

pubmed/17438021. 2.3.3

[49] Keith R Laws, Verity C Leeson, and Tim M Gale. The effect of masking on picture naming. *Cortex*, 38(2):137–147, January 2002. ISSN 00109452. doi: 10.1016/S0010-9452(08)70646-4. URL http://dx.doi.org/10.1016/S0010-9452(08)70646-4. 6.2.3

[50] Mia Liljeström, Annika Hultén, Lauri Parkkonen, and Riitta Salmelin. Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, 30(6):1845–56, 2009. ISSN 1097-0193. doi: 10.1002/hbm.20785. URL http://www.ncbi.nlm.nih.gov/pubmed/19378277. 1.1.1

[51] Bradford Z Mahon and Alfonso Caramazza. Concepts and categories: a cognitive neuropsychological perspective. *Annual Review of Psychology*, 60:27–51, January 2009. ISSN 0066-4308. doi: 10.1146/annurev.psych.60.110707.163532. URL http://www.ncbi.nlm.nih.gov/pubmed/18767921. 1.1.2, 2.3.3

[52] Bradford Z Mahon and Alfonso Caramazza. What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, 15 (3):97–103, March 2011. ISSN 1879-307X. doi: 10.1016/j.tics.2011.01.004. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3056283&tool=pmcentrez&rendertype=abstract. 1.1.2

[53] Bradford Z Mahon, Shawn C Milleville, Gioia a L Negri, Raffaella I Rumiati, Alfonso Caramazza, and Alex Martin. Action-related properties shape object representations in the ventral stream. *Neuron*, 55(3):507–20, August 2007. ISSN 0896-6273. doi: 10.1016/j.neuron.2007.07.011. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2000824&tool=pmcentrez&rendertype=abstract. 4.4.2

[54] Jeremy R Manning, Michael R Sperling, Ashwini Sharan, Emily a Rosenberg, and Michael J Kahana. Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *The Journal of Neuroscience*, 32(26): 8871–8, June 2012. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.5321-11.2012. URL http://www.ncbi.nlm.nih.gov/pubmed/22745488. 1.2.1

[55] Ksenija Marinkovic, Rupali P Dhond, Anders M Dale, Maureen Glessner, Valerie Carr, and Eric Halgren. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38:487–97, 2003. URL http://linkinghub.elsevier.com/retrieve/pii/S0896627303001971. 1.1.1, 1.1.1, 1.2.1

[56] Alex Martin. The representation of object concepts in the brain. *Annual Review of Psychology*, 58:25–45, 2007. URL http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.psych.57.102904.190143?ai=s1&af=R. 1, 1.1.2

[57] Alex Martin and L L Chao. Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11(2):194–201, April 2001. ISSN 0959-4388. URL http://www.ncbi.nlm.nih.gov/pubmed/11301239. 1.1.3

[58] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-min Chang, Vicente L

Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–5, 2008. ISSN 1095-9203. doi: 10.1126/science.1152876. URL http://www.ncbi.nlm.nih.gov/pubmed/18511683. 1.1.4, 2.1.2, 2.3.2, 2.3.4, 3, 3.1, 3.3.2, 6.1.1

[59] Fermín Moscoso del Prado Martín, Olaf Hauk, and Friedemann Pulvermüller. Category specificity in the processing of color-related and form-related words: an ERP study. *NeuroImage*, 29(1):29–37, January 2006. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2005.07.055. URL http://www.ncbi.nlm.nih.gov/pubmed/16246594. 2.3.2

[60] Hellen E Moss, Jenni M Rodd, Emmanuel A Stamatakis, Peter Bright, and Lorraine K Tyler. Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*, 15(5):616–27, 2005. ISSN 1047-3211. doi: 10.1093/cercor/bhh163. URL http://www.ncbi.nlm.nih.gov/pubmed/15342435. 2.3.3

[61] Brian Murphy, Partha Talukdar, and Tom M Mitchell. Selecting corpus-semantic models for neurolinguistic decoding. *First Joint Conference on Lexical and Computational Semantics*, pages 114–123, 2012. URL http://www.aclweb.org/anthology/S/S12/S12-1019.pdf. 1.1.4, 2.4, 5.4

