

Abstract neural representations of language  
during sentence comprehension

Evidence from MEG and Behaviour

---

Sophie Luise Arana

**Funding Body** This research was funded by the Max Planck Society for the Advancement of Science ([www.mpg.de/en](http://www.mpg.de/en)).

**International Max Planck Research School (IMPRS) for Language Sciences** The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at [www.mpi.nl/imprs](http://www.mpi.nl/imprs).

**The MPI series in Psycholinguistics** Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at [www.mpi.nl/mpi-series](http://www.mpi.nl/mpi-series).

ISBN: 978-94-92910-30-1

Cover design: Julia Misersky

Printed by: Ipskamp printing

© Sophie Luise Arana, 2021

All rights reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author.

# Abstract neural representations of language during sentence comprehension

-

Evidence from MEG and Behaviour

Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,  
volgens besluit van het college voor promoties  
in het openbaar te verdedigen op

vrijdag 11 februari 2022  
om 12.30 uur precies

door

Sophie Luise Arana

geboren op 30 april 1990  
te Berlijn (Duitsland)

*Promotor*

Prof. dr. P. Hagoort

*Copromotoren*

Dr. J.M. Schoffelen

Dr. M. Rabovsky (Universit t Potsdam, Duitsland)

*Manuscriptcommissie*

Prof. dr. U. Noppeney

Dr. S.L. Frank

Prof. dr. N. Weisz (Paris Lodron Universit t Salzburg, Oostenrijk)

Abstract neural representations of language  
during sentence comprehension

-

Evidence from MEG and Behaviour

Dissertation

to obtain the degree of doctor  
from Radboud University Nijmegen  
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,  
according to the decision of the Doctorate Board to be defended in public on

Friday, February 11, 2022  
at 12.30 pm

by

Sophie Luise Arana

born on April 30, 1990  
in Berlin (Germany)

*Supervisor*

prof. dr. P. Hagoort

*Co-supervisors*

dr. J.M. Schoffelen

dr. M. Rabovsky (University of Potsdam, Germany)

*Manuscript Committee*

prof. dr. U. Noppeney

dr. S.L. Frank

prof. dr. N. Weisz (Paris Lodron University of Salzburg, Austria)







# Contents

Chapter 1	Introduction	3
Chapter 2	Sensory modality-independent activation of the brain network for language	17
Chapter 3	MVPA does not reveal neural representation of hierarchical phrase structure during reading	43
Chapter 4	Measuring compositional sentence meaning through behaviour	91
Chapter 5	Neural dynamics of combinatorial sentence processing	117
Chapter 6	Discussion	145
	Bibliography	153
Chapter A	Appendix	175



# Introduction

If you have ever gotten lost in a fiction story or were transported by a poem, you have experienced a powerful aspect of the human language: creativity. We can create meaning beyond what individual words can express, through combining them in ever new and infinite ways. This allows us to communicate ideas and events that never have or even could become reality. Take this stanza from Emily Dickinson's poem "Could I but ride indefinite". By creatively combining concepts, that by themselves are somewhat void of meaning, she is able to communicate an elusive feeling such as a state of carefree freedom.

...

*I said But just to be a Bee  
Upon a Raft of Air  
And row in Nowhere all Day long  
And anchor off the Bar*

The seemingly unbounded flexibility in our use of words is enabled by its counterpart, namely, structure and rules. For example, sentences in a language usually adhere to a fixed word order (e.g. subject verb object) and this word order does not vary arbitrarily from verb to verb. We don't say "John loves Mary" but then say "kissed John Mary". Such regularities exist at all levels of language: syntax, semantics, phonology and morphology. These rules allow us to immediately understand and use newly encountered words. For example, when I first encountered the verb "appen" (engl.: "to app") in the Netherlands, it took me a second to realise it was a verbing of the messaging application "WhatsApp". Then, I was immediately able to use the novel verb, asking whether my roommate

had “*apped* with our landlord lately” or complaining about someone “*apping* away all day”. The morphosyntactic rules of verb order and inflection allowed me to do so.

As humans, we constantly extract regularities from our environment and in the form of “abstract” or invariant mental representations they start to shape our behaviour (Brette 2019). One example for human behaviour revealing abstract mental representations is our ability to recognise a spoken word across a variety of speakers, accents, and noise conditions. Beyond speech, we extract sensory modality-independent patterns, forming concepts and categories based on a variety of inputs (Murphy 2004). Beyond single words and concepts, we can recognise general relations between entities in the world independent of the exact nature of those entities. All these abstract mental representations allow us to quickly generalise our knowledge to new situations (Lake et al. 2019) and learn fast from limited experience (Tenenbaum et al. 2011). Therefore, humans are optimally suited to recognise and apply linguistic regularities. In fact, language itself is likely a manifestation of this sensitivity to structure.

Further, we can distinguish between the theoretical notion of a mental representation on the one hand and its manifestation in the form of the brain signal, a neural representation, on the other hand. While behavioural evidence for the former is an underlying assumption and motivation for the work in this thesis, evidence for the latter can be regarded as the distant objective. The conditions a neural signal must meet to be termed a “neural representation” are somewhat debated (see Box 1). The experiments presented in this thesis should be considered a first step in establishing neural representations through identifying a correspondence relation between neural activity and abstract dimensions of language. Beyond correspondence, there is evidence for a causal link between those abstract dimensions and human behaviour (i.e. resulting in both functional and erroneous but rule-based behaviour), which I will briefly review below. However, whether each neural correlate can similarly be linked to behaviour is not explicitly addressed in this thesis. Therefore, whenever I conclude that some neural pattern “encodes” or “represents” a property of the stimulus, this should be taken to mainly imply that it “corresponds”.

For the language researcher, identifying which abstract dimensions modulate brain activity, can be useful for theorising about and investigating cognitive and neural processing theories. This is nicely exemplified by research on neural

representations within auditory cortex (Obleser and Eisner 2009, King et al. 2018): For example, having established neuronal tuning for both category-specific voice-onset times, e.g. distinguishing /b/ from /p/, as well as sub-phonetic details within a category, e.g. how prototypical /b/ sounds, Fox and colleagues were able to show which physiological mechanisms at the implementational level can generate this pattern of representation (Fox et al. 2020). The reason why these leaps in understanding have occurred only recently is partly due to methodological advances such as multivariate pattern analyses (MVPA), that allow us to specify the representational content encoded by the brain (Norman et al. 2006, Guggenmos et al. 2018). In the past decade, there has been a surge of new studies that have relied on MVPA to reveal neural representations beyond low-level features of the stimulus and across all cognitive domains, e.g. visual features of objects (Cichy et al. 2019), semantic features (Chan et al. 2011, Simanova et al. 2010, Simanova et al. 2014, Deniz et al. 2019), event structure (Frankland and Greene 2015) as well as abstract features of space (Bellmund et al. 2018), actions (Tucciarelli et al. 2015), physics (Schwettmann et al. 2019), magnitude (Luyckx et al. 2019, Sheahan et al. 2021) and rule learning (Reverberi et al. 2012). In this thesis, I apply MVPA to reveal the neural correlates underlying various abstract dimensions of language. Before I provide a detailed description of the linguistic dimensions of interest, I will briefly introduce the general concept of MVPA as well as the specific techniques applied throughout this thesis.

## BOX 1: Neural Representations

The term representation is used widely but inconsistently within the cognitive neuroscience literature. At one extreme, researchers assume the brain to "represent" a stimulus as soon as brain activity is reliably elicited by that stimulus (Kriegeskorte and Diedrichsen 2019). According to this definition, any brain signal that is correlated with a given stimulus feature and invariant to its exact physical manifestation, would constitute an abstract neural representation. At the other extreme, researchers debate whether abstract stimulus properties can be mapped onto neural representations at all. (Brette 2019).

*Box continues on next page ...*

No two brain responses of the same person and in response to the same stimulus are ever exactly identical. Therefore, one fundamental issue with this mapping is to prove a neural signal to be truly invariant given variation in both external stimuli as well as internal brain states

Assuming, that the invariance condition can be approximately satisfied, the minimum necessary condition for establishing a neural representation is some form of correspondence between a neural signal and some external state or stimulus property. This notion of representation through correspondence is purely statistical and potentially only meaningful from the perspective of the researcher, rather than the system itself.

Furthermore, a representation does not necessarily need to be implemented through sustained neuronal activity only but could rely on so called "activity-silent" encoding, likely implemented through a multitude of physiological processes (Stokes et al. 2015, Wasmuht et al. 2018, Fitz et al. 2020). For a correspondance to be scientifically meaningful, researchers often argue that evidence for a causal relationship between behaviour and neural correlate is necessary (Grootswagers et al. 2018). For example, if a given behaviour is a function of the neural representation of the environment, then the weaker the encoding of this representation in the brain the more the behaviour should degrade as well. Furthermore, a neural representation may be falsely evoked (i.e. a visual illusion) and thereby produce erroneous behaviour. Finally, according to an even more stringent definition, correspondance and behavioural relevance do not suffice to license the term "neural representation". Instead, a neural representation additionally needs to be accessible by other brain areas higher up in the processing hierarchy (Baker et al. 2021). In other words, a neural activation pattern only constitutes a neural representation if it directly serves as input to brain processes, rather than being a theoretical description imposed by the researcher. This latter condition is rarely explicitly tested within empirical reports.

## 1.1 Multivariate pattern analyses

In the classical, univariate analysis approach, differences in brain activity between conditions are estimated by treating each measurement variable (e.g. a sensor, time point or a voxel) as an independent piece of data. MVPA, in contrast, takes into account distributed patterns of brain activity across multiple variables. Therefore, MVPA can leverage information from both activation strength but also signal variability across sensors, which is not picked up by univariate approaches (Hebart and Baker 2018). Furthermore, the MVPA approach is non-directional. Both more and less activation within a sensor can provide information, whereas in univariate analyses fine-grained differences in signal direction are often cancelled out during averaging procedures. These properties of MVPA make it a suitable tool for identifying “spatial population codes”, i.e. distributed groups of neurons that are tuned to specific stimulus properties (Averbeck et al. 2006). Although such neurons exist at the narrow spatial scale of cortical columns, early success of MVPA in the domain of vision suggested that they are nonetheless detectable with non-invasive neuroimaging techniques such as fMRI (Kamitani and Tong 2005). This is likely possible because of random local variations in the distribution of those tuned populations, which translate into weak biases within individual fMRI voxels and thus can be picked up on by combining information across multiple voxels (but see also Op de Beeck 2010). Overall the technique has proven to be able to discriminate neural activity patterns, even in the absence of mean activation differences (Jimura and Poldrack 2012, Fahrenfort et al. 2017, Mur et al. 2009, Gwilliams and King 2020).

