Justus-Liebig-Universität Gießen
Experimentelle Sensomotorik

Representational Geometry of categorical perception of animate and inanimate objects

Zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften an der Fakultät 06: Psychologie und Sportwissenschaften der Justus-Liebig-Universität Gießen vorgelegte Dissertation

von

Friederike Seyfried

28.08.2019

Friederike Seyfried

Supervisors:	Prof. Dr. Mathias Hegele
	Prof. Dr. Alexander Goesmann

Contents

Abstract

Visual perception begins with simple sensory information and is transformed in the brain into increasingly complex representations of recognizable objects. In order to recognize existing items, concepts - mental representations of things that exists - are necessary. These concepts are organized into classes called categories that allow humans to organize their perception, generalizing between similar items. Visual perception is only one way of accessing conceptual representations, another way is language. While language can be perceived visually as written language, its fundamental form is auditory and it also uses a different (lexical) format. Despite these differences, conceptual organization for vision and language comprehension appears to be similar. It has even been claimed that conceptual representations are independent of the modality and format they are accessed through.

I investigated modality-independent conceptual representations for two modalities (visual and auditory) and two formats (pictorial and lexical) using behavioral and fMRI data and using the category of animacy as a window into conceptual organization. I carried out a behavioral experiment and two fMRI experiments expanding upon previous research by directly comparing the two modalities and formats in within-subject designs, by including a larger part of the animacy spectrum and a perceptual tasks. Categorical perception emerged in the behavioral experiment in the absence of a task suggesting categorical organization of the stimulus set, but, while animacy did seem to play a role in participants' judgments, other factors such as the distinction between man-made and natural items and possibly real-life co-occurrence of objects also seemed to play a role.

In the brain, candidate regions for modality-independent convergence zones were observed, for example the bilateral superior and middle temporal lobes and supramarginal gyri, but also some regions that have not been strongly associated with modality-independent processing like the pre-SMA.

Directly testing for categorical representations for both specific modalities and in a modality-independent way revealed that categorical processing was not as present as previous research suggested. In most previous studies investigating this topic, a task was used that involved categorical processing which might have led to an overestimation of categorical processing effects. By varying the conceptual demands of the task, future research can investigate how top-down vs. bottom-up aspects influence both behavioral similarity judgments and brain activity.

CHAPTER 1

Introduction

1.1 From senses to seals

Consider the photograph on the cover of this thesis. It is a cookie which is supposed to be in seal shape. Most people, when asked, are able to recognize at least that they are seeing a cookie and that the cookie has the shape of some kind of animal. Having never seen this cookie previously, some mechanism allows us to make sense of what is in the end a beige blob on a dark background. This process of recognition happens automatically and without effort, yet it is only possible because of the complexities of visual perception:

1.1.1 Visual perception

Visual perception is one of the most studied fields of cognitive neuroscience. It is beyond the scope of this text to describe all aspects of it, but I will give a short overview of what we know about how primates (mostly, humans) perceive the world around them with their eyes.

Object recognition is a very important part of visual perception. Only because humans can recognize objects based on the multi-dimensional visual input the eye is exposed to, they can also interact with objects. The transformation from multidimensional input to object representation happens in ventral brain regions including parts of the occipital lobe and the ventral temporal cortex in what is often called the ventral visual pathway (Mishkin and Ungerleider, 2002; Mishkin et al., 1983; Felleman and Van Essen, 2005). Object recognition is achieved despite the fact that objects appear in various positions and lighting situations and the fact that the eye is constantly moving which makes the input to the retina highly variable.

During perception of a real world object, the input goes through various stages in which it is transformed into the object percept. The input to the retina is altered and made less complex and variable in the first stage of visual processing in the early visual cortex (V1). V1 neurons are specialized for boundaries and edges (Chaudhuri and Albright, 1997). In V1 the visual input is transformed into a sparse representation (Vinje and Gallant, 2000) that already contains possible object boundaries. In V2, slightly more complex structures such as curvatures are processed (Connor et al., 2007). In V4, the receptive field is even larger and bigger fragments of boundaries containing more visual angles are represented (Pasupathy and Connor, 2001). See 1.2 for an overview of where the ventral visual stream (and the dorsal visual stream) are located in the brain. The ventral pathway is classically described as emerging from early visual areas (V1 and V2) and from there extending into V4 and IT (Connor et al., 2007). However, more recent research indicates that object processing involves brain regions previously not thought to be part of the ventral visual pathway (for example V3) and that the ventral visual stream is more correctly thought of as being an interconnected network (see Kravitz et al., 2013, for a review).

In posterior IT, fragments of V4 such as curve and edge fragments are bound together to form contours (Brincat and Connor, 2004). It is hypothesized that in posterior IT, single neurons might represent very common combinations of shapes, such as the typical two-dimensional contour of a quadruped not because it represents quadrupeds, but because it represents this typical combination of edge and curve fragments. Three-dimensional coding of shapes seems to also be processed in the interior temporal cortex (IT) (Yamane et al., 2008). IT is also thought to represent conceptual information. In the ventral occipitotemporal stream, representations are thus getting increasingly invariant to position and larger in scale and aspects of stimuli that are processed get increasingly complex (Cohen et al., 2000). It is this entire process that enables us to do what I described using the seal cookie example: to transform a somewhat shapeless beige blob on a dark green background into a percept.

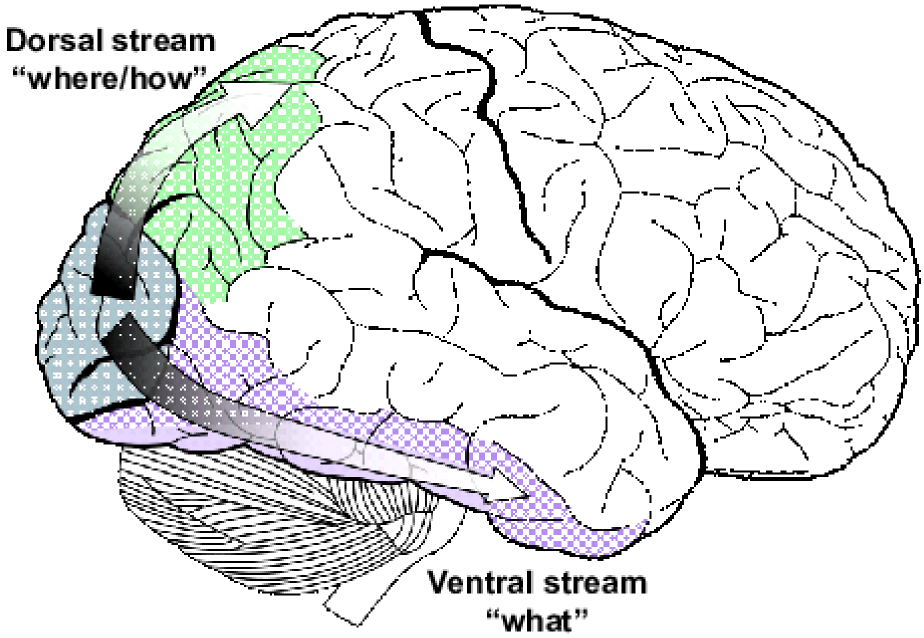


Figure 1.1: The dual-stream model of visual perception in humans from (Chinellato and del Pobil, 2016). The ventral stream is depicted in purple here. Early visual processing starts in the occipital lobe. From there, two pathways emerge. The dorsal visual pathway (here in green) is related to action planning. The ventral visual pathway, also called "What" - Stream, is associated with object recognition (Goodale and Milner, 2004).

1.1.2 Language comprehension: spoken language

The visual modality is not the only way to perceive objects. Everyone who reads this thesis, has inevitably perceived objects through a completely different modality: through language. Language comes in two main flavors: as spoken language and as written language. Spoken language, perceived through the ear, is the more ancient and fundamental way to comprehend language. Similarly to visual object recognition, the auditory signal undergoes a series of transformations until larger linguistic chunks such as phonemes (letter sounds of a given language), words or sentences are perceived. Speech processing poses unique challenges to the auditory perception system. Language sounds can vary a lot in actual spoken language - think about the variation of the /l/ sound in between Standard German and the Cologne dialect - but they still have to be perceived as the same phoneme - a problem that seems somewhat comparable to the issues of variability of input in the visual modality.

Auditory perception starts with the spoken language stream reaching the ear in the form of sound waves. As revealed by studies in rhesus monkey, auditory processing, similar to visual processing, is separated into two streams: one extending into the parietal cortex dedicated to spatial processing, the other one localized in the temporal lobe dedicated to the recognition of auditory objects (Rauschecker and Tian, 2000). fMRI studies indicate that human auditory processing is similarly organized (Rauschecker, 1998).

The auditory signal, after being transmitted to the brain, is first processed independently of whether this signal is speech or some other sounds. The two cochleas encounter sound waves which are transmitted to the cochlear nuclei in the brainstem. From there, the information is transmitted to the superior olivary complex and then the superior colliculi in the midbrain and the medial geniculate in the thalamus. From the thalamus, the auditory signal reaches the auditory cortex in the cerebral cortex (Recanzone and Sutter, 2008).

In the auditory cortex, the signal is processed by three regions, the core auditory cortex, the belt and the parabelt regions which seem to be more variable in humans compared to macaques (Recanzone and Sutter, 2008). The belt regions are where the dual stream architecture of auditory processing seems to originate (Rauschecker and Tian, 2000). Early processing of both monkey vocalization and human speech seems already to be carried out in the belt regions, which in humans are located in the lateral superior temporal gyrus (Rauschecker and Tian, 2000; Scott et al., 2000).

Spoken language processing seems to be carried out in the temporal (also called antero-ventral or ”What”-stream) (Rauschecker and Scott, 2009), but the postero-dorsal stream is thought to also play a role, for example in the link between speech comprehension and action (in this case speech production) (Bornkessel-Schlesewsky et al., 2015). It has been suggested that the antero-ventral stream processes speech items after initial spectrotemporal analysis of the auditory signal in more purely auditory regions in a hierarchical fashion: increasingly complex and large speech items (phonemes, syllables, words and finally conceptual relations) are processed in increasingly anterior regions of the antero-ventral stream (Bornkessel-Schlesewsky et al., 2015).

A model of the two streams in auditory perception in humans can be seen in 1.2.

1.1.3 Language comprehension: written language

labelLangPerc

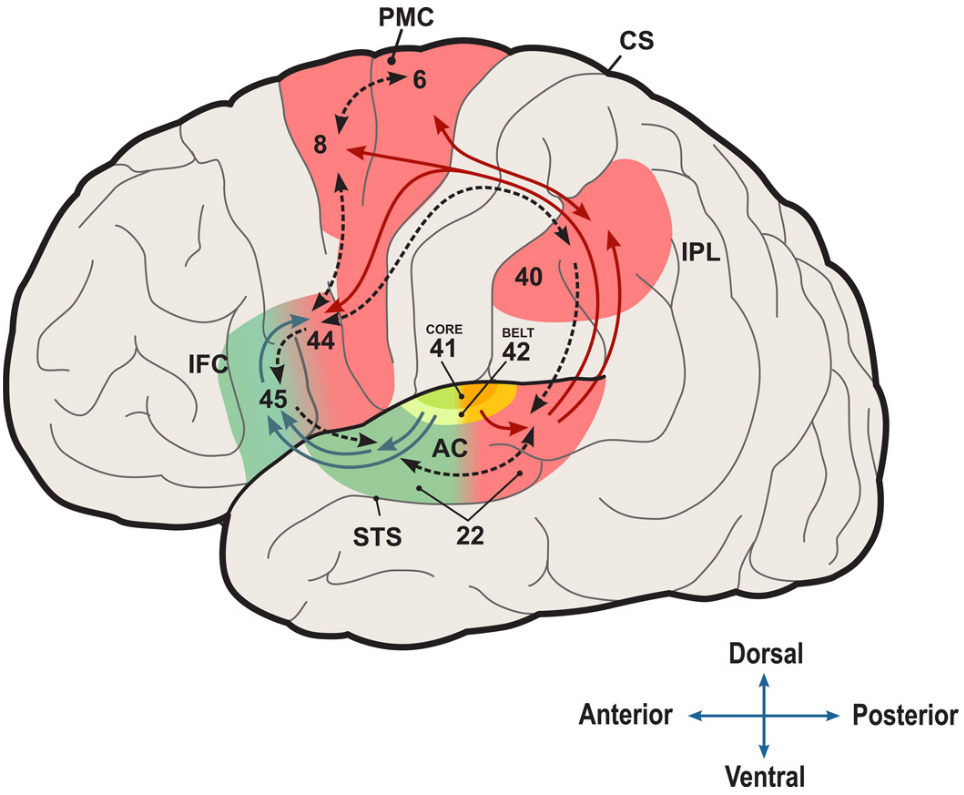


Figure 1.2: The dual-stream model of language perception in humans from Bornkessel-Schlesewsky et al. (2015). The antero-ventral stream (in green) is associated with increasingly complex auditory objects and, in its more anterior part, with conceptual processing related to auditory signals. The postero-dorsal stream (in red) is associated with sequence processing (Bornkessel-Schlesewsky et al., 2015).

Language perception in its earliest form is an auditory process. Humans learn to speak before they learn to read and reading, while wide-spread in today's world, is a new cultural technique that only became common since the arrival of modernity. Reading comprehension is partially related to visual perception: Reading is in its first stages a visual process which is combined with linguistic processes to enable written language comprehension. Reading is achieved by the brain by processes represented in the dual visual system 1.1 (Dehaene et al., 2005), with the focus - as for object recognition - on the ventral occipitotemporal areas. Word perception seems to be a specialized form of object perception. Similarly to object recognition, reading and word recognition occurs in a gradient-shaped way throughout the occipitotemporal stream with early areas being attuned to edges and curves (necessary to recognizing letters) and more downstream areas being attuned to more complex parts of words such as singles letters, letter pairs and finally entire words (Cohen et al., 2000).

Letter part processing first occurs in V1 and V4 (Dehaene et al., 2005). More complex responses to the full letters independent of variations in scale and orientation have been observed in a region anterior to V4. This region has been called the visual word form area (VWFA) (Cohen et al., 2000). It does not only seem to process single letters but also word forms and is a particularly important part of the ventral occipitotemporal visual stream for reading. It develops its fine-tuning towards visual word forms in all literate adults independent of the alphabet they learn: it even can serve as a marker of literacy (Dehaene-Lambertz et al., 2018). The relevance of this area for recognizing and reading words has been known since the 19th century (Dejerine, 1892) and early neuroimaging (Nobre et al., 2002).

It deserves a mention here, that the role of the Visual Word Form Area with regard to its importance in reading is not uncontroversial. There have been observations that put a question mark on the idea that the VWFA is where complex word fragments such as letters up to whole words are processed. Price and Devlin (2003) reviewed several functional imaging as well as patient studies concerning the VWFA in 2003 and called into question its unique function for letter and word fragment processing. The question is discussed in more detail by Dehaene and Cohen (2011) when they revisit their own Visual Word Form hypothesis .

The dorsal occipitoparietal visual stream is thought to also contribute to reading, especially in non optimal conditions when reading requires more effort due to degraded input. The visual word form area is specialized for reading under ideal conditions, whereas the dorsal visual pathway plays a role when the written input is presented in a not ideal format, for example with an angle or in the non-preferred hemifield. Under non-optimal conditions, letter by letter reading strategies which are processed by the dorsal visual pathway have to be employed (Cohen et al., 2008). The recognized words and word parts processes by the visual system then have to interface with the linguistic system in the brain in order to enable language processing of the input. Anatomically, the connections between the VWFA and traditional language areas are not yet well understood, but it is plausible that in humans (as opposed to the nonlinguistic nonhuman primates) exclusive connections have formed. Early studies in the arcuate fasciculus (Catani et al., 2004) and the inferior fronto-occipital fasciculus (Catani et al., 2002) point in that direction.

1.1.4 Concepts and categories in the mind

Language (whether written or spoken) and visual perception are channels that give humans sensory access to the things that exist in the world. The content of the senses is however not meaningful in itself. Letters are just black edges on white pages and spoken words are just sound waves with different frequencies. How does the brain form meaning from what are physically very simple stimuli? If humans were not able to recognize and categorize the things they perceive with their senses, their world would likely be unmanageable, a ”blooming, buzzing confusion” as James (1890) wrote and as is picked up on by Oakes and Rakinson in their book title Rakison and Oakes (2003). Because of concepts and categorization, the world is ordered. In this chapter, I will briefly discuss the terms ’concept’ and ’category’ because they are not always used in the same way across different disciplines:

’Concept’ is a term that is hard to pin down. Its meaning depends on the field in which it is used. In cognitive science, the most used idea of a concept is, that it is a psychological item, represented in the mind (Pinker, 2007; Carey, 2009). This idea is frequently referred to as the representational theory of mind (Margolis, 1994).

Concepts and categories are closely related to each other. Concepts refer to something, they are the mental representations of meaning (Margolis, 1994). Categories are the types of things that they refer to, classes or types or groups (Margolis, 1994). Typical categories include things of substance, quantities, qualities, relations between objects and abstract concepts. Here, we will deal with concrete objects of substance (Carey, 2009). Categories are thought to underlie many aspects of cognition in allowing recognition of repeated classes of objects in the world during perception, being able to communicate ideas and abstracting from concrete reality (Kiefer and Pulvermüller, 2012).

Another term often used in the line of research dealing with concepts and categories is the term ’semantic’ or ’semantic knowledge’. Semantics, in philosophy, is the study of meaning and the structure of meaning, but in neuroscience it often is used purely in association with word meaning. Since this text is about more than word meaning, I will prefer the term conceptual over semantic. I can’t completely avoid the mixing of terms that occurs since the term ’semantic knowledge’ is also frequently used. Wherever possible, the terms ’semantic knowledge’ or ’semantic memory’ will be skipped and I will use conceptual knowledge. In other works, for example by Taylor et al. (2011) the terms were used interchangeably which is also an option, but also points to the fact that the distinction is often not clear. Kiefer and Pulvermüller (2012) opted for using ’semantic’ over ’conceptual’. Meaning can be conveyed through more than just words, after all. The view that ’semantics’ is associated with specific words also reflects the fact that word semantics can be somewhat distinct from conceptual knowledge. Words have meaning but those meanings are language specific and not all languages divide the conceptual space in the same way. An often used example for this phenomenon is the German word ”Politik” which in English is translated as ’policy’, ’politics’ or ’governance’ depending on what exactly the meaning to be conveyed is.

1.2 The Neuroscience of Concepts and Categories

Perceptual processes of the modalities studied in this work are quite different and highly specialized. Conceptual processing on the other hand might convergence to the point of being modality-independent. How does the brain encode concepts and which organizational principles apply? Before the advent of neuroimaging, researchers have tried to map conceptual knowledge and category knowledge onto the brain. In fact, categories were an early interest of neuropsychology since some neurodegenerative diseases and lesions in the brain seem to lead to selective impairment in the processing of specific categories but not others. When the connection between specific areas of the brain and deficits in conceptual knowledge was first noticed is hard to say, but earliest reports seem to date back to the 1920s where Head described semantic aphasia (Head, 1921), the 1970s when Luria investigated deficits brought about by lesions to the occipitotemporal regions (Luria and Hutton, 1977) and to the 1980s when Warrington and Shallice and Warrington and McCarthy described patients with semantic dementia after herpes simplex encephalitis or aphasia (Warrington and Shallice, 1984; Warrington and McCarthy, 1987).

Early research into the representation of categories and conceptual knowledge in the brain followed two lines of reasoning: (1) Researchers tried to localize specific categories to brain regions. (2) Researchers attempted to identify brain regions that underlie a conceptual knowledge store and regions that supported retrieval of conceptual knowledge. We would probably ask slightly different questions today, trying less to find the location or certain categories and conceptual knowledge in specific brain regions and thinking more in terms of systems processing specific conceptual content, but the candidate regions identified during this early phase of research into conceptual knowledge nevertheless remain part of the neural system that we think of underpinning conceptual processing today.

One part of the brain is strongly linked to conceptual processing, is the bilateral temporal lobe. Several temporal regions have been implicated in conceptual processing, most notably the anterior temporal lobe (ATL), often thought to be a hub for amodal semantic processing (for a review see Patterson et al., 2007), the inferior temporal lobe (IT) and the posterior (middle) temporal lobe (pMTG) (Noonan et al., 2013; Mollo et al., 2016). Outside of the temporal lobe, the lateral occipital cortex (LOC) has been cited as a candidate region where visual and tactile conceptual processing converge (Beauchamp, 2005) and the (ventral) lateral prefrontal cortex (VLPFC) is thought to be involved in retrieval of conceptual knowledge (Martin, 2007). Apart from single regions involved with some particular category or with conceptual processing as such, different and diverse parts of the brain seem to be involved in the processing of specific concepts or even just objects, for example action words seem to elicit activity in areas of the motor cortex associated with the extremity the action is performed with (mouth for "kiss", leg for "kick", arm for "hit" (Hauk and Pulvermüller, 2004). The dorsal visual stream, which is thought to support vision-to-action mapping instead of object recognition nevertheless has been observed to be active during object recognition of specific categories of objects, namely tools. Tools are closely connected with the actions typically performed by them. Thus, conceptual processing seems to involve many different regions of the brain, depending on what the content of the concept happens to be, while some areas are associated with storing and retrieving concepts as such.

Different theories have emerged as to what the role of these different brain regions in conceptual processing might be, as well as how conceptual processing might be organized in the brain in general. I will divide them into distributed theories, convergence zone theories and hub theories and introduce them in this order.

1.2.1 Distributed theories

Conceptual knowledge encompasses all of human knowledge that is general (not related to personal events) and explicit (not how to walk or play the piano). Concepts are also accessible through various modalities, you can eat a seal cookie, see it, hear or see the word 'seal cookie' or hear the crumbs from it and it is usually clear it is a cookie. Activity elicited by conceptual processing is situated in different areas of the human brain, depending on what the content of the concepts is. Finally, humans have to acquire concepts during their development and since they can only acquire knowledge through their senses and different concepts might be acquired majorly through one sense or the other (how often have you smelled a door knob?). All these facts make it a plausible idea, that conceptual knowledge is distributed throughout the brain and depending on what the concept is, different areas fire together to support a conceptual recognition or retrieval process.

There are different levels of distributed theories. Martin (2007) suggests that especially categories are encoded in different brain areas. Some categories such as faces and houses are linked to specific regions of the brain that represent those categories, i.e. the Fusiform Face area and the Perirhinal Place Area (which is not only associated with places but also houses), respectively. Also, broader categories are associated with specific brain regions (tools with the regions pMTG, IPS and VMPC and animates with the lateral fusiform gyrus and the pSTS) (Martin, 2007). Evidence for more fine-grained distributed conceptual knowledge has also been observed in studies, where very specific activity was observed in response to narrow concepts such as those denoted by action words such as kiss, kick or hit (Hauk et al., 2004) which elicited activity in areas related to those extremities.

Apart from different categories or types of concepts, distributed models of conceptual knowledge also relate to the different modalities that concepts can be accessed through. It is possible to approach the concept of a cookie through many different senses and it is not clear if the concept of 'cookie' is always the exact same independent of whether we eat one, see one or talk about one. Some theories suggest that conceptual knowledge depends on the sensory-motor experience we had when first encountering the concept (Martin, 2007). Distributed theories are supported by the pattern of loss that is observable during Alzheimer's disease which is characterized by widespread atrophy of neurons throughout larger regions, not by focal loss (Kiefer and Pulvermüller, 2012). Semantic dementia is linked specifically to the loss of categorical knowledge caused by the loss of neurons beginning in anterior temporal regions. Kiefer and Pulvermüller (2012) view the pattern of impairment associated with semantic dementia as evidence for distributed representations of conceptual knowledge, since it progresses from more specific categorical knowledge to the loss of more general knowledge and from knowledge about finer categories (i.e. say specific types of birds) to more overarching categories (birds or animals).

A central part of distributed theories and an argument against any kind of amodal representation is the theory of embodied cognition, sometimes also called grounded cognition. According to Barsalou, one of the most eminent researchers who developed the ideas surrounding embodied cognition, grounding cognition rejects the idea that there are amodal states that influence cognition settling cognition instead in bodily states (Barsalou, 2008). Embodied cognition is much broader than conceptual knowledge, but here I will deal with the influence it had on the thought about conceptual knowledge only.

Embodied cognition for conceptual knowledge assumes that concepts are the result of accumulated knowledge acquired through sensory-motor processing (Ralph et al., 2016). The idea that concepts are represented in patterns similar to those active when they were acquired is well supported (Martin, 2007). The focus of investigation seems to have been on studying concepts related to action both using neuroimaging and neuropsychological methods (Aziz-Zadeh and Damasio, 2008). An open question remains how abstract concepts can be represented in the brain in an embodied way (Kiefer and Pulvermüller, 2012) or how metaphorical language such as "kick the bucket" could be embodied (Aziz-Zadeh and Damasio, 2008).

1.2.2 Hub theories

Several lines of data are not in support of distributed only conceptual knowledge. Most notably, damage to specific brain areas seems to damage conceptual processing as such, not only one category or in one modality (Ralph et al., 2016). One area, whose damage leads to impairment of conceptual processing independent of modality or category, is the anterior temporal lobe (ATL) (Patterson et al., 2007). In a particular neurodegenerative condition, semantic dementia, the ATL atrophies bilaterally. Patients with ATL are increasingly unable to perform semantic tasks, that involve different modalities and different categories depending on the severity of the atrophy in ATL. In addition to that, there is evidence that computationally (Ralph et al., 2016), a hub is necessary to perform the conceptual tasks that humans are capable of.

The hub is supposed to be involved in the processing of conceptual knowledge in an amodal way, for all categories and all modalities of access, while specific attributes of some categories and semantic knowledge associated with specific modalities (i.e. color for vision) are still represented in a distributed way throughout the brain. This hypothesis is often referred to as the "hub and spoke" hypothesis or model (Patterson and Ralph, 2015; Patterson et al., 2007; Lambon Ralph et al., 2010).

1.2.3 Covergence zone theories

Convergence zone theories resemble hub theories to some degree with some important differences. Convergence zones were first introduced by Damasio (1989), as regions of the brain that process conceptual knowledge independent of modality. The main differences to hub-theories is that convergence zones presuppose more than one of such zones whereas there is usually just one 'hub' (Patterson et al., 2007). Convergence zones are also less broad compared to the hub. There could be convergence zones where visual and motion knowledge for example for tools convergences whereas in other zones different aspects of knowledge are merged (Patterson et al., 2007). The convergence zones are viewed as pointers that bind together different aspects from different modalities and formats for one category (Ralph et al., 2016). Convergence zones according to researchers who argue for them, allow conceptual knowledge to be increasingly integrated and abstracted away from the modalities through which conceptual knowledge is activated (Binder and Desai, 2011). The location of these convergence zones is assumed to be in more divergent regions than just the ATL. Binder and Desai propose that both temporal and parietal regions are involved in the processing of conceptual knowledge (Binder and Desai, 2011). In a study using crossmodal decoding Fairhall and Caramazza also identified the posterior MTG & STS regions are involved in modality-independent representation of conceptual content (Fairhall et al., 2011).

1.3 Open questions

Concepts and categories are necessary to understand the percepts of the different sensory modalities, but it is not clear how tied they are to specific modalities. There are several lines of research that point towards the existence of amodal conceptual representations in the brain, i.e. the existence of modality-independent deficits after atrophy in the ATL (Lambon Ralph et al., 2010) or the existence of convergence zones that seem to process conceptual knowledge independent of modality (Binder and Desai, 2011). The question if conceptual representations are modality-independent has been asked before, but there are several patterns in this field of research, that limit the observations that have been made so far. I will give an overview on two aspects that deserve some attention: (1) The role of task in conceptual and categorical processing and (2) The modalities that have been studied. Finally, I will introduce the category distinction that I used as a window into conceptual processing between the visual and the linguistic modality: namely animacy.

1.3.1 The influence of task on conceptual processing

When investigations of the neural substrate of conceptual processing were restricted to studying patients, many different tasks were developed to study the impairments caused by lesions in different regions of the brain (i.e. picture naming, judging if an item is presented in the typical color, or deciding whether two objects perform the same function Hillis and Caramazza, 1995; Gainotti and Silveri, 1996; Riddoch and Humphreys, 1987; Caramazza and Mahon, 2003). With the arrival of neuroimaging methods, researchers started to investigate conceptual processing in healthy human adults. The earliest studies looking at unimpaired human adults using either fMRI or in the early days predominantly PET, date back to the 90s (Damasio et al., 1996; Vandenberghe et al., 1996; Mummery et al., 1998). These early studies very often used written word and picture stimuli and asked participants to perform explicit semantic tasks, for example to judge how large stimuli were (Vandenberghe et al., 1996), whether a stimulus had a certain color or occured in a similar environment as another stimulus (Mummery et al., 1998), asked participant to match words to pictures (Perani et al., 1999) or asked participants to judge whether a stimulus belonged more to a selection of two other items (the so-called Pyramid and Palm tree test, in which participants are presented with the stimulus of a Pyramid and have to judge whether the pyramid belongs more to a fir tree or a palm tree Chee et al., 2000; Visser et al., 2012). Researchers also started to investigate spoken and written words and the convergence of conceptual processes using different types of tasks for these two modes of representation for language early on. Booth et al. (2002a) studied how both crossmodal tasks (for example judging whether two written words rhymed or how spoken words were spelled) influenced activity in the brain. They observed that crossmodal tasks lead to increased activation in supramarginal and angular gyri. In a similar experiment, Booth et al. (2002b) compared a semantic task (whether two words were related as e.g. "chair" and "table" are) and a language-related task (whether two words rhymed). They observed modality-independent activity in the frontal regions, as for example the left inferior frontal gyrus . In several studies, the task used simply asks participants to detect an oddball stimulus. In a study by Connolly et al. (2012), that investigated the representation of different animal concepts, each animal stimulus was presented in three different forms (i.e. three different photographs of a ladybird) and participants had to press a button when the third stimulus was not the same animal, referred to as "rabbit, rabbit, tarantula". Most tasks used center the stimuli used in specific ways, namely by focusing on semantic features such as size or context or by asking to judge if stimuli belonged to the same category (e.g. furniture).