[62] Kenneth A Norman, Sean M Polyn, Greg J Detre, and James V Haxby. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10 (9):424–30, September 2006. ISSN 1364-6613. doi: 10.1016/j.tics.2006.07.005. URL http://www.ncbi.nlm.nih.gov/pubmed/16899397. 2.1.1

[63] Paul L Nunez and Richard B Silberstein. On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? *Brain Topography*, 13(2):79–96, January 2000. ISSN 0896-0267. URL http://www.ncbi.nlm.nih.gov/pubmed/11154104. 2.2.3

[64] Mark Palatucci, Dean a Pomerleau, Geoffrey Hinton, and Tom M Mitchell. Zero-shot learning with semantic output codes. In *Neural Information Processing Systems*, pages 1410–18. Citeseer, 2009. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.5348&rep=rep1&type=pdf. (document), 1.1.4, 2, 2.1, 2.1.1, 2.1.2

[65] Tiina Parviainen, Päivi Helenius, and Riitta Salmelin. Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. *Cerebral Cortex*, 15(7):1054–63, July 2005. ISSN 1047-3211. doi: 10.1093/cercor/bhh206. URL http://www.ncbi.nlm.nih.gov/pubmed/15563727. 1.1.1, 1.2.1

[66] Ernesto Pereda, Rodrigo Quian Quiroga, and Joydeep Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77(1-2):1–37, 2005. ISSN 0301-0082. doi: 10.1016/j.pneurobio.2005.10.003. URL http://www.ncbi.nlm.nih.gov/pubmed/16289760. 6.2.2

[67] Francisco Pereira, Tom M Mitchell, and Matthew Botvinick. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, 45(1 Suppl):S199–209, March 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.11.007. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?

artid=2892746&tool=pmcentrez&rendertype=abstract. 2.1.1, 3.3.2, 6.2.1

[68] David C Plaut. Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cognitive Neuropsychology*, 19(7):603–639, 2002. doi: 10.1080/02643290244000112. URL http://www.informaworld.com/index/G1GCF2DN7UGPPUN8.pdf. 1.1.2

[69] Sean M Polyn, Vaidehi S Natu, Jonathan D Cohen, and Kenneth A Norman. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756): 1963–6, 2005. ISSN 1095-9203. doi: 10.1126/science.1117645. URL http://www.ncbi.nlm.nih.gov/pubmed/16373577. 1.2.1

[70] Cathy J Price. The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191(1):62–88, March 2010. ISSN 1749-6632. doi: 10.1111/j.1749-6632.2010.05444.x. URL http://www.ncbi.nlm.nih.gov/pubmed/20392276. (document), 2, 2.2, 2.2.3, 2.2.5, 2.3.4

[71] Friedemann Pulvermüller. Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5(12):517–24, December 2001. ISSN 1364-6613. URL http://www.ncbi.nlm.nih.gov/pubmed/11728909. 1

[72] Friedemann Pulvermüller, Risto J Ilmoniemi, and Yury Shtyrov. Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17 (6):884–92, June 2005. ISSN 0898-929X. doi: 10.1162/0898929054021111. URL http://www.ncbi.nlm.nih.gov/pubmed/15969907. 1.1.4

[73] Friedemann Pulvermüller, Yury Shtyrov, and Olaf Hauk. Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2):81–94, 2009. URL http://www.sciencedirect.com/science/article/pii/S0093934X08001673. 2.3.2, 5.7.2, 6.2.3

[74] Timothy T Rogers, Peter J Nestor, and Karalyn Patterson. Where you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8:976–987, 2007. URL http://www.nature.com/nrn/journal/v8/n12/full/nrn2277.html. 1

[75] Indrayana Rustandi. *Predictive fMRI Analysis for Multiple Subjects and Multiple Studies*. PhD thesis, Carnegie Mellon University, 2010. 1.1.4