In this thesis I combine MVPA with magnetoencephalography (MEG). MEG has been shown to be equally sensitive to spatially distributed neural codes at the cortical column-level, likely due to random variation introduced by the irregular folding of the cortex that translate to the signal detected at the MEG sensors (Cichy et al. 2015). MEG is a non-invasive neuroimaging technique, that records magnetic fields produced by electrical currents of large populations of pyramidal neurons (approx. 50,000) activated synchronously. Because MEG allows to measure this fast neural firing activity with high temporal resolution, it is ideally suited for investigating fast and highly dynamic processes such as language comprehension. Furthermore, given theoretical and anatomical constraints about how different brain sources project to MEG sensors we can reconstruct the spatial distribution

of neural activity at the cortex with a resolution of 2-3 mm (Hämäläinen et al. 1993).

By applying MVPA to source-reconstructed MEG data, we are able to leverage the multivariate nature of the neural code along both temporal and spatial dimensions. In addition, MVPA can be applied in a searchlight approach, i.e. estimating the multivariate signal based on a moving window that spans only a subset of available source locations (Kriegeskorte et al. 2006) and time points (Su et al. 2012) at each step. This approach increases the sensitivity to focally distributed neural codes within contiguous macroscopic brain regions and can additionally reveal complex parallel distributed processes.

Both multivariate and univariate approaches can provide insight into abstract neural representations, through the notion of discriminability or dissimilarity. On a theoretical level, the two approaches are often contrasted in that the former targets information content and the latter activation, which has consequences for inferences drawn from either method (Allefeld et al. 2016, Hebart and Baker 2018). This does not mean, however, that univariate approaches cannot provide any insight into the nature of neural representations at all. For example, neural adaptation observed in priming paradigms has been taken to indicate underlying representations of invariant stimulus properties that are similar across target and prime. Event-related potentials (ERPs) such as the mismatch negativity, observed in response to a deviant auditory event in a sequence (e.g. oddball paradigm), have been exploited to identify which abstract features of sound stimuli are discriminated within primary auditory cortex (Paavilainen 2013, Näätänen et al. 2001). In both cases, the neural signal responds to a dissimilarity relation between two events. Throughout this thesis, I also rely on the notion of dissimilarity by applying two MVPA techniques specifically, classification and representational similarity analysis (RSA).

Classification of brain data involves training a supervised learning algorithm to discriminate different experimental conditions given a multivariate neural activation pattern. When training a classifier on MEG data, the multivariate patterns to be discriminated may be described by a high-dimensional vector containing the strength of the magnetic field (or a reconstructed cortical source). An algorithm may either find a decision boundary, i.e. discriminative algorithm, or may generate a probabilistic model of the data points, i.e. generative algorithm, within this high-dimensional space, against which new multivariate patterns can



be compared. Both generative (e.g. Gaussian Naive Bayes) and discriminative algorithms (e.g. support vector machines) have been shown to successfully classify stimuli based on brain data (Grootswagers et al. 2017, Guggenmos et al. 2018).

Whether information about the stimulus property of interest is present in the data is usually evaluated based on the accuracy of the classification, i.e. how many trials can be successfully labeled based on either the generative model or a decision boundary. Importantly, for the purpose of this thesis I am not interested in maximal accuracy of the classification. Even low accuracies can be seen as evidence for a neural correlate as long as performance is above the assumed or empirically estimated chance level, e.g 50% for a two-class classification problem. Importantly, some caution is warranted when assessing classification accuracies. Because the parameters of the learning algorithm usually outnumber the amount of data points, there will always be some dimension of neural activity according to which trials can be separated. We call this overfitting. In order to prevent overfitting, classifiers are best trained in a cross-validated fashion, i.e. they are trained on one part of the data and tested on another. Only when classification accuracy is estimated through cross-validation can the inference be made that classification relies on discriminable neural activity patterns that are invariant to the exact conditions used to train the classifier. Finally, Even when training is cross-validated, the standard error of accurate classification can be inflated in neuroimaging studies with small sample sizes (Varoquaux 2017). Therefore, in this thesis, accuracy will always be evaluated against the empirical chance level estimated by classification on permuted labels.

Once a measure of neural representational content is established in terms of discriminability (such as decoding accuracy) we can compare it against theoretical models of mental representation using representational similarity analysis (RSA). When comparing against theoretical models, we usually prefer a higher-dimensional description of the representational content. Instead of discriminating between two classes only, we can compute the degree of dissimilarity between all item pairs. Pairwise dissimilarity could be extracted by training multiple classifiers, but a metric based on direct comparison of the neural patterns (e.g. Euclidean distance or Pearson correlation) is usually more efficient.

It is important to note, that while multivariate approaches might allow for more high-dimensional characterisation of representational content, valid inferences still crucially depend on the exclusion of potentially confounding factors.

In the same way, that univariate analyses relying on binary contrasts need to control for orthogonal covariates, MVPA results can equally be misleading if not properly controlled (Popov et al. 2018). The standard cautionary tale is of a study, that reported discrimination of cognitive states during movie watching based on simultaneously recorded fMRI data. While they achieved extremely high accuracy in discriminating cognitive states related to the content of the movie, their decoders relied highly on signal stemming from voxels outside the brain which were likely related to physiological noise such as laughter-induced head motion (Sona et al. 2007).

Finally, we can leverage multivariate information not only along the dimension of measurement channels but also along the dimension of subjects. In the standard analysis approach, individual subject topographies are mapped onto a common anatomical source space. When targeting abstract representation, however, we usually assume the neural activation patterns to be idiosyncratic to a given subject. Ignoring these idiosyncrasies can reduce sensitivity to neural representations (Haxby et al. 2020). We can use multivariate approaches to estimate transformations of individual subject data such that they can be aligned to a common representational space instead. Such a transformation sharply increase our sensitivity to fine-grained stimulus-specific representational content to the degree that more explicit models of linguistic representations can be tested (Huizeling et al. 2020). In chapter 2 we show how multivariate canonical correlation analysis applied to MEG data can reveal abstract, word-specific fluctuations in the brain signal.

## 1.2 Abstract properties of Language

### Abstraction beyond low-level perceptual features

Abstract mental representation can be described at any hierarchical level of processing, starting at the level of low-level perception and progressing to multimodal and eventually complex structural abstraction (Gilead et al. 2020). The most basic form of abstraction is simply the recognition of external stimuli, invariant to their exact physical instantiation. Such invariant recognition occurs both within and across modalities. For example, people can recognise speech sounds beyond the individual variation across speakers and accents. We assume this ability to rely on abstract representations of language-specific sound categories (Maye et al. 2008), which are learned and highly flexible, e.g. people can quickly adapt to novel accents based on systematic phonetic transformations. We also learn abstract rules about how sounds may be combined together (morphosyntactic rules). For example, children at the age of 3 and beyond will produce overgeneralisations of morphological rules leading to errors like “she falled me” (e.g. Tomasello 2000). Beyond the level of individual speech sounds and syllables, entire words are recognised in terms of abstract representations. For example, when driving a car we can be alerted about an upcoming crossroad by either reading the word “stop” on a stop sign or by other passengers yelling “stop”. The abstract representation of the word “stop” and its behavioural consequences are modality-independent, since they can equally be activated through a visual or an auditory signal. Therefore, comparing neural activation across sensory modalities can provide insights into processing of abstract stimulus features. For example, in the past, researchers investigating abstract semantic information, have argued that a classifier trained on one modality and tested on another, must pick up on abstract semantic features, rather than physical aspects of the stimulus (Simanova et al. 2014, Deniz et al. 2019). In Chapter 2 we characterise the spatiotemporal dynamics of sensory modality-independent neural processing.

### Abstraction of syntactic structure

Beyond the invariant recognition of sounds and words, we also form abstract representations at the level of a sentence. For example, people tend to repeat the phrasal structure of sentences they were recently exposed to (Bock 1986), a

phenomenon called structural priming. This behaviour suggests that we extract the abstract structural configurations of sentences across varying lexical and semantic instantiations. Such structural configuration is crucial for sentence comprehension, because it can provide cues with respect to the hierarchical relations of words: Words can be grouped together into phrases (e.g. “the woman chases the cat”) and phrases in turn can be nested within other phrases (e.g. [the woman [who owns a dog] chases the cat]). Such hierarchical phrase structures can define sentence meaning according to non-adjacent dependencies between words (e.g. “a woman chasing”) rather than purely sequential meaning (e.g. “a dog chasing”). To extract hierarchical phrase structure from sequential input we rely on both semantic information but also on abstract syntactic information such as word class (e.g. nouns and verbs) and syntactic rules (e.g. word order). We are sensitive to these abstract syntactic regularities already from a young age. For example, children as young as 2 years old will start to use a newly learned word in the rule-conform word order even when first encountering it in the wrong order (Akhtar 1999).

Prior neuroimaging work has identified several brain areas, that activate in response to structured language (e.g. sentences as opposed to word lists) and are hence promising candidates for a locus of abstract neural representations of sentence structure. These areas include the left inferior frontal gyrus and left anterior and posterior temporal cortex as well as the inferior parietal cortex (Friederici 2011, Hagoort and Indefrey 2014, Hagoort 2017, Hultén et al. 2019). Although the exact functional role of each of these regions is still somewhat debated, most studies show all or at least one of them to activate in a modality-independent manner (Bemis and Pylkkänen 2013, Uddén et al. 2019), to activate in response to both syntactic and semantic dimensions of the stimulus (Fedorenko et al. 2012, Fedorenko et al. 2018 Matchin et al. 2019) and to increase in activation with increasing structural demands such as more nested phrases (Pallier et al. 2011, Nelson et al. 2017, Brennan et al. 2012). Only few neuroimaging studies have directly probed which aspects of sentence structure are represented in the brain during sentence processing. Some evidence for abstract phrase structure comes from studies investigating the neural effects of structural priming (Noppeney and Price 2004 Segal et al. 2013, Boudewyn et al. 2014). The scarcity of neural data, however, does not sufficiently support the complete invariance to semantic

features and information structure, which have been suggested to be confounding factors in earlier psycholinguistic studies (Ziegler et al. 2019).

## Abstraction at the syntax-semantic interface

Syntactic structure can rarely be considered independent from semantics (Jackendoff 2003, Goldberg 2006). In fact, semantic cues can determine the structural interpretation of a sentence even in the absence of clear syntactic structure. This is illustrated in sentences containing structurally ambiguous prepositional phrases, such as “The woman saw the dog with binoculars”. The syntactic cues in the sentence license two possible structural interpretations, one attaches the prepositional phrase to the verb (“The woman with binoculars”), the other to the immediately preceding noun (“The dog with binoculars”). Regardless of the ambiguity, most people have a preference to interpret the “binoculars” as an instrument to the verb “saw” based on semantic information and word knowledge.