A task that is often used in studies investigating crossmodal conceptual processing is a variation of a n-back task in which participants are asked to judge whether a stimulus previously presented (presented n - times back) is the same or different from the stimulus last viewed. This task can be varied in many ways to elicit different forms of processing. In a study investigating the effect of task and modality on the processing of tool and animal concepts, Noppeney and colleagues used a variation of the n-back tasks, for example asking participants whether stimuli within a block was identical as stimuli, whether they were performed similar actions (i.e. flying) or whether they were similar in size (Noppeney et al., 2005). This task engaged both memory and semantic processing.

Asking participants to judge the size of an object is another task that is used (Mahon et al., 2007; Fairhall et al., 2011; Bracci et al., 2017). While size seems to be a frequent semantic feature that tasks focus on, some studies guide the attention of participants and their processing demands to features or associations of concepts in general. In a study by Shinkareva et al. (2011) looking at tools and dwellings in picture and written-word form, participants were asked to come up with a number of associated words for the stimulus set before fMRI data was obtained and were instructed to think about these associated concepts (for example, "cold", "knights" and "stones" for castle) when they were presented in the scanner. In a study by Marques et al. (2008), participants were asked to verify whether a statement about a concept was true or false (i.e. "Snake rolls up", "Snake is long", "Saw has teeth", "Saw cuts trees") for features related to either visual form ("is long", "has teeth") or a feature related to motion ("rolls up", "cuts trees"). Another approach is to direct the attention of the participants directly towards categories, i.e. have them name what class a stimulus belongs to. This type of task was used for example by Devereux and colleagues (Devereux et al., 2013).

In some studies, the task is unrelated to the stimulus but instead directed towards a different item. Several studies make use of a fixation cross depicted on screen to give participants a point to focus on. Participants are instructed to fixate the cross throughout the study and are then asked to press a button when the fixation cross changes color (Jacques et al., 2016; Bell et al., 2009; Kriegeskorte et al., 2008b). This has the advantage of controlling participants' attention on the part of the screen where the stimulus is depicted while also not using a task directly related to the stimulus. It also minimizes the influence of top-down processing of the stimuli and puts the focus on feedforward mechanisms (Kriegeskorte et al., 2008b). In some task stimuli are presented in two modalities and participants are asked to judge whether the same concept is presented, i.e (Hocking and Price, 2009). It is important to consider the influence that tasks can have on the processing that the brain has to carry out and to not overlook how - apart from just perceiving and processing the stimulus as such - tasks can direct neural activity in a specific direction.

The role of task has been explicitly investigated since the beginning of this line of research. Price et al. (1996) compared passive viewing of pictures with explicitly naming pictures in a PET study in order to separate the processes at work during the recognition of objects from the processes required when naming an object (such as phonological processing and selection of the appropriate word form). However, in many cases the task emphasizes the semantic and categorical aspects of the stimuli. To my knowledge only five studies have attempted to study the influence of task on the amodal processing of stimuli. In 2010, van Doren and colleagues studied the influence of conscious vs. subliminal perception of both written words and pictures on activity in two regions that are associated with the amodal processing of words and pictures, namely the IFS (inferior frontal sulcus) and the OTS (occipito-temporal sulcus. They collected fMRI data while participants saw words and pictures that were forward and backward masked to ensure that only 50% of stimuli were consciously identified (whereas the others were only subconsciously perceived). Participants indicated whether they had perceived a stimulus with a button press. In a second fMRI experiment, they combined this task in which participants also performed an episodic memory retrieval task in which they were asked to identify via key press whether a newly presented stimulus had already been presented in the preceding fMRI sessions. They observed overlapping activity in the OTS, the intraparietal sulcus, inferior frontal junction, the middle third of the inferior frontal sulcus (IFS) for pictures and words that participants were able to identify. The OTS and the IFS activity also predicted performance in the memory task for words and pictures and dynamic causal modeling revealed that IFS influences the modality-specific systems by interacting with OTS and not directly with the modality-specific systems. They authors concluded that the OTS and IFS are involved in amodal processing of word and picture meaning independent of the semantic demands of the task (Van Doren et al., 2010).

Wiggett and colleagues investigated the influence task has on the functional organization of animacy and inanimacy in respectively ventrolateral and ventromedial visual brain regions. They presented pictures of animals from four categories (mammals, fishes, birds and reptiles/amphibians) and objects from four inanimate categories (vehicles, furniture, musical instruments, tools) while participants were asked to match stimuli at different levels of analysis (they either had to indicate via button press whether two stimuli were the same exemplar, from the same intermediate category (if they were both a horse or a car), or from the same general category i.e. both were mammals). They did not observe that task had an effect on the division between ventrolateral and ventromedial visual areas for animate and inanimate objects and concluded that the previously reported functional specialization in these areas did not depend on top-down demands introduced by different types of task (Wiggett et al., 2009).

In 2005, Devlin and colleagues specifically investigated how task modulates activity related to tool and animal words in the posterior fusiform gyrus. Participants read words and performed either a perceptual (word length) or a semantic (was an item manmade or an animal) task. They observed that category-selective patterns of activity in the fusiform gyrus were only present for the word stimuli when participants performed a semantic task (Devlin et al., 2005).

There are also non-fMRI studies that have investigated the role of task in conceptual processing. Similar to the fMRI study carried out by Devlin and colleagues, Chan and colleagues looked at semantic processing of spoken and written words (both animals and manmade objetcs)using intracranial electroencephalography in the anteroventral temporal lobe. Chan and colleagues observed cateogory-specific responses across several different tasks with differing amounts of semantic processing costs (size, familiarity and concreteness judgments and a memory-task) for both written words and spoken words across different measures related to synaptic activity such as local field potential) and in unit firing and single firing measures (Chan et al., 2011).

Mollo and colleagues used magnetoencephalography to study both manmade objects and animals at two different levels of analysis, a specific single stimulus level (violin-violin) and a general conceptual level (musical instrument - musical instrument). They observed differences in oscillatory activity for specific vs general tasks for animals and manmade objects (Mollo et al., 2016).

To my knowledge, apart from the five articles, no other studies have been reported that attempt to investigate the influence of task on semantic processing. The results of the studies I summarized is hard to reconcile - in part because there is no single line of research that investigates this topic and the differences of method. Some studies come to the conclusion that task demands do not influence semantic processing whereas in others studies, a modulation by task is observed.

Since the majority of studies has used a conceptual task, it is possible that amodal representations of concepts have been systematically overestimated. Explicit conceptual tasks might lead to top-down activation of concepts as well as categories, an activation which might not be present if a task is used that is purely perceptual. This study attempts to investigate modality-independent category effects using tasks that do not direct the attention of participants towards conceptual knowledge in any way.

1.3.2 Stimulus modalities and stimulus formats

One of the goals of this thesis is to compare conceptual organization between different ways in which concepts can be accessed. My project used written words and spoken words and pictures to access concepts. The two language stimuli can be experienced through two modalities: visually, as written words but also aurally as spoken words. Vision and auditory perception are two modalities associated with different sensory organs and processing streams in the brain (see chapters 1.1.1 and 1.1.2). Written words are perceived through the visual system initially, but they are presented in a different format (in a lexical format) than pictures, which are in pictorial format. The combinations of modality and format chosen here are not the only ones that can be studied and have been studied. Written words are sometimes referred to as 'visual language' or 'language in visual form' because writing is first perceived with the eyes. There is, however, another type of language that is visually processed, namely sign language. Waters et al. (2007) have studied conceptual processing when concepts are referred to via British Sign Language (BSL): in a 2007 they presented pictures, written English words, BSL words and fingerspelled words and discussed the mid-fusiform gyrus as an interface between sign language, pictures and written words. In a study using pictures as stimuli and different tasks that either focused on the sign language structure of the word for the picture, the semantic category of the object or physical aspects of the object, Waters et al. (2007) observed differences in the involvement of right and left STC between deaf and hearing participants.

Apart from the fact that language can be processed either visually or aurally, there is also processing of auditory signals that is not lexical. There is a significant line of research that looks into processing of concepts that are conveyed through sound, for example of animals or tools in use. Beauchamp et al. (2004) investigated how information conveyed via nonlinguistic sounds and pictures converged in temporal regions using fMRI and an experiment where both pictures of animals and manmade objects and recordings of the sounds animals and manmade objects make (for example the blow of a hairdryer) were presented. Man et al. (2012) used a similar approach in 2012 presenting silent videos or sound recordings of tools, for example chainsaws, without visual input in alternating blocks while measuring fMRI data.

Apart from non-linguistic sound, another important sensory channel through which humans activate concepts is touch. Amedi and colleagues investigated the convergence of touch and visual perception by presenting models of manmade objects, animals and models of vehicles as well as textures which were all handed to the participant by the experimenter, while the participant lay in the scanner (Amedi et al., 2002). Information conveyed through touching objects seems to also converge with information conveyed visually and aurally in temporal regions (for a review of this topic see Beauchamp, 2005). Recently, crossmodal decoding has been used by Man et al. (2015), to further study where information is represented independent of sensory modality in a study using sight, sound and touch. An interesting aspect of experiments using models of objects that are given to participants to touch, is that there seems to be a difference in the activity elicited and the informational representation between pictures and real objects, (see Culham, 2017, for an overview of the effect of presenting real objects in the scanner). How different perceptual processes lead to the activation of one conceptual representation has been studied using modalities that are linguistic (i.e. sign language) and non-linguistic (real world objects, real world object sounds). Most studies looking at this question use modalities that share some aspects, for example being linguistic or both being both non-linguistic. There have been comparisons made between language and non-language conceptual processing (Devereux et al., 2013; Costanzo et al., 2013) and these studies, together with previously cited studies suggest that modalitity-independent processing is likely. Different sensory modalities such as touch, vision and sound seem to converge in the brain even though the initial sensory processing is entirely different, pointing towards at least some degree of abstraction instead of entirely separate processing streams for different modalities.

1.3.3 (In)Animacy as a window into conceptual organization

The different concepts investigated in the studies described in 1.3.2 come from many different categories, including tools, vehicles and animals but most studies did not explore specific category distinctions systematically: One particular categorical distinction is between animates and inanimates. An advantage of using animacy as the window into conceptual organization between different modalities and formats, is that it is a particularly well studied categorical distinction. It has been argued that this distinction might even be to some degree innate (Caramazza and Mahon, 2003). At least, this category is observed to develop early in human life, lost late during neurodegenerative diseases and it plays a role in many different cognitive processes such as visual perception and language comprehension (Capitani et al., 2010).

There is plenty of evidence for animate behavior shaping the psychology of visual perception. Humans react differently to animate motion, compared to the motion of inanimates, easily perceiving the slight differences between the two (Zwickel et al., 2011). It has even been suggested that the difference in motion (self-guided, initiated by other forces or no motion at all) is the foundation of animacy distinctions (Mandler, 2004).Animate forms also seem to elicit specific behavior. Infant humans learn the distinction between animates and inanimates quickly (Hegele and Seyfried, 2018). They seem specialized in perceiving mammal and mostly human faces from very early on (Haxby et al., 2000; Morton and Johnson, 1991; Johnson and Rakison, 2006). Mammal body shape is almost as quickly perceived, but even the motion and shape of far removed animate entities such as spiders or insects, is perceived instantly when present. Humans even perceive the behavior of animates when no animate is present, attributing rustling in a forest to humans or wild animals, tapping doors to robbers or kidnappers and endowing storms and waves with a consciousness and a will. As the famous Heider and Simmel movies using animations demonstrate, even simple forms such as triangles can be viewed as animate agents by humans (Heider and Simmel, 1944). It could be said, that human perception is built to perceive animacy even when animacy is not present. Animates are also more easily and stably remembered than inanimates (Bonin et al., 2013).

Visual perception is very common in mammals. Language is rarer and arguably unique to humans though many animals from dolphins to rats to nonhuman primates, communicate by means of vocalization. It is impossible to know how language developed in humans and what the content of the first truly linguistic utterances was, however, as far as we understand nonhuman communication, referring to animates (either predators, prey or conspecifics) seems to be what animals do when they vocalize (Fedurek and Slocombe, 2011). It is unsurprising that human language also has a particular way of talking about animate objects. Animates, especially humans, are the ideal subjects that act upon objects. Exceptions to that rule in many languages require a special linguistic marker (a phenomenon known as 'Differential Object Marking'). The degree to how marked such utterances are and how they are marked differs from language to language (Aissen, 2003), the phenomenon was first observed and described by Bossong (1985). In languages without differential object marking (as for example German), animacy also plays into syntactic constructions (Grewe et al., 2007). Linguistics has an entire animacy hierarchy (see 1.3 which classifies how likely an entity is to take subject role in a sentence and which starts out from the 1st person, followed by familiar 2nd persons, to other human conspecifics, to mammals, other animals and inanimate objects, even to things that are not even objects (Silverstein, 1976; Comrie, 1981; Zaenen et al., 2004).

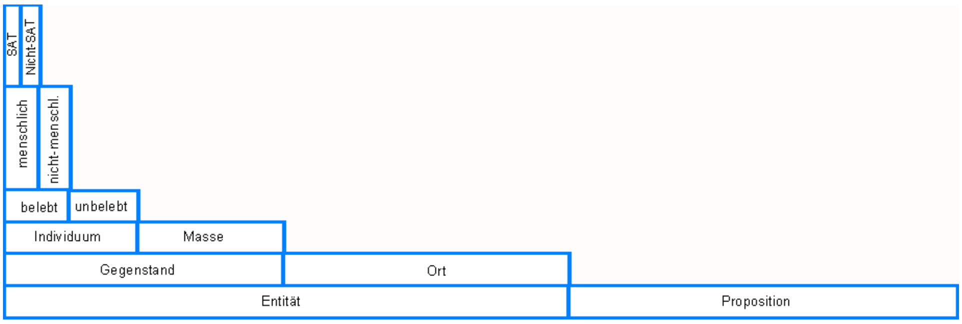


Figure 1.3: The animacy hierarchy (also empathy hierarchy but see Lockwood and Macaulay (2012) for a review of different aspects of this hierarchy) as depicted by Christian Lehmann based on (Comrie, 1981)

In linguistic and psycholinguistic studies, animacy is often assumed to be a graded property. In biology, being animate (being living) is a binary categorization. There are some areas, where debate exists whether a thing is "living" and what exactly the definition of a living being is (for example around viruses - are they living or not?), but the distinction between living and nonliving things is still binary. An item is either living or not. It seems to be the case however, that animacy as it appears in language, is graded (Radanović et al., 2016). In a study of both linguistic, semantic and biological animacy, Radanović and colleagues came to the conclusion, that linguistic animacy is most antropocentric, thus in language humans are often treated differently. In biology, animacy is binary. Semantic animacy lay between those two, with humans and animals being treated more similarly across two languages (Radanović et al., 2016).

Animacy is operationalized in experiments in different ways. Most experiment cannot cover the entire spectrum of (in)animacy, so specific categories are picked: As Radanovic noted, in many studies animates were represented by animals and inanimates by tools (to name a few different studies using that approach: Borghesani et al. (2016); Devlin et al.

(2005); Fairhall et al. (2011); Noppeney et al. (2005); Rundle et al. (2018); Simanova et al. (2012). Other studies have used more diverse stimuli sets, including a larger spectrum of the animacy continuum: Bell and Ungerleider used faces and body parts as animate stimuli and objects and places in the environment as inanimates (Bell et al., 2009) and Mahon and Caramazza used animals and tools but also nonmanipulable objects (Mahon et al., 2009). Finally, several studies by Connolly and colleagues decidedly investigate finer sub-categorization within the realm of visual animal stimuli (i.e. (Connolly et al., 2012; Guntupalli et al., 2016; Sha et al., 2015).

The animacy spectrum is an ideal candidate to study how conceptual knowledge is organized across modalities. There is vast research in language and visual perception that points towards animacy shaping aspects of behavior as well as neural processes. Animacy is also a very prominent category and at the same time contains many nuances that allow studying subcategories in detail. Finally, while animacy does play a role in both visual perception and language comprehension, it is not clear that therefore this category would shape both modalities in the same way. Animacy in visual perception is often thought to be influenced by the similarity of the body of animals (especially mammals and there especially conspecifics) whereas in language body similarity should not play a role. Languages also differ in how animacy is marked grammatically, so - given that there are over 6000 languages spoken today - it is possible that animacy in language is organized quite differently depending on which language a person speaks. The human body is not subject to variation in that way. Using animacy as the test case to study modality independent conceptual organization provides me with a category central to human cognition. Animacy plays an important role in shaping both visual perception and language comprehension. It also offers a rich space containing many subcategories that can be investigated in detail.

1.3.4 Research question

Humans make sense of the world by recognizing percepts from different modalities and formats, i.e. visual and auditory or pictorial and lexical as representing certain concepts that are grouped into larger categories. Amodal conceptual representations that are independent of modality and format, have been proposed (Caramazza and Mahon, 2003). The evidence for amodal conceptual representations might, however, be systematically overestimated due to the use of tasks that center conceptual processing that have been used in the majority of the previous research (see chapter 1.3.1). Another limitation of the field is that the modalities used to study amodal conceptual representations are often limited to just two modalities (i.e.written words and pictures Noppeney et al., 2005; Devereux et al., 2013 or i.e. sounds and silent videos Man et al., 2012) and that some modalities that are used to study amodal conceptual representations do share at least part of their processing stages. Pictures and written words are different, but they share initial visual perception. It has not been tested, how amodal conceptual representations and categorical perception is modulated by testing 1) more than one modality and format and 2) by excluding tasks that prime participants for conceptual processing and by 3) sampling a wider spectrum of the animacy continuum. To expand the current knowledge on amodal conceptual representations specifically addressing these three caveats, I carried out a behavioral as well as two fMRI studies. I studied two modalities (visual and auditory perception), two formats (pictorial and lexical items), 6 categories covering the animacy spectrum (humans, mammals, insects, plants, tools and furniture) using tasks that did neither center specific attributes of the items nor were explicitly linked to category knowledge, but instead were purely perceptual.

CHAPTER 2

General methods

This chapter gives a short overview over the different methods I used. There are 3 main methods: (1) Representational Similarity Analysis (2) Computational lexicography (3) MRI and fMRI ,which was used to obtain data from the brains of human participants. Each experimental chapter will have its own method section in which a more detailed account of the specific methods will be given. This is mainly an overview to facilitate understanding.

2.1 Representational Similarity Analysis

A method to investigate information content contained in data patterns is representational similarity analysis (RSA) (Kriegeskorte et al., 2008a). This method can be applied to a variety of data, for example to behavioral data, fMRI data and to computational modeling data, making it possible to compare these different methods. Representational similarity analysis is built on a long tradition of mathematical similarity analysis (Kriegeskorte et al., 2008a). The basic idea of it is to calculate representational dissimilarity matrices (RDMs) based on the data associated with specific stimuli, conditions or mental states. The data pattern associated with a given a stimulus or condition is taken to be the representation. Pairs of data patterns are made comparable by calculating the distance between them which is entered into the RDM. This is done for each stimulus pair, yielding a diagonal matrix with the diagonal being the comparison of the stimuli with themselves (dissimilarity = 0) and identical lower and upper triangles that contain the comparison of the nonidentical stimuli (Kriegeskorte et al., 2008a). The RDM contains the distance between the conditions or stimuli and thus captures the information content in the sense that it captures which aspects of the conditions that are compared are highlighted and which are not (Kriegeskorte et al., 2008b). The RDM also can be calculated for a variety of neuroscientific methods, for example single cell recordings, fMRI spatiotemporal patterns or electrophysiological measurements but also for behavioral data or computational methods. RDMs calculated based on different data can again be compared with each other since the method abstracts from the specific channels and resolution of the different kinds of data (Kriegeskorte et al., 2008a). The possibility to compare vastly different methods makes RDMs a powerful tool to compare how information content is represented across different neuroscientific methods - for example to compare how behavioral judgments are related to the information represented in a linguistic corpus. In this work, I use representational similarity analysis to study behavioral measurements and relatedness measures obtained from a linguistic corpus analysis.

The steps undertaken when data is analyzed using representational similarity analysis are very similar across the different methods. First, for each stimulus or condition, the data pattern is estimated. How exactly this happens depends on the method at hand. In the second step, all stimuli pairs are compared to each other by calculating a distance between the patterns associated with them. All dissimilarity scores of all stimuli pairs are form the RDM. The next and third step consists of a comparison of several RDMs, for example the RDM of a dataset and a model or the RDMs of two datasets. In the 4th step, the (dis)similarity between the RDMs is statistically quantified using bootstrapping. The step by step process is described in Figure 2.1.

2.2 Linguistic corpus analysis - computational lexicography

Lexicography, the study of the semantic and syntactic relationships between words within a lexicon of a given language, has long been a part of the field of linguistics. In recent years, due to the arrival of computers, lexicography has shifted from the analogue techniques of lexicons and card boxes to the use of computer based lexicons. The amount

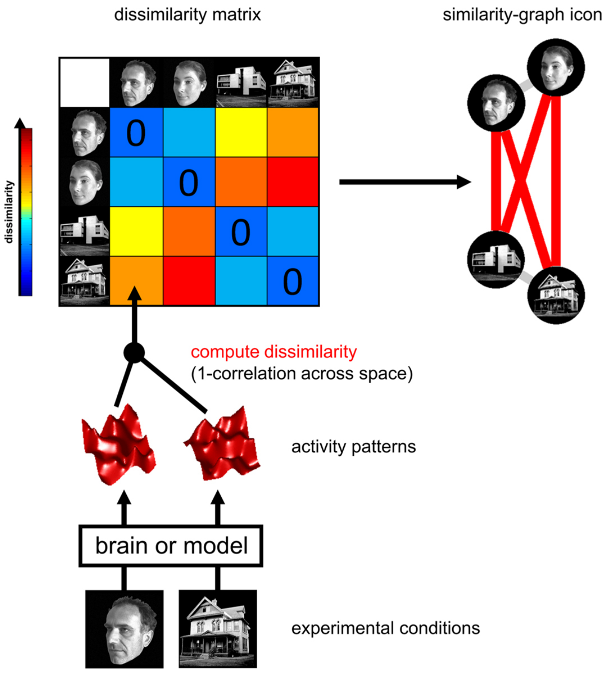


Figure 2.1: How RDMs are calculated from Kriegeskorte et al. (2008a). For each stimulus pair, a dissimilarity score is calculated as a distance between the two data patterns that form the pair. The dissimilarity value is entered into a matrix, that is indexed by the stimulus set (if the stimulus set includes i.e. 10 pictures, there would be a 10x10 diagonal matrix). The diagonal represents the dissimilarity of each stimulus with itself, which is 0. The other cells of the matrix represent the dissimilarity between other stimulus pairs. This process can be carried out for different kinds of data including behavioral data, fMRI data and data obtain from computational modeling.

of textual evidence that can be taken into consideration has also increased in recent years. With the advent of the internet, the availability of large text materials and increasing computational power, lexicons could perform more and more sophisticated calculations over increasingly large databases of words, giving scientists powerful tools to study language and specifically the relations between words. For my work, the most important aspect of it is, that it allows me to study how concepts associated with words are related to each other. I carried out my experiments in German, thus I needed a lexical database in German. I opted to use GermaNet (http://www.sfs.uni-tuebingen.de/GermaNet/) a database that is based on Princeton WordNet (https://wordnet.princeton.edu/), one of the largest lexical databases.

GermaNet divides the lexical space into different concepts that are connected by semantic relations (Henrich, 2010). One semantic concept is represented by a synset, which is a set of words that have almost exactly the same meaning and are thus synonyms (Henrich, 2010). The synsets provide the meaning associated with words in this corpus (Hamp and Feldweg, 1997). GermaNet sorts the synsets according to how semantically related they are and can provide information about many semantic relationships, for example part-whole relations or hyperonymy (Hamp and Feldweg, 1997; Henrich, 2010). Its focus on the meaning of words made it suitable for the goals of this work, namely studying conceptual representations in behavioral data. The database is divided into nouns, adjectives and verbs, but I only investigated nouns. The underlying data for this database are printed German lexica such as the Duden, Thesauri (though they are less common in German than in the English speaking world) and other printed monolingual German dictionaries. The main goal in using this database was to calculate semantic relatedness between words. I therefore used the semantic relatedness API that is implemented in the database to compute distances between the stimulus words I chose. There are different relatedness measures available in this implementation. I opted for the relatedness measure "path". It is computed by the formula $sim(s_1, s_2) = \frac{MAXSHORTESTPATH - length(s_1, s_2)}{MAXSHORTESTPATH}$ where MAXSHORTESTPATH is the longest possible "shortest" path between any two synsets in GermaNet. It is a measure of the distance between two synsets (meanings associated with words) which makes it comparable to similarity judgments.

2.3 Magnetic Resonance Imaging and functional Magnetic Resonance Imaging

A lot can be said and has been said about the contributions of MRI and fMRI to neuroscience. It is impossible - without writing several textbooks - to explain every detail of this method which allows us to study the brain structure and activity of both animals and humans while they are still alive. Since the first functional fMRI study was performed in 1991 (Belliveau et al., 1991), the method has become ubiquitous. MRI and fMRI combines knowledge from many different fields: physics, which brought us the use of the magnetic field and the phenomenon of spin, vascular biology, which studied where blood vessels in the brain are located and how they dilate, cellular neurobiology which studied how the work of neurons is related to blood flow in the brain and how this blood flow again manifests in the fMRI signal and statistics which discovered the tools to make sense of the data we obtain. I will give an overview to enable understanding of the experimental chapters that use fMRI and chapter 4 which describes the acquisition of the dataset. I will mainly follow the seminal textbook by Huettel et al. (2014).

2.3.1 The physical basis of MRI

The name ”magnetic resonance imaging” already suggests that the magnetic force plays a role in this method. An MRI scanner is a giant coil magnet that has a static magnetic field that is measured in Tesla. In cognitive neuroscience, 1.5 or 3 Tesla scanner are frequently used. Any substance that is not diamagnetic (essentially nonmagnetic with no field of its own) that enters an MRI scanner, will react to this static field. This is why it is important to remove certain objects such as keys or credit cards before having an MRI scan taken. Ferromagnetic substances, such as iron, is drawn into the magnet. A strong magnetic field such as the 3T field of a Siemens Prisma scanner, also influences human tissue, specifically the nuclei of hydrogen which consist of one proton. Hydrogen is common in water which is common in the human body including the brain. Protons have a quantum mechanical property called spin. Leaving out what spin actually is in quantum mechanics, it is easiest to picture that the protons literally spin around an axis. This axis aligns with the axis of the magnetic field in a scanner. Once the head of a person enters the field of a scanner, all its protons’ axes align their spin with the field. Because the proton is charged, the spin leads to there being a magnetic moment in the direction of the axis with a specific frequency called Lamor frequency. Some protons spin parallel to the axis of the magnetic field and some spin in the opposite direction which requires more energy. This state of proton spin can be excited by inducing energy at the Lamor frequency which leads to some parallel spinning protons adopting the high-energy spin direction. Once the energy stops, the nuclei relax into the low energy state. The excitation pulse pushes the net magnetization from the axis of the static magnetic field into the traverse plane. The relaxation after the energy pulse stops has two parts : the magnetization in the direction of the scanner field recovers (T1 recovery) and the traverse magnetization is lost (T2 decay). The MRI signal is the transverse magnetization of all voxels (Huettel et al., 2014).

How does MRI create an image using the magnetic forces on nuclei? In order to scan an entire brain in it, a slice of it is excited by a pulse. MR images are always acquired in slices that image one part of the head. Within the slice, phase and frequency gradients are applied in order to be able to capture the x- and y-dimensions of the slice in space. Different pulse sequences are used to created images that are T1 or T2 weighted.

T1 weighted sequences are typically used for structural scans that display the different tissues in the brain (white matter, grey matter, cerebrospinal fluid, bones etc.) whereas T2*[1] weighted images are used for functional imaging and the BOLD contrast (Huettel et al., 2014). The images used here were acquired using a T1 contrast in the case of structural images, and a T2* contrast in case of the functional images.

2.3.2 fMRI and the BOLD contrast

Functional MRI does not measure differences in brain tissue, but instead differences in brain activity, making it possible to locate where activity happens in the brain and connect this activity with different functions. Brain activity comes from the coordinated activity of neurons, which is electric. There are different types of electric activity of neurons that can be measured. The most well known is probably the 'firing' of neurons, which describes the neuron's action potential. Apart from firing, neuronal activity also produces other measurable electric effects, for example local field potentials (LFPs) which result from differences in ionization in the space between neurons due to neuronal activity. The electric activity generated by neurons can be measured using electroencephalographic methods but it is too small to be detected by MRI scanners. Increases in neural activity however also increase blood flow in the brain which leads to more oxygen being delivered to the brain regions in which the activity is located. fMRI is able to capture the difference between oxygenated and deoxygenated blood which have different magnetic properties. The signal used in fMRI studies depends on this difference and is called Blood-oxygenation-level-dependent (BOLD) contrast. Oxygenated hemoglobin is diamagnetic because it does not have any unpaired electrons, whereas deoxygenated hemoglobin has free electrons which makes it paramagnetic. The paramagnetic deoxygenated hemoglobin leads to faster T2* relaxation because it disturbs the magnetic field leading to faster dephasing of the traverse magnetization.