[76] Indrayana Rustandi, Marcel Adam Just, and Tom M Mitchell. Integrating multiple-study multiple-subject fMRI datasets using canonical correlation analysis. *Proceedings of the MICCAI 2009*, 2009. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.148.401&rep=rep1&type=pdf. 1.1.4, 5.7.3

[77] Riitta Salmelin. Clinical neurophysiology of language: the MEG approach. *Clinical Neurophysiology*, 118(2):237–54, 2007. URL http://linkinghub.elsevier.com/retrieve/pii/S1388245706013976. 1, 1.1.1, 1.1.1, 2.2, 2.3, 5.7.2

[78] Elisabet Service, Päivi Helenius, Sini Maury, and Riitta Salmelin. Localization of syn-

tactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, 19(7):1193–205, July 2007. ISSN 0898-929X. doi: 10.1162/jocn.2007.19.7.1193. URL http://www.ncbi.nlm.nih.gov/pubmed/17583994. 1.1.1

[79] Svetlana V Shinkareva, Robert a Mason, Vicente L Malave, Wei Wang, Tom M Mitchell, and Marcel Adam Just. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, 3(1):e1394, 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0001394. URL http://www.ncbi.nlm.nih.gov/pubmed/18167553. 1, 2.3.2, 2.3.4

[80] Svetlana V Shinkareva, Vicente L Malave, Robert a Mason, Tom M Mitchell, and Marcel Adam Just. Commonality of neural representations of words and pictures. *NeuroImage*, 54(3):2418–25, February 2011. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.10.042. URL http://www.ncbi.nlm.nih.gov/pubmed/20974270. 1.1.2

[81] Irina Simanova, Marcel van Gerven, Robert Oostenveld, and Peter Hagoort. Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PloS One*, 5(12):e14465, January 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0014465. URL http://www.ncbi.nlm.nih.gov/pubmed/21209937. 1.1.2

[82] Robert D Steel, James H Torrie, and David A Dickey. *Principles and Procedures of Statistics*. McGraw-Hill, New York, 1960. 2.1.3

[83] Gustavo P Sudre, Dean a Pomerleau, Mark Palatucci, Leila Wehbe, Alona Fyshe, Riitta Salmelin, and Tom M Mitchell. Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage*, 62(1):451–463, May 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.04.048. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811912004442. 1.1.4, 2

[84] Patrick Suppes, Bing Han, Julie Epelboim, and Zhong-Lin Lu. Invariance of brain-wave representations of simple visual images and their names. *Proceedings of the National Academy of Sciences*, 96(25):14658–63, 1999. URL http://www.pnas.org/cgi/content/abstract/96/25/14658. 1.1.4

[85] Jakub M Szewczyk and Herbert Schriefers. Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368:208–21, January 2011. ISSN 1872-6240. doi: 10.1016/j.brainres.2010.10.070. URL http://www.ncbi.nlm.nih.gov/pubmed/21029726. 6.2.3

[86] Antti Tarkiainen, Päivi Helenius, Peter C Hansen, Piers Cornelissen, and Riitta Salmelin. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122:2119–32, November 1999. ISSN 0006-8950. URL http://www.ncbi.nlm.nih.gov/pubmed/10545397. 1.1.1, 1.2.1

[87] Michael J. Tarr and Yi D. Cheng. Learning to see faces and objects. *Trends in Cognitive Sciences*, 7(1):23–30, January 2003. ISSN 1879-307X. URL http://www.ncbi.nlm.nih.gov/pubmed/12517355. 1.1.2

[88] Samu Taulu and J Simola. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, 51(7):1759–68, April 2006. ISSN 0031-9155. doi: 10.1088/0031-9155/51/7/008. URL http:

//www.ncbi.nlm.nih.gov/pubmed/16552102. 1.2.2

[89] Samu Taulu, Matti Kajola, and Juha Simola. Suppression of interference and artifacts by the Signal Space Separation Method. *Brain Topography*, 16(4):269–75, January 2004. ISSN 0896-0267. URL http://www.ncbi.nlm.nih.gov/pubmed/15379226. 1.2.2, 5.4