On the other hand, syntactic dependencies (e.g. subject, object) as defined through word order and morphology provide cues with respect to the semantic relation between words. Given the order of words in “John loves Mary”, we identify “John” as the lover and “Mary” as the one being loved. Such relational assignment results in abstract representations at the syntax-semantic interface. For example, the abstract notion of an agent generalises across several semantic instantiations (“the one who loves”, “the one who eats”, “the one who breaks”) and syntactic structures (the subject in active sentences “the woman ate the vegetable” or object in passive sentences “the vegetable was eaten by the woman”). Evidence for abstract representations of relational information comes from behaviour in both adults and children (Rissman and Majid 2019): Young children can use the transitive argument structure of novel verbs (e.g. “the bunny is blicking the frog”) to correctly assign agent and patient roles, i.e. pointing to a picture illustrating the correct role assignment (Noble et al. 2011). Again, only few studies have probed neural correlates of abstract relational roles such as agent and patient in the brain. Some evidence for a neural instantiation comes from fMRI studies suggesting that representations of agent and patient may be encoded in superior temporal cortex across both language (Frankland and Greene 2020) and visual tasks (Wang et al. 2016) and these encodings have consequence for downstream processing of sentence meaning (Frankland and Greene 2015). In Chapter 3 and 5 we investigate

two different types of abstract structure emerging at the syntax-semantics interface, namely, prepositional phrase structure and thematic role assignment.

### 1.3 Outline of thesis

The ability to form invariant mental representations that generalise across exemplars is what enables humans to learn a language quickly and use it productively and creatively. Many language behaviours (e.g. speech errors, priming in production) provide cues about which abstract mental representations likely guide our language learning and comprehension. Furthermore, novel multivariate analysis approaches for neuroimaging data allow us to investigate the distributed neural code underlying such abstract representations. While much knowledge has been gained about neural representation of modality-specific abstraction (e.g. speech sound categorisation), fewer research has targeted higher-level linguistic abstraction, i.e. abstraction that is independent of sensory modality and targets regularities beyond the word level. In this thesis, I investigate the neural correlates underlying different levels of linguistic abstraction and reveal their spatiotemporal dynamics using MEG in combination with various MVPA techniques.

**Chapter 2** focuses on sensory modality-independent language processing. Previous studies have reported certain brain areas to similarly increase in activation both when words are presented in written and spoken form. Because these studies rely on averaging, however, they only capture generic components of the neural response. Word-specific fluctuations of the neural signal, on the other hand, are difficult to capture due to differences in position and orientation of neuronal sources across individual subjects. I apply multiset canonical correlation analysis to transform individual subject's neural data to recover word-specific signal. I investigate the spatiotemporal dynamics of modality-independent processing by comparing word-specific neural signals across subjects either reading or listening to sentences.

In **Chapters 3-5** I turn to abstract representations of linguistic structure at the sentence-level. During sentence comprehension we need to not only process each word individually, but also find meaningful combinations of all words according to both semantic and syntactic rules.

For example, although sentences are perceived serially, their meaning is often defined according to hierarchical dependencies of individual words and

phrases. Whether such hierarchical sentence structure is automatically encoded during passive tasks such as language comprehension has been debated (Frank et al. 2012). In **Chapter 3**, I investigate the neural correlates underlying implicit structural interpretation of syntactically ambiguous sentences. Specifically, we trained classifier to distinguish whether a person was reading a verb-attached or a noun-attached prepositional phrase based on the brain signal evoked as they read the final, semantically disambiguating word of the phrase.

In addition to the structural interpretation of a sentence, comprehension usually requires the computation of relational roles, i. e. who did what to whom. As a consequence we need to represent a sentence's meaning across multiple dimensions, namely, a collection of multiple thematic role-filler assignments. It is difficult, however, to construct a quantitative model of such high-dimensional meaning. One approach that has been proven successful to quantify other high-dimensional stimuli, such as visual scenes, is representation by similarity. In **Chapter 4** I tested whether human similarity judgments of transitive sentences would capture combinatorial sentence meaning. For this I collected similarity judgments from a large sample (n=200) using a multiple arrangement task. Finally, I suggest a possible use case for sentence-level similarity judgments as a behavioural benchmark for computational models of language processing.

Having shown that people are sensitive to relational information in simple transitive sentences, we then investigated the spatio-temporal dynamics of brain activity underlying the abstract processing of event structure in **Chapter 5**. So far, ERP data has been compatible with multiple alternative cognitive models of sentence processing. For example, it is not clear whether combinatorial processes, such as those establishing role-filler assignment, are reflected in the brain signal immediately (Rabovsky et al. 2018) or only after word-level processing has been completed (Brouwer et al. 2017). By explicitly modelling event structure of sentences and relating those models to MEG-recorded neural data using RSA, I investigate the dynamics of combinatorial processing with high temporal resolution.

In **Chapter 6** I summarise the results of this thesis and discuss some of its implications for future research.





## Sensory modality-independent activation of the brain network for language

The meaning of a sentence can be understood, whether presented in written or spoken form. Therefore, it is highly probable that brain processes supporting language comprehension are at least partly independent of sensory modality. To identify where and when in the brain language processing is independent of sensory modality, we directly compared neuromagnetic brain signals of 200 human subjects (102 males) either reading or listening to sentences. We used multiset canonical correlation analysis to align individual subject data in a way that boosts those aspects of the signal that are common to all, allowing us to capture word-by-word signal variations, consistent across subjects and at a fine temporal scale. Quantifying this consistency in activation across both reading and listening tasks revealed a mostly left-hemispheric cortical network. Areas showing consistent activity patterns included not only areas previously implicated in higher-level language processing, such as left prefrontal, superior and middle temporal areas, and anterior temporal lobe, but also parts of the control network as well as subcentral and more posterior temporal-parietal areas. Activity in this supramodal sentence-processing network starts in temporal areas and rapidly spreads to the other regions involved. The findings indicate not only the involvement of a large network of brain areas in supramodal language processing but also that the linguistic information contained in the unfolding sentences modulates brain activity in a word-specific manner across subjects.

## 2.1 Introduction

Language can be realized in different modalities: amongst others through writing or speech. Depending on whether the sensory input modality is visual or auditory, different networks of brain areas are activated to derive meaning from the stimulus. Besides different brain circuits being recruited to process low-level sensory information, differences in linguistic features across sensory modalities prompt a differential activation of brain areas involved in higher-order processing as well. For instance, speech is enriched with meaningful prosodic cues, but also requires coarticulated signals to be parsed into individual words. Written text has the advantage of instantaneous availability of full information compared to the temporally unfolding nature of speech. These differences are paralleled in the brain's response, and thus the sensory modality in which language stimuli are presented determines the dominant spatiotemporal patterns that will be elicited (Hagoort and Brown 2000).

Regardless of low-level differences, the same core message can be conveyed in either modality. Therefore, language processing models of the past and present (Geschwind 1979, Hagoort 2017, Hickok and Poeppel 2007) include not only early sensory (up to 200ms) processing steps, but also contain late (200 - 500 ms), more abstract, and supposedly supramodal processing steps. While early processing is largely unimodal, and supported by brain regions in the respective primary and associative sensory areas, later processes (for instance lexical retrieval and integration) that activate several areas within the temporo-frontal language network are assumed to do so independent of modality.

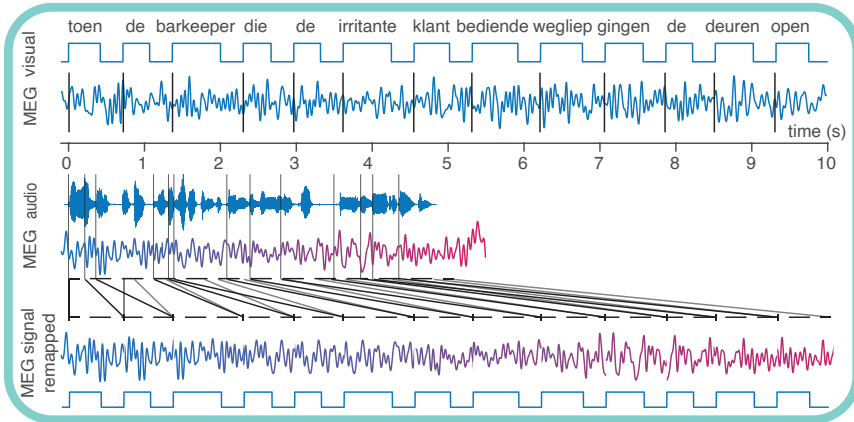
In order to gain insight into the location and timing of brain processes representing this latter, higher order processing of the linguistic content, researchers so far relied on carefully manipulated experimental conditions. As a result, our current understanding of how the brain processes language across different modalities reflects a large variety in tasks (semantic decision task (Chee et al. 1999), error detection task (Carpentier et al. 2001, Constable et al. 2004), passive hearing/listening (Jobard et al. 2007), size judgment (Marinkovic et al. 2003) and stimulus material (words (Chee et al. 1999), sentences (Bemis and Pykkänen 2013), and stories (Berl et al. 2010, Deniz et al. 2019, Regev et al. 2013)). Despite this wealth of experimental findings and resulting insights, an important interpretational limitation stems from the fact that the majority of studies employ modality specific

low-level baseline conditions (tone pairs and lines, spectrally rotated speech and false fonts, non-words, white noise (Lindenberg and Scheef 2007) to remove the sensory component of the processing. It is difficult to assess in how far such baselines are comparable across auditory and visual experiments. Recent fMRI work has demonstrated sensory-modality independent brain activity by directly comparing the BOLD response across visual and auditory presentations (Deniz et al. 2019, Regev et al. 2013). Yet, fMRI signals lack the temporal resolution to allow for a temporally sufficiently fine-grained investigation of the response to individual words.

Few studies used magnetoencephalography (MEG) to study supramodal brain activity and all are based on event-related averaging (Marinkovic et al. 2003, Bemis and Pykkänen 2013, Papanicolaou et al. 2017, Vartiainen et al. 2009). Averaged measures capture only generic components in the neural response. While generic components make a large contribution to the neural activity measured during language processing, there also exist meaningful variability in the neural response that is stimulus-specific and robust (Ben-Yakov et al. 2012). A complete analysis of the supramodal language network needs to tap into these subtle variations as well.

Here, we overcome previous limitation by achieving a direct comparison without relying on modality-specific baseline conditions, leveraging word-by-word variation in the brain response. Using MEG signals from 200 subjects, we performed a quantitative assessment of the sensory modality independent brain activity following word onset during sentence processing. The MEG data forms part of a large publicly available dataset (Schoffelen et al. 2019), and has been used in other publications (Lam et al. 2016, Schoffelen et al. 2017, Hultén et al. 2019, Lam et al. 2018). We identified widespread left hemispheric involvement, starting from 325 ms after word onset in the temporal lobe and rapidly spreading to anterior areas. These findings provide a quantitative confirmation of earlier findings in a large study sample. Importantly, they also indicate that supramodal linguistic information conveyed by the individual words in sentence context leads to subtle fluctuations in brain activation patterns that are correlated across different subjects.

## A Temporal alignment procedure



## B Cross-validated multiset canonical correlation analysis

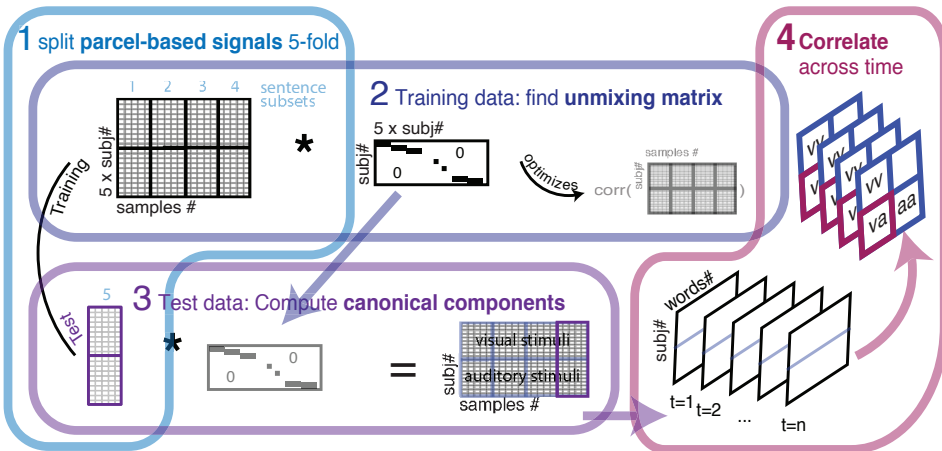


Figure 2.1.: Analysis pipeline.

Figure caption on next page.

**Figure 2.1.:** (A) Temporal alignment procedure. MEG signals of auditory and visual subjects differed in length due to different presentation rates. To achieve alignment between signals of auditory and visual subjects, auditory signals were epoched into overlapping segments. Each segment's first sample corresponds to the auditory word onset but each segment's length depends on the duration of the equivalent visual stimulus. Segments were then concatenated in original order to recover signal for the full sentence length. This way, the neural response to each word is fully taken into account in further comparisons, including in the case of short words for which stimulus late processing coincided with the next word presentation. (B) Starting point for the multi-set canonical correlation analysis were parcel-based neural signals for all subjects, consisting of five spatial components each. 1. Signals for all sentence trials were split into five subsets and for cross-validation one subset of sentences was left out as test data, while the remaining four subsets served as training data. 2. Based on the training dataset only an unmixing matrix was found, per parcel, defining the linear combination of the five spatial components so that the correlation across sets (subjects) and time samples were maximized. The cross-covariance was computed between all subjects' spatial components and across time collapsing over sentence trials. 3. The projection was applied to the test data to compute canonical variables for the left out sentence trials (purple outline) for all subjects. Steps 2 and 3 were repeated for all folds until each sentence subset had been left out once and the resulting canonical variables were concatenated until the entire signal was transformed. 4. Canonical variables were epoched according to word onsets and for each time point a subject-by-subject correlation matrix was computed across words. Correlation between cross-modal subjects (pink outline) were interpreted as quantifying supramodal activation.

## 2.2 Methods

### Subjects

A total of 204 native Dutch speakers (102 males), with an age range of 18–33 years (mean of 22 years), participated in the experiment. In the current analysis, data from 200 subjects were included. Exclusion of four subjects was due to technical issues during acquisition, which made their datasets not suitable for our analysis pipeline. All subjects were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological, developmental, or language deficits. The study was approved by the local ethics committee (CMO, the local “Committee on Research Involving Human Participants” in the Arnhem–Nijmegen region) and followed the guidelines of the Helsinki declaration. All subjects gave written informed consent before participation and received monetary compensation for their participation.

### Experimental Design

The subjects were seated comfortably in a magnetically shielded room and presented with Dutch sentences. From the total stimulus set of 360 sentences six subsets of 120 sentences were created. This resulted in six different groups of subjects who were presented with the same subset of stimuli although in a different (randomized) order with some overlap of items between groups. Within each group of subjects half of them performed the task in only the visual, the other half in only the auditory modality. In the visual modality, words were presented sequentially on a back-projection screen, placed in front of them (vertical refresh rate of 60 Hz) at the center of the screen within a visual angle of 4 degrees, in a black mono-spaced font, on a grey background. Each word was separated by an empty screen for 300 ms and the inter-sentence interval was jittered between 3200 and 4200 ms. Mean duration of words was 351 ms (minimum 300 ms and maximum 1400 ms), depending on word length. The median duration of whole sentences was 8.3 s (range 6.2 - 12 s). Auditory sentences had a median duration of 4.2 s (range 2.8 - 6.0 s) spoken in a natural pace. The duration of each visual word was determined by the following quantities: (i) the total duration of the audio-version of the sentence/word list (audiodur), (ii) the number of words in the sentence (nwords), (iii) the number of letters per word (nletters), and (iv) the

total number of letters in the sentence (sumnletters). Specifically, the duration (in ms) of a single word was defined as:  $(n\text{letters} / \text{sumnletters}) * (\text{audiotur} + 2000 - 150 * n\text{words})$ . In the auditory task the stimuli were presented via plastic tubes and ear pieces to both ears. Before the experiment, the hearing threshold was determined individually and the stimuli were then presented at an intensity of 50 dB above the hearing threshold. A female native Dutch speaker recorded the auditory versions of the stimuli. The audio files were recorded in stereo at 44100 Hz. During the post processing the audio files were low-pass filtered at 8500 Hz and normalized so that all audio files had the same peak amplitude, and same peak intensity. All stimuli were presented using the Presentation software (Version 16.0, Neurobehavioral Systems, Inc). Sentences were presented in small blocks, of five sentences each, along with blocks containing scrambled sentences, which were not used here. See Lam et al. (Lam et al. 2016) for more details about the stimulus material used. In order to check for compliance, 20% of the trials were randomly followed by a yes/no question about the content of the previous sentence/word list. Half of the questions addressed the content of the sentence (e.g. Did grandma give a cookie to the girl?) whereas the other half, addressed one of the main content words (e.g. Was the word 'grandma' mentioned?). Subjects answered the question by pressing a button for 'Yes' / 'No' with their left index and middle finger, respectively.

## MEG Data Acquisition & Structural imaging

MEG data were collected with a 275 axial gradiometer system (CTF). The signals were analog low-pass-filtered at 300 Hz and digitized at a sampling frequency of 1,200 Hz. The subject's head was registered to the MEG-sensor array using three coils attached to the subject's head (nasion, and left and right ear canals). Throughout the measurement, the head position was continuously monitored using custom software (Stolk et al. 2013). During breaks the subject was allowed to reposition to the original position if needed. Participants were able to maintain a head position within 5 mm of their original position. Three bipolar Ag/AgCl electrode pairs were used to measure the horizontal and vertical electrooculogram and the electrocardiogram.

A T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) pulse sequence was used for the structural images, with the following parameters: volume TR = 2300 ms, TE = 3.03 ms, 8 degree flip-angle, 1 slab, slice-matrix size =

256 × 256, slice thickness = 1 mm, field of view = 256 mm, isotropic voxel-size = 1.0 × 1.0 × 1.0 mm. A vitamin-E capsule was placed as fiducial behind the right ear to allow a visual identification of left-right consistency.

## Preprocessing

Data were bandpass filtered between 0.5 and 20 Hz, and epoched according to sentence onset, each epoch varying in length, depending on the number of words within each sentence. Samples contaminated by artifacts due to eye movements, muscular activity, and superconducting quantum interference device jumps were replaced by NaN before further analysis. Since all sentences had been presented in random order, we reordered sentences for each subject to yield the same order across subjects. Subsequently, the signals of the auditory subjects were temporally aligned to the signals of the visual subjects, ensuring coincidence of the onset of the individual words across modalities (Figure 2.1A). This alignment was needed to accommodate for differences in word presentation rate. The alignment was achieved by first epoching the auditory subject's signals into smaller overlapping segments. Each segment's first sample corresponded to one of the word onsets as annotated manually according to the audio file while each segment's length depended on the duration of the visual presentation of the corresponding word. Finally, all segments were concatenated again in the original order. By defining segments that were longer than the corresponding auditory word duration, the neural response to each word is fully taken into account and matched to the visual signal, even in the case of short words where the response partly coincided with the next word presentation. MEG data were then downsampled to 120 Hz.

## Source Reconstruction

We used linearly constrained minimum variance beamforming (LCMV) (Van Veen et al. 1997) to reconstruct activity onto a parcellated cortically constrained source model. For this, we computed the covariance matrix between all MEG-sensor pairs, as the average covariance matrix across the cleaned single trial covariance estimates. This covariance matrix was used in combination with the forward model, defined on a set of 8,196 locations on the subject-specific reconstruction of the cortical sheet to generate a set of spatial filters, one filter per dipole location. Individual cortical sheets were generated with the Freesurfer



package (Dale et al. 1999, version 5.1) ([surfer.nmr.mgh.harvard.edu](http://surfer.nmr.mgh.harvard.edu)), coregistered to a template with a surface-based coregistration approach, using Caret software (Van Essen et al. 2001) (download [here](#) and [here](#)), and subsequently downsampled to 8,196 nodes, using the MNE software (Gramfort et al. 2014) ([martinos.org/mne/stable/index.html](http://martinos.org/mne/stable/index.html)). The forward model was computed using FieldTrip's singleshell method (Nolte 2003), where the required brain/skull boundary was obtained from the subject-specific T1-weighted anatomical images. We further reduced the dimensionality of the data to 191 parcels per hemisphere (Schiffelen et al. 2017). For each parcel, we obtained a parcel-specific spatial filter as follows: We concatenated the spatial filters of the dipoles comprising the parcel, and used the concatenated spatial filter to obtain a set of time courses of the reconstructed signal at each parcel. Next, we performed a principal component analysis and selected for each parcel the first five spatial components explaining most of the variance in the signal.

## Multi-set Canonical Correlation Analysis

Multi-set canonical correlation analysis (MCCA) (Parra 2018), (de Cheveigné et al. 2018) was applied to find projections of those five spatial components that would transform the subject-specific signals so as to boost similarities between them. Canonical correlation analysis (CCA) is a standard multivariate statistical method often used to investigate underlying relationships between two sets of variables. Classically, canonical variates are estimated by transforming the two sets in a way that optimizes their correlation. We applied a generalized version of the classical approach (MCCA) (Kettenring 1971), which extends the method to multiple sets, here multiple subjects. In our case, we find linear combinations of the five spatial components for each of two subjects, so that the correlation across time between those subjects is maximized. Since we have more than two subjects, we find for each subject its own linear combination, which maximizes the correlation across time between all subjects from both modality groups (auditory and visual stimulation). Following Parra we obtained the optimal projection as the eigenvector with the largest eigenvalue of a square matrix  $D^{-1}R$ , where  $R$  and  $D$  are square matrices:

$$R = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{pmatrix}, D = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{N,N} \end{pmatrix}$$

Where  $a^{lk}$  are cross-covariance matrices between subject pairs and D contains only the diagonal blocks of within subject covariances (Parra 2018). In our case the cross-covariance matrices are of size 5-by-5 containing the cross-covariance between all five spatial components for a given subject (pair). The cross-correlation is computed across time points for each sentence and subsequently averaged across sentences. It is important to note here, that the canonical variates resulting from the optimal projection do not reflect sentence averages anymore but have the same temporal resolution as the original source signals. CCA is prone to overfitting and known to be unstable (Dinga et al. 2019). For reliable CCA estimates the number of samples should be much larger than the number of features, i.a. a sample-to-feature ratio of 20/1 is recommended (Stevens 2012). We estimated the canonical variables over concatenated data, which included between 756 and 1453 samples per sentence compared to only five features (spatial components) which provides a decent sample-to-feature ratio. Further, we estimated our canonical variables out-of sample using 5-fold cross-validation to limit overfitting. We randomly split all sentences into five subsets, estimating projections on 96 sentences and applying them to the 24 left out sentences (Figure 2.1 B).

## Statistical Analysis

As per the study design, the subjects were assigned to one of six stimulus sets. Different groups of subjects were presented with different sets of sentences. Since MCCA relies on commonalities across datasets, we could only combine data from subjects who received the exact same stimulation. We therefore applied MCCA for each subgroup of subjects who listened to or saw the same stimuli separately. Initially, we constrained our analysis to the first set of 33 subjects (henceforth exploratory dataset). After applying the projection to the data we computed a time-resolved Pearson correlation between all possible subject pairings. To this end, we first epoched the resulting canonical components according to individual word onsets and selected only content words (nouns, adjectives & verbs) for subsequent steps. Before computing the correlation we subtracted the mean

across samples. For each pair of subjects, we computed the correlation between two sets of observations, i.e. a pair of vectors with each data point reflecting the subject-specific neural signal for each of the individual words (lexical items), at a given time point relative to word onset, and at a given cortical location. Correlation coefficients of cross-modality pairings, that is correlations between subjects reading and subjects listening to the sentences are interpreted as capturing supramodal processing. We used a permutation test with clustering over time and space (parcels) for family-wise error rate correction for statistical inference (Maris and Oostenveld, 2007), using 1000 randomizations of the epoched words. To this end, we randomised word order for the source-reconstructed parcel time series of the auditory subjects to test for exchangeability of the exact word pairing across sensory modalities. By destroying the one-to-one mapping of individual lexical items, the null distribution allowed for a distinction between individual item specific shared variance, and shared variance due to a more generic response. We also computed modality-specific responses as a quality check of the analysis pipeline given the well known spatiotemporal activity patterns of early sensory brain areas. For this, we averaged correlation across either only pairs of subjects reading or only pairs of subjects listening. These correlations were not constrained to content words but computed across all words. For statistical inference, we again used a permutation test with the same parameters as described earlier. This time, however, we randomised word order for the source-reconstructed parcel time series of both auditory and visual subjects, thereby destroying the one-to-one mapping of individual items within both modalities.

Finally, we analysed the remaining sets of subjects (confirmatory dataset) using the analysis pipeline described earlier. We evaluated the overlap in the results across all six subgroups using information prevalence inference (Allefeld et al. 2016). Prevalence inference allows formulation of a complex null hypothesis (i.e., that the prevalence of the effect is smaller than or equal to a threshold prevalence, where the threshold can be realised by different values). For each of the six sets of data, we obtained spatial maps of time-resolved supramodal correlations, as well as 1000 permutation estimates after word order shuffling (see above). We used the smallest observed average correlation across subgroups as the second-level test statistic. We then tested the majority null hypothesis of the prevalence of the effect being smaller than or equal to a threshold prevalence. For this, we computed the largest threshold such that the corresponding null hypothesis could still be

rejected at the given significance level,  $\alpha$ . This was done after concatenating the minimum statistic from all parcels and time points, using the maximum statistic to correct for multiple comparisons in time and space (parcels). For each parcel, we evaluated the highest threshold at which the prevalence null hypothesis could be rejected at a level of  $\alpha = 0.05$  (see Figs. 2.6 and 2.7 for cortical maps showing thresholds averaged and per time point).

To ensure that MCCA as a preprocessing step did not artificially increase correlations between cross-modality subjects, we conducted an additional control analysis on the exploratory dataset. For this, we additionally tested the observed correlation patterns computed on all words (both function and content words) against a null distribution obtained by permuting sentence order 500 times and, importantly, doing this before MCCA. This permutation was not fully unconstrained, because we aimed at aligning sentences across modalities with the same number of words to avoid loss of data and to preserve ordinal word position. Thus, we did a random pairing between sentences with the same number of words after binning the sentences according to their word count. Sentences consisting of 9, 14, or 15 words were infrequent, with fewer than five occurrences each. After each permutation, we performed the temporal alignment between sensory modalities (aligning the word on-sets), followed by cross-validated MCCA and computation of the time-resolved correlations of cross-modal subject pairs. Due to the long computation time of the canonical variates, we created this null distribution for the exploratory data only. Results from this additional, conservative permutation test can be found in Figure 2.5.

## 2.3 Results

### Modality-specific activation

We first quantified the similarity between different subjects' brain response within only the exploratory dataset (33 subjects) by correlating word-by-word fluctuations in brain activity between all possible pairs of subjects. Averaging the correlations across those subject pairings for which subjects were stimulated either in the same sensory modality, or each in a different modality, allowed us to evaluate the modality-specific brain response and the supramodal response, respectively. As displayed in Figure 2.2, early sensory cortical areas only show correlated activity for the group of subjects receiving the stimuli in the corresponding sensory modality, for the visual (red), and auditory (blue) modalities. We found that MCCA is a crucial analysis step in order to reveal meaningful inter-subject correlations. Only after MCCA does cortical activity in visual and auditory areas become significantly correlated (cluster-based permutation test,  $p = 0.001$  for both) across those subjects performing the task in the visual or auditory domain respectively (Figure 2.2A).

### Supramodal activation patterns

We averaged between-subject correlations over all cross-modal subject pairings as a metric for supramodal activity. We observed significant supramodal correlated activation patterns in mostly left-lateralized cortical areas (Fig. 2.3; cluster-based permutation test,  $p = 0.001$ ). The effect has a large spatial and temporal extent, becoming apparent as early as 250 ms and lasting until 700 ms after word onset. Parcels in middle superior temporal gyrus (STG) contribute to the effect at the earliest time points, followed by the posterior and anterior part of the STG and about 50 ms later the anterior temporal pole. Supramodal correlated activation in ventral temporal cortex follows a similar temporal and spatial pattern, with supramodal correlations starting out more posterior around 292 ms and evolution towards the middle anterior temporal lobe at 308 ms. Other areas that express supramodal activity at relatively early time points are medial prefrontal cortex and primary auditory cortex (250 ms), followed by subcentral parietal regions and supramarginal gyrus at around 300 ms, and finally dorsolateral frontal cortex (DLFC, 325 ms). By the time 375 ms have passed, the entire orbito-frontal cortex,

anterior and DLFC as well as inferior frontal gyrus (IFG) show strong supramodal subject correlation. Supramodal activation in the frontal lobe further extends towards posterior regions including pre- and postcentral gyrus. At around 400 ms supramodal subject correlation in the anterior temporal pole reaches its peak. In addition to the lateral cortical areas, correlated activity also extends to left dorsal and ventral anterior cingulate cortex (ACC) as well as left fusiform gyrus. The spatio-temporal patterns of supramodal activation described so far are robust, also when ordinal word position and MCCA overfitting were controlled for in the statistical evaluation (Fig. 2.5; cluster-based permutation test,  $p = 0.002$ )

## Prevalence Inference

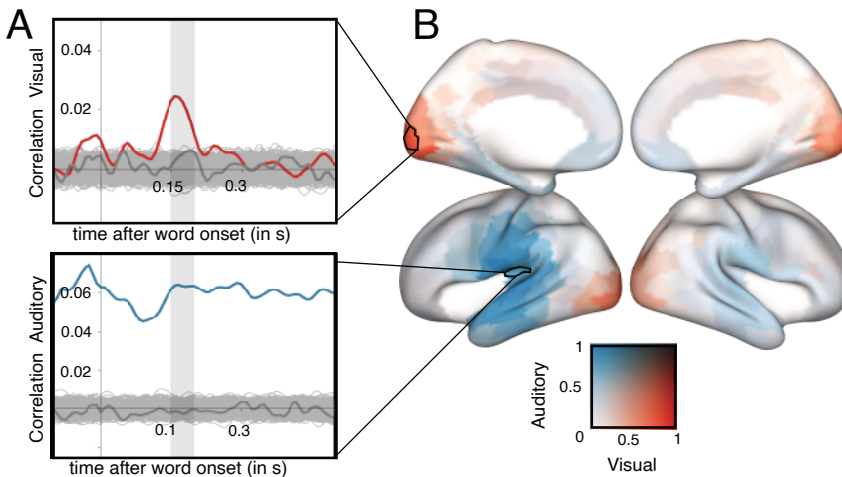
Our confirmatory analysis combined over all six datasets and tested whether the spatiotemporal patterns observed in the exploratory dataset would generalize to the population. For those parcels at which the global null hypothesis could be rejected, we infer that at least in one of the datasets an effect of supramodal processing was present (Fig. 2.4). In addition, we evaluated the majority null hypothesis of whether, in the majority of subgroups in the population, the data contains an effect (threshold  $> 0.5$ , significant parcels under the majority null hypothesis outlined in black in figure 2.4B).

The global null hypothesis (no information in any set of subjects in the population) could be rejected at a level of  $\alpha = 0.05$  in on average 40 parcels per time point (between 325 and 617 ms after onset, std. = 31.48). For those parcels for which the largest bound  $\gamma_0$  is larger than or equal to 0.5, we can infer that in the majority of datasets, the activity patterns were similar across subjects, independent of modality. This majority null hypothesis could be rejected (at a level of  $\alpha = 0.05$ ) in 90% of parcels that also showed a global effect (Fig. 2.4, black outline; see also Fig. 2.8 for results in the right medial hemisphere). For parcels at which the global null hypothesis could be rejected, the average largest lower bound  $\gamma_0$  at which the prevalence null hypothesis can be rejected is shown in Figures 2.6 and 2.7. Compared to the temporal pattern of the largest nominal suprathreshold cluster from the cluster-based permutation test conducted on the exploratory dataset, the effect became significant in the majority of datasets later in time and was less long lasting (325 - 608 ms). Given this time span, the majority null hypothesis was rejected in on average 42% of those parcels contributing most to the largest cluster. The orbitofrontal cortex and IFG showed an involvement

in supramodal processing in both analyses, but the effect there was much more temporally sustained in the exploratory dataset. In addition, according to the exploratory dataset, supramodal activation of STG occurred almost 100 ms earlier as compared to IFG. Based on the confirmatory dataset, however, supramodal correlated activation in IFG and STG appeared almost simultaneously. Finally, the exploratory analysis revealed supramodal activation in primary and premotor areas extending over the entire left dorsolateral surface, of which only the most ventral parcels close to the Sylvian fissure were significantly supramodal in the majority of datasets. Thus, the spatial extent of the effect was partly reduced for prevalence inference compared to the cluster-based permutation approach on the exploratory data. Nevertheless, widely overlapping anatomical regions were indicated by both analyses, encompassing dorsolateral frontal gyrus and the middle and superior parts of the temporal lobe at first, and the inferior frontal and orbitofrontal cortex as well as anterior temporal lobe later.

## 2.4 Discussion

Our aim was to quantify similarities of the brain response across reading and listening at a fine temporal scale. To this end we correlated word-by-word fluctuations in the neural activity across subjects receiving either auditory or visual stimulation. We identified a widespread left-lateralized brain network, activated independently of modality starting 325 ms after word onset. Importantly, dividing our large study sample into six subsets, we could directly quantify the consistency and generalizability of these activity patterns. The spatial distribution of the supramodal activation is in line with the known involvement of left hemispheric areas, including parts of left temporal cortex, left inferior parietal lobe, as well as prefrontal cortex (Vigneau et al. 2010, Chee et al. 1999, Constable et al. 2004, Braze et al. 2011, Liuzzi et al. 2017, Lindenberg and Scheef 2007, Spitsyna et al. 2006, Homae et al. 2002). The involvement of both STG and IFG fits predictions from the Memory, Unification and Control model (MUC), in which activity reverberating within a posterior-frontal network (Baggio and Hagoort 2011, Hagoort 2017) is thought to be crucial for language processing. According to the MUC model temporal and parietal areas support the retrieval of lexical information, while unification processes are supported by inferior frontal cortex. Bidirectional communication (Schiffelen et al. 2017) between these areas is facilitated by white

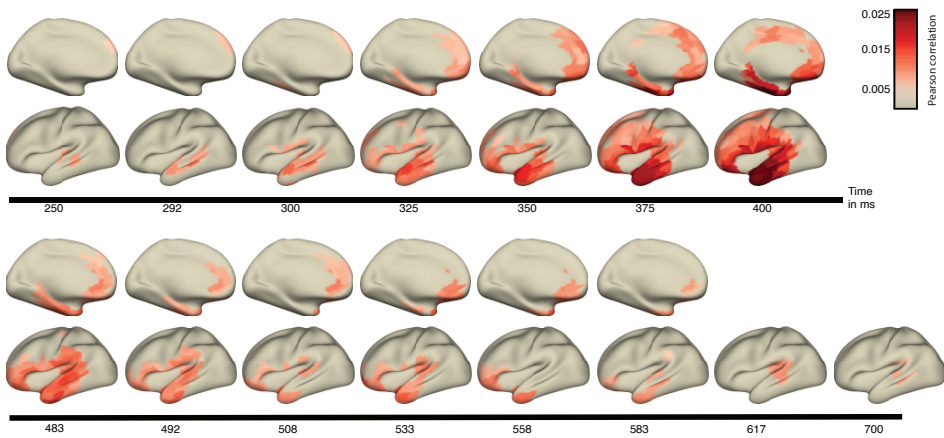


**Figure 2.2.: Specificity of the within-modality correlated activity patterns.**

(A) Time-resolved correlation values averaged across all visual subject pairings for a parcel in left primary visual cortex (upper panel) and all auditory subject pairings for a parcel in left primary auditory cortex (lower panel), before (dark grey line), and after MCCA (blue and red lines). Light grey lines show recomputed correlation values for 1000 random permutations of word order across subjects. Notably, signals of auditory subjects highly correlate even before word onset. This is likely due to a more varied distribution of information in the auditory signal caused by the continuous nature of auditory stimulation and as a result differing time points at which individual words become uniquely recognizable. The MCCA is blind to the stimulus timing and will thus find canonical variables that yield maximal correlations at any timepoint if possible. (B) Cortical map of the spatial distribution of correlations, comparing within modality visual subject pairs (red) with auditory subject pairs (blue). Correlation strength is expressed as the Pearson correlation coefficient averaged over a time window from 150 to 200 ms post word-onset and normalized by the maximum value of that window.

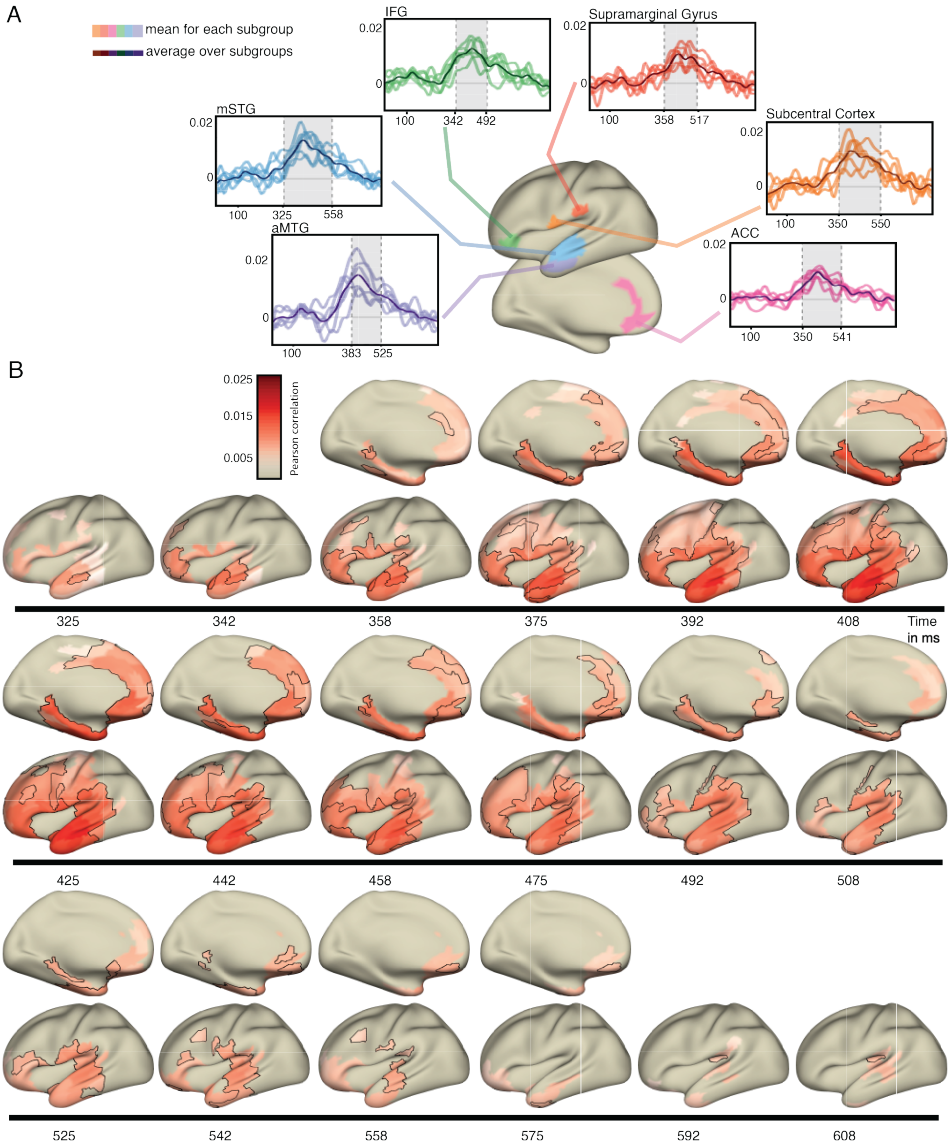
matter connections. We observe that temporal areas are supramodally activated at earliest time points and sustain activation for the longest compared to other regions. Over time, supramodal activation spreads from middle and posterior left STG to the anterior temporal pole. This rapid progression of activity from posterior to anterior regions mirrors previous observations (Marinkovic et al. 2003, Vartiainen et al. 2009), adding to those findings a direct quantitative comparison of the supramodal brain activity.





**Figure 2.3.: Supramodal correlated activity patterns.**

Time-resolved spatial maps of supramodal correlated activity patterns (averaged over all possible cross-modal subject pairings) in the left hemisphere. Medial views of the brain surface are depicted in the first and third row, lateral views in the second and fourth row. Color codes for strength of correlation. Colored parcels were most strongly correlated between cross-modal subject pairs (nonparametric permutation test, corrected for multiple comparisons).



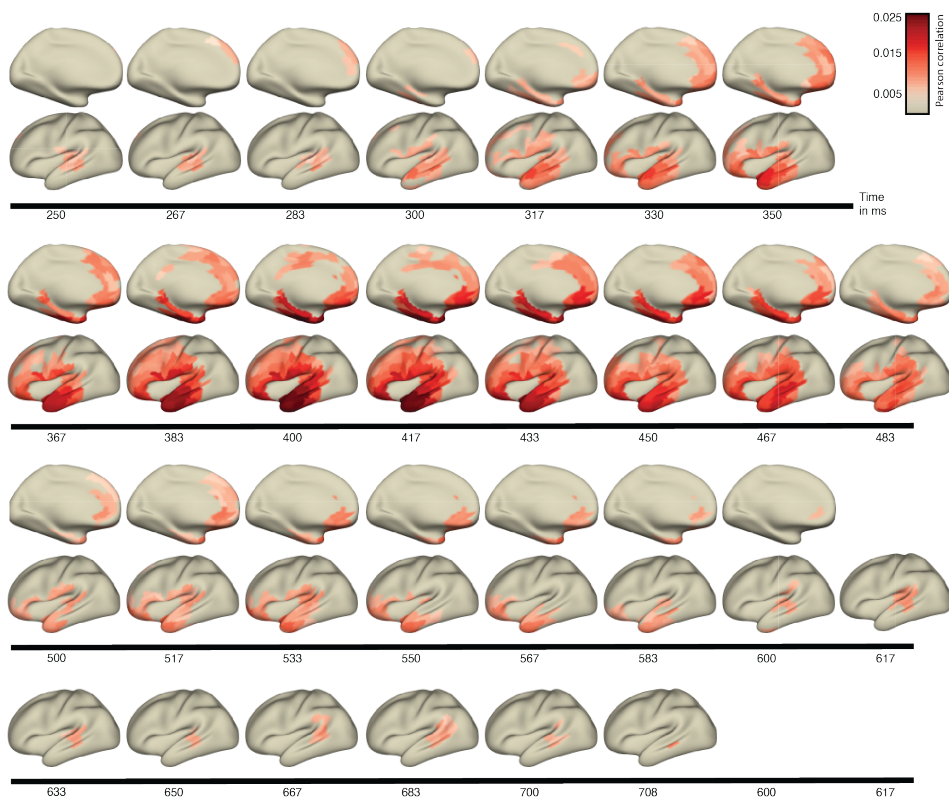
**Figure 2.4.:** Supramodal correlated activity patterns consistent across the majority of datasets.

Figure caption on next page.

**Figure 2.4.:** Supramodal correlated activity patterns of word-specific activity consistent across the majority of datasets. (A) Averaged correlation time courses (mean over all possible cross-modal subject pairings) are shown for selected parcels in inferior frontal Gyrus (IFG, green), supramarginal Gyrus (red), Subcentral Cortex (orange), Anterior cingulate cortex (ACC, pink), anterior middle temporal Gyrus (aMTG, purple), and middle superior temporal gyrus (mSTG, blue). Time courses are shown for each dataset individually (light-colored lines) as well as averaged (dark lines). Grey shaded areas mark statistically significant time points. (B) Time-resolved spatial maps of cross-modal correlations in the left hemisphere. Medial views of the brain surface are depicted in the first, third and fifth row, lateral views in the second, fourth and sixth row. For those parcels that were part of the largest nominal suprathreshold cluster tested on only the exploratory dataset, the mean correlation over all six datasets is shown. Color codes for strength of correlation. In addition, the parcels at which the majority null hypothesis according to prevalence inference could be rejected are outlined in black.

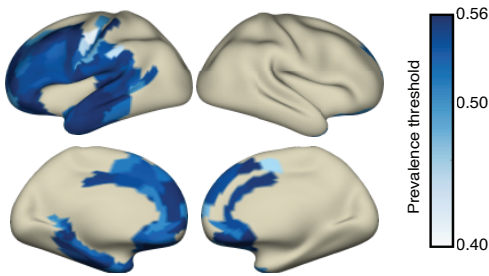
## Beyond the core language network and the single word level

We observed modality-independent activity in dorsal frontal cortex, in addition to more widely reported inferior parts of the frontal cortex (Jobard et al. 2007, Lindenberg and Scheef 2007, Constable et al. 2004, Marinkovic et al. 2003, Homae et al. 2002, Michael et al. 2001). This could be due to us using linguistically rich sentence material, of varying syntactic complexity, as opposed to single words (Chee et al. 1999, Marinkovic et al. 2003, Booth et al. 2002, Liuzzi et al. 2017, Vartiainen et al. 2009) or short phrases (Bemis and Pylkkänen 2013, Carpentier et al. 2001, Braze et al. 2011). Indeed, discrepancies with respect to frontal lobe involvement in modality-independent processing seem to mainly arise from differences in stimulus material and task demands (Braze et al. 2011). A recent meta-analysis has identified that more complex syntax robustly activates dorsal parts of the left IFG (Hagoort and Indefrey 2014). Further, a previously published analysis of these MEG data showed DLFC to be sensitive to sentence progression effects (Hultén et al. 2019). Two previous fMRI studies using narratives (Deniz et al. 2019, Regev et al. 2013) add to the debate. Regev et al. correlated BOLD responses evoked by different modalities. They report supramodal activation in the left frontal lobe, extending beyond inferior frontal regions. Deniz et al. study modality-independent brain areas by modelling semantic features of the stimulus in one modality and used the model to predict the BOLD signal in the other modality. They report BOLD signals in prefrontal cortex to be well predicted across modalities. In sum-



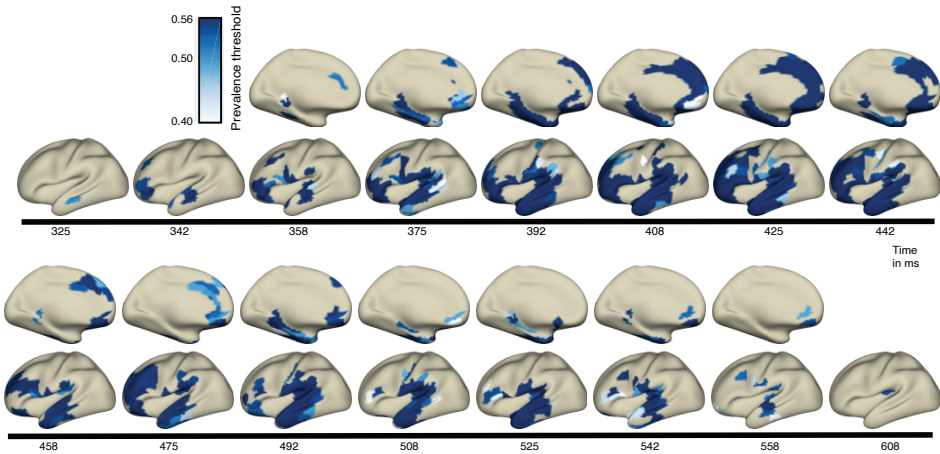
**Figure 2.5:** Significant supramodal correlated activity patterns as assessed by an additional permutation test.

In an additional significance test, we shuffled the sentence order 500 times prior to MCCA, controlling for the possibility that MCCA as a preprocessing step may artificially increase correlations between subjects (through overfitting). Importantly, this permutation was not fully unconstrained, since we aimed at aligning sentences across modalities with the same number of words, to avoid loss of data and to preserve ordinal word position. Thus, we did a random pairing between sentences with the same number of words, after binning the sentences according to their word count. Sentences consisting of 9, 14 or 15 words were infrequent, with fewer than 5 occurrences each. After each permutation, we performed the temporal alignment between sensory modalities (aligning the word onsets), followed by cross-validated MCCA and computation of the time-resolved correlations of crossmodal subject pairs. Due to the long computation time of the canonical variates, we created this null distribution for the exploratory data only. In the figure, color codes for strength of correlation. Colored parcels were most strongly correlated between cross-modal subject pairs (corrected for multiple comparisons).



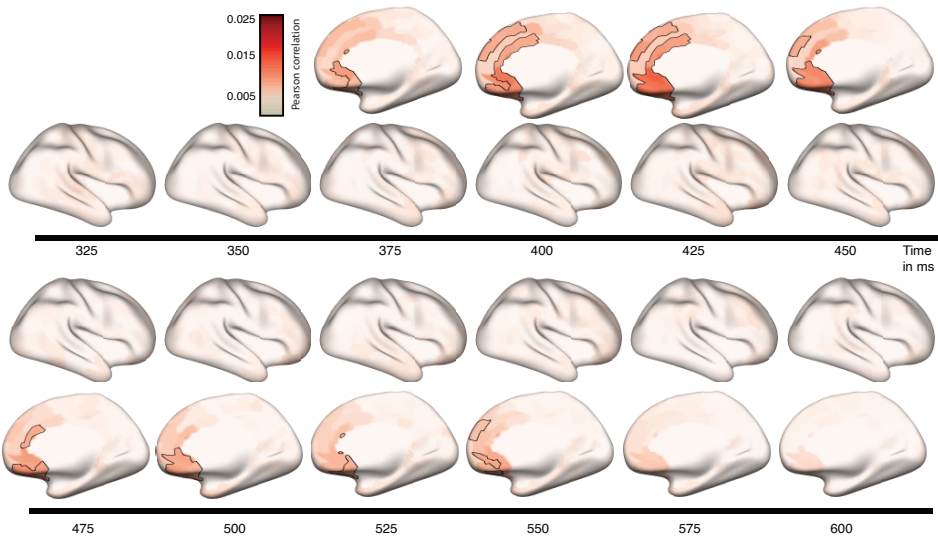
**Figure 2.6: Cortical map of maximum threshold  $\gamma_0$ .**

For those parcels at which the global null hypothesis could be rejected, the mean (over time) maximum threshold is plotted, for which the null hypothesis can be rejected ( $\alpha = 0.05$ ). Given the sample size of six datasets, the number of second-level permutations and a significance level of  $\alpha = 0.05$  the maximally possible threshold that can be reached is 0.5633.



**Figure 2.7.: Cortical map of prevalence threshold  $\gamma_0$ .**

For those parcels at which the global null hypothesis could be rejected, the maximum threshold is plotted, for which the null hypothesis can be rejected at a level of  $\alpha = 0.05$ .



**Figure 2.8.: Time-resolved spatial maps of cross-modal correlations for the right hemisphere.**

The average correlation over all six datasets is shown. Color codes for strength of correlation. In addition, the parcels at which the majority null hypothesis according to prevalence inference could be rejected are outlined in black.

mary, while complex stimuli consistently activate prefrontal areas beyond inferior frontal cortex, the exact stimulus features which cause this supramodal activation are still debated.

Some previous studies, using narratives and fMRI, report supramodal activation not to be restricted to the left hemisphere (Deniz et al. 2019, Regev et al. 2013, Jobard et al. 2007). It could be that the previously observed bilateral involvement is due to differences in context-based semantic processing during narratives, as compared to the processing of isolated sentences in our experiment. Menenti and colleagues have specifically contrasted BOLD activity in response to sentences presented within a neutral or a local context. The authors indeed reported right frontal cortex to be more sensitive to local discourse context as compared to its left-hemispheric homolog (Menenti et al. 2009). Further research is needed to determine whether this effect of presence and absence of narrative thread similarly affects lateralisation of brain activity in MEG.

Even though mostly restricted to the left hemisphere, our results also implicate extra-linguistic areas in supramodal processing. Specifically, we find bilateral supramodal activation within ACC. The ACC is a midline structure, forming part of a domain-general executive control network supporting language processing (Hagoort 2017, Cattinelli et al. 2013). It is sensitive to statistical contingencies in the language input and thus might play a role in mediating learning and adaptation in response to predictive regularities in both local experimental as well as global environment (Weber et al. 2016). It should be noted that deep sources are normally poorly detectable in MEG (Hillebrand and Barnes 2002) and we thus consider any interpretations with respect to the midline structures as tentative.

## Supramodal orthography-phonology mapping

We observed supramodal activation in post-central and subcentral gyrus, as well as supramarginal gyrus, which coincides temporally with supramodal activation of primary auditory cortex. Activity in supramarginal gyrus has been repeatedly elicited by cross-modal tasks (Sliwinska et al. 2012), such as rhyming judgments to visually presented words (Booth et al. 2002), for which conversion between orthographic and phonological representations is likely needed. At the same time, post- and subcentral areas partly span articulatory motor and somatosensory areas for the mouth and tongue. Together, the supramodal activation of these areas

suggests that retrieval of phonetic and articulatory mappings is not limited to speech perception only but also occurs during passive reading.

## Leveraging word-by-word variability of the neural response

Neuroelectric brain signals exhibit strong moment-to-moment variability. While some of this variability is related to the experimental stimulation, and therefore associated with specific cognitive activity, some of it is unrelated, ongoing neural activity. By applying MCCA across subjects we reduced this type of noise and made subtle word-by-word fluctuations in the MEG signal interpretable. Comparing neural activity across subjects is challenging due to differing position or orientation of neuronal sources relative to the MEG sensors. We used parcellated MEG source reconstruction in combination with exact temporal alignment of individual sentences across subjects. This allowed for the extraction of signal components that are shared across subjects, thus reducing the intersubject spatial variability, which is commonly observed in more traditional (for instance, dipole fitting) procedures (Vartiainen et al. 2009). MCCA thus allowed us to more directly investigate time-resolved inter-subject correlations and move beyond event-related averages (Marinkovic et al. 2003). Importantly, our analysis approach allows us to conclude that the identified supramodal activity is word-specific. Our findings therefore go beyond showing a general activation of those areas as compared to baseline but rather reveal consistent word-by-word fluctuations of activation within the recruited areas.

## Latency of supramodal processing

The temporal alignment procedure, as a necessary preparation step for the MCCA procedure, followed by the estimation of time-resolved intersubject correlations, focused on common signal aspects that are exactly synchronized across subjects. The differences in sensory modality specific characteristics of the input signal require dedicated processing with likely different processing latencies, which may also lead to latency differences in the activation of supramodal areas. For example, Marinkovic and colleagues report shorter reaction times during the visual task, yet found earlier activity peaks for the auditory task in corresponding early sensory cortex and left anterior temporal lobe (Marinkovic et al. 2003). In contrast, other work observed earlier anterior temporal lobe activation for visual, compared to



auditory stimulation (Bemis and Pylkkänen 2013). Our results indicate a certain degree of overlap across modalities in the temporal window within which supramodal cortical areas are activated. It is possible, that we observed more temporally extensive activation, for instance, related to unification processes, because we used longer sentences. In addition, any overlap may have been amplified as a necessary consequence of the MCCA procedure. Evidently, correlations between signals from auditory subjects were boosted with less temporal specificity compared to visual subjects (Figure 2.2). This observation was unexpected and may be due to more continuous stimulation in the auditory experiment. As the sound of a spoken word unfolds, the timing at which it becomes uniquely recognizable will vary across word. Thus, the distribution of information in the auditory signal is much more varied as compared to the visual. MCCA will pick up on any common relationship across subjects regardless of timing. In our specific application, projections were estimated on concatenated data, effectively making the method blind to word onset boundaries.

In conclusion, this study provides direct neurophysiological evidence for sensory modality independent processes supporting language comprehension in multiple left hemispheric brain areas. We identified a network of areas including domain general control areas as well as phonological mapping circuits over and above traditional higher-level language areas in frontal and temporal-parietal regions, by quantifying between-subject consistency of their respective word-specific activation patterns. These consistent activation patterns were word-specific, and thus likely reflect more than just generic activation during language processing. Finally, we show that alignment of individual subject data through MCCA is a promising tool for investigating subtle word-in-context specific modulations of brain activity in the language system.



## MVPA does not reveal neural representation of hierarchical phrase structure during reading

During comprehension, the meaning extracted from serial language input can be described by hierarchical phrase structure. Whether our brains explicitly encode hierarchical structure during processing is, however, debated. In this study we recorded Magnetoencephalography (MEG) during reading of structurally ambiguous sentences to probe neural activity for representations of underlying phrase structure. 10 human subjects were presented with simple sentences, each containing a prepositional phrase that was ambiguous with respect to its attachment site. Disambiguation was possible based on semantic information. We applied multivariate pattern analyses (MVPA) to the MEG data using linear classifiers as well as representational similarity analysis to probe various effects of phrase structure building on the neural signal. Using MVPA techniques we successfully decoded both syntactic (part-of-speech) as well as semantic information from the brain signal. Importantly, however, we did not find any patterns in the neural signal that differentiate between different hierarchical structures. Nor did we find neural traces of syntactic or semantic reactivation following disambiguating sentence material. These null findings suggest that subjects may not have processed the sentences with respect to their underlying phrase structure. We discuss methodological limits of our analysis as well as cognitive theories of "shallow processing", i.e. in how far rich semantic information can prevent thorough syntactic analysis during processing.

## 3.1 Introduction

Although we perceive language mainly in a sequential fashion (e.g. by reading word by word) we need to take into account information beyond the sequential order to fully comprehend its meaning. For example, in a sentence like “The woman who owns two dogs chases the cat” we understand that the woman is the one chasing, not the dogs. This knowledge can be expressed through hierarchical, structured relationships between the words. Specifically, words can be grouped into constituents (e.g. “Who owns two dogs” and “The woman chases the cat”) and constituents in turn can be nested into higher-level phrases, as shown in 3.1. The resulting nested phrase structure then fully describes the important conceptual units and their relationships with each other. Thus, hierarchical phrase structure also directly relates to thematic role assignment (the woman being assigned the agent role of the chasing action).

### 3.1 [[The woman [who owns the dogs]] chases the cat]

This type of structured meaning is to a large degree determined by syntax. As seen above, syntactic aspects like word order, function words (here: the relative pronoun ‘who’) as well as morpho-syntactic features such as number agreement provide cues with respect to the word-phrase relationships. Semantic information (e.g. animacy) or even just semantic association itself can also guide how structure should be assigned. In the above example, syntactic cues, however, override simple semantic association between the lemmas “dog” and “chase”. In theory, hierarchical descriptions can be applied to all linguistic levels of the stimulus during language processing (e.g. syntactic, semantic and phonological structure) (Jackendoff 2003).

How hierarchical phrase structure building is neurally encoded as we process language is still an open question. In fact, some have even disputed its neural and psychological reality during language use altogether (Frank et al. 2012). Some recent evidence for the reality of hierarchical phrase structure building comes from neuroimaging studies that assess its consequences on memory load (Nelson et al. 2017, Pallier et al. 2011) and production (Giglio et al. (in prep)). For example, Pallier et al. varied linguistic constituent size while keeping overall sentence length constant and identified brain regions whose activity parametrically increased with the size of the constituents (larger constituents thought to result in higher memory demands and stronger neural activity) (Pallier et al. 2011). Following a

similar approach, Nelson et al. modelled neural activity according to a hierarchical phrase-structure model and found it to explain more variance when fitted to intracranial data as compared to alternative models that were based on transition probabilities only (Nelson et al. 2017). This is in line with behavioral evidence, demonstrating that humans prefer a hierarchical interpretation over a linear one, for example when interpreting ambiguous noun phrases, such as “second blue ball” (Coopmans et al. 2021). At the same time, there are several studies demonstrating that reading times can often be sufficiently accounted for by sequential-structure models (Frank and Bod 2011), casting doubt on how pervasive the construction of hierarchical structure during language processing really is.

In early psycholinguistic experiments, hierarchical structure building has been measured through reading time behaviour for structurally ambiguous sentences. One example for such ambiguity is prepositional phrase attachment. Prepositional phrases (PPs) in sentence-final position (examples 3.2 & 3.3) are structurally ambiguous with respect to their attachment to the main clause. For example, a prepositional phrase can be interpreted as noun-attached as in sentence 3.2 (a cop with the revolver) or as verb-attached as in sentence 3.3, in which case it modifies the verb (seeing with binoculars). In contrast to other structurally ambiguous stimuli such as garden-path sentences, different prepositional phrase attachments do not involve different word forms or function words. Hence, any disambiguation cannot depend on syntactic information. Still, human readers are able to assign a unique meaning to such structurally ambiguous sentences with ease, relying on world knowledge to connect the semantic information provided by both the prepositional phrase itself with its preceding context in the most plausible way (e.g. revolvers are likely to be carried by cops and binoculars are likely instruments for seeing.). Note that sentence-final prepositional phrases are not rare or non-canonical. For example, in the structurally annotated TIGER corpus (see methods for details) we found about 43% of all prepositional phrases to be structurally ambiguous.

3.2 The spy saw the cop with the revolver.

3.3 The spy saw the cop with the binoculars.

Originally, structurally ambiguous sentences had been shown to lead to prolonged reading times at the disambiguating word (e.g. noun-attached PPs being read more slowly than verb-attached PPs). Based on these findings, Frazier had

proposed sentence comprehension to rely on an initial structural interpretation of the sentence driven by syntactic cues only and following certain rules such as the minimal attachment principle. According to the minimal attachment principle, the preferred structure is always the more shallow one (i.e. the one resulting in a minimal amount of nested dependencies). Therefore, according to minimal attachment the verb-attached reading of the PP is preferred already when encountering the preposition. In the case of a noun-attached phrase, subsequent words thus leads to the need for post-hoc structural reanalysis and as a consequence longer reading times (Rayner et al. 1983, Frazier and Rayner 1982). Frazier's early theory was quickly overturned in favour of a parallel (or cascading) processing model (McClelland and Kawamoto 1986, Van Den Brink and Hagoort 2004, Pulvermüller et