The differences in magnetic properties of the hemoglobin make the BOLD contrast possible. The blood containing both oxygenated and deoxygenated hemoglobin is not in the brain itself but transported to it through a network of blood vessels in the head. It was discovered in 1990 that (1) the different magnetic properties of oxygenated and deoxygenated blood makes it possible to image blood vessels and (2) that it also can capture differences in brain activity (Ogawa et al., 1990). The BOLD response follows

---

[1] T2* and T2 both describe the decay of the traverse magnetization, but T2* describes the actual decay including effects of field inhomogeneities of the scanner, whereas T2 describes a more ideal decay

a typical course which can be described by the hemodynamic response function (HRF). It has a brief initial dip after which the signal amplitude rises to its peak at around 4-6s to fall to the initial level unless the brain region is further stimulated by an event. The HRF can be viewed in figure 2.2

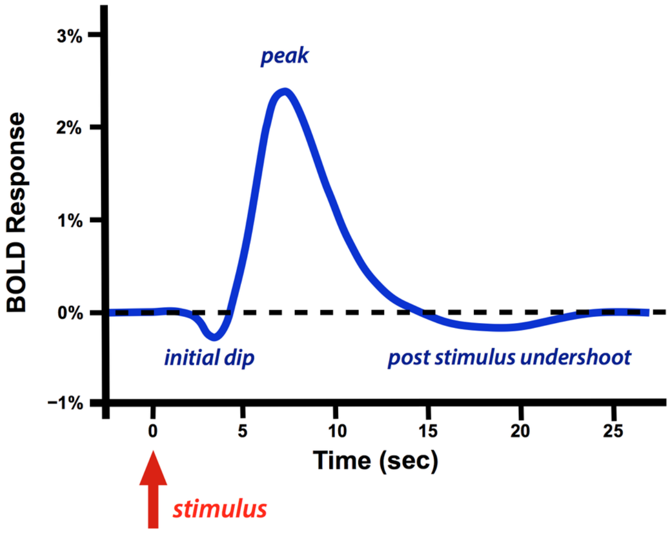


Figure 2.2: The canonical hemodynamic response function (HRF), taken from http://mriquestions.com/does-boldbrain-activity.html). The BOLD response reaches its peak at around 4-6s after which it levels off again to a slight post stimulus undershoot until the next stimulation occurs.

fMRI studies report effects that are, in essence, blood flow related, but in many cases - in my case, too - the blood flow isn't what the studies are interested in. Instead, fMRI studies often investigate neural activity for which the blood flow is a proxy. How related blood flow and neural activity actually are is not completely understood, but it has been a very active field of research. The BOLD response has been observed to correlate with more direct measures of neural activity for example local field potentials which reflect intracortical processing and input to cortical regions (Logothetis et al., 2001; Logothetis, 2008).

2.3.3 Analyzing fMRI data

2.3.3.1 Preprocessing

The images generated by the scanner need to undergo a series of modifications to be analyzed. The raw data usually contains several sources of variance that are not related to the experimental manipulation that are reduced by preprocessing steps before analysis. How I carried out each step for the data I collected is described in 4, but why and in what way fMRI data is preprocessed before analyzing it deserves a deeper explanation. One important step during preprocessing is to correct field inhomogeneities by calculating a field map (*field map correction*). The field map is calculated from two gradient echo sequences that are acquired at different times. The difference between the two sequences can be used to calculate the voxel displacement map which can be used to correct for the field inhomogeneities that caused the displacement. The functional images are also realigned which accounts for the motion in the scanner (*realignment*). fMRI experiments require participants to be very still but it is almost impossible to not make involuntary movements when being scanned. Apart from that, respiratory and metabolic processes also induce movement. The images are fitted together in all 3 axes of space by translation and rotation parameters being fitted. One functional image is acquired in slices consecutively. Due to that process of acquisition, the slices of one TR are not all acquired at the same time. This is accounted for during *slice time correction.* In order to achieve a higher spatial resolution and to match functional and structural scans of one participant, functional scans are co-registered (*coregistration*) with the structural scans. This is somewhat similar to the realignment process, only more complex because functional images compared to structural images are blurry and have a different voxel size. In order to make results comparable and because all human brains are different, the functional images are also normalized to a template brain. Anatomical landmarks in the brain are used to match functional scans to the template in this step called *normalization.* In most cases, the functional data is also smoothed using a Gaussian kernel to improve the signal-to-noise ratio (*smoothing*) (Penny et al., 2011).

2.3.3.2 The General linear model: Activation mapping

Most fMRI studies are still analyzed using the general linear model $Y = \beta X - \epsilon$. In this formula, Y stands for the time series of one voxel in the brain, X is a design matrix based on how the experiment was carried out, ϵ is a term for noise. The term β, often referred to as beta, contains the parameter weights for each regressor in the design matrix that are estimated. The design matrix is constructed based on the events in the experiment. The experimenter knows when what kind of event was shown in each experimental condition and for how long. All events in the experiment are modeled as a box - function. The box - function containing all events is convolved with a model of the

HRF 2.2 to obtain a model of how a voxel that is reacting to the type of event shown would react over time. A linear regression is carried out for each voxel that was scanned using the ordinary least square method. The estimated weights of different conditions once estimated can be compared in a statistical test (either a t-test or an F-test) by attaching contrast weights to them (Huettel et al., 2014; Penny et al., 2011).

CHAPTER 3

Conceptual organization in human similarity judgments

3.1 Introduction

Categorization is thought to support perception in providing meaning for what we perceive - to help us "make sense of the senses". The store of meanings of both objects in the world and symbols such as languages or street signs is called semantic knowledge. Semantic knowledge seems to be organized along specific dimensions. One particularly important dimension for human categorization seems to be the distinction between animates (humans and animals) and inanimates (objects). Humans acquire this distinction early in their life (Poulin-Dubois et al., 1996; Rakison and Poulin-Dubois, 2001) and this distinction guides a plethora of behaviors 1.3.3. Another important distinction between items lies between so called "natural" and man-made objects (Rosch et al., 1976).

Within these larger dimension that drive the organization of semantic knowledge, there is further hierarchical organization based on specific features. The animates are further subdivided into humans and animals, the animals into mammals and insects and fish and so on. Depending on personal knowledge and experience, the world of existing things can be subdivided differently between individuals (for example, depending on your knowledge you might perceive spiders as insects or group whales together with sharks). The process of sorting things into these groups is called categorization and the groups or classes are called categories.

Categories like human vs. animal or the larger distinction between animates and inanimates play a role in visual perception as well as audition and language. More generally, there are many different ways to perceive an item (one can see it, hear it spoken about, read about it, certain things can be smelled or experienced through touch), yet there is an impression of being able to perceive the same thing as I attempted to illustrate with my seal cookie example in the introduction. The meaning of percepts has been thought of as being independent of any modality, being amodal. New insights however point towards cognition and specifically things like word meaning being "embodied" for a longer explanation of the ideas of embodied cognition see **??** The ideas of distributed semantics put the theory of amodal concepts into question: If the meaning of symbolic items such as words and of pictures of real world objects is grounded in the sensory experiences that occurred when encountering them, how can semantic knowledge be modality-independent or even amodal? On the other hand, there is also evidence for considerable overlap between modalities.

In order to study how comparable the organization of concepts is between language and visual perception, I carried out a Similarity Judgment study using words and pictures denoting the same concepts. I focused on specific aspects of semantic knowledge, especially the animate-inanimate distinction, but also on the animacy continuum by choosing concepts from 6 categories, 3 increasingly animate (insects, mammals and humans) and 3 increasingly inanimate (plants, tools and furniture). Participants carried out a Multi-Arrangement experiment in both formats in which they sorted 60 concepts in both formats according according to their perceived similarity. I then analyzed the data using representational similarity analysis.

3.2 Materials and Methods

3.2.1 Participants

20 students (13 female), native language German, participated after giving written informed consent.

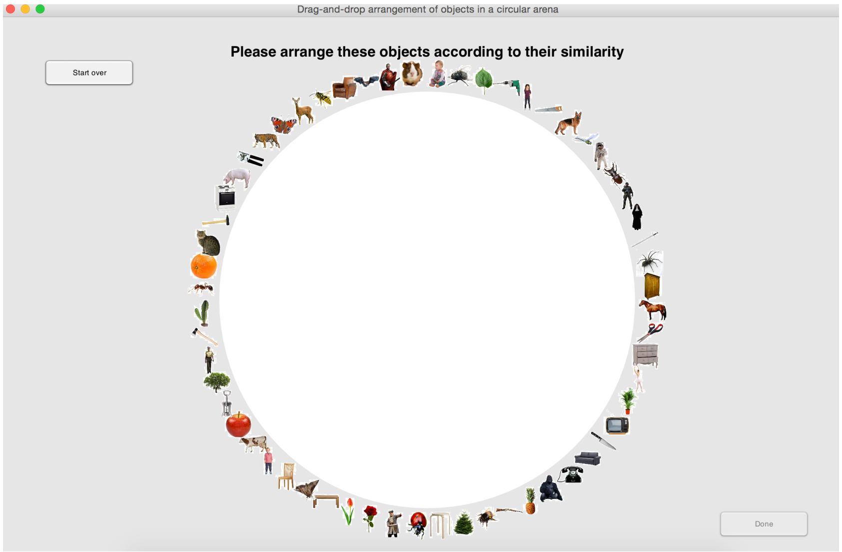
3.2.2 Stimulus Material

I selected 60 concepts from 6 categories that spanned the animacy continuum (humans, mammals, insects/crustacea, plant parts, tools/weapons, furniture). The concepts were rendered in two formats, as color photographs and written words. The photographs and words all had a white background. The written words were created using the gimp software (https://www.gimp.org/). The font used was Arial, the words were centered on 300*300 pixel square images filling out the entire square with an edge of. The photographs were presented in the center of white square images, 150*150 in size. In order to ensure comparability between the categories, we controlled word frequency, grapheme count and visual complexity of the written words measured in number of pixels. I controlled visual complexity of the photograph stimuli measured in number of non-white pixels per image. There was no difference in visual complexity for either words ($F = 0.256$; $p = 0.935$) or photographs ($F = 0.306$; $p = 0.907$). There was also no difference in either logarithmic or non-logarithmic word frequency between categories (non-logarithmic frequency: $F = 1.698$, $p = 0.151$, logarithmic frequency: $F = 2.189$; $p = 0.0689$). There was also no difference in grapheme count ($F = 0.417$; $p = 0.835$).

3.2.3 Procedure

In order to be able to collect dissimilarity data from 60 stimuli and therefore $(60*60-60)/2 = 1770$ stimuli pairs in a reasonable amount of time, we used the multi-arrangement method introduced by Kriegeskorte and Mur (2012). This experimental setup makes it possible to collect dissimilarity values from a large stimulus set where using pair-wise comparisons would be unfeasible. In its basic form, stimuli are presented around a white circle on a computer screen called "arena". Participants are instructed to sort the stimuli on the arena by moving them into it using drag and drop. The closer together the stimuli are, the more similar they are in the participants' mind. Figure 3.1 depicts the first trial for the picture format. This sorting task is repeated over several trials, the exact amount of which depends on the collected evidence for each stimulus pair. The first trial has the participants sorting through the entire stimulus set. Once all stimuli have been placed on the arena, the next trial starts. In all trials following the first, a subset of stimuli is selected using the "lift the weakest" algorithm that selects the stimulus pairs with the least evidence for their dissimilarity. The evidence depends on the closeness of stimuli, with the idea being that the closer stimuli are on the arena presented on the screen, the less specific participants are in making smaller judgments about how similar or dissimilar a group of stimuli is. The Lift-the-Weakest algorithm thus allows the participant to zoom into a subset ($n >= 3$) of stimuli and make finer distinctions between stimuli (for example, in a set containing animals and plants, there could be finer distinctions made between wild and domesticated animals which might be lost in the larger scale of the distinction between animals and plants).

Our experiment instead of presenting one stimulus set, presented two sets that contained the same concepts but referred to either by words or by pictures. I chose to present the two tasks in a mixed form instead of carrying out two blocks of sorting words and pictures. The experiment started with two trials in which participants sorted through the entire word and the entire picture stimulus set. All participants started with the the written word stimuli to avoid the picture information biasing the word sorting. They carried out the sorting task for all written words and in the second trial for all the photographs. Following these first two trials, participants carried out the sorting task for subsets of both word and picture stimuli depending on stimulus pair evidence according to the Lift-the-Weakest algorithm (see figure3.3). This meant, that if one format had less evidence for its stimulus pairs, trials would be carried out only in that format. The mixed presentation after the initial two trials was chosen in order to avoid one of the formats exerting influence over the other. Every trial could be either a picture or a word trial and would contain a unique stimulus set, presenting a unique sorting task for the participant. This style of presentation was chosen with the intent to keep the focus on the sorting based on similarity of the stimulus set at hand. On average, participants carried out 24.05 word trials and 48.75 picture trials. There were significantly more picture trials carried out than word trials ($p < .001$). The difference in trial number could be due to the picture stimulus set containing a richer set of features compared to the word stimuli.



Please arrange these objects according to their similarity

Figure 3.1: Example of one possible first trial for the pictorial format. All 60 stimuli are grouped around the white circular arena . The instructions ”Please arrange these objects according to their similarity”, are displayed on top of the arena.

Please arrange these objects according to their similarity

Figure 3.2: Participants were instructed to arrange the items based on how similar they intuitively felt to them. Items could be moved into the arena using drag and drop. The closer two items were sorted together, the more similar they felt.

Please arrange these objects according to their similarity

Figure 3.3: In the trials following the first two, subsets of items were presented using the "Lift-the-Weakest" algorithm so that participants could sort the subsets with more detail. The algorithm selects the stimulus pairs with the least current evidence starting from the pair with the least amount of evidence. Here, the algorithm has selected mostly plant items that can be sorted into groups using the entire space of the arena.

The experiment was self-timed in that participants could decide how much time they spend sorting through one trial. The experiment ended when either all stimulus pairs reached sufficient evidence weight or when the experiment had been running for longer than 1h30min. Evidence for a stimulus pair depends on the distance between the two stimuli on the screen. It is assumed that larger distance increases the signal-to-noise ratio, since smaller distances mean less precise differences in distance. Evidence weight is calculated as the square of the signal-to-noise ratio of a given stimulus pair (Kriegeskorte, 2014). Sufficient evidence weight is reached at evidence weight $w = 0.5$. This is based on the evidence utility function of the evidence weight $u(w) = 1 - e^{-wd}$ which saturates at around w= 0.5 (Kriegeskorte, 2014). For a depiction of the evidence utility function see 3.4

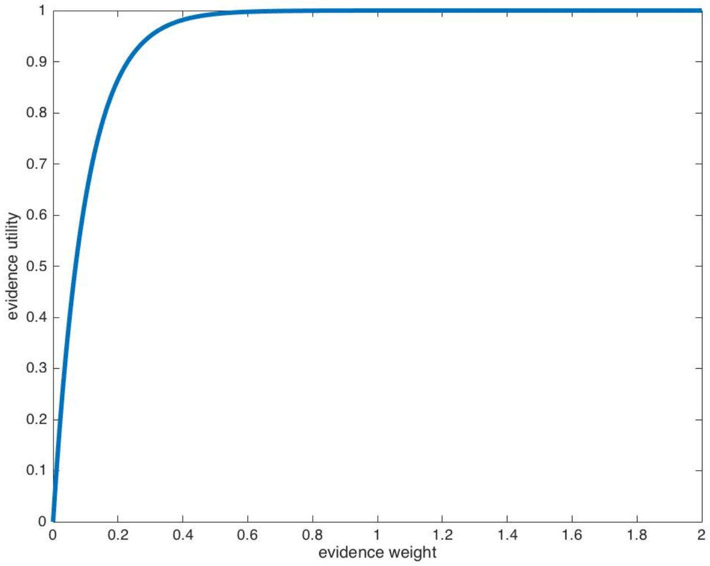


Figure 3.4: Evidence utility saturation curve. At around evidence weight of 0.5 the evidence utility for a stimulus pair levels off.

This time condition was introduced since in some cases, single stimulus pairs failed to reach the evidence weight threshold of 0.5 due to scaling[1] reasons, which lead to the experiment going on infinitely in theory. The evidence saturation curve can be seen in figure 3.4. Participants were instructed to sort both words and photographs according to how similar they felt stimuli were based on intuition.

3.2.4 Statistical Analysis

The data was analyzed using Matlab and the RSA toolbox by (Nili et al., 2014). representational dissimilarity matrices (RDMs) were calculated for each format for each individual subject based on the combined data from all trials each participant performed. Individual RDMs were then averaged to obtain one averaged RDM for both formats. I tested which models best explained the dissimilarities as observed by the the participants. Based on the hypothesis that participants would sort the items according to category, I used a categorical model to test the data. I also hypothesized that particular perceptual aspects of the items might play a role, so I also used models that were based

---

[1] When only 3 pictures or words were left in the arena, they were scaled up to such a degree that not enough difference in distance could be made in order for the last pairs to meet the evidence weight threshold

on word length and visual features of the photographs, i.e. color. Another aspect that I hypothesized to play a role in dissimilarity perception is place or occurrence. I used another model that grouped items based on whether they frequently occurred together to test the data obtained for written words. This model was based on relatedness measures derived from a corpus of German texts, namely GermaNet. I calculated relatedness scores using the relatedness measure ”path” implemented in the Germanet software for all 1770 unique word pairs and created a model RDM based on the relatedness scores. I also tested the picture data with this model. Finally, to compare how similar conceptual structures for the two formats were, I directly compared the RDMs of the written word and picture format. Statistical testing for to assess whether a model RDM is significantly related to a data RDM is carried out using a one-sided rank test (Nili et al., 2014). The different models are also compared with regard if they differ in how they relate to the data RDM using a two-sided rank test where false-discovery rate is also controlled (Nili et al., 2014).

3.3 Results

In order to ensure comparability between the two conditions I first calculated whether evidence utility was comparable for both and whether both conditions actually reached a satisfying evidence utility. The results for both evidence weight and evidence utility for words and pictures can be seen in figure 3.5.

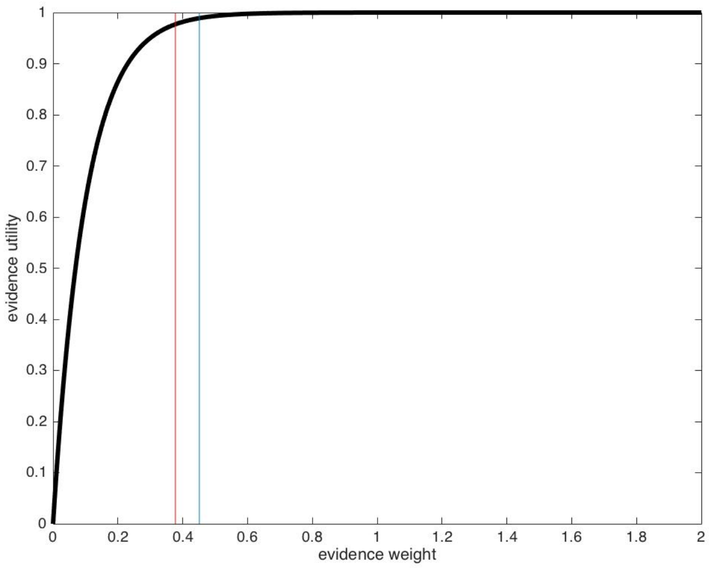


Figure 3.5: Figure 2.
Evidence utility saturation curve with mean evidence weights for pictures (in blue) and words (in red). Due to the time limit we set in the experiment, not all stimulus pairs reached the evidence weight of 0.5. Words reached a mean evidence weight of 0.377 and pictures of 0.4513. Evidence utility for these evidence point reached 0.9771 and 0.9890 respectively.

3.3.1 The models

I tested the data with 5 models that focused on different aspects of the stimulus set. Three models were categorical in nature. The first model was based on the animacy of the items, 30 of which were animate and 30 of which were inanimate. The second model instead focused on the specific categories we had created the stimulus set with, namely humans, mammals, insects, plants, tools and furniture. Finally, the last categorical model combined animacy and the 6 categories to form a combination model with one overarching category (animacy) and 6 subcategories. I also used two models that were derived from data obtainable about our chosen stimuli. I created one model of visual similarity between stimuli based on the photographs we used as picture stimuli. This model was included because I hypothesized that participants might use visual similarity to sort through the dataset especially when they sorted the photographs. Visual similarity was calculated based using the ssimval algorithm implemented in matlab which calculates how similar two images are based on both luminance and contrast measurements (Wang et al., 2004). The final model focused on the linguistic aspects of the stimulus set. I calculated the relatedness measure between the word pairs to test whether linguistic aspects such as co-occurence or interchangeability played a role. All five models can be viewed in figure 3.6.

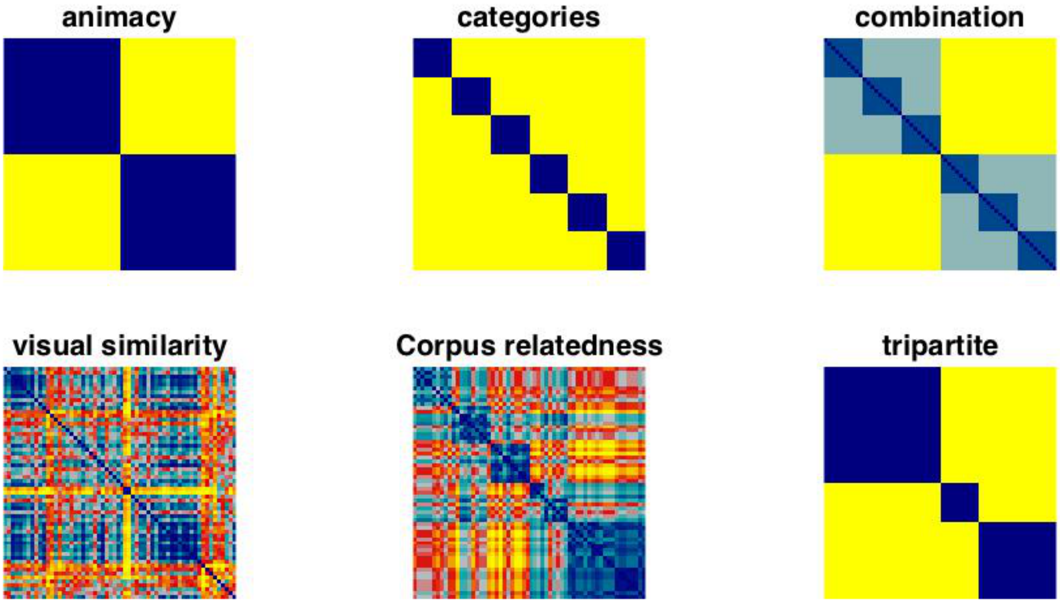


Figure 3.6: The five models used to test the behavioral data obtained in this first RSA experiment. **1.** Animacy: Animacy was modelled as a binary distinction. Humans, Mammals and Insects were modelled as animate whereas Plants, Tools and Furniture were modelled as inanimate. **2.** Categories: In this model, all 6 categories were modelled as distinct leading to the six clusters seen in the model RDM. **3.** Combination model: in this model the animacy model and the category model were combined in one model to account for the possbility that participants used both aspects to sort stimuli. **4.** Visual Similarity: The 4th model RDM represents the visual similarity between stimuli which was calculated based on pixel-size distances. **5.** Semantic Relatedness: This model RDM was calculated based on relatedness scores obtained from the GermaNet corpus for each stimulus pair. **6.** The tripartite category model RDM is based on the observation that human categorization is not binary but instead distinguishes between living things, especially animals, man-made objects and nonliving objects and plants that are neither (Cree and McRae, 2003) .

3.3.2 The words

The result RDM of the picture stimulus set can be seen in figure 3.7. Participants sorted the words into the 6 categories. They also seem to have sorted all animal word stimuli as closer together while the other animate category (humans) were not sorted to be as similar and have sorted tools and furniture together while the 3rd inanimate category (plants) were not grouped with the two other inanimate categories. Plants and tools seem to have been perceived as particularly dissimilar.

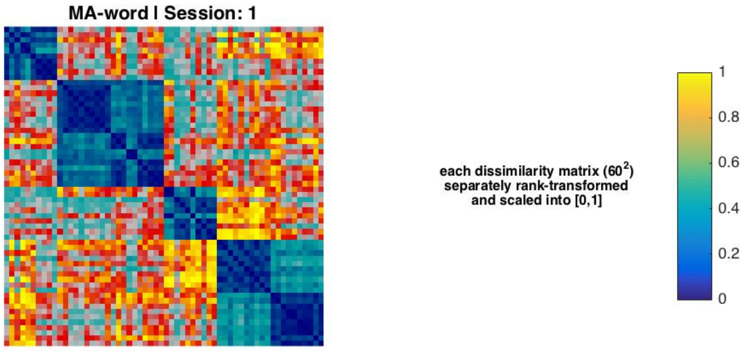


Figure 3.7: Average RDM for the word condition for all 20 participants. Blue colors indicate more similar stimuli, whereas yellow and red colors indicate dissimilarity. The results indicate a categorical structure to human similarity judgments regarding words with humans forming a cluster in the left upper corner. Lower on the diagonal a larger cluster of animals is visible which is subdivided into a mammal and an insect cluster. Even lower on the diagonal, a cluster for plants is visible. In the lower right corner, the two inanimate categories, tools and furniture, form a cluster with two clearly defined sub-clusters.

In order to further explore the dissimilarity between the word stimuli, I visualized the data as hierarchical tree using hierarchical cluster analysis (Johnson, 1967) which assumes a hierarchical category structure. The tree reveals that words were sorted into two basic categories, one containing furniture and tools the other containing insects, mammals, humans and plants. In the firsts category, words were sorted into two categories that were part of the experimental design, namely furniture and tools. The second cluster was further subdivided into two clusters, one containing all animals and the other one humans and plants. The animal cluster is further subdivided into a cluster containing the insects and the other the crustaceae and the mammals. The other cluster is subdivided into humans and mammals. In figure 3.8, the hierarchical tree for the word stimuli can be observed.

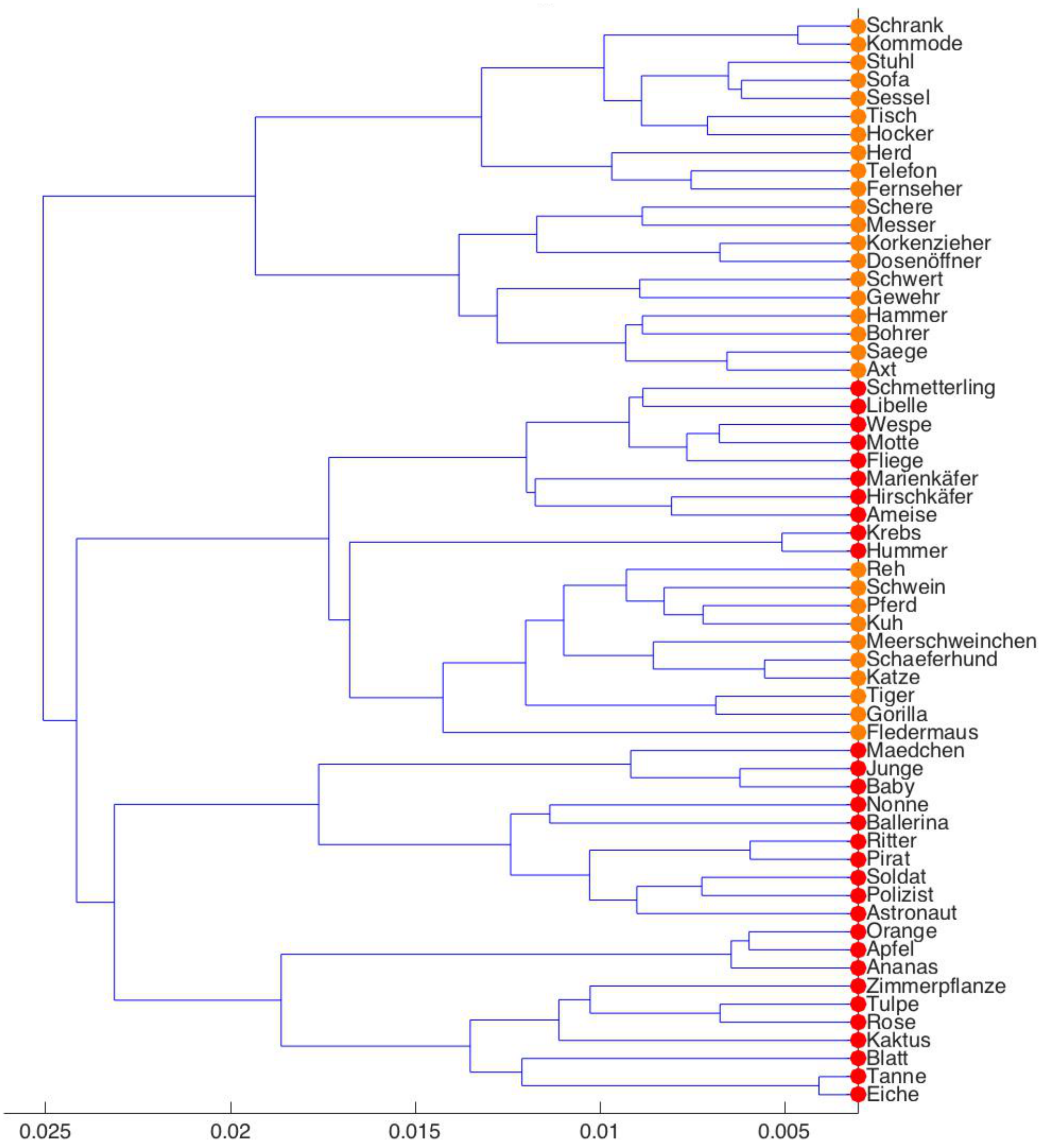


Figure 3.8: Dendrogram for the word condition. The hierarchical tree visualization presupposes a hierarchical category structure. The main distinction appears to be between inanimates and animates. The Inanimate group subdivides into two subgroups: furniture and tools. The animate group is further subdivided into a group containing both animal categories and a second group containing a bit surprisingly plants and humans.

I compared the models with the RDM for the word condition in order to test which of the 6 models would represent the trends in the data. The results of this statistical analysis can be observed in figure 3.9. All six models were significantly correlated with the data RDM. Also, all six models stayed below the noise ceiling which is visualized as the horizontal grey bar on top of the six vertical model bars. The noise ceiling indicates the performance of the assumed true model given the noise in the data. Its upper bound is given by the average correlation of the average RDM (the data RDM) with the individual subjects' RDMs (upper bound = 0.3463) and the lower boundary is calculated by correlating an average RDM leaving each subjects' data RDM out and then correlated the left out individual RDM with the average RDM without this subjects' RDM (lower bound = 0.2591). The model that explained the behavioral data in the word condition the least was the visual similarity model that was based on features of the picture stimulus set (p = 0.0319, r = 0.034). The binary animacy model showed a higher correlation with the data (r = 0.1492, p = 0.000001) followed by the 6-category model (r = 0.1756 , p = 0.000001). The three models that were most highly correlated with the data were the tripartite model (r =0.1839 , p = 0.000001), the corpus relatedness measure model (r =0.2068 , p = 0.000001) and the combination model which combined animacy and categories (r = 0.2187, p = 0.000001).

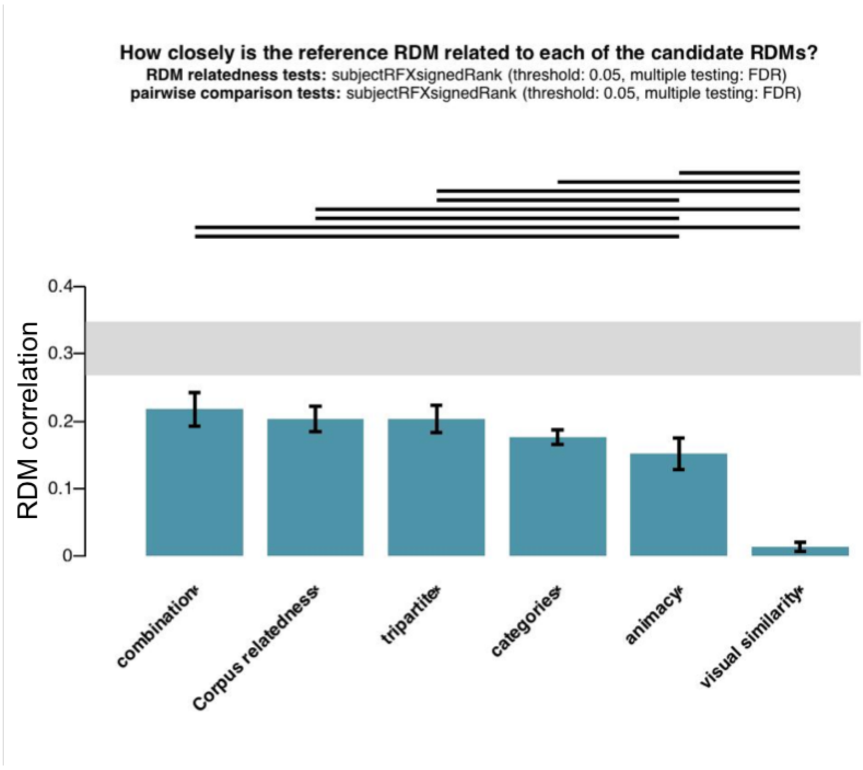


Figure 3.9: Statistical Results of the word condition. 6 models were used to test the data, two representing specific aspects of the stimulus set, namely visual similarity of the pictures and corpus relatedness of the words and four being based on the categorical information in the stimulus set: a binary animacy model, a model representing the 6 categories and a model representing 3 broader categories and finally a combination of animacy and category model. The combination model was most highly correlated with the data as was the corpus relatedness model and the tripartite category model, followed by the models representing categories, animacy and finally visual similarity.

3.3.3 The pictures

The result RDM of the word stimulus set can be seen in figure 3.10. As for the word stimuli, several clusters are apparent to the eye. On the leftmost upper corner the human stimuli are clustered together, followed on the diagonal by a cluster for the two animal categories. Lower down on the diagonal lies a cluster of the plant stimuli, followed by the object categories tools and furniture. The animals seem to have been sorted particularly far from the human stimuli as indicated by the yellow cells.

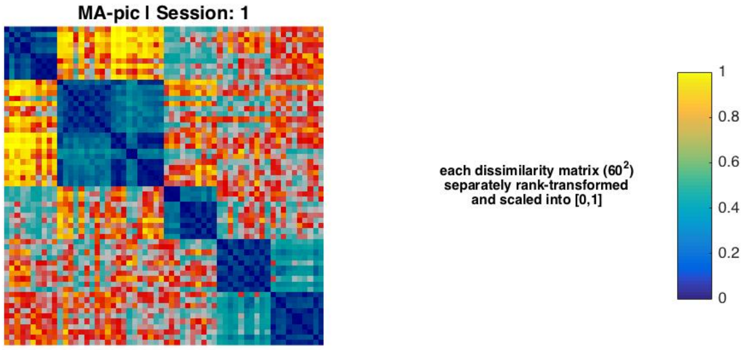


Figure 3.10: Average RDM for the picture condition. In the picture condition, a clear category structure is visible, similarly to the word condition. 4 cluster are observable: a human cluster in the upper left corner, followed by a cluster containing both animal categories. The next cluster on the diagonal is the plant cluster. In the lower right corner, the two artifact categories form a cluster which is also clearly subdivided into tools and furniture.

Just as in the word condition I also calculated a hierarchical cluster analysis for the picture stimuli. In figure 3.11, the hierarchical tree for the picture stimuli can be seen. Interestingly, humans and plants form a group. The other four categories, the two animal categories and the two man-made object categories tools and furniture also appear together. The human and plant category is further subdivided into the categories humans and plants. The larger division further subdivides into man-made objects which further subdivides into tools and furniture and animals, which interestingly further subdivides into the crustaceae category and the other animals that further split into mammals and insects.

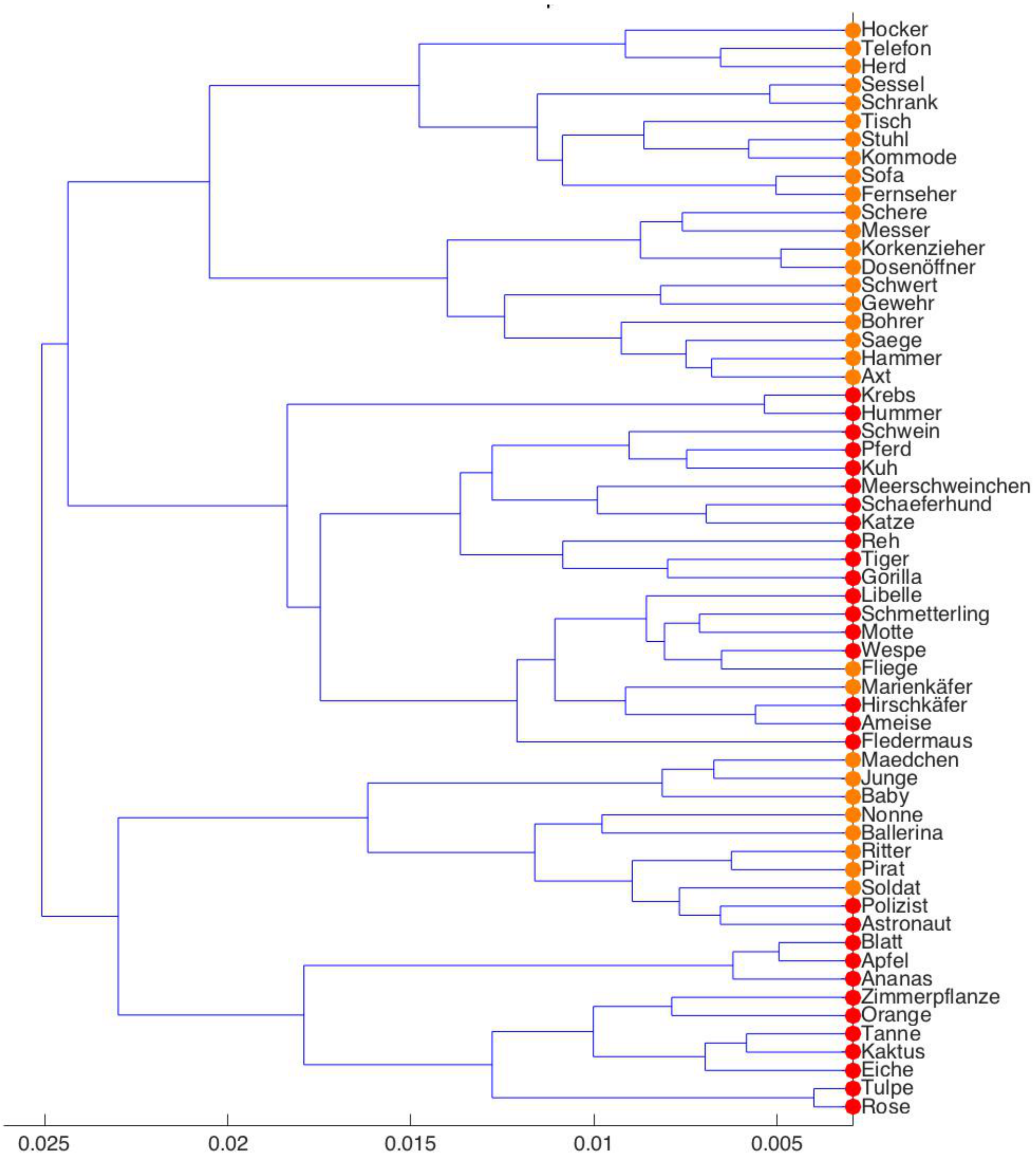


Figure 3.11: Dendrogram for the picture condition. The dendrogram for the picture condition also displays a subdivision between animates and inanimates. As in the word condition, the inanimate group is subdivided into tools and furniture. The animate group subdivides into an animal group and again the human and plant group which was also present in the word condition.

I compared the models with the RDM for the word condition in order to test which of the 6 models would represent the trends in the data. The results of this statistical analysis is visualized in 3.12.

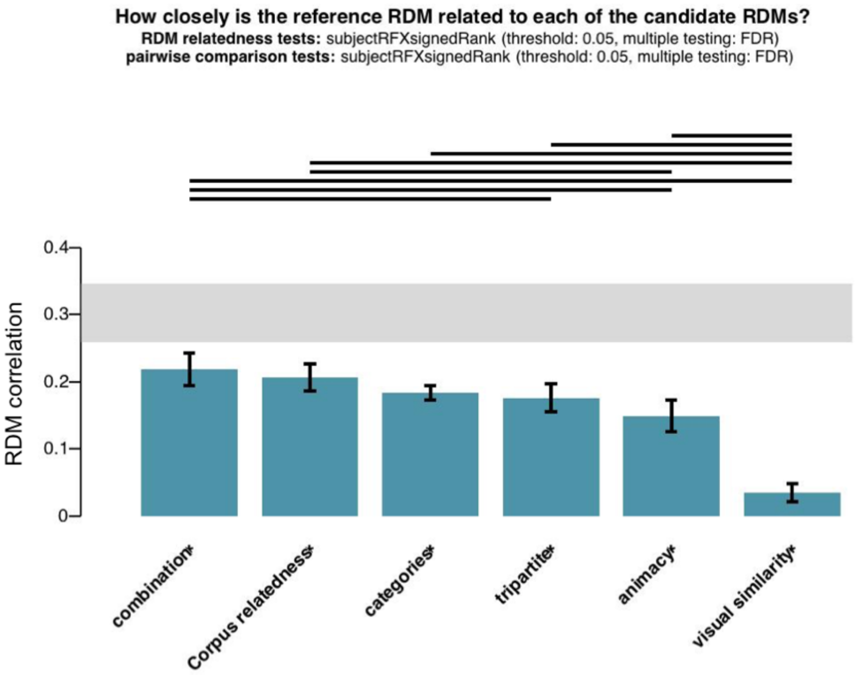


Figure 3.12: Statistical Results of the picture condition. The model that best explained the data, was, similarly to the word condition, the combination model that unites animacy and category distinctions. The corpus relatedness model performed almost as well, followed by the category model, the tripartite model and the animacy model. Notably, the visual similarity model did not perform well in the picture condition either, despite being based on low level aspects of the stimuli.

As for the word condition, all six models significantly correlated with the data and no model reach the noise ceiling (upper bound = 0.3479 , lower bound = 0.268) The model that showed the highest correlation to the data was, as it was in the word condition the combination model (r = 0.2179, p = 0.000001) followed by the corpus relatedness model (r = 0.2035, p = 0.000001). For the picture condition the third best performing model is a different one from the word condition, namely in the case of the picture condition it is the category model (r = 0.2032, p = 0.000001), followed by the tripartite model (r = 0.1769, p = 0.000001). The animacy model (r = 0.1518, p = 0.000002

) and - interestingly given that the visual similarity model is based on aspects of the picture stimuli - the visual similarity model (r = 0.0136, p = 0,0319) are the two lowest performing models.

3.3.4 The similarity of words and pictures

To quantify how similar the data for words and pictures was, I compared the two using one as the model RDM and the other as data RDM. The two RDMs were highly correlated. When using the picture data RDM as model RDM, it significantly correlated with the word data RDM with a correlation coefficient of r = 0.2652 which was just below the lower bound of the noise ceiling (lower bound = 0.3479, upper bound = 0.2689). The opposite test, using the word data RDM as model RDM and correlating it with the picture data lead to a correlation that reached into the lower bound of the noise ceiling (lower bound =0.2591, upper bound = 0.3463) r = 0.2672. The correlation between the two data sets thus is higher than any of the model RDMs I used.

3.4 Discussion

Perceived similarity between German words and pictures is highly comparable. The RDMs for both formats were highly similar, even more than any format RDM was to any of the tested models. This result points towards similar organizational principles for both formats in human perception. This result is in line with previous research (Tyler and Moss, 2001; Bright et al., 2004; Devereux et al., 2013). It is also in line with the hypothesis, that at least at some point during perception and semantic processing, format plays less of a role in categorization. Human categorization did however show some differences between the formats for some aspects. As can be observed in figures 3.11, and 3.8 sorting was quite similar between formats. The results show the same intriguing pattern, namely that the main division is between natural things (humans, mammals, insects and plants) and man-made things (furniture and tools). This is in line with previous research as the man-made vs. natural distinction is a quite important categorical distinction (Rosch et al., 1976; Warrington and Shallice, 1984; Iordan et al., 2015). What is puzzling, however, is that the group of natural things is not divided or sorted by animacy exactly (although the animacy hierarchy contributes to the pattern of results as the analysis indicates), but instead, two groups emerge, one containing animals, the other humans and plants.

It is very plausible, that the two formats' semantic organization differs in some aspects, as they are perceived in a different way, especially since the two formats have distinct properties. The identity of a concept presented in the picture format is recognized based on shape, color and texture, which are visually processed. In the word format, processing is first carried out by the visual perception system as well, but the language comprehension system is also involved. Shape or color and other visual properties of the words presented do not play a role in concept recognition. Instead, words are recognized based on their letter sequence. Linguistic properties play a role in word reading, such as the length of the word or its frequency. In order to test, which aspects drove human judgment of item similarity, I carried out another analysis comparing the data RDMs with models that emphasized different aspects of the stimulus set.

I observed that, both words and pictures are sorted based on a mixture of category boundaries. Both larger boundaries such as the distinction between animates and inanimates and single categories such as mammals, insects or tools play a role. Perceived similarity correlates with corpus based relatedness scores and is not particularly influenced by visual similarity in either format. It is interesting, that visual similarity seemed to play no great role in either format. The model performed better for the picture format, but even there it does not reflect the categorization judgments made as well as models that focused on semantic properties. Interestingly, the model derived from corpus path relatedness did not outperform the combined categorical model and performed only slightly better than the tripartite model. While this model is able to include greater variance between single concepts, there might be an even better approach to capture how natural concepts are used, namely corpus derived measures that only focus on distributional features of words and construct word meaning and word concept similarity based on co-occurence data (see Pereira et al., 2013). GermaNet on the other hand is based on word semantics and relations between meanings of words (see also chapter 2.2).

Another interesting aspect of the data that as far as I am aware cannot currently be tested is real-world object co-occurence scores. The current dataset contains similarity judgments of both words and pictures. To my knowledge, there are currently no co-occurence or relatedness scores of objects or images. It is possible that models based on real-world co-occurence of objects in the experience of humans might explain the data - especially the similarity judgments of pictures - to an even greater extent and reveal other principles that guide participants in sorting these stimuli. One particular option that could in my opinion be promising is the scene grammar approach by Võ et al. (2019). On visual inspection of the data, it seems that one principle that drove similarity judgments, was co-occurence in real world circumstances based on participants' experiences. Humans were grouped with tools, furniture and sometimes houseplants, whereas insects and mammals were grouped together separate. This seems to indicate that apart from categorical boundaries and semantic relatedness, real - world co-occurence played a role. Testing this hypothesis based on a visual grammar model is not possible as of now. Another way to test, if this is a driving force of similarity judgment could be to test other participant groups for example farmers, to see if they group farm animals in a different way based on their lived experience.

A linguistic semantic based corpus such as GermaNet does not seem to capture aspects of real - life co-occurence patterns such as humans and plants either because they don't reflect category differences that are connected to word semantics. There is a growing number of computational models that acquire word vectors based on corpus data (Mikolov et al., 2013; Pennington et al., 2014; Devlin et al., 2018). Apart from the scene grammar by Võ et al. (2019), these corpus based word vectors are also a promising avenue to test how co-occurence influences similarity judgments.

One limitation of this study is the fact that words and pictures were presented in a mixed way to all participants. Both formats' conceptual organization appeared to be very similar. This effect could be due to participants being primed to perceived them as similar because they made their similarity judgments on both formats in one experiment. Also, all participants gave their judgments on both formats in one session. An interesting follow-up experiment in order to study how much the effect is influenced by the proximity of the two formats, would be to separate the two formats more and more. Participants could carry out the sorting tasks in separate blocked sessions or on different days. My hypothesis is that this proximity might lead to a slight decrease in similarity between the formats, but not change the overall pattern of results.

3.5 Conclusion

Similarity Judgments for pictures and words highlight similar categories, focusing on animacy and finer category subdivisions such as mammals, insects or tools. This supports the idea that semantic knowledge follows similar organizational principles for different formats, in this case written language and visual perception. It also does not contradict the idea that semantic knowledge could be stored independent of format at least in parts of the brain. In order to further investigate how different parts of the brain deal with different semantic knowledge accessed through different formats, further studies investigating and comparing language and visual perception of concepts in the brain are needed.

CHAPTER 4

The dataset: stimuli, procedure and fMRI data

4.1 Participants

Twenty participants (13 female, mean age 23.4 years, SD = 2.5) took part in the experiment. All participants were right handed, had normal or corrected-to-normal vision and no psychological or neurological impairments. They were financially compensated or received course credit for their participation. All participants gave informed written consent prior to the experiment in accordance with the ethical standards laid down in the Declaration of Helsinki. The experimental protocols were approved by the local ethics committee

4.2 The stimulus set

The stimuli consisted of 42 objects from six different semantic categories (humans, mammals, insects, plants, tools and furniture). All 42 objects were presented in 2 input modalities (visual or auditory) and in 2 input formats (lexical or pictoral) i.e. as pictures (visual pictorial), written words (visual lexical) and spoken words (auditory lexical), yielding a total number of 126 stimuli for the experiment. The 42 picture stimuli were colored photographs of isolated real-world objects on a gray background (RGB: 128, 128, 128). Spoken words were recorded by a female speaker and standardized in terms of rise-time, fall-time, peak intensity and peak amplitude using a custom Praat script (Boersma and Van Heuven, 2001). I controlled several variables to ensure that differences in stimulus perception were due to categorical differences and not due to featural differences between single categories of stimuli. For the object picture stimuli I controlled that there were no differences between the categories in terms of visual complexity. Visual complexity was measured as number of pixels. We found no difference in number of pixels between the categories (Anova : $F= 1.445$, $p= 0.232$). For the word stimuli I compared both logarithmic and non-logarithmic frequency measures were obtained from the Celex corpus (http://celex.mpi.nl/). We observed no difference between categories for either frequency nor logarithmic frequency (frequency: $F= 1.445$, $p= 0.232$; log frequency: $F = 1.725$, $p = 0.154$) These measures controlled word frequency for both written and spoken words. However, the Celex corpus is of written text. Word frequency measures are usually obtained from written corpora because there is more written text readily available and it is easier to analyze written text. A corpus of spoken language would first have to be transcribed or at least single words would have to be recognized in order to compute word frequency.

We also controlled for word length in three ways: grapheme count, phoneme count and length of recorded audio. Phoneme count did not differ for the three categories ($F=$ 1.365; $p = 0.261$) and neither did the length of the recorded audio file for each stimulus ($F = 0.381$; $p=$ *0.86*) or the grapheme count: (F=0.563, $p = 0.728$).

The stimulus set was first tested in a behavioral experiment which is described in chapter 3. We used the same stimuli but reduced their number to be able to conduct this experiment in the scanner in a reasonable amount of time. The stimuli that were discarded where concepts that are refered to by particularly long words in German, for example 'ladybug' (Marienkäfer) in order to make stimuli more comparable in terms of word length. Specifically the category of insects and tools contained longer words and even though the groups were not significantly different, we aimed to make the 6 categories even more similar in terms of their surface properties.

4.3 The experimental procedure

The experiment was controlled using Presentation software (Neurobehavioral Systems, Albany, USA). A camera (MRC Systems MR 12-m-i) was used to monitor participants' gaze in order to ensure that stimuli were fixated. The 126 stimuli were presented in a rapid event-related design. Visual stimuli were presented foveally for a duration of 650 ms on a constantly visible uniform gray background centered on the screen. A grey fixation cross was presented to stabilize participantsǵaze on the center of the screen. Participants viewed the visual stimuli spanning a visual angle of about 7°. on a screen (Nordic Neurolab AS 1920x1080, 60Hz Monitor) placed behind the scanner bore via a surface mirror. Viewing distance was 140 cm in total. Auditory stimuli were presented with their natural length ranging from 427 to 917 ms. They were presented via fMRI compatible headphones (MR Confon, HP AT01). Each trial duration was set to three seconds. Each stimulus was presented exactly once per run.

In addition, the sequence included 18 catch trials, in which participants either saw a visual stimulus being flipped 45° clockwise, or a pure tone (1000 Hz sine wave, duration: 100ms) interrupted the presentation of the spoken word stimulus (always 350 ms after stimulus onset). Participants were instructed to press a button whenever they detected one of these rare events. Catch trials were modeled as covariates of no interest in the fMRI data analysis. The sequence also included 36 null trials with no stimulus presented (four of them at the beginning, and 32 randomly interspersed within the sequence). The trial-onset asynchrony was three seconds; the stimulus-onset asynchrony was either three seconds or a multiple of that duration when null trials occurred in the sequence. The trials (including 126 stimulus presentations and 32 interspersed null trials) occurred in random order (no sequence optimization). For each run, a different random sequence of stimuli was generated with the only constraint of not more than two consecutive null trials. One run lasted approximately nine minutes. Participants saw 6 runs on 2 sessions on two different days, altogether 12 sessions. Before the Experiment, participants were instructed both in writing and verbally to focus on the middle of the screen, and observe carefully the objects they would perceive visually and aurally. They were also instructed to respond with a button press to a flipped trial or a tone in or after a spoken word. The original instruction text is included in the Appendix 8.

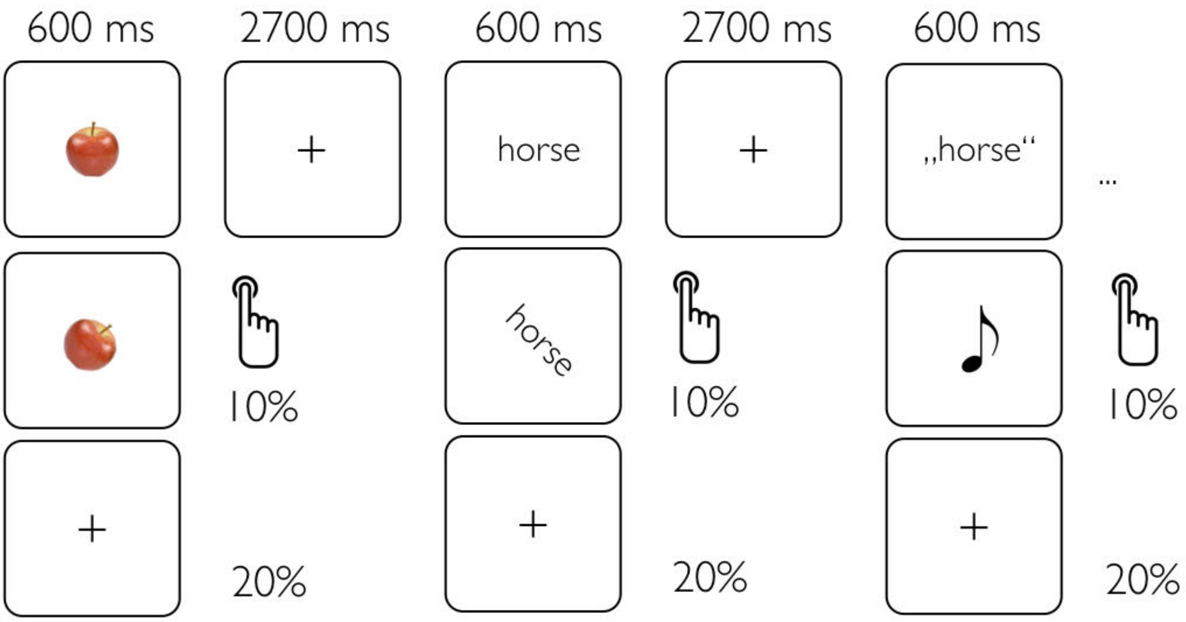


Figure 1. Time course of the fMRI experiment. Participants either viewed objects, read written words or listened to spoken words. 18 times in a session, a test trials was presented. 6 randomly selected spoken words were interrupted or followed by a sinusoidal tone, 6 randomly selected written words and pictures flipped by 45°. Participants had to indicate that they perceived these events by pressing a button.

4.4 fMRI data acquisition

Functional and anatomical MRI data were acquired at the Bender Institute of Neuroimaging (Giessen, Germany) using a 3T PRISMA Scanner (Siemens, Erlangen Germany) and a 64-channel head coil. Head motion was minimized using cushions. A gradient-echo field map was measured before the functional runs to receive information about inhomogeneities in the static magnetic field. For functional imaging, a total of 3190 volumes were registered on average per person. A T2*-weighted gradient-echoplanar imaging (EPI) sequence was used with 38 axial slices covering the whole brain [slice thickness: 3 mm; 0.75 mm gap; descending slice order; echo time (TE): 30 ms; TR: 2.1 s; flip angle: 78°; field of view: 192 mm; matrix size: 64*64 mm; voxel size: 3.0*3.0*3.0 mm]. Functional imaging was preceded by a field mapping sequence with identical voxel size, resolution and field of view. Flip angle was set to 90°, and TR was 1000ms. For the magnitude image, TE was set to 10ms while for the phase image, TE was set to 12.46 ms. Structural images consisting of 176 sagittal slices were acquired using a fat-suppressing T1-weighted magnetization-prepared, rapid-acquisition gradient echo sequence (matrix size: 256* 256 mm; field of view: 240 mm; TE: 3.53 ms; TR: 1880 ms; voxel size: 0.94*0.938*0.938 mm).

4.5 fMRI data preprocessing

Data preprocessing was done using SPM 12 (Penny et al., 2011). It included (1) motion correction and B0 unwarping using field maps (combining the step of applying field maps to correct for field inhomogeneities and the realignment step)(2) slice time correction, (3) corregistration of functional and structural scans (4) normalization to the MNI-ICBM 152 brain template (MNI, Montréal, Canada), (5) spatial smoothing using a Gaussian kernel of 9 mm full-width half-maximum, (6) temporal high-pass filtering with a cutoff of 128 s/0.008Hz. All figures containing fMRI-data were created using Mango and overlaid on the MNI-ICBM 152 brain template (MNI, Montréal, Canada Mazziotta et al., 2001).

4.6 Behavioral measures of participants

4.6.1 Task performance

We included a perceptual task in our experiment for two reasons: (i) To ensure that participants were constantly paying attention to the middle of the screen (ii) To make the experiment more interesting for participants who without a task would just lie in the scanner and perceive the stimuli. The task I chose was purely perceptual and did not incite the participants to think about the meaning of the words and pictures they perceived. The reasoning behind choosing the task we included in our experiment and the influence different tasks have on brain activity is discussed in chapter 1.3.1. As described in Chapter 4.3, each run included 18 catch trials (6 for each modality) that participants were instructed to respond to with a button press. In general, participants were able to perform this task. Per run, participants missed less than one catch trial (mean = 0.4937; sd = 1,160). However, there were single sessions in which participants missed more than 5 catch trials, possibly indicating that the participant was not paying attention. To ensure the quality of the fMRI Data, these runs were marked and data from the eye camera and motion Data was considered in order to come to a conclusion as to whether to exclude these runs. This concerned 5 runs.

False alarms occured more frequently than misses, especially during the first session often indicating that the participant had not understood the task and pressed the button for every appearance of an image or a written word. When this was observed, the experimenters re-instructed the participants to only press the button when they heard a tone or when the image actively flipped. Since false alarms were an issue for several of the first participants, instructions were adjusted to include that the image had to actively flip while the participant was observing it and not be inherently flipped already. This change of instructed led to a decrease of false alarms. If all cases of false alarms are included participants on average had 2.4895 false alarms per run (sd = 10,6303). When all sessions with more than 40 false alarms (which was taken as a sign of misunderstanding of the instructions) were excluded as outliers, we observed 0.6304 false alarms per run on average (sd = 2.226).

4.6.2 Motion during the Scans

One source of artifacts in fMRI data is the movement of participants during the scanning procedure. fMRI experiments require the participant to lie as still as possible during a considerable time span (in our experiment about 1:30h) and this can prove stressful and uncomfortable. Participants tend to move either voluntarily or unvoluntarily. Therefore, fMRI data has to be checked for motion artifacts in order to ensure that observed activity is due to the BOLD signal measured in a voxel and not due to excessive motion. I took several steps before analyzing the data to control the data quality in terms of participant movement. I used the spm toolbox tsdiffana (http://imaging.mrc-cbu.cam.ac.uk/imaging/DataDiagnostics) to visualize variability in motion for single volumes for each run for each participant. This gave me a first impression of which volumes might be problematic and included too much noise due to participant movement. Then I used the spm toolbox ArtRepair (Mazaika et al., 2011) to further analyze variability in motion, calculate a threshold for the overall experiment of how much motion would be acceptable and to repair Volumes that were affected by excessive movement. I then repaired affected volumes using the threshold 0.5 indicated by the ArtRepair toolbox. The repair algorithm linearly interpolates from neighboring volumes to the correct volumes where movement exceeded the threshold (Mazaika et al., 2011).

CHAPTER 5

First univariate fMRI analysis: Modality-independent conceptual processing

5.1 Introduction

Questions concerning the nature of conceptual knowledge are at the heart of cognitive science and cognitive neuroscience. In addition to being of interest in and of themselves, concepts are assumed to support a wide variety of mental processes, ranging from language comprehension and production to visual perception and action planning. Previous research has led to divergent views on the organization of conceptual knowledge, with opposing perspectives centered largely around the question of whether there are segregated representational semantic systems for different sensory input modalities in the brain, or whether preconceptual, modality-specific processing converges in an amodal conceptual knowledge system (e.g. Patterson et al., 2007; Martin, 2007; Pulvermüller and Fadiga, 2010).

A key assumption of approaches positing unitary conceptual representations in the brain, is that these representations can be accessed independent of modality. For example, from this perspective, the word 'pirate' and the picture of a pirate activate overlapping conceptual content, and this may be implemented neuroanatomically via a semantic 'hub' Patterson et al. (2007). Accordingly, many studies on the neural correlates of conceptual representations have focused on precisely this comparison of visually presented words and pictures (e.g. Bright et al., 2004; Devereux et al., 2013; Vandenberghe et al., 1996; Chao et al., 1999). These studies typically show common activations in a distributed semantic processing system shared by both visually presented words and pictures comprising regions in the middle and inferior temporal cortex, and the inferior frontal gyrus. These heteromodal perisylvian association areas have been dubbed 'hub territory' by Pulvermüller and colleagues (e.g. Pulvermüller, 2013; Pulvermüller and Fadiga, 2010) in recognition of the fact that they have been implicated in general semantic processing in a number of studies, while none of them can be viewed as a unique hub. As argued by Binder, Desai and colleagues (Binder et al., 2009; Binder and Desai, 2011), they are high-level convergence zones for supramodal semantic processing.

However, the comparison between pictures and written words does not constitute an orthogonal comparison of sensory input modalities, but rather a comparison of lexical and non-lexical stimulus material within the same modality, i.e. vision. Thus, while the above-mentioned studies provide insights into converging semantic processing of lexical and non-lexical visual input, they cannot provide evidence for cross-modal conceptual representations in the brain.

Far fewer studies have directly compared the processing of lexical material in the auditory and visual modalities, i.e. spoken and written words, which share their verbal input format, but differ in input modality (auditory vs visual, e.g. Vartiainen et al., 2009), or the processing of lexical material presented auditorily (spoken words) with that of visually presented pictures, a stimulus set that does not overlap with respect to either sensory input modality or lexicality (e.g. Costanzo et al., 2013). To the best of our knowledge, only one neuroimaging study has attempted to identify the neural correlates of conceptual representations by varying both lexicality and modality for a set of concrete objects (but see also von Stein et al., 1999) for an EEG-based approach to this question). Handjaras and colleagues (Handjaras et al., 2016) used fMRI to examine the perception of objects as pictures, written words and spoken words, and found evidence for amodal conceptual processing. However, the main research question of this study was whether conceptual processing differs between sighted and congenitally blind individuals. Accordingly, stimulus modality (visual vs. auditory) and format(lexical vs. pictorial) was tested as a between-subjects factor in this study with four participant groups including 5 participants seeing or hearing each stimulus category. In addition, Handjaras et al. (2016) used a property-generation task in order to fully map the semantic network associated with amodal conceptual stimulus processing. Previous research, however, has shown that the spread of activation within a network encoding conceptual knowledge depends on the amount of semantic processing that participants had to engage in during and/or after stimulus presentation for written words, spoken words, and pictures e.g. (Vandenberghe et al., 1996). Thus, given the limitations of the study by Handjaras et al. (2016) both in terms of the experimental design and task, it is difficult to conclude on the basis of this study that conceptual knowledge in the brain relies on a distributed modality-independent representational system. At the very least, using semantic tasks such as the property-generation task might artificially inflate the extent of abstract object property processing over that engaged in during mere perception. Consequently, it remains largely unknown whether, and if so, in which neural networks, multimodal stimuli converge, particularly in the absence of tasks that require participants to engage in explicit semantic processing such as superordinate picture naming or property generation.

To shed further light on this question, the present fMRI experiment employed a direct, within-participants comparison of items in picture, written word and spoken word form in the context of an indirect (i.e. non-semantic) task. To avoid top-down expectations about the categories involved in the experiment, I used a perceptual task rather than a conceptual or semantic task.

5.2 Methods

The stimulus set, experimental procedure, imaging parameters and data preprocessing methods are described in detail in chapter 4 which describes the dataset in detail. Here, I will only describe the analysis of the fMRI data.

5.2.1 Data Analysis

For each participant, we modeled one separate predictor for each experimental condition, resulting in three explanatory variables (EVs) of interest, namely visually presented pictures (vis_pic), visually presented words (vis_word), and aurally presented words (spo_word). The spoken words were modeled as being 0.97s long each. For each condition, the EV modeled the 42 stimuli in each run. I also included four EVs of no-interest to the GLM, modeling catch trials (one per input format) and nulltrials. Six motion parameters and scanner drift were included as covariates. For group-level analyses, parameter estimates were assessed with a random-effects model. For the baseline contrasts as well as the conjunction analyses, Z- statistic images were generated using a Z-statistics threshold of 4.44 and a FWE-corrected cluster probability threshold of $P = 0.001$. The clusters and peaks were labeled using the AAL atlas (Rolls et al., 2015; Tzourio-Mazoyer et al., 2002). Only regions that covered more than 4% of the cluster were named and labels were written down in order of how much of the cluster they covered. I looked at overlapping activity for pairs of modality or format using conjunction analyses. In order to identify amodal clusters responding to all objects regardless of input modality, I calculated a conjunction analysis over all three baseline contrasts: ([vis_pic>Null], [vis_word > Null] , [spo-word > Null]).

5.3 Results

I analyzed the data in three ways. First we look at three contrasts: Spoken Words > Written Words and Pictures, Written Words > Spoken Words and Pictures and Pictures > Spoken Words and Written Words. I then carried out 4 conjunction analyses. I first tested for the overlapping activity of all possible pairs of contrasts, looking at conjunctions of Spoken Words > Rest and Written Words > Rest, Written Words > Rest and Pictures > Rest and the two maximally different modalities Spoken Words > Rest and Pictures > Rest. Finally, I carried out a conjunction analysis of all three modality conditions > Rest. I carried out all analyses using spm12, including the the build in method in spm12 based on (Nichols et al., 2005).

5.3.1 Results of the 3 modality and format contrasts

I first compared activity for all spoken word stimuli with written words and pictures. This contrast elicited activity in 4 clusters. The first cluster spanned 7843 voxels and several brain regions in both hemispheres: the LIFG, the left temporal pole, the right calcarine sulcus and cuneus, the left angular gyrus, the left middle cingulate gyrus and the left middle temporal gyrus. The cluster had three peaks, all three in the left middle temporal gyrus. A second cluster of 3451 voxels was observed in the right hemisphere. The cluster lay in the regions of right temporal pole, the right superior frontal gyrus, the hippocampus, the middle temporal pole, angular gyrus, medial orbitofrontal gyrus,

Table 5.1: Spoken Words compared to Written Words and Pictures

Region	Cluster Size	Cluster	MNI coordinates			Voxel	Voxel T48
		pFWE-corr				pFWE-corr	
			x	y	z		
LIFG (BA44), L temporal pole, R calcarine S, R cuneus, L angular G, R precuneus, L mid cingulate G, L cuneus, L lingual G, R lingual G, L insula, L sup occipital G	7943	0.000					
L MTG			-60	-34	14	0.000	21.03
L Heschl's G			-48	-16	2	0.000	20.58
L rolandic operculum			-42	-31	11	0.000	16.56
R sup temporal pole, R sup frontal G, R hippocampus, R mid temporal pole, R angular G, R medial frontal orbital G, RIFG (BA47, 45), RMTG, RSTG	3451	0.000					
R rolandic operculum			63	-19	11	0.000	20.32
R Heschl's G			57	-10	2	0.000	19.21
R MTG			63	-25	5	0.000	18,87
L sup medial frontal G, R mid cingulate G, R ant cingulate R, R SMA, R sup frontal G, L SMA	722	0.000					
R SMA			6	20	56	0.001	6.1
L mid cingulate G			-3	17	32	0.002	6.02
L ant cingulate G			3	29	20	0.007	5.59
R mid frontal G, R sup frontal G	108	0.045					
R sup frontal G			24	44	17	0.002	6

RIFG (pars orbitalis and triangularis) and it had three peaks in the rolandic operculum and two in the superior temporal gyrus. The next cluster with 722 voxels covered regions in both hemispheres. It lay in the left medial superior gyrus, the right middle cingulate gyrus, the right anterior cingulate gyrus, the left middle cingulate gyrus, the right SMA, the right superior frontal gyrus and the left anterior cingulate gyrus. The three peaks were observed in the right SMA and the left anterior and left middle cingulate gyrus. The last cluster included 108 voxels and lay in the right middle frontal gyrus and the right superior frontal gyrus, with one peak also lying in the right superior frontal gyrus. A summary of the results can be seen in table 5.1.

Secondly, I compared activity for Written Words with the two other modalities, Spoken Words and Pictures. This contrast elicited activity in two clusters. The first, 334 clusters, lay in the left hemisphere, in the left fusiform gyrus, the left inferior occipital gyrus, the left inferior temporal gyrus with three peaks in the left calcarine sulcus, the left lingual gyrus and the left inferior occipital gyrus. The second cluster included 143 voxels and was located in the right hemisphere. It spanned the right fusiform gyrus, the right lingual gyrus, the right inferior occipital gyrus and the right middle occipital gyrus. One peak was located in the left inferior temporal gyrus. The detailed summary of these results can be seen in table 5.2.

Table 5.2: Written Words compared to Spoken Words and Pictures

Region	Cluster Size	Cluster pFWE-corr	MNI coordinates			Voxel pFWE-corr	Voxel T48
			x	y	z		
L Fusiform, L inf occipital, LITG, L calcarine S, L lingual G, LMTG, L mid occipital G	334	0.000					
L inf occipital G			-21	-97	-7	0.000	8.52
L inf occipital G			-30	-91	-10	0.000	7.59
LITG			-42	-64	-10	0.010	5.48
R Fusiform, R Lingual G, R inf occipital G, R mid occipital G, R calcarine S	143	0.018					
R inf occipital G			33	-88	-4	0.000	6.6

Table 5.3: Pictures compared to language stimuli

Region	Cluster Size	Cluster pFWE-corr	MNI coordinates			Voxel pFWE-corr	Voxel T48
			x	y	z		
R mid occipital G, L hippocampus, R postcentral G, L inf occipital G, R fusiform, R inf occipital G, L sup occipital G, LITG, L mid occipital G, L fusiform G, R lingual G, R inf parietal G, R sup occipital G, RITG, RMTG	5413	0.000					
R inf occipital G			33	-82	-13	0.000	16.07
L inf occipital G			-42	-79	-4	0.000	14.64
R fusiform			33	-52	-13	0.000	14.62

The third contrast compared Pictures with the two other modalities, Spoken Words and Written words. There was only one large cluster with 5413 voxels. The cluster reached both hemispheres and lay in the regions of the right middle occipital gyrus, the left hippocampus, the right postcentral gyrus, the left inferior occipital gyrus, the right fusiform gyrus, the right inferior occipital gyrus and the left superior occipital gyrus. There were three peaks in the right inferior occipital gyrus, the left inferior occipital gyrus and the right fusiform gyrus. See table 5.3 for a summary of the results.

5.3.2 Results of the 3 modality and format contrasts exlusively masked

In order to test for activity that was exclusive to one modality, I looked at each modality compared to the baseline and masked this contrast with the two other contrasts for the two other modalities. I will report only two of the three modalities here. Written words > Null exclusively masked by Spoken Words > Null and Pictures > Null did not elicit activation in any clusters at a threshold of $p < 0.001$. Spoken words compared to baseline and masked by the two contrasts Written words > baseline and Pictures > baseline elicited activation in 3 clusters. The First cluster (2672 voxels) was located in the left hemisphere, in the temporal lobe. I also looked at the 500 most activated voxels of this cluster. These 500 most active voxels were located in the left temporal lobe, in the areas of the superior temporal pole, the rolandic operculum, the insula, the Heschl's Gyrus, the middle temporal lobe and the superior temporal lobe. The three peaks were observed in the left superior temporal lobe and the cluster also extended into the temporal pole, the rolandic operculum, the insula, the Heschl's gyrus, the middle temporal lobe. The cluster also reached the frontal lobe, namely the left inferior frontal gyrus. The second cluster observed for spoken words > baseline masked by the two other clusters spanned 923 voxels and was located in the right hemisphere, also mainly in the temporal lobe. The 500 most active coxels spanned the regions of the temporal pole, the insula, Heschl's gyrus, the superior and middle temporal gyrus, but also the postcentral gyrus and RIFG. The three peaks, similarly to the 1st cluster, were located in the right superior temporal gyrus. The cluster included the right superior temporal pole, the right insula, the right postcentral gyrus, the Heschl's gyrus and two areas in the right inferior frontal gyrus, the pars triangularis and orbitalis. The final cluster had 292 voxels and was located in the Cerebellum with all three peaks also in the cerebellum.

The contrast "Written Words > baseline masked with the two other contrasts compared to baseline did not elicit activity in any cluster in the brain. The contrast Pictures > Baseline masked by the two word contrasts did however elicit activity in 3 clusters. The first (237 voxels) was located in occipital and parietal areas of the right hemisphere. There were three peaks, two in the right middle occipital gyrus and one in the right cuneus. The cluster also lay in the right superior parietal gyrus, the angular gyrus and the right middle occipital gyrus. The second cluster, including 133 voxels, was located in the right temporal lobe, including the parahippocampal gyrus, the right amygdala, the right hippocampus and a part of the cerebellum. The three peaks were located in the right fusiform gyrus, the right lingual gyrus and the right amygdala. The smallest cluster with 128 voxels was located in the left occipital and parietal lobe. The three peak locations lay in the occipital gyrus (two in the superior occipital gyrus and one in the middle occipital gyrus). The cluster also extended to the superior parietal gyurs and the precuneus.

5.3.3 Results of the pairwise conjunctions

I first looked at the overlapping activity of spoken words and written words. I observed three 5 clusters for this conjunction. The first cluster had 6219 voxels and lay in the regions of the left precentral gyrus, the LIFG (pars opercularis), the right middle frontal

Table 5.4: Conjunction of all language stimuli: spoken words and written words

Region	Cluster Size	Cluster pFWE-corr	MNI coordinates			Voxel pFWE-corr	Voxel T48
			x	y	z		
L precental G, LIFG (BA44), R mid frontal G, R SMA, R precental G, L sup medial G, L mid frontal G, L pallidum, L sup temporal pole, L ant cingulate G, RIFG (BA44), RITG, R amygdala, R insula, RIFG(BA47), L/R sup frontal G, R putamen, R sup medial frontal G, R putamen	6219	0.000					
L precental G			-42	2	53	0.000	8.54
L SMA			-6	14	50	0.000	8.33
R SMA			6	17	56	0.000	8.04
Cerebellum	1210	0.000					
Cerebellum			33	-64	-28	0.000	7.32
Cerebellum			27	-67	-49	0.000	7.06
Cerebellum			-33	-61	-28	0.001	6.23
LMTG, LSTG	502	0.000					
LMTG			-51	-58	8	0.000	6.07
LSTG			-63	-40	11	0.013	4.79
LIPG			-51	-46	-1	0.069	4.22
L angular G, L IPG, L mid occipital G, L SPG	280	0.001					
LIPG			-33	-55	47	0.000	6.78
L cingulate, R mid and post cingulate	194	??					
R post cingulate G			3	-34	26	0.019	5.28
L mid cingulate G			-6	-7	32	0.504	4.03
Conjunction Visual							
L Fusiform, L inf occipital, LITG, L calcarine S, L lingual G, LMTG, L mid occipital G	334	0.000					
L inf occipital G			-21	-97	-7	0.000	8.52
L inf occipital G			-30	-91	-10	0.000	7.59
LITG			-42	-64	-10	0.010	5.48
R Fusiform, R Lingual G, R inf occipital G, R mid occipital G, R calcarine S	143	0.018					
R inf occipital G			33	-88	-4	0.000	6.6

gyrus, the RIFG (pars triangularis), the right supplementary motor area, the right precentral gyrus, the left medial superior frontal gyrus and it had three peaks, in the left precentral gyrus and the left and right SMA. In order to look at the focus of the activation in this cluster, I looked at the 500 most activated voxels of this cluster. These voxels lay in the right supplementary motor area, the left medial superior frontal gyrus and the middle cingulate gyrus. The second cluster had 1210 voxels and lay in the Cerebellum for some reason. The third cluster (502 voxels) was located in the left superior and middle temporal gyrus with three peaks in the middle and superior and again middle temporal gyrus. The next cluster included 280 voxels and was located in parietal and occipital and temporal areas. The peak was located in the inferior temporal gyrus. Finally, the fifth cluster was located in the cingulate gyrus in both the right and left hemisphere and had two peaks in the right posterior cingulate gyrus and the left middle cingulate gyrus. The summary of the linguistic conjunction results can be seen in table 5.4.

I also calculated the overlap of activity between the contrasts written words > implicit baseline and pictures > implicit baseline. I observed overlapping activity in 5 clusters: The first and largest cluster (7395 voxels) spanned a variety of regions in both hemispheres, the left hippocampus, the middle occipital gyrus, the middle temporal gyrus, the left postcentral gyrus, the right fusiform gyrus and the left inferior temporal gyurs.

The three peaks were all located in occipital regions, in the right inferior occipital gyrus, the left inferior occipital gyrus and the middle occipital gyrus. Since this cluster was also rather large, I looked at the 500 most active voxels of this cluster. These lay in two sets of regions, teh R fusiform gyrus, right lingual gyrus, right superior occipital gyrus, right inferior occipital gyrus and in the left hemisphere in similar regions, in the left fusiform gyrus, the inferior occipital gyrus, the left inferior temporal gyrus and the lingual gyrus. The next largest cluster (2165 voxels) was observed frontal areas as well as some basal ganglia. It contained 2165 voxels and included the right precentral gyrus, the right inferior frontal gyrus (pars opercularis, triangularis and orbitalis), the right superior frontal gyrus, the pallidum and the amygdala. The 500 most active voxels were also investigated for this cluster and the areas spanned by these 500 voxels also lie in frontal and basal ganglia areas: the RIFG (pars opercularis and triangularis), the middle frontal gyrus, the precentral gyrus and the putamen and pallidum. The three peaks were observed in the middle frontal gyrus, the right insula and the right precentral gyrus. The third largest cluster contained 1829 voxels, and was located in the left frontal gyrus and also had voxels in the putamen. It spanned the regions of the LIFG (pars opercularis, triangularis and orbitalis), the putamen and the left postcentral gyrus. The three peaks were observed in the left precentral gyrus and the LIFG, pars triangularis. The 500 most active voxels were rather spread out, but activity from these were also located in right frontal areas, in the right middle frontal gyrus, the right precentral gyrus, the right pars triangularis and the right insula. The next cluster was also observed in frontal parts of the brain. The cluster, containing 964 voxels, spanned the right supplementary motor area, the right and left medial superior frontal gyrus, the right middle cingulate gyrus, the left anterior cingulate gyrus and the left superior frontal gyrus. The three peaks were observed in the left SMA, the right SMA and the right middle cingulate gyrus. The smallest cluster was observed in the cingulate gyrus with 121 voxels spanning the left middle cingulate gyrus, the left posterior cingulate gyrus, the right middle cingulate gyrus and the right posterior cingulate gyrus. The one peak was also observed in the right posterior cingulate gyrus. The results of the conjunction of the 2 visual contrasts can be seen in table 5.5.

Finally, the overlapping activity between the two maximally different modalities (spoken words (auditory and linguistic) and pictures (visual and nonlinguistic) was calculated. This conjunction elicited activity in 7 clusters. The largest contained 4164 voxels and was located in the cerebellum and the frontal cortex. The peaks were all located in the cerebellum. I looked at the 500 most active voxels of this cluster and these 500

Table 5.5: Conjunction of pictures and written words : all visual stimuli

Region	Cluster Size	Cluster pFWE corr	MNI coordinates			Voxel pFWE corr	Voxel T48
			x	y	z		
L Fusiform, L inf occipital, LITG, L calcarine S, L lingual G, LMTG, L mid occipital G	334	0.000					
L inf occipital G			-21	-97	-7	0.000	8.52
L inf occipital G			-30	-91	-10	0.000	7.59
LITG			-42	-64	-10	0.010	5.48
R Fusiform, R Lingual G, R inf occipital G, R mid occipital G, R calcarine S	143	0.018					
R inf occipital G			33	-88	-4	0.000	6.6

voxels were split into two clusters, one in the frontal lobe and one in the cerebellum. The first subcluster of these 500 voxels was located in the R IFG (pars opercularis, orbitalis and triangularis), the precentral gyrus, the putamen and the superior frontal gyrus with the three peaks located in the right insula, the R IFG (pars opercularis) and the right middle frontal gyrus. The second cluster was located in the cerebellum with all three peaks in the cerebellum. The second largest cluster (1237 voxels) was observed in left frontal areas, in the left superior frontal gyrus, the LIFG (pars orbitalis, opercularis and triangularis), the left precentral gyrus and the amygdala. The three peaks were located in the left precental gyrus, the left insula, the pars triangularis. The 500 most activated voxels were located in similar areas. The second largest cluster had 1050 voxels and was located in left and right frontal areas, the right SMA, the left medial superior frontal gyrus, the right medial superior gyrus, the right middle cingulate gyrus and the left superior gyrus. The three peaks were located in the left SMA, the right SMA and the right middle cingulate gyrus. The next cluster contained 551 voxels and was located in temporal and parietal areas in the right hemisphere, namely the right middle temporal gyrus, the right inferior temporal gyrus, the right superior parietal gyrus, the right supramarginal gyrus and the right superior temporal gyrus and the right angular gyrus. The three peaks were located in the right inferior parietal, the right superior temporal and the right middle temporal gyrus. The next cluster was observed in left temporal regions, in the left middle temporal and superior temporal gyrus. The three peaks were observed in the left middle temporal and one in the left superior temporal gyrus. The results are summed up in table 5.6.

5.3.4 Results of conjunction of all three modalities

The analysis revealed overlapping activity in 7 clusters in frontal, temporal, parietal and to a smaller extend occipital brain regions. The first cluster was located in the

Table 5.6: Conjunction of spoken words and pictures

Region	Cluster Size	Cluster pFWE-corr	MNI coordinates			Voxel pFWE-corr	Voxel T48
			x	y	z		
R precental, RIFG (BA44,45, 47), R sup frontal G, L sup occipital G, R pallidum, R post cingulate G, L cuneus,	4164	0.000					
Cerebellum			33	-61	-28	0.000	7.35
Cerebellum			-33	-61	-28	0.000	7.18
Cerebellum			-30	-64	-49	0.000	7.14
L sup frontal G, LIFG (BA47, 45, 44), L precental, L amygdala, L olfactory B, L sup temporal pole, L mid frontal G, L putamen, L insula	1237	0.000					
L precental G			-42	2	53	0.000	8.54
L insula			-27	26	2	0.001	6.35
L IFG (BA45)			-51	17	29	0.002	6.05
L/R SMA, L/R sup med frontal G, L/R ant- mid cingulate, L/R sup frontal G,	1050	0.000					
L SMA			-6	14	50	0.000	8.24
R SMA			6	17	53	0.000	8.21
R mid cingulate G			12	23	32	0.049	4.96
RMTG, RITG, RSTG R sup & inf parietal G, R supramarginal G, R angular G,	551	0.000					
R IPG			36	-55	44	0.000	6.64
R STG			51	-43	17	0.012	5.43
R MTG			54	-52	8	0.012	5.42
L MTG, L STG	328	0.000					
L MTG			-51	-58	8	0.000	7.07
L MTG			-39	-64	17	0.413	4.14
L STG			-63	-40	11	0.447	4.1
L mid frontal G, L sup frontal G	151	0.014					
L sup frontal G			-30	53	-1	0.061	4.89
L mid frontal G			-39	50	23	0.264	4.34
L mid frontal G			-30	53	29	0.492	4.05
L/R mid-post cingulate G	123	0.030					
R mid cingulate G			3	-31	29	0.003	5.81
R sup occipital G, R precuneus, R sup parietal G, R cuneus	118	0.034					
R cuneus			18	-67	38	0.038	5.05
R cuneus			21	-70	29	0.149	4.57
R sup parietal G.			15	-73	50	0.397	4.16

right frontal lobe and basal ganglia regions. It contained 1548 voxels and lay in the R IFG, the right precentral gyrus, the right superior frontal gyrus, the right frontal superior temporal pole and the right rolandic operculum. It also included regions lying in subcortical structures, namely the putamen and the amygdala. One peak was also located in the right insula. The 500 most active voxels of these cluster were also located in frontal areas, in the RIFG, the middle frontal gyrus and the putamen and pallidum in the basal ganglia. The second cluster that showed overalaping activity for all three modalitites included 1161 voxels and was located in the left hemisphere. It included the regions of the L IFG, the left precentral gyrus, the left superior frontal gyrus and middle frontal pole and the left insula. It also ranged into structures belonging to the basal ganglia such as the left putamen. The 500 most active voxels were also located in left frontal areas, the LIFG, the middle frontal gyrus, the left precentral gyrus and the superior frontal gyrus. The third largest cluster was observed in the Cerebellum with all peaks in the cerebellum. The next cluster was observed in the left occipital and parietal areas, in the left angular gyrus, the left inferior parietal gyrus and the left middle occipital gyrus. The one peak was located in the right inferior parietal gyrus. A next cluster was observed in the left middle frontal gyrus, with three peaks, two of which were also located in the left middle frontal gyrus and one in the superior frontal gyrus. The next cluster was observed in the left and right cingulate gyrus. This cluster included 123 voxels and spanned the left middle cingulate gyrus, the left posterior cingulate gyrus, the right posterior cingulate gyrus and the right middle cingulate gyrus. The peak was observed in the right middle cingulate gyrus. The smallest cluster was located in the right superior occpital gyrus, right precuneus and right superior parietal gyrus and included 118 voxels. The peak was located in the right cuneus. The third cluster lay in regions in both hemispheres and included 945 voxels. It lay in the left and right SMA, the left and right superior medial frontal gyrus, the left and right middle and anterior cingulate gyrus and the left superior frontal gyrus. A summary of the results of the conjunction of all three modalities can be seen in table 5.7 and in figure 5.3.4.

The fourth cluster included 508 voxels, lying in the right hemisphere, namely in the right middle and superior temporal gyrus, the right superior and inferior parietal gyrus and the right angular and supramarginal gyrus. The fifth cluster, including 305 voxels covered the left middle and superior temporal gyrus. The sixths cluster was observed in the left parietal and occipital lobe, covering the angular, superior and inferior parietal and middle occipital gyrus. The seventh and last cluster was also observed in the left

Table 5.7: Conjunction of spoken words and written words and pictures

Region	Cluster Size	Cluster pFWE-corr	MNI-coordinates			Voxel pFWE-corr	Voxel T48
			x	y	z		
R precentral G, R IFG (BA44, 45, 47), R pallidum, R sup frontal G, R sup temporal pole, R amygdala, R rolandic operculum							
R insula			27	26	2	0.000	6.58
R mid frontal G			42	-1	56	0.001	6.19
R IFG (BA45)			36	32	23	0.003	5.87
L sup frontal G, LIFG (BA44, 45, 47), L putamen, L precentral G, L sup temporal pole, L mid frontal G, L insula	1161	0.000					
L precentral G			-42	2	53	0.000	8.54
L precentral G			-36	2	41	0.001	6.15
L IFG (BA45)			-51	17	29	0.002	6.05
Cerebellum	1103	0.000					
Cerebellum			33	-64	-28	0.000	7.32
Cerebellum			27	-67	-46	0.000	6.87
Cerebellum			-33	-61	-28	0.001	6.23
L/R SMA, L/R sup medial frontal G, L/R mid cingulate G, L/R ant cingulate G, L sup frontal G	945	0.000					
L SMA			-6	14	50	0.000	6.66
R SMA			6	17	56	0.000	6.55
R mid cingulate G			12	23	32	0.049	4.96
R MTG, R supramarginal G, R sup parietal G, R STG, R angular G, R inf parietal G	508	0.000					
R inf parietal G			36	-55	44	0.000	5.69
R STG			51	-40	17	0.037	4.58
R STG			66	-40	17	0.055	4.33
L MTG, L STG	305	0.000					
L MTG			-51	-58	8	0.000	7.07
L STG			-63	-40	11	0.447	4.1
L MTG			-45	-46	-4	0.590	3.94
L angular G, L inf parietal G, L mid occipital G, L sup parietal G	215	0.003					
L inf parietal G			-33	-55	47	0.000	6.7
L mid frontal G, L sup frontal G	145	0.017					
L sup frontal G			-30	53	-1	0.061	4.89
L mid frontal G			-39	50	23	0.266	4.34
L mid frontal G			-30	53	29	0.692	3.47
L/R mid-post cingulate G	117	0.035					
R post cingulate G			3	-34	26	0.019	5.28

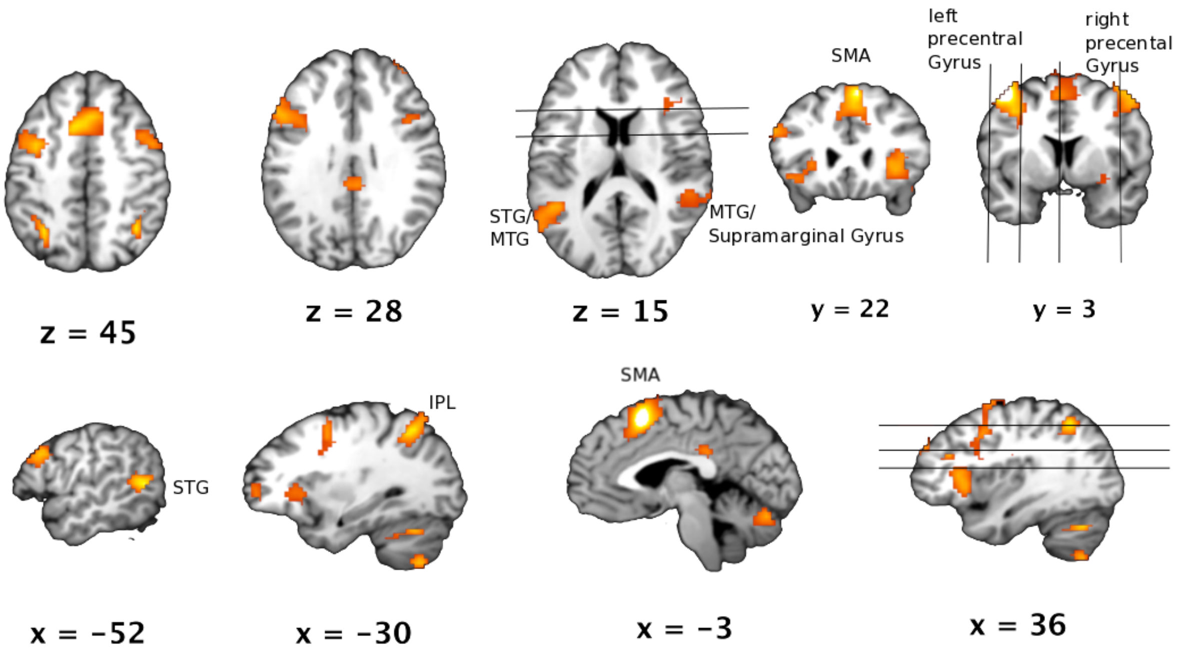


Figure 5.1: Overlap of Activity for the three modalities used here: spoken words, written words and pictures. The conjunction analysis for the three modalities revealed an overlap of activity in the left and right precentral gyrus, the SMA, the left superior and middle temporal gyrus and the right middle temporal and supramarginal gyrus,

hemisphere and covered the middle and superior frontal gyrus. Finally, an eighth cluster was observed in the left and right middle and posterior cingulate gyrus

5.4 Discussion

The present fMRI study undertook the first direct, within-participants comparison of spoken words, written words and pictures using a non-semantic task. Participants were asked to perceive concepts presented in the three modalities while in some cases performing a perceptual task unrelated to conceptual processing. I observed activity in various brain regions for each modality when compared to the other two modalities. I also observed activity overlapping for pairs of modalities in vast networks of brain regions and finally were also able to view activity for three modalities in a network of brain regions. I will discuss the implications of the activity observed here, going from the single contrasts, to pairwise conjunctions to the conjunction of all three modalities. The central result is the overlapping activity of all three modalities. Clusters observed here are candidate regions for amodal processing. The network of regions that responded to all three modalities comprised premotor cortex, the paracingulate gyrus, the pre-SMA, bilateral posterior superior and middle temporal gyri, the bilateral lateral occipital gyrus/superior parietal lobule and the bilateral middle and inferior frontal gyri. These regions were identified as part of a candidate network where semantic knowledge independent of modality might converge. In the following, I discuss these regions in turn, beginning with temporal cortex and moving on to prefrontal and occipital areas. After discussing the main result - the overlap of activities for all three modalities - I will discuss the implications of the results from the conjunctions of all three pairs of modalities and in the single contrasts.

5.4.1 Conjunction of all three modalities

5.4.2 Posterior bilateral middle and superior temporal cortex and inferior parietal cortex

I observed cross-modally overlapping activity in bilateral posterior middle and superior temporal regions. These findings are consistent with the results of previous studies that examined related contrasts, though with a more restricted stimulus set. For example, Costanzo and colleagues compared spoken words and visually presented objects and observed an overlap of activity for these two stimulus types in middle temporal/parietal and middle frontal regions (Costanzo et al., 2013). However, the overlap in activity was more left-lateralized in their study in comparison to the present results. Two further recent studies have implicated the pMTG and pSTG in the representation of amodal conceptual knowledge via multivariate pattern analysis. Man et al. (2012) compared pictures of objects and the sounds they typically make (e.g. a video of a bell and a sound of a bell) and were able to crossmodally decode objects in the right posterior superior temporal region, as well as in inferior frontal and premotor cortex. Converging results were obtained by Fairhall et al. (2011), who used representational similarity analysis to contrast two visual input modalities (written words and pictures) and observed cross-modal overlap in the left pMTG/pSTG. Along with these previous findings, the present results are consistent with the assumption that posterior temporal regions constitute high-level convergence zones which perform supramodal semantic processing (Binder et al., 2009; Binder and Desai, 2011). Specifically, they are thought to represent abstract conceptual knowledge and to bind representations from multiple modalities. This contrasts with the assumed role of ventrolateral frontal areas, which are typically associated with the retrieval of this stored information (Martin, 2007; Binder and Desai, 2011). We indeed observed overlapping ventrolateral frontal activation for all of our stimulus types, namely in the bilateral inferior frontal gyri. This is particularly interesting, since our paradigm did not include any prompt to explicitly process semantic information but instead only required participants to detect infrequent, deviant events in a continuous stream of multimodal stimuli.

5.4.3 Bilateral lateral occipital cortex

Overlap of activity for the three stimulus presentations modes used here was also observed in the lateral occipital cortex, which has been mainly associated with object recognition (Grill-Spector and Weiner, 2014). It has also been implicated in the multi-sensory integration of visual and tactile, but not auditory object information (Beauchamp, 2005; Amedi et al., 2002). This is in contrast with our results, where LOC activation for aurally presented stimuli overlapped with that for visually presented words and pictures. It has been suggested that LOC activation for tactile stimuli is a result of mental imagery (Beauchamp, 2005). It is possible that the LOC activity observed in our study is also at least partially the result of mental imagery, since stimulus presentation was completely randomized for the three modalities and participants encountered pictures very frequently during the experiment. For tactile stimuli it has also been shown that LOC can be activated without imagery being present (Pietrini et al., 2004), but it remains an open question whether this is also the case for auditory stimuli.

5.4.4 Superior and medial prefrontal regions

The largest cluster observed in the present study was found in the superior frontal cortex, extending to medial prefrontal regions. We assume that the medial prefrontal cluster lies in the pre-SMA, as it is located dorsally within the frontomedial wall and lies anterior to the anterior commisure line (see the anatomical definition by Picard and Strick, 2001). This cluster is thus posterior to the dorsomedial prefrontal region (DMPFC) described by Binder and colleagues (Binder et al., 2009; Binder and Desai, 2011) as having been observed in a number of studies on supra-modal conceptual processing. According to Binder and colleagues, DMPFC plays an important role in the self-initiated retrieval of semantic information, as well as being linked to systems for the processing of affective and reward information.

In a recent review of supplementary motor area involvement in language processing, (Hertrich et al., 2016) summarize the functional role of the pre-SMA as cognitive control managing complex sequencing, ambiguity resolution and task switching. As most of the studies showing a correlation between pre-SMA activation and semantic processing that were reviewed by Hertrich et al. (2016) either required language production or an overt semantic task, they are not directly comparable to the present study. However, a semantic priming experiment by Ulrich et al. (2013) revealed priming effects in the pre-SMA for both masked and unmasked priming conditions, thus suggesting a potential involvement of the pre-SMA in preconscious semantic processing. Nevertheless, Ulrich et al. (2013) study also employed a lexical decision task, which may have played a role in engendering this activation.

An alternative possible interpretation of the pre-SMA activation is suggested by the observation that, even though activation in this cluster overlapped across all stimulus modalities, examination of the percent signal change per stimulus type suggests that it was particularly pronounced for the spoken words (see figure 5.4.4). Interestingly, the pre-SMA is a key area in a network for the cognitive control of primate vocalisations, including spoken language production in humans (for a review, see Loh et al., 2016). In contrast to the SMA, which is more directly linked to motor control, the pre-SMA appears to be linked particularly to the voluntary production of speech. In the context of the

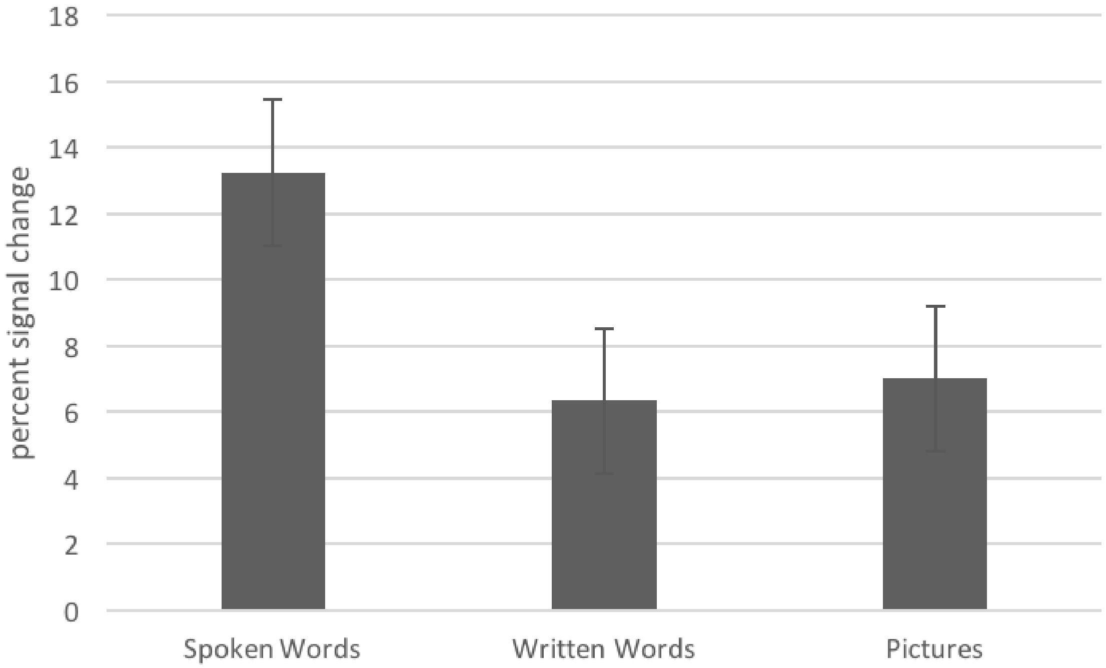


Figure 5.2: Percent signal change in the SMA for the three modalities. Percent signal change was highest for the spoken word modality and lower for the two other modalities with written words displaying the lowest percent signal change in the SMA region.

present study, the pre-SMA activation for spoken words could thus be viewed as reflecting mirror-type activity for spoken word perception. Speculatively, the activation overlap in this region between all three modalities might be taken to indicate that participants had the tendency to internally vocalize the written word and picture stimuli. As argued above for visual imagery in the presence of picture stimuli, the frequent encounter of spoken words may have encouraged this process. In addition to the pre-SMA, the vocalization network involves the SMA proper and mid-cingulate cortex.

5.4.5 Pairwise conjunctions

To test for overlapping activity for the two language conditions (Spoken Words and Written Words), I carried out a conjunction analysis for these the two contrasts Spoken Words > baseline and Written Words > baseline. I observed overlapping activity in 5 clusters: One located in the middle frontal gyrus, the SMA and the precentral gyrus; one in Cerebellum, one in the left superior and middle temporal gyrus, one in the left angular gyrus, left inferior parietal gyrus and left middle occipital gyrus and the final smallest cluster in the left and right cingulate gyrus.

The pattern of overlapping activity observed here is at least in part in line with previous results about where spoken and written language processing convergences in the cortex. There are two areas that are associated with the processing of language independent of the modality it is expressed in (visually in reading written language and auditorily in hearing spoken language), Broca's area or more broadly the left inferior frontal gyrus (LIFG) and the left posterior temporal gyrus and sulcus (STG and STS) (Regev et al., 2013; Spitsyna et al., 2006). This is true both for words (Booth et al., 2002a; Marinković, 2004) and sentences (Spitsyna et al., 2006; Lindenberg and Scheef, 2007). We do observe overlapping activity for both Written Words and Spoken words in the left middle and superior temporal gyrus which is in line with these previous results. We do not observe overlapping activity for the two linguistic conditions in a cluster in the LIFG (even though one cluster does extend to the LIFG and the RIFG without the focus of the activity lying there). This could be due to the fact that this region is associated with syntactic processing. In the present study, only single words that did not require syntactic processing were used.

In addition to activity in areas that have been associated with language processing for a long time such as the left temporal lobe, I also observed activity in areas that are not part of the so called "classical language network" (Tremblay and Dick, 2016). The activity I observed in the left and right SMA and the medial frontal lobes as well as in the cingulate gyrus could be interpreted as being part of the Multiple Demand system in the Human brain (Duncan, 2010; Fedorenko et al., 2013) reflecting a general increase in cognitive load. Activity in the left and right SMA and in the cingulate cortex could alternatively also be interpreted as reflecting internal vocalization. This interpretation is supported by the fact that the activity was particularly high for spoken words which could have lead to internal vocalization of the words participants just heard.

I also tested which activity overlapped for written words and pictures - the two visual conditions in our experiment. I observed 5 clusters for this conjunction analysis. The first cluster was located in left and right temporal and occipital areas with the focus of the activation lying in the fusiform gyrus. The second cluster on the other hand was located in frontal and basal ganglia areas. The second largest cluster covered a similar area in the right hemisphere. There also were two clusters located in the medial frontal lobe and the cingulate gyrus. The overlap between language and visual perception has been studied frequently using written words and pictures. A PET study by Vandenberghe and colleagues (1996) observed a network of regions active for both pictures and words in the left superior occipital gyrus, the middle and inferior temporal cortex as well as the IFG (Vandenberghe et al., 1996). This network of regions has also been observed in studies following Vandenberghe and colleagues (Bright et al., 2004; Krieger-Redwood et al., 2015) that used conjunction analysis, but also in studies using MVPA methods as for example (Fernandino et al., 2016; Shinkareva et al., 2011).

Apart from regions that have been identified as part of a general semantic network (Fernandino et al., 2016), I also observed activation in medial frontal regions, for example in the cingulate gyrus and the left and right SMA. It is possible that this activity too can be attributed to silent vocalization. However, activity in the SMA was not as high for written words compared to spoken words. Alternatively, this activity could again reflect the involvement of the multiple demand system in our task. The overlap between semantic processing for words and pictures and for sound and visual perception has long been studied. In 1996 Vandenberghe and colleagues used PET to look at activity that was common for cross-modal semantic tasks performed on written word and picture stimuli. They observed activity in the superior occipital cortex, the middle and inferior temporal gyrus and the IFG (Vandenberghe et al., 1996). They used two stimuli sets that were to be visually perceived though. The maximally great contrast between language and pictures is between spoken language and pictures. Sounds and pictures provide a similar test case for maximally different modalities: Beauchamp and colleagues identified the posterior superior temporal sulcus (pSTS) and posterior middle temporal gyrus (pMTG) as a site for integration of visual and auditory information about animals and man-made objects (Beauchamp, 2005) based on studies using pictures and sound recordings of animals and objects such as hairdriers. In a study by Visser and Lambon Ralph (2011) a conjunction analysis of environmental sound, auditory speech and picture semantic processing lead to increased activation in the inferior temporal gyrus and the cerebellum, while a contrast using all three modalities used in this study lead to increased activation in the left and right superior temporal gyrus, the left and right inferior temporal gyrus, the left and right inferior frontal gyrus, the right anterior cingulate gyrus, the left occipital lobe and the left putamen.

Finally, in the study that is perhaps closest to our study, Costanzo et al. (2013) compared spoken word and picture processing using fMRI and EEG and observed a left-lateralized network of regions where semantic processing between pictures and spoken words overlapped. The network included the left fusiform and middle temporal gyrus, the angular and supramarginal gyri, the intraparietal sulcus and the superior parietal lobule.

The network of regions I observed, was more extended than in any of the previous studies but did include many areas that have already been observed as integration sites for visual picture and auditory spoken word processing. I observed overlapping activity in the left and right frontal lobes with a focus on the RIFG and LIFG and the left precentral gyrus, in the left and right cerebellum as well as in temporal and parietal areas. However, the activity I observed was not left lateralized and was much more spread out than in previous studies. We suspect that this has to do with the task used in previous studies. Costanzo et al. (2013) for example used a categorization task in which participants had to indicate whether or not a stimulus represented a real existing object and if so, whether it was an animate or inanimate stimulus. I used no task related to semantic processing in order to tap into non-explicit semantic activity that occurs during object recognition. Task does seem to play a role in how semantic activity is spread (see Ralph et al., 2016, for a review of how semantic control is manipulated by task demands) but to our knowledge there is only one study that has explicitly tried to manipulate task demands during semantic processing across formats (written words and pictures) (Van Doren et al., 2010).

5.4.6 Single Modalities

In order to understand activity related to the processing of only one modality and format, I looked at the voxels active for the combination of one modality and format when directly compared with the activity for the other two combinations when compared to the baseline. For the contrast Spoken Words > Written Words and Pictures, I observed three clusters, one located in the left temporal lobe and temporal pole, one located in the right temporal lobe and one located in the right and left cerebellum. The two largest clusters are located in the bilateral temporal lobes, temporo-parietal areas and the LIFG and RIFG. The superior temporal cortex and the temporal plane are thought to be involved in the processing auditory signals (Hickok and Poeppel, 2007), (Rauschecker and Scott, 2009), (Bornkessel-Schlesewsky, 2015). With increasing complexity of a speech stimulus, the superior and middle temporal gyri become involved in processing the pieces of language presented (Davis and Johnsrude, 2003), but see also chapter 1.1.2 and figure 1.2. The activity of both the left and the right hemisphere cluster in the temporal region was concentrated on the two temporal pole areas, the rolandic operculum, the insula and the Heschl's gyrus. These are involved in auditory processing for both speech and non-speech sounds. The activity here can best be interpreted as reflecting auditory processing of speech sounds and immediately following steps of processing speech sounds such as phonological analysis that is not included in the processing of written words or pictures.

For the contrast Pictures > Spoken Words and Written Words, I observed 3 clusters in occipital, parietal and temporal regions. Activity in these areas has been associated with the perception of pictures, specifically with pictures of objects (Grill-Spector, 2003). Similarly to the activity observed in the auditory spoken word condition, the activity observed for the pictures is most probably related to visual processing as well as object-picture specific processing.

There was no activity for Written Words > Spoken Words and Pictures. This is unsurprising given that written words occupy a middle position between language perceived auditorily and objects viewed visually.

5.4.7 Activity in the Cerebellum

We observe Clusters in the Cerebellum in 4 of our analyses: Spoken words > Written Words and Pictures, the conjunction of Spoken Words and Written Words, the Conjunction of Spoken Words and Pictures and the Conjunction of all three modalities. The presence of the Cerebellum in this contrast and the three Conjunction analyses was somewhat surprising to us, since the Cerebellum is neither part of the classical language network (Tremblay and Dick, 2016) nor traditionally considered part of the network of brain regions that process semantic information nor has it been implicated as a convergence zone for two or more modalities. The Cerebellum has long been thought to be only involved in motor processing, but especially due to the work of Leiner et al. (1993), the involvement of the Cerebellum in cognitive and emotional and social processes has been studied. Our results seem to suggest that the activity in the Cerebellum that was observed is primarily influenced by the Spoken Language > Baseline contrast. It has been observed that involvement of the Cerebellum in language is right lateralized (lateralized linguistic cerebellum (Mariën et al., 2013). We considered it therefore possible, that the activity in the Cerebellum which covered both hemispheres of the Cerebellum, was right-lateralized for the language contrasts and either left-lateralized or evenly distributed for the processing of pictures. Cognitive activity in the cerebellum has been linked to temporal aspects of stimuli (for language for example sequential processing of syntax, rhythm or acoustic pitch (Argyropoulos and Muggleton, 2012), predictions and the storage of forward models (Argyropoulos, 2016; Lesage et al., 2017) or the general regulation of accuracy, speed and consistency (Guell et al., 2014). It has been observed that depending on whether the cerebellum is involved in motor, cognitive, social or emotional processing, different parts of the cerebellum are involved that connect to different parts of the neocortex. In our data, larger parts of the cerebellum seem to be involved which makes it hard to trace back cerebellum activity to one particular aspect of our experiment and signal is strong for at least two of the three modalities (with signal being somewhat less strong for written words in general) which makes it hard to trace back cerebellum activity to one modality. The cerebellum also has not been linked to passive perception of single words or single objects. Instead it has been more frequently linked to the processing of sequential stimuli such as linguistic sentences or visual sequences or language production such as writing, but its involvement with many cognitive tasks has become a focus of study in more recent time. I tentatively conclude that the cerebellum might have a role to play in perception of single items.

5.4.8 Limitations

One caveat pertaining to the interpretation of the present results is that pure overlap of activity as indicated by the conjunction analysis does not allow us to analyze the underlying information content of the regions activated. Further analyses using multivariate techniques, e.g. representational similarity analysis (RSA), could help to clarify the (dis)similarity between the information processed for the different modalities, and thereby provide further insights into the role of the regions observed here in the processing of conceptual information. Furthermore, the fMRI protocol used here was not optimized to detect activity in the anterior temporal lobe, a region that is assumed to play an important role in amodal semantic processing in some theories, possibly in the form of an amodal conceptual hub (e.g. Patterson et al., 2007; Ralph et al., 2016 but see Binder and Desai, 2011, for a different view). In comparison to other regions, the ATL suffers from susceptibility-induced signal dropout. Thus, standard fMRI doesn't always reliably capture signal from this region (see Visser et al., 2010b, for a discussion and meta-analysis) for discussion and a meta-analysis). As the present study used a standard EPI sequence for functional imaging, the absence of ATL activity in our results cannot be viewed as evidence against the importance of the ATL for amodal conceptual processing.

5.5 Summary and Conclusion

Our results are consistent with a network of convergence zones in which semantic knowledge is processed independently of modality (Binder, 2016), rather than with the existence of a single semantic hub as suggested by (Patterson et al., 2007; Pulvermüller and Fadiga, 2010). The novel contribution of this experiment lies in the direct, within-participants comparison of spoken words, written words and pictures, and the use of a perceptual task. Its results thus indicate that conceptual processing involves a distributed network of supramodal convergence zones even in the absence of an explicit semantic task.

CHAPTER 6

Second fMRI - analysis: Categorical processing

6.1 Introduction

The first fMRI analysis I carried out tested which regions were involved in amodal conceptual processing for the two modalities (visual and auditory) and the two formats (lexical and pictorial). The stimulus set not only features different modalities and formats, but also different categories of objects, namely three animate categories: humans, mammals and insects/crustacea and three inanimate categories: plants, tools and furniture. Previous research has studied how the animate-inanimate distinction is represented in the brain for different modalities and formats. It has been observed that separate brain regions, especially in temporal areas, seem to represent animate and inanimate objects in the cortex of monkeys (Bell et al., 2009). The areas in which animates and inanimates are represented in humans can be further subdivided into regions that represent specific categories such as faces, body parts or place-selective regions (Bell et al., 2009; Kanwisher et al., 1997; Epstein and Kanwisher, 1998).

Categorical effects for animate and inanimate objects (often represented as animals and tools) in ventral temporal regions have been observed in studies using picture stimuli

(Martin et al., 1996; Chao et al., 1999; Ishai et al., 2000; Martin and Chao, 2001). Early studies by Grabowski et al. (1998) and Damasio et al. (1996) focused on lexical processing using pictures of famous people, animals and tools that participants had to name while PET or fMRI data was acquired. They observed activity in distinct areas for naming the objects from these different categories: the middle frontal gyrus responded most to naming persons, the precentral gyrus responded most to naming tools, whereas the LIFG was active for all naming acts.

Written words have been used to study how the membership in the category of animacy influences functional organization in the brain. An early study, using PET, looked at written words using different tasks focusing on either color of the object, the location the object is typically found in, or number of syllables of the words. The authors observed that artifacts led to more activation in the left posterior middle temporal gyrus and the parahippocampal gyrus, whereas the reverse contrast (natural living items > artifacts) did not lead to any significant activation (Mummery et al., 1998). In other studies, (i.e. Perani et al., 1999) differential effects for living and nonliving things were observed, but other studies (Devlin et al., 2001; Price and Devlin, 2003) did only observe common activation for animate and inanimate objects. Noppeney et al. (2005) directly compared activity when seeing pictures and reading words that showed animals and tools using different tasks, an explicit semantic task. Participants had to judge whether the items performed the same action or were the same size) and an implicit semantic task (an n-back task asking whether the item was the same tool or animal). They observed category specific effects for picture stimuli in the ventral stream independent of task (Noppeney et al., 2005), but category specific effects for both picture and word modality in dorsal areas that were greater for the explicit semantic task.

Using fMRI, Rundle and colleagues 2018 also observed that animacy modulated activation in ventral temporal areas using both tool and animal written words and a task that requested that participant thought about the items they saw using specific questions, i.e. ”Was the last item sharp?” (Rundle et al., 2018).

As is the case for pictures, some studies cast doubt on whether conceptual organization follows the animate-inanimate distinction, or, at least, they cast doubt on how the stability with which this pattern can be observed: Marques et al. (2008) studied written sentences with animate and inanimate nouns and did not observe effects of category (which they call domain) but only effects of specific features of their stimuli. Taken together, the different approaches that have been used studying both spoken and written words and pictures, suggest that category specific effects exist, but in both pictures and written words, some studies do not observe this functional specialization. While there is support for some category specific regions (for example faces and places), for other categories it is less clear where in the brain they might be represented.

Functional specialization patterns for animate and inanimate items have also been studied using spoken words, but much less frequently. In a study using blind and sighted individuals, Mahon et al. (2009) directly compared the functional organization of animal and tool pictures and spoken words and observed that the medial-lateral distinction for these two categories was similar in sighted participants who observed pictures and in sighted and blind participants who listened to spoken words while performing a size judgment task. In both blind and sighted participants the response to living stimuli was larger in lateral ventral temporal cortex and to nonliving in medial ventral temporal cortex. Using intracranial EEG, Chan and colleagues also observed category-specific effects for spoken and written words in the anterior ventral cortex across different measures of synaptic and unit firing activity, independent of task (Chan et al., 2011). To my knowledge, these two studies are the only ones that tested category-specific functional specialization using spoken words, but they support an observation that has been made for other modality-format combinations, namely higher activity for tools in medial temporal areas vs. higher activity for animals in lateral temporal areas.

One particularity of studies investigating categorization is that animacy is often represented by tools and animals (i.e. Noppeney et al., 2005; Mechelli et al., 2006; Vandenberghe et al., 1996; Fairhall et al., 2011). Some studies also included a wider spectrum of categories, for example vegetables or vehicles (Bright et al., 2004) or musical instruments (Hocking and Price, 2009). Animacy, however, is complex and has been described as a hierarchy 1.3 instead of a binary distinction with humans being most animate, followed by mammals, other animals, plants, tools, manmade objects and natural objects.

Wiggett and colleagues carried out an fMRI study that used a diverse stimulus set including a wider spectrum of the animacy continuum. They tested four animate categories: mammals, birds, fish and reptiles/amphibians and four inanimate categories: tools, vehicles, instruments and furniture and observed lateral activation in the fusiform gyrus for their animate category and medial activity in response to the inanimate category (Wiggett et al., 2009). This result is in line with previous observations, that animates and inanimates seem to be represented in different parts of the temporal cortex. One open question about the functional organization of different categories in the ventral stream is whether there are distinct areas for specific categories, or whether there exists a gradient of categories of which the already identified areas are just a part (Op de

Beeck et al., 2008). Including a spectrum of categories makes it possible to address this question for the three combinations of modality and format included.

Finally, categorical perception across modalities has been studied using explicitly semantic tasks (Noppeney et al., 2005; Connolly et al., 2012; Mahon et al., 2007; Bracci et al., 2017). To my knowledge, there is only one fMRI study that (among other tasks) studied categories in the brain for written words and pictures using a pure observation task (Chao et al., 1999). My study refrained from using a categorization task opting instead to instruct participants to merely perceive the stimuli attentively. This allows me to investigate categorical perception in absence of further semantic processes. This study expands upon previously carried out research that either used a limited part of the animacy continuum and/or compared only two modalities or formats while also focussing on the part of categorization that occurs during perception without further instruction directing the participants towards the semantic content of the stimuli.

Functional specialization patterns for several categories that seem to generalize across different modalities have been identified, especially the lateral-medial organization of animate and inanimate categories in the fusiform gyrus. Tools have also been associated with action-related areas in the dorsal stream, for example the posterior superior temporal gyrus. I hypothesize that I would observe the lateral-medial functional specialization for animate and inanimate stimuli for all three types of stimuli. Since this study used categories with different positions on the animacy continuum, it is at least possible to test whether there exists a gradient of animacy in ventral temporal regions across different modalities and formats.

6.2 Materials and Methods: Analysis

One separate predictor for each category (humans, mammals, insects, plants, tools, furniture) for each modality (spoken words, written words and pictures) was modeled resulting in 18 explanatory variables (EVs) of interest. The spoken words were modeled as being 0.97s long each. We also included four EVs of no-interest to the GLM, modeling catch trials (one per input format) and nulltrials. Six motion parameters and scanner drift were included as covariates. I looked at different category contrasts for individual modality-format combinations as well as for all collapsed. I used combined contrasts in which human and mammal stimuli were used to represent animates and tools and furniture as a representation of inanimates. Insects and plants occupy a grey zone and thus were left out of both the animate and the inanimate larger category. For group-

Table 6.1: Animate pictures (mammals and humans) compared to inanimate pictures (tools and furniture)

Region	Cluster Size	Cluster	MNI coordinates			Voxel	Voxel T48
		p-FWE-corr				p-FWE-corr	
			x	y	z		
R Fusiform G, R inf occipital G, R sup occipital G, T mid temporal G, R lingual G, R calcarine G	660	0.000					
R lingual G			18	-88	-7	0.000	9.65
R mid occipital G			45	-76	2	0.002	5.42
R inf occipital G			45	-79	-7	0.009	5.07
L inf occipital G, L fusiform G, L inf temporal G	484	0.000					
L mid occipital G			-24	-97	-1	0.000	7.46
L mid occipital G			-48	-82	5	0.001	5.73
L inf temporal G			-39	-49	-16	0.131	4.11

level analyses, parameter estimates were assessed with a random-effects model. For the baseline contrasts, Z- statistic images were generated using a Z-statistics threshold of 4.44 and a FWE-corrected cluster probability threshold of P = 0.001. The clusters and peaks were labeled using the AAL atlas (Tzourio-Mazoyer et al., 2002). Only regions that covered more than 4% of the cluster are named and labels are reported in order of how much of the cluster they covered.

6.3 Results

Animacy is often represented by animals and tools, but both animates and inanimates contain a larger spectrum of concepts. I looked at contrasts that included more than one category in both object classes. Animates were represented by both human and mammal stimuli and inanimates by both tool and furniture stimuli. I tested the contrasts Animates > Inanimates and Inanimates > Animates for all three modality-format combinations and across modality-format combinations. Two contrasts led to increased activation. In the picture modality, the contrast Animates > Inanimates led to increased activation in bilateral occipitotemporal regions, including the inferior occipital gyrus and the fusiform gyrus. The results are displayed in table 6.1 and in figure 6.3.

The other contrast that led to significantly increased activation was Inanimates > Animates for spoken words. There was increased activation in the left middle and superior occipital gyrus (see 6.3 in panel b as well as in table 6.2). The other contrasts, including the contrast that looked at (in)animacy across all three modality-format combinations led to no significantly increased activation.

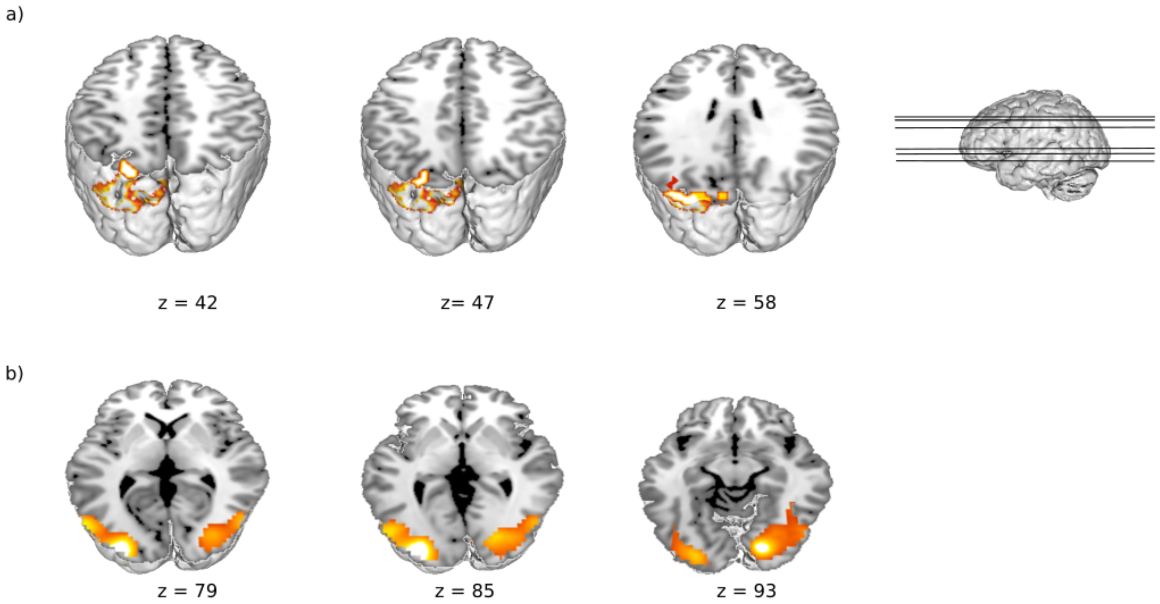


Figure 6.1: Results of the two contrasts that reached significance: Animate > Inanimate in the Picture modality (panel a) and Inanimates > Animates for the Spoken word stimuli panel b). a) The contrast Animates > Inanimates in the picture stimuli activated the bilateral fusiform gyrus and middle occipital gyrus. b) The contrast Inanimates > Animates activated the left superior and middle occipital gyrus.

Table 6.2: Inanimate spoken words (tools and furniture) compared to animate spoken words (Humans and Mammals). MNI coordinates and local maxima are reported.

Region	Cluster Size	Cluster	MNI coordinates			Voxel	Voxel T48
		p-FWE-corr				p-FWE-corr	
			x	y	z		
L mid occipital G, L sup occipital G	120	0.041					
L sup occipital G			-18	-94	29	0.133	4.31
L mid occipital G			-30	-91	29	0.477	3.84
L sup occipital G			-15	-88	41	0.563	3.76

6.4 Discussion

I investigated how (in)animacy is represented in the brain for different modality-format combinations: pictures, written words, and spoken words using 6 categories occupying different positions on the animacy continuum. I observed clusters in only two: spoken words and pictures. The written words did not lead to any increased activation. Because no contrast led to significant activation in two modality-format combination, I was not able to compare the functional specialization of the modalities or formats tested for in this study.

6.4.1 Animacy in the picture modality

The contrast Humans +Mammals > Tools + Furniture led to increased activation in the right fusiform gyrus and the left inferior occipital lobe. The reverse contrast however, did not lead to any increased activation. The cluster for animate objects in the right fusiform gyrus is in line with many previous studies that observed animacy related activity in the fusiform gyrus (Chao et al., 1999; Mahon et al., 2007). In most studies, activity was however not only observed for the contrast of Animates > Inanimates, but also for the reverse contrast in the picture domain. In a post-hoc analysis of this data, I specifically investigated activity in the fusiform gyrus using a fusiform gyrus ROI taken from the AAL atlas (Tzourio-Mazoyer et al., 2002). Looking specifically at the bilateral fusiform gyrus did not change the results, as there was no activity observed in this ROI for the contrast Inanimates >Animates for pictures either. Since there is no significant activation for Inanimate > Animate, there is nothing meaningful to say about the lateral - medial pattern for animate and inanimate objects that has been observed in the ventral pathway.

It is possible, that the lack of activation in the Inanimates > Animates contrast was due to a lack of signal. This is supported by an early study by (Chao et al., 1999) who carried out a similar experiment as I did using picture stimuli. They used different tasks to study category-specific effects including a passive viewing task. For the passive viewing task they used 1732 stimuli in total, 432 per category, which is a lot more than what I used. In the current study, each category was only presented 168 (7 items per category*2 for the combination of mammals and humans and tools and furniture*12 for each run) times. To further elucidate whether there were not enough stimulus repetitions to elicit significant activation for the Inanimate > Animate contrast, several approaches can be taken, namely using different tasks, more repetitions or a blocked design:

When using a more explicit semantic task, Chao et al. (1999) used fewer stimulus repetitions. This is a common approach. Choice of task can be a factor leading to the lack of activation for the Inanimate > Animate contrast. To name just a few examples, Bright et al. (2004) studied category-specific effects for pictures and words using animals, fruits, vegetables and tools while asking participants to categorize the stimuli they perceived into semantic classes . Gerlach et al. (1999) studied category-effects in pictures using object-decision tasks asking participants to press a button when the object they saw was natural or man-made or known or new and similarly in Gerlach et al. (2000). There are also studies that have addressed the question of task relevance for category-effects: Wiggett et al. (2009) studied animacy effects in the visual cortex using stimulus-related judgments on different levels. The three levels they used were (1) General: 'Are both stimuli mammals?' (2) Intermediate: 'Are both stimuli horses?' and 3) Stimulus: 'Is it the same stimulus?'. A similar approach was taken by Noppeney et al. (2005) who studied animacy effects for pictures and words asking participants to perform either an implicitly or explicitly semantic task. In the implicit semantic task, participants were asked to judge whether the two previous stimuli were identical; in the explicit semantic task, they were asked to judge whether they performed a similar action (could fly) or had a similar size. It seems that research that did observe category specific effects used tasks that drew the attention towards conceptual attributes of the stimulus set explicitly. The lack for the Inanimate > Animate contrasts could be due to either a lack of power or a lack of engagement due to the perceptual task. I suggest, that it might be a combination of both. An effect of inanimates has been observed in a viewing task, in that case, however, more stimulus repetitions were used (Chao et al., 1999). To my knowledge, in all other previous studies investigating categorical processing of animate and inanimate objects, a semantic task was used. It is a possibility, that category processing is engaged to a lesser degree during perception when active involvement of conceptual knowledge is not a task requirement. Using a perceptual task only might have led to participants only engaging with the stimuli in a superficial manner without accessing enough categorical information, at least in some instances. In several studies, the task also involves keeping categories of previously observed stimuli in memory (i.e. Amedi et al., 2002; Pietrini et al., 2004; Proklova et al., 2016). Therefore, the functional specialization pattern described so far is to some degree due to working memory processes.

To be able to compare neural localization patterns for animates and inanimates across modalities and formats, two approaches can be followed in future studies. One alternative is to include more stimulus repetitions. One issue with that might be that the experiment was already long enough to be carried out on two days. Because of that, future studies could use a semantic task that is not an n-back task, to ensure larger engagement of categorical processes during perception. As I have explained in 1.3.1, there is a variety of tasks that have been used. It is an open question, which tasks lead to most pronounced category effects. Varying the task used over a series of experiments could elucidate the relationship between task and category specific activity.

6.4.2 Category specific effects in auditory lexical stimuli

For the spoken words, one contrast led to increased activation: Inanimates > Animates. The other tests did not lead to significant results. The results make it impossible to compare category-specific functional specialization for pictures, written words and spoken words. However, the effect for inanimate stimuli in the spoken word modality an interesting result as category specific effects have not been studied as frequently in spoken words as in written words or pictures. Mahon and colleagues, while observing activity for both living and nonliving items, observed a larger cluster of activity for spoken words referring to nonliving artifacts in ventral medial temporal regions (Mahon et al., 2009) which is not present here.

It might be considered surprising that occipital areas are active in response to spoken words. I am not aware of any study reporting occipital activation for semantic processing of single nouns when either spoken language was studied alone or in combination with another modality. However, similar coordinates have been reported in processing single spoken nouns in studies investigating common neural correlates for different languages: Correia and colleagues studied single spoken words denoting animals across two languages (Dutch and English) and were able to crosslinguistically decode word meaning in comparable occipital areas (Correia et al., 2013). In a similar study, investigating the meaning of single nouns in written form across English and Portuguese, similar coordinates in occipital areas were also reported by Buchweitz and colleagues (Buchweitz et al., 2012). These studies indicate that a relationship between occipital areas and single word meaning are not completely implausible, especially since these occipital regions were specifically reported to be related to word meaning processing in different formats, only in this case in two different languages.

The categorical organization of concepts in the brain has also been studied looking only at language stimuli. In most cases, the task participants were asked to engage in, was explicitly semantic in nature. Early studies such as (Mummery et al., 1998) used a task in which participants had to match a target word to two other words and decide which word the target word was more similar too based on either color or location where the concept was observed. Other studies such as (Perani et al., 1999) used a same-different matching task in their study which compared written words to pictures. The importance of task when studying category effects for language was explicitly investigated by (Devlin et al., 2005). The authors tested semantic as well as perceptual decision tasks for manmade objects and animals represented by their written name and only observed category specific effects when participants engaged in the semantic decision task. When considering this result together with similar results in the study by Chao et al. (1999), it becomes clear, that task choice contributes to category specific activations in both picture and word format.

6.4.3 Limitations: mixed presentation, task and stimulus repetitions

The results of the experiment did not allow me to compare category-specific effects in different modalities or formats, as no contrast led to significant activation for either. In fact, in the written word modality, no contrast led to any clusters of activity. I have identified three possible reasons for this: (1) not enough stimulus repetitions (2) the choice of task (3) the effect of mixed stimulus representation.

To my knowledge, only one study has investigated category-specific effects for picture and written word format using a purely perceptual task. Chao et al. (1999) used different tasks, namely picture naming, picture-word matching and viewing and observed similar effects across tasks. However, they used more stimulus repetitions in the viewing condition than in the other two conditions. It is possible that the lack of activity especially in the written word condition, but also in the spoken word and pictures, was due to the fewer repetitions of one category. This is also due to the fact that I decided to look at three modality-format combinations and 6 categories in one experiment in a within participant design. One possibility is to include fewer categories (for example only animals and tools).

Related to the first possible limitation of this study is another difference to previous studies: the task. In most studies using a univariate approach to category-specific effects, a semantic task was used which might lead to overestimation of the effect of stimulus related bottom-up conceptual processing. One task that is often used is a type of n-back task where participants are asked to judge whether has the same category as the n-previously observed stimulus (Visser et al., 2010a). This type of task requires to hold at least one stimulus in working memory and compare it to the current stimulus on some level. This task therefore tests a different aspect of categorization related to memory processes which was not the objective of the current study. Another task that has been used is a straightforward categorization task where participants are asked to judge if the current stimulus is i.e. an animal or a vehicle (see Devlin et al., 2001; Visser and Lambon Ralph, 2011; Devereux et al., 2013). Such a task might be more suitable here.

Finally, in most studies that investigated more than one modality or format, they were presented in blocks. Mahon and colleagues or Chao and colleagues even could be said to have carried out separate experiments for pictures and spoken words (Chao et al., 1999; Mahon et al., 2009). In other cases smaller blocks of one modality were presented (Simanova et al., 2012; Vartiainen et al., 2009; Noppeney et al., 2005). In this study, stimuli were completely randomized across modality, format and category which might have led to a perceptual task switching effect decreasing activation for the individual categories. This might hold especially true for the written words modality. Written words led to the lowest activity of the three modality-format combinations, which might be related to the fact that they were less salient than the pictures (colorful photographs) and the spoken words which were perceived through a different sensory channel.

In order to carry out the comparison of category-specific effects across modalities and formats, a follow up experiment should take these three aspects into consideration. In my opinion, the best approach is to carry out three separate experimental sessions for each modality instead of randomizing across modality, while increasing stimulus repetition. With that, the continuum of animate and inanimate categories and possibly the perceptual task can be preserved. Alternatively, a semantic categorization task could also be introduced on specific task trials that are not analyzed to direct the attention of participants towards semantic processing.

6.5 Conclusion

Category-specific effects for animate and inanimate objects (often represented by animals and tools) have been observed in ventral temporal regions for pictures, spoken words and written words. In this study, I used a wider continuum of animate stimuli (humans and mammals) and inanimate stimuli (tools and furniture) to represent a spectrum of animacy. I also used a purely perceptual task that was not analyzed and mainly was used to help participants pay attention. I observed increased activation in the fusiform gyrus and the left inferior occipital gyrus for the contrast Animates (Humans and Animals) >

Inanimates (Tools and Furniture) in the picture modality and increased activation for the contrast Inanimates (Tools and Furniture) > Animates (Humans and Mammals) in the modality of spoken words. The reverse contrasts did not lead to increased activation in either modality and for the written word modality no contrast led to significant activation. I hypothesize that this could be due to three aspects of the study (1) the lack of a semantic task in a design that (2) used comparatively few stimulus repetitions and (3) was randomized across modality and category. In order to improve the design for future studies, I suggest to abandon the complete randomization and carry out three separate scanning sessions, one for each modality, with increased stimulus repetition and possibly introducing a semantic task on task trials that are not analyzed. This approach conserves the wide spectrum of categories, the within participant comparison of modalities and formats, and the perceptual task which I believe are strengths of the study.

CHAPTER 7

Discussion

In the previous three chapters, 3 experiments have been presented that test if conceptual knowledge is represented in a similar way for three combinations of the two modalities visual and auditory perception and two formats lexical and pictorial: written words (visual lexical), spoken words (auditory lexical) and pictures (visual pictorial). Humans can access representations of meaning through various modalities that use different sensory organs and neural pathways, yet the meaning of the percepts seems to be - at least at some point - refer to the same concept. Some structural principles of conceptual organization are also shared between different modalities and formats. The importance of animacy is well-documented for both the visual and the linguistic format, again suggesting that conceptual organization might at least in part be shared between both formats and modalities (see Chapter 1.3.3). I used behavioral as well as fMRI methods to ask how conceptual representations are different and similar in both human behavior and the brain. In this chapter, I will summarize the observations from both methods as well as discuss limitations and a future outlook. The more detailed discussions of my experimental findings can be seen in the individual discussion chapters of each experiment. Each experiment contributed to our understanding of how conceptual knowledge is represented between modalities and formats either in human behavior or in the brain.

7.1 Behavioral conceptual representations

Conceptual representations in human behavioral judgment are highly similar for written words and pictures, which while both visually perceived initially, differ in that one of them is also linguistic. In fact, the model that best explained the data from each format, was the data RDM from the other format. I also compared the behavioral results with 6 models that captured different aspect, either of the categorical structure present in the stimulus set as we designed it, or aspects of the stimulus set such as visual features. For pictures, the best fitting model was the combination model which included both the animate-inanimate distinction as well as the 6 categories. This was also the best performing model for the written words condition. The second best model for both conditions was the corpus relatedness model which represented linguistic measures of how similar concepts are to each other. Overall the pattern of results for the model to data comparison were also similar for both conditions, further supporting that the two datasets contained a similar conceptual structure. Both formats seem to be organized by larger conceptual categories such including animacy and subclasses of animates and inanimates such as humans and mammals and tools and plants.

It is interesting to note that a model derived from linguistic measures performed second best for the lexical format as well as the pictorial format. That could indicate that judgments about pictures are influenced by linguistic aspects, possibly because objects (and pictures of objects) are linked to their linguistic names very intimately. It has been suggested that categorical organization is influenced by naming processes that occur during language learning in infancy (Xu et al., 2005). Visual aspects of stimuli seemed to not influence similarity judgments, neither for the pictorial nor the lexical format. It is also interesting to note, that all 6 models for both formats seemed to not capture all features revealed by the cluster analysis I carried out. Apparently, in the case of the word condition, the plant category was group closely with the human category. This seems to be easy to explain, since plants are cultivated and eaten by humans - yet a classical categorical model of (in)animates is not organized in this way. In the picture category, the situation is similar. This pattern suggests that apart from animacy other underlying factors might drive perceptual judgements in these two formats. It is possible that co-occurence of objects in everyday life play a role.

Ideally, I could test the data of both formats against a model capturing how frequently objects co-occur in daily life. To my knowledge, however, there is no broad enough dataset that measures how often object pairs such as boys and apples are experienced together. The closest to this idea seems to be the ”scene grammar” studied by Võ et al. (2019) which so far has not covered the breadth of concepts I have studied but that might be a target though for future use with dataset once it captures enough everyday objects. This could shed further light on what drives perceptual judgments and also possibly point towards differences between conceptual representations between the two formats. Behavioral measures such as perceptual judgments do not allow me to draw further conclusions about how conceptual representations for written words and pictures converge in the brain. Based on the models tested and the direct comparison, conceptual geometry appears highly similar and driven by similar principles.

7.2 Modality-indepedent convergence zones in the brain

To further study how conceptual processing is carried out for different modalities and formats, I used fMRI which allows us to make inferences about brain regions. I compared three modalities, adding spoken words to the two visual modalities that were part of the behavioral experiment making slight adjustments to the concepts used previously. The three modalities included in this experiment allow us to study specific subparts of the modality space in detail: pictures and written words are both visually perceived, while spoken and written words are both linguistic with pictures and spoken words being maximally different. Activation for each modality-format combination when compared to the other two combinations is in line with our current knowledge of how sensory processing is carried out on each individual modality: auditory cortex activation for the spoken words (Okada et al., 2010) and activation in occipito-temporal areas for pictures (Grill-Spector, 2003). The written words condition when compared with the other two did not lead to any activation which also is explicable since both visual as well as language information is present int the two comparison contrasts.

Comparing pairs of modalities or formats reveals brain regions involved in processing of conceptual knowledge shared between modalities and formats and thus candidate regions for convergence zones. Overlapping activity between written and spoken word (identical lexical format but different modality) was observed in the left frontal region extending towards and touching the LIFG and RIFG and located in the the superior motor area (SMA) and pre-SMA and superior and middle temporal regions. The SMA activity is somewhat surprising given that it is not part of the classical language network. It has however, been associated with vocalization (Loh et al., 2016). For the two visual modalities, written words and pictures, the gros of the activity was located in temporal and occipital areas with focus on the fusiform gyrus which is not unexpected. Interestingly we also observed activity in the SMA. In this case, it is less clear if that activity could be attributed to silent vocalization, but given the presence of written words it remains a possibility. When comparing the two maximally different conditions, spoken words (lexical auditory) and pictures (pictorial visual), we observed an extended network of regions with focus on the left and right frontal lobes and precental gyri. Interestingly, this activity was bilateral, which is surprising considering that language processing is left-lateralized. This might indicate that conceptual processing resources that are shared between pictures and spoken words are more distributed than the classical language region.

Testing for overlapping regions for all three modalities allows me to study which networks are involved when different types of language processing and pictures are compared: Activity was located in superior and middle bilateral temporal regions, right frontal areas and LIFG areas as well as the SMA. Again, it is notable that the activity is bilateral despite even more linguistic input being included possibly pointing towards the activity being less linguistically driven in nature. The SMA effect seems to be influenced at least in large part by the spoken word condition but possibly also by the written words. Notably absent from any analysis of overlapping activity, is the anterior temporal lobe which is one of the candidate regions for modality independent processing. This region is close to air-filled cavities in the head next to the brain, which makes it susceptible to signal loss. The pulse sequences used in this experiment were not especially designed to capture ATL activity, so the absence of this region could simply be to too great signal loss. This is not the only explanation as the ATL is also associated sometimes with processing of concepts on a larger scale for example text processing.Another view of the ATL is that it does not subserve actual conceptual knowledge but instead is involved in the integration of meaning across modalities, which could be viewed as a twist on the ATL-as-hub hypothesis (Kiefer and Pulvermüller, 2012). If that is the case, in a similar experiment as I carried it out but optimized for ATL signal, I would except ATL activation. As it is, the present results do not count against a role of the ATL as a modality-independent hub. The results do support a more distributed network of modality-independent regions though of which the ATL, if its absence is due to signal loss, would just be one.

The results are also in line with the behavioral observations made in chapter 3 given that both point towards a significant degree of overlap between modalities even including spoken words which are perceived quite differently in their initial stages. While this analysis of fMRI data is not specifically focused on how conceptual structure is organized, it nevertheless supports the theory that there is significant overlap between modalities even when they involve an entirely different sensory organ and a different processing pathway in the brain.

7.3 Conceptual representations in the brain

As a next step, I studied how different categories are represented in the brain: Unfortunately, this analysis only led to two significant activations, one for the comparison of inanimate objects to animate objects in the spoken word modality and the other one of animate objects compared to inanimate objects in the picture modality. The opposite contrasts did not lead to significant activation in either pictures or spoken words and in the written word modality, no comparison lead to activation. The two activations itself are in line with previous research but the lack of activation for other category contrast is surprising. There are previous studies that used a similar approach and previous fMRI results were very much in line with the literature. Quality control did not observe major issues, for example excessive motion or spikes. These observations taken together indicate that the dataset is not substantially flawed.

There are a few reasons why the experimental design lead to category knowledge not contributing to the overall fMRI signal as it was the case in similar previous studies: Many studies looking at categories were limited to one modality (Martin et al., 1996; Mummery et al., 1998; Mahon et al., 2009), sometimes two (Kriegeskorte et al., 2008b; Costanzo et al., 2013) and very rarely three (Chan et al., 2011). This reduces the amount of stimuli that have to be shown, increasing how often a category can be observed by the participants, which increases the signal. I believe that the inclusion of three modalities makes the dataset much more interesting since a direct comparison like this has not been attempted, but it does lead to the scanner time being spread between more stimuli that have to be shown. A second choice I made ties into this: I chose to include a spectrum of categories that represent a larger spectrum of the animacy hierarchy, whereas in previous research that investigated animacy, the animacy spectrum was represented by only 2 categories (very frequently tools and mammals Noppeney et al., 2005; Fairhall et al., 2011; Simanova et al., 2012; Rundle et al., 2018 but see Wiggett et al., 2009, for a study that includes a larger spectrum of items). While those two stimulus types certainly are animate and inanimate and represent a part of that continuum, as I explained in 1.3.3, animacy is more of a spectrum. In including 6 categories, this study is more inclusive of the gradations of the animacy hierarchy as depicted in figure 1.3 but it again limits the time that each category can be presented, the number of repetitions and thus decreases the signal.

In many cases (Visser et al., 2010b; Visser and Lambon Ralph, 2011; Christian Gerlach, 2009; Bell et al., 2009), stimulus presentation was also blocked (one modality or format or category is presented in a block followed by a block of another modality, format or category). I opted for an event-related design that was fully randomized. A blocked by category design might have increased the signal for any specific modality. Finally, I chose to not include a category specific task that puts the attention of participants on conceptual knowledge. Many other studies have done that and the importance of tasks can be directly observed in Chao et al. (1999). One study that, except for the analysis method used, was quite similar was carried out by Devereux et al. (2013) who compared two formats (written words and pictures) and 6 categories (without the focus on animacy) using fMRI. The authors did observe categorical organization for both formats that was very comparable, which suggests that in our design categorical organization should also be observable. There are however, differences in the experimental design: a blocked approach instead of an event-related approach and a task that asked participants to specifically name the words. Taken together, it is plausible that processing of category knowledge did not generate a strong enough signal to be compared except in the two comparison were I did observe activity.

7.4 Outlook

Despite it creating many issues, it is important that research studying modality-independent conceptual processing is broadened to include (1) more than two modalities and formats and (2) more than 2 categories. The same questions I asked, could possibly be answered using separate experiments as well as a blocked design and different tasks. The choice of categories to include seems especially relevant. Very often only mammals and tools are used to represent animacy but as described in figure 1.3, animacy is a graded category. Within single one single modality and format, the graded nature of the animacy category distinction has been investigated (Connolly et al., 2012) but to my knowledge a comparison of modalities and how the animacy hierarchy is represented across them has not been done. The dataset I collected would have been ideal to carry out this comparison, were it not for the possible lack of signal in categorical processing. In the future, to answer a similar question, other aspects of the experimental approach taken here could be modified systematically to encourage categorical processing. Category effects have been observed to be present in perceptual stages of processing for example for object recognition but there seems to be a task effect here as well, in that again the tasks used possibly directed attention towards object categories (Gerlach et al., 2000, for a study investigating the picture modality using two different categorization tasks and an object decision task).

If the number of modalities (3) and the number of categories included (6) is to be taken constant, it would be possible to either adopt a blocked design or vary the tasks used. How tasks influence conceptual processing has been studied (i.e. Wiggett et al., 2009; Van Doren et al., 2010) have studied but as far as I know not yet looking into it using three modalities. As described in 1.3.1, many different tasks have been used in this line of research and it is plausible that not all tasks create the same effect. It would be interesting to investigate, how much any category effect within a single modality as well as between modalities is driven by task choice. In systematically varying the amount of conceptual processing that is required by a task, future studies can investigate how task processing influences categorical processing across modalities.

7.5 Conclusion

Visual perception and language comprehension both lead to the recognition of the meaning of concepts. Conceptual knowledge organization follows specific patterns of organization along categorical lines, one of the most prominent category distinction being between animates and inanimates. Several lines of research suggest that conceptual representations are organized along very similar lines between different modalities and might be independent of modality. Using behavioral and brain imaging methods, I studied both conceptual organization and brain activity of different modalities. The results support both shared conceptual organization in behavioral data as well as neural convergence zones that support conceptual processing across language and visual processing. Animacy does contribute significantly to conceptual organization, although other aspects such as possibly real life co-occurence of objects also play a role. My research also suggests that categorical processing might be more spurious than previously thought especially when modalities and categories are presented in random order and no task with explicit conceptual content is present. Future research is needed to investigate how categorical perception is influenced by task demands and how task effects modulate similarities and differences between modalities. Nevertheless, the dataset I collected provides a rich sample of categories and modalities and can be used in various ways to inform the scientific community about modality independent conceptual processing.

CHAPTER 8

Appendix

8.1 Instructions

Instruktionen

Lieber Teilnehmer,

danke dass du an unserem Experiment „In-Animacy“ teilnimmst! Es handelt sich um ein visuelles und auditorisches fMRT-Experiment - du wirst sowohl Bilder hören und Wörter lesen als auch gesprochene Wörter über Kopfhörer hören. Deine Aufgabe ist es die Bilder und Wörter (gesprochen wie geschrieben) wahrzunehmen und genau zuzuhören beziehungsweise hinzusehen. Gelegentlich wirst du einen lauten Ton folgend auf ein gesprochenes Wort hören oder ein Bild wird sich plötzlich drehen. Wenn du einen Ton oder eine Bilddrehung wahrnimmst, drücke bitte mit deinem linken Zeigefinger den Knopf auf der Tastatur.

Das Experiment läuft wie folgt ab : Nachdem wir dich bequem auf der Scannerliege platziert und in den Scanner gefahren haben, werden wir kurz mit dir kommunizieren um zu überprüfen, dass es dir im Scanner gut geht. Dann beginnt das Experiment mit einigen vorbereitenden Scans. Schließlich beginnt das eigentliche Experiment. Du wirst dann auch über den Kopfhörern die Geräusche des Scanners hören. Das Experiment beginnt mit einem schwarzen Bildschirm, wenn der Bildschirm grau wird, dann geht es gleich los und du solltest dich auf die Bilder und Wörter, die präsentiert werden, konzentrieren. Ein einzelner Durchlauf des Experiments dauert ca 10 min, danach kannst du kurz eine Pause machen. Es werden, in zwei Sessions, 12 Durchläufe durchgeführt.

Du kannst das Experiment jederzeit abbrechen wenn es z.B. dir schlecht geht oder du Schmerzen hast und musst dazu keinen Grund angeben. Dazu kannst du den Alarmknopf benutzen, den wir dir vor dem Experiment übergeben.

Instructions for the fMRI experiment in German

Bibliography

Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language Linguistic Theory*, pages 1–50.

Amedi, A., Jacobson, G., Hendler, T., Malach, R., and Zohary, E. (2002). Convergence of Visual and Tactile Shape Processing in the Human Lateral Occipital Complex. *Cerebral Cortex*, pages 1–11.

Argyropoulos, G. P. and Muggleton, N. G. (2012). Effects of Cerebellar Stimulation on Processing Semantic Associations. *The Cerebellum*, 12(1):83–96.

Argyropoulos, G. P. D. (2016). The cerebellum, internal models and prediction in 'non-motor' aspects of language: A critical review. *Brain and Language*, 161(C):4–17.

Aziz-Zadeh, L. and Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris*, 102(1-3):35–39.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59(1):617–645.

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15(2):145–153.

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004). Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron*, pages 1–15.

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B. H., and Ungerleider, L. G. (2009). Object Representations in the Temporal Cortex of Monkeys and Humans as Revealed by Functional Magnetic Resonance Imaging. *Journal of Neurophysiology*, 101(2):688–700.

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weiskoff, R. M., Cohen, J. M., Vevea, T. J., Brady, B., and Rosen, R. (1991). FunctionalMapping. *Science*, pages 1–4.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, pages 1–13.

Binder, J. R. and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11):527–536.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, 19(12):2767–2796.

Boersma, P. and Van Heuven, V. (2001). Praat, a system for doing phonetics by computer. *Glot international*, pages 1–7.

Bonin, P., Gelin, M., and Bugaiska, A. (2013). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Memory & Cognition*, 42(3):370–382.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002a). Functional Anatomy of Intra- and Cross-Modal Lexical Tasks. *NeuroImage*, 16(1):7–22.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002b). Modality independence of word comprehension. *Human Brain Mapping*, 16(4):251–261.

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

Bornkessel-Schlesewsky, I. (2015). Two routes to actorhood: lexicalized potency to act and identification of the actor role. *Frontiers in Psychology*, pages 1–21.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., and Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends in Cognitive Sciences*, 19(3):142–150.

Bossong, G. (1985). *Empirische Universalienforschung : differentielle Objektmarkierung in den neuiranischen Sprachen*. Narr, Tübingen.

Bracci, S., Daniels, N., and Op de Beeck, H. (2017). Task Context Overrules Object- and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex*, 84:81–12.

Bright, P., Moss, H., and Tyler, L. K. (2004). Unitary vs multiple semantics: PET studies of word and picture processing. *Brain and Language*, 89(3):417–432.

Brincat, S. L. and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7(8):880–886.

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., and Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language*, 120(3):282–289.

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2010). WHAT ARE THE FACTS OF SEMANTIC CATEGORY-SPECIFIC DEFICITS? A CRITICAL REVIEW OF THE CLINICAL EVIDENCE. *Cognitive Neuropsychology*, 20(3-6):213–261.

Caramazza, A. and Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8):354–361.

Carey, S. (2009). *The Origin of Concepts*. Developmental Cognitive Neuroscience. OUP USA.

Catani, M., Howard, R. J., Pajevic, S., and Jones, D. K. (2002). Virtual in Vivo Interactive Dissection of White Matter Fasciculi in the Human Brain. *NeuroImage*, 17(1):77–94.

Catani, M., Jones, D. K., and ffytche, D. H. (2004). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1):8–16.

Chan, A. M., Baker, J. M., Eskandar, E., Schomer, D., Ulbert, I., Marinkovic, K., Cash, S. S., and Halgren, E. (2011). First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. *Journal of Neuroscience*, 31(49):18119–18129.

Chao, L. L., Haxby, J. V., and Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, pages 1–7.

Chaudhuri, A. and Albright, T. D. (1997). Neuronal responses to edges defined by luminance vs. temporal texture in cacaque area V1. Vis. 1997-4271. *Visual Neuroscience*, pages 1–14.

Chee, M. W. L., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J., and Chee, M. (2000). Overlap and Dissociation of Semantic Processing of Chinese Characters, English Words, and Pictures: Evidence from fMRI. *NeuroImage*, 12(4):392–403.

Chinellato, E. and del Pobil, A. P. (2016). *The Visual Neuroscience of Robotic Grasping Achieving Sensorimotor Skills through Dorsal-Ventral Stream Integration*. Springer.

Christian Gerlach (2009). Category-specificity in visual object recognition. *Cognition*, 111(3):281–301.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., and Michel, F. (2000). The visual word form area. *Brain*, pages 1–17.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., and Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *NeuroImage*, 40(1):353–366.

Comrie, B. (1981). *Language universals and linguistic typology : syntax and morphology*. Blackwell, Oxford :.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H., and Haxby, J. V. (2012). The Representation of Biological Classes in the Human Brain. *Journal of Neuroscience*, 32(8):2608–2618.

Connor, C. E., Brincat, S. L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17(2):140–147.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte, M. (2013). Brain-Based Translation: fMRI Decoding of Spoken Words in Bilinguals Reveals Language-Independent Semantic Representations in Anterior Temporal Lobe. *Journal of Neuroscience*, 34(1):332–338.

Costanzo, M. E., McArdle, J. J., Swett, B., Nechaev, V., Kemeny, S., Xu, J., and Braun, A. R. (2013). Spatial and temporal features of superordinate semantic processing studied with fMRI and EEG. *Frontiers in Human Neuroscience*, pages 1–13.

Cree, G. S. and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132(2):163–201.

Culham, J. (2017). Neuroimaging reveals the human neural representations for visually guided grasping of real objects and pictures. *Journal of Vision*, 17:383.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1(1):123–132.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., and Damasio, A. (1996). A Neural Basis for Lexical Retrieval. *Nature*, pages 1–1.

Davis, M. H. and Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *Journal of Neuroscience*, pages 1–9.

Dehaene, S. and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6):254–262.

Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9(7):335–341.

Dehaene-Lambertz, G., Monzalvo, K., and Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology*, 16(3):e2004103–34.

Dejerine, M. J. (1892). Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. . *Mem.Soc.Biol*, pages 1–30.

Devereux, B. J., Clarke, A., Marouchos, A., and Tyler, L. K. (2013). Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects. *Journal of Neuroscience*, 33(48):18906–18916.

Devlin, J., Chang, M.-W., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*, pages 1–14.

Devlin, J. T., Rushworth, M. F. S., and Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia*, 43(1):69–74.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, E. H., Jalal Fadili, M., and Tyler, L. K. (2001). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia*, pages 1–22.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4):172–179.

Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, pages 1–4.

Fairhall, S. L., Anzellotti, S., Pajtas, P. E., and Caramazza, A. (2011). Concordance between perceptual and categorical repetition effects in the ventral visual stream. *Journal of Neurophysiology*, 106(1):398–408.

Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, pages 1–6.

Fedurek, P. and Slocombe, K. E. (2011). Primate Vocal Communication: A Useful Tool for Understanding Human Speech and Language Evolution? *Human Biology*, 83(2):153–173.

Felleman, D. J. and Van Essen, D. C. (2005). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, pages 1–47.

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., and Binder, J. R. (2016). Heteromodal cortical areas encode sensory-motor features of word meaning. *Journal of Neuroscience*.

Gainotti, G. and Silveri, M. C. (1996). Cognitive and Anatomical Locus of Lesion in a Patient with Category-specific Semantic Impairment for Living Beings. *Cognitive Psychology*, pages 1–36.

Gerlach, C., Law, I., Gade, A., and Paulson, O. B. (1999). Perceptual differentiation and category effects in normal object recognition. *Brain*, pages 1–12.

Gerlach, C., Law, I., Gade, A., and Paulson, O. B. (2000). Categorization and category effects in normal object recognition A PET Study. *Neuropsychologia*, pages 1–11.

Goodale, M. A. and Milner, A. D. (2004). Separatevisuaplathwaystorperceptionandaction. *Trends in Neurosciences*, pages 1–6.

Grabowski, T. J., Damasio, H., and Damasio, A. (1998). Premotor and Prefrontal Correlates of Category-Related Lexical Retrieval. *NeuroImage*, pages 1–12.

Grewe, T., Bornkessel-Schlesewsky, I., Zysset, S., Wiese, R., von Cramon, D. Y., and Schlesewsky, M. (2007). The role of the posterior superior temporal sulcus in the processing of unmarked transitivity. *NeuroImage*, 35(1):343–352.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, 13(2):159–166.

Grill-Spector, K. and Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Publishing Group*, 15(8):536–548.

Guell, X., Hoche, F., and Schmahmann, J. D. (2014). Metalinguistic Deficits in Patients with Cerebellar Dysfunction: Empirical Support for the Dysmetria of Thought Theory. *The Cerebellum*, 14(1):50–58.

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., and Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex*, 26(6):2919–2934.

Hamp, B. and Feldweg, H. (1997). Germanet - a lexical- semantic Net for German. *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 1–7.

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Marotta, G., and Pietrini, P. (2016). How concepts are encoded in the human brain: A modality independent, category-based cortical organization of semantic knowledge. *NeuroImage*, 135(C):232–242.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, pages 1–7.

Hauk, O. and Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, 21(3):191–201.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, pages 1–11.

Head, H. (1921). Aphasia: An Historical Review: THE HUGHLINGS JACKSON LECTURE FOR 1920. *Brain*, 43(4):390–411.

Hegele, M. and Seyfried, F. (2018). A Case for Sensorimotor Simulation as a Basis for the Animate–Inanimate Distinction in Infancy. *Journal of Motor Learning and Development*, 6(s1):S154–S168.

Heider, F. and Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, pages 1–18.

Henrich, V. (2010). GernEdiT - The GermaNet Editing Tool. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation LREC*, pages 1–8.

Hertrich, I., Dietrich, S., and Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience and Biobehavioral Reviews*, 68:602–610.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing . *Nature Reviews Neuroscience*, pages 1–10.

Hillis, A. E. and Caramazza, A. (1995). Cognitive and Neural Mechanisms Underlying Visual and Semantic Processing: Implications from "OpticAphasia". *Journal of Cognitive Neuroscience*, pages 1–26.

Hocking, J. and Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain and Language*, 108(2):89–96.

Huettel, S. A., Song, A. W., and McCarthy, G. (2014). *Functional Magnetic Resonance Imaging*. Sinauer.

Iordan, M. C., Greene, M. R., Beck, D. M., and Fei-Fei, L. (2015). Basic Level Category Structure Emerges Gradually across Human Ventral Visual Cortex. *Journal of Cognitive Neuroscience*, 27(7):1427–1446.

Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed Neural Systems for the Generation of Visual Images. *Neuron*, pages 1–12.

Jacques, C., Witthoft, N., Weiner, K. S., Foster, B. L., Rangarajan, V., Hermes, D., Miller, K. J., Parvizi, J., and Grill-Spector, K. (2016). Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia*, 83(C):14–28.

James, W. (1890). *The Principles of Psychology*. Cosimo Classics. Cosimo, Incorporated.

Johnson, C. R. and Rakison, D. H. (2006). Early Categorization of Animate/Inanimate Concepts in Young Children with Autism. *Journal of Developmental and Physical Disabilities*, 18(2):73–89.

Johnson, S. C. (1967). Hierarchical Clustering Schemes. *Psychometrika*, pages 1–14.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, pages 1–10.

Kiefer, M. and Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *CORTEX*, 48(7):805–825.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1):26–49.

Krieger-Redwood, K., Teige, C., Davey, J., Hymers, M., and Jefferies, E. (2015). Conceptual control across modalities : graded specialisation for pictures and words in inferior frontal and posterior temporal cortex. *Neuropsychologia*, 76(C):92–107.

Kriegeskorte, N. (2014). Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, pages 1–13.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008a). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:1–28.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. A. (2008b). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, 60(6):1126–1141.

Lambon Ralph, M. A., Sage, K., Jones, R. W., and Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, 107(6):2717–2722.

Leiner, H. C., Leiner, A. L., and Dow, R. S. (1993). Cognitive and language fundions of the human cerebellum. *Trends in Neurosciences*, pages 1–4.

Lesage, E., Hansen, P. C., and Miall, R. C. (2017). Right Lateral Cerebellum Represents Linguistic Predictability. *Journal of Neuroscience*, 37(26):6231–6241.

Lindenberg, R. and Scheef, L. (2007). Supramodal language comprehension: Role of the left temporal lobe for listening and reading. *Neuropsychologia*, 45(10):2407–2415.

Lockwood, H. T. and Macaulay, M. (2012). Prominence Hierarchies. *Language and Linguistics Compass*, 6(7):431–446.

Logothetis, N., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, pages 1–8.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878.

Loh, K. K., Petrides, M., Hopkins, W. D., Procyk, E., and Amiez, C. (2016). Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neuroscience and Biobehavioral Reviews*, pages 1–13.

Luria, A. R. and Hutton, J. T. (1977). A modern assessment of the basic forms of aphasia. *Brain and Language*, pages 1–23.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., and Caramazza, A. (2009). Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron*, 63(3):397–405.

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., and Martin, A. (2007). Action-Related Properties Shape Object Representations in the Ventral Stream. *Neuron*, 55(3):507–520.

Man, K., Damasio, A., Meyer, K., and Kaplan, J. T. (2015). Convergent and invariant object representations for sight, sound, and touch. *Human Brain Mapping*, 36(9):3629–3640.

Man, K., Kaplan, J. T., Damasio, A., and Meyer, K. (2012). Sight and Sound Converge to Form Modality-Invariant Representations in Temporoparietal Cortex. *Journal of Neuroscience*, 32(47):16629–16636.

Mandler, J. M. (2004). The conceptual foundations of animals and artifacts. *Creations of the Mind: Theories of Artifacts and their Representation*, pages 499–505.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, pages 1–17.

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C. H. S., Beaton, A., Desmond, J., De Witte, E., Fawcett, A. J., Hertrich, I., Küper, M., Leggio, M., Marvel, C., Molinari, M., Murdoch, B. E., Nicolson, R. I., Schmahmann, J. D., Stoodley, C. J., Thürling, M., Timmann, D., Wouters, E., and Ziegler, W. (2013). Consensus Paper: Language and the Cerebellum: an Ongoing Enigma. *The Cerebellum*, 11(1):457–25.

Marinković, K. (2004). Spatiotemporal Dynamics of Word Processing in the Human Cortex. *The Neuroscientist*, 10(2):142–152.

Marques, J. F., Canessa, N., Siri, S., Catricalà, E., and Cappa, S. (2008). Conceptual knowledge in the brain: fMRI evidence for a featural organization. *Brain Research*, 1194:90–99.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58(1):25–45.

Martin, A. and Chao, L. L. (2001). Semantic memory and the brain: structure and processes . *Cognitive Neuroscience*, pages 1–8.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, pages 1–4.

Mazaika, P. K., Hoeft, F., Glover, G. H., and Reiss, A. L. (2011). HBMMazaika2009.ppt. In *Human Brain Mapping*, pages 1–1.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Goualher, G. L., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1412):1293–1322.

Mechelli, A., Sartori, G., Orlandi, P., and Price, C. J. (2006). Semantic relevance explains category effects in medial fusiform gyri. *NeuroImage*, 30(3):992–1002.

Mikolov, T., Corrado, G., Chen, K., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv*, pages 1–12.

Mishkin, M. and Ungerleider, L. G. (2002). Contributions of striate input to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioral Brain Research*, pages 1–21.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object visionand spatial vision:twocorticalp hways. *Trends in Neurosciences*, pages 1–4.

Mollo, G., Karapanagiotidis, T., Bernhardt, B. C., Murphy, C. E., Smallwood, J., and Jefferies, E. (2016). An individual differences analysis of the neurocognitive architecture of the semantic system at rest. *Brain and Cognition*, 109(C):112–123.

Morton, J. and Johnson, M. H. (1991). CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition. *Psychological Review*, pages 1–18.

Mummery, C. J., Patterson, K., Hodges, J. R., and Price, C. J. (1998). Functional Neuroanatomy of the Semantic System: Divisible by What? *Journal of Cognitive Neuroscience*, pages 1–13.

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, 25(3):653–660.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS computational biology*, 10(4):e1003553–11.

Nobre, A. C., Allison, T., and McCarthy, G. (2002). Word recognition in the human inferior temporal lobe. *Nature*, pages 1–4.

Noonan, K. A., Jefferies, E., Visser, M., and Lambon Ralph, M. A. (2013). Going beyond Inferior Prefrontal Involvement in Semantic Control: Evidence for the Additional Contribution of Dorsal Angular Gyrus and Posterior Middle Temporal Cortex. *Journal of Cognitive Neuroscience*, 25(11):1824–1850.

Noppeney, U., Price, C. J., Penny, W. D., and Friston, K. J. (2005). Two Distinct Neural Mechanisms for Category-selective Responses. *Cerebral Cortex*, 16(3):437–445.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings Natl Acad Sci USA*, pages 1–5.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., and Hickok, G. (2010). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex*, 20(10):2486–2495.

Op de Beeck, H. P., Haushofer, J., and Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience*, 9(2):123–135.

Pasupathy, A. and Connor, C. E. (2001). Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation. *Journal of Neurophysiology*, pages 1–15.

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976–987.

Patterson, K. and Ralph, M. A. L. (2015). *Chapter 61. The Hub-and-Spoke Hypothesis of Semantic Memory.* Elsevier Inc.

Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *ArXiv*, pages 1–12.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images.* Elsevier.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., and Cappa, S. (1999). Word and picture matching] a pet study of semantic category effects. *Neuropsychologia*, pages 1–14.

Pereira, F., Botvinick, M., and Detre, G. (2013). Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments. *Artificial Intelligence*, 194(C):240–252.

Picard, N. and Strick, P. L. (2001). Imaging the premotor areas . *Current opinions in Neurobiology*, pages 1–10.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, Y. C., Cohen, L., Guazzelli, M., and Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings Natl Acad Sci USA*, pages 1–6.

Pinker, S. (2007). *The language instinct.* Harper Perennial Modern Classics.

Poulin-Dubois, D., Lepage, A., and Ferland, D. (1996). Infants' Concept of Animacy. *Cognitive Development*, pages 1–18.

Price, C. J. and Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, 19(3):473–481.

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. G., and Friston, K. J. (1996). The neural regions sustaining object recognition. *The Royal Society*, pages 1–8.

Proklova, D., Kaiser, D., and Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of Cognitive Neuroscience*, 28(5):680–692.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9):458–470.

Pulvermüller, F. and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, pages 1–10.

Radanović, J., Westbury, C., and Milin, P. (2016). Quantifying semantic animacy: How much are words alive? *Applied Psycholinguistics*, 37(06):1477–1499.

Rakison, D. H. and Oakes, L. M., editors (2003). *Early Category and Concept Development : Making Sense of the Blooming, Buzzing Confusion.* Oxford University Press, USA, New York, NY.

Rakison, D. H. and Poulin-Dubois, D. (2001). Developmental Origin of the Animate-Inanimate Distinctions. *Psychological Bulletin*, pages 1–20.

Ralph, M. A. L., Jefferies, E., Patterson, K., and Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Publishing Group*, 18(1):42–55.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, pages 1–6.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.

Rauschecker, J. P. and Tian, B. (2000). Mechanisms and streams for processing of ‘‘what’’ and ‘‘where’’ in auditory cortex. *Proceedings of the National Academy of Sciences*, pages 1–7.

Recanzone, G. H. and Sutter, M. L. (2008). The Biological Basis of Audition. *Annual Review of Psychology*, 59(1):119–142.

Regev, M., Honey, C. J., Simony, E., and Hasson, U. (2013). Selective and Invariant Neural Responses to Spoken and Written Narratives. *Journal of Neuroscience*, 33(40):15978–15988.

Riddoch, M. J. and Humphreys, G. W. (1987). A case of integrative visual agnosia. *Brain*, pages 1–32.

Rolls, E. T., Joliot, M., and Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122(C):1–5.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D. M., and Boyes-Braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychologiy*, pages 1–58.

Rundle, M. M., Coch, D., Connolly, A. C., and Granger, R. H. (2018). Dissociating frequency and animacy effects in visual word processing : An fMRI study. *Brain and Language*, 183:54–63.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, pages 1–7.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., and Connolly, A. C. (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*, 27(4):665–678.

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., and Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*, 54(3):2418–2425.

Silverstein, M. (1976). Hierarchy of features and ergativity. In Dixon, R. M. W., editor, *Grammatical Categories in Australian Languages*, pages 112–171. Australian National University, Canberra.

Simanova, I., Hagoort, P., Oostenveld, R., and van Gerven, M. A. J. (2012). Modality-Independent Decoding of Semantic Information from the Human Brain. *Cerebral Cortex*, 24(2):426–434.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., and Wise, R. J. S. (2006). Converging Language Streams in the Human Temporal Lobe. *Journal of Neuroscience*, 26(28):7328–7336.

Taylor, K. I., Devereux, B. J., and Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes*, 26(9):1368–1401.

Tremblay, P. and Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, 162(C):60–71.

Tyler, L. K. and Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, pages 1–9.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1):273–289.

Ulrich, M., Hoenig, K., Grön, G., and Kiefer, M. (2013). Brain Activation during Masked and Unmasked Semantic Priming: Commonalities and Differences. *Journal of Cognitive Neuroscience*, 25(12):2216–2229.

Van Doren, L., Dupont, P., De Grauwe, S., Peeters, R., and Vandenberghe, R. (2010). The amodal system for conscious word and picture identification in the absence of a semantic task. *NeuroImage*, 49(4):3295–3307.

Vandenberghe, R., Price, C. J., Wise, R. J. S., Josephs, O., and Frackowiak, R. S. G. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, pages 1–3.

Vartiainen, J., Parviainen, T., and Salmelin, R. (2009). Spatiotemporal Convergence of Semantic Processing in Reading and Speech Perception. *Journal of Neuroscience*, 29(29):9271–9280.

Vinje, W. E. and Gallant, J. L. (2000). Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science*, pages 1–5.

Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., and Ralph, M. A. L. (2010a). The inferior, anterior temporal lobes and semantic memory clarified: Novel evidence from distortion-corrected fMRI. *Neuropsychologia*, 48(6):1689–1696.

Visser, M., Jefferies, E., Embleton, K. V., and Lambon Ralph, M. A. (2012). Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing: Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes. *Journal of Cognitive Neuroscience*, pages 1–13.

Visser, M., Jefferies, E., and Lambon Ralph, M. A. (2010b). Semantic Processing in the Anterior Temporal Lobes: A Meta-analysis of the Functional Neuroimaging Literature. *Journal of Cognitive Neuroscience*, pages 1–12.

Visser, M. and Lambon Ralph, M. A. (2011). Differential Contributions of Bilateral Ventral Anterior Temporal Lobe and Left Anterior Superior Temporal Gyrus to Semantic Processes. *Journal of Cognitive Neuroscience*, pages 1–11.

Võ, M. L.-H., Boettcher, S. E., and Draschkow, D. (2019). ScienceDirect Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, 29:205–210.

von Stein, A., Rappelsberger, P., Sarnthein, J., and Petsche, H. (1999). Synchronization Between Temporal and Parietal Cortex During Multimodal Object Processing in Man. *Cerebral Cortex*, pages 1–14.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *Transactions on image processing*, pages 1–14.

Warrington, E. K. and McCarthy, R. A. (1987). Categories of Knowledge. *Brain*, pages 1–24.

Warrington, E. K. and Shallice, T. (1984). Category specific semantic impairments. *Brain*, pages 1–25.

Waters, D., Campbell, R., Capek, C. M., Woll, B., David, A. S., McGuire, P. K., Brammer, M. J., and MacSweeney, M. a. (2007). Fingerspelling, signed language, text and picture processing in deaf native signers: The role of the mid-fusiform gyrus. *NeuroImage*, 35(3):1287–1302.

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

Xu, F., Cote, M., and Baker, A. (2005). Labeling Guides Object Individuation in 12-Month-Old Infants. *Psychological Science*, pages 1–6.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, 11(11):1352–1360.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., OConnor, M. C., and Wasow, T. (2004). Animacy Encoding in English: why and how. Technical report, Proceedings of the 2004 ACL Workshop onDiscourse Annotation, Barcelona, July 2004.

Zwickel, J., Hegele, M., and Grosjean, M. (2011). Ocular tracking of biological and nonbiological motion: The effect of instructed agency. *Psychonomic Bulletin & Review*, 19(1):52–57.

List of Figures

List of Tables

Acknowledgments

First of all, I want to thank Prof. Dr. Mathias Hegele for giving me the opportunity to do and to finish this PhD even with all the problems of people being on other continents and foreign administrations. I also want to thank Prof. Dr. Alexander Goesmann for taking on this project and Prof. Dr. Katja Fiehler & Prof. Dr. Jörn Munzert for being part of my committee. I also loved being part of the IRTG which was made possible by you. It's an amazing community that I hope many other PhD candidates can join and learn from. Thank you to everyone in the BION, but especially Carlo and Bertram for answering any questions about fMRI data and physics and sequences and statistics.

One of the biggest mistakes a PhD candidate can make is thinking they are alone: I am very lucky that I have friends who were with me: First of all, thank you, Jochen. Without you, this would not have been possible in too many ways to count. For Ghüs, for Churchill, for listening, for not giving up even when I very much wanted to, for forever fighting the darkness with every seal rescue video you could find. Jing, you an inspiration to me every day! Now thanks to you, I will have my thesis story too. Anya, you are the most pragmatic, tenacious, truly helpful person I know. Whenever the US gets too weird, I remember that Germany doesn't have you. I look forward to lots of elegant dirt wine.

I am also very lucky in having an amazing family: I want to thank my parents who had to endure much with me during the entire process. I hope you can visit me in ever more unlikely places. I have to thank my grandparents, especially Opa Engelbert! My brothers: Raffael, for music and birds and caterpillars and for getting texts from remote airports when I am in remote airports. Valentin, you can make it through this, whatever it is.

Ping, who gave me a chance when I needed it and a push when I needed it. Franziska, without you, who knows if I would still be in science. Thank you also to all my colleagues in State College: Lindsey, Marissa, Laura, Ben and Chun-Ting. I feel at home in the States thanks to you! Special thanks have to go to Kelly who made me feel welcome before I was even there.

Thank you to all my (ex)-colleagues in Gießen: Caro, Raphael (for answering every last question I had about this process), Adam, Julia, Sanja (for challenging me), Falko (for rescuing the data especially), Lisa, Amalaswintha, Michael, Johannes, Tim, Chris, Heiko, Lisa, Christine, Hermann, Mathias R. (and I bet I forgot someone). I never thought I'd end up in a sports department but I did and I am so much better for it.

DECLARATION

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an anti-plagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen "Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis " in carrying out the investigations described in the dissertation

State College PA, August 28, 2019

Friederike Seyfried