[90] Samu Taulu, Juha Simola, Matti Kajola, Liisa Helle, Antti Ahonen, and Jukka Sarvas. Suppression of uncorrelated sensor noise and artifacts in multichannel MEG data. In *International Conference in Biomagnetism*. Internation Conference in Biomagnetism, 2012. 5.4

[91] Leslie G Ungerleider and Mortimer Mishkin. Two cortical visual systems. In D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, editors, *Analysis of Visual Behavior*, pages 549–586. The MIT Press, Cambridge, MA, 1982. 4.4.2

[92] Mikko A Uusitalo and Risto J Ilmoniemi. Signal-space projection method for separating MEG or EEG into components. *Medical & Biological Engineering & Computing*, 35(2): 135–40, March 1997. ISSN 0140-0118. URL http://www.ncbi.nlm.nih.gov/pubmed/9136207. 1.2.2

[93] Johanna Uusvuori, Tiina Parviainen, Marianne Inkinen, and Riitta Salmelin. Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cerebral Cortex*, 18(2):456–66, February 2008. ISSN 1460-2199. doi: 10.1093/cercor/bhm076. URL http://www.ncbi.nlm.nih.gov/pubmed/17566066. 1.2.1

[94] Miranda van Turennout, Timothy Ellmore, and Alex Martin. Long-lasting cortical plasticity in the object naming system. *Nature Neuroscience*, 3(12):1329–34, December 2000. ISSN 1097-6256. doi: 10.1038/81873. URL http://www.ncbi.nlm.nih.gov/pubmed/11100155. 1.2.1

[95] Johanna Vartiainen, Tiina Parviainen, and Riitta Salmelin. Spatiotemporal convergence of semantic processing in reading and speech perception. *Journal of Neuroscience*, 29 (29):9271–80, 2009. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.5860-08.2009. URL http://www.ncbi.nlm.nih.gov/pubmed/19625517. 1.1.1, 1.1.1, 1.2.2, 2.3, 2.3.2

[96] Johanna Vartiainen, Mia Liljeström, M. Koskinen, H. Renvall, and Riitta Salmelin. Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. *Journal of Neuroscience*, 31(3):1048–58, January 2011. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3113-10.2011. URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3113-10.2011. 1.2.1, 2.3.4

[97] Wei Wang, Gustavo P Sudre, Yang Xu, Robert E Kass, Jennifer L Collinger, Alan D Degenhart, Anto I Bagic, and Douglas J Weber. Decoding and cortical source localization for intended movement direction with MEG. *Journal of Neurophysiology*, 104(5):2451–61, November 2010. ISSN 1522-1598. doi: 10.1152/jn.00239.2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20739599. 5.4

[98] Jill Weckerly and Marta Kutas. An electrophysiological analysis of animacy effects in the

processing of object relative sentences. *Psychophysiology*, 36:559–570, 1999. 6.2.3

[99] Thomas Witzel, Rupali P RP Dhond, Anders M Dale, and Eric Halgren. Spatiotemporal cortical dynamics underlying abstract and concrete word reading. *Human Brain Mapping*, 28(4):355–62, April 2007. ISSN 1065-9471. doi: 10.1002/hbm.20282. URL http://www.ncbi.nlm.nih.gov/pubmed/16944493http://scholar.google.com/scholar?q=intitle:Spatiotemporal+Cortical+Dynamics+Underlying+Abstract+and+Concrete+Word+Reading#0. 2.3

[100] Taeko N Wydell, Tiina Vuorinen, Päivi Helenius, and Riitta Salmelin. Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience*, 15(7):1052–62, October 2003. ISSN 0898-929X. doi: 10.1162/089892903770007434. URL http://www.ncbi.nlm.nih.gov/pubmed/14614815. 1.1.1

[101] Jinyin Zhang, Gustavo P Sudre, Xin Li, Wei Wang, Douglas J Weber, and Anto I Bagic. Clustering linear discriminant analysis for MEG-based brain bomputer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(3):221–231, February 2011. ISSN 1558-0210. doi: 10.1109/TNSRE.2011.2116125. URL http://www.ncbi.nlm.nih.gov/pubmed/21342856. 5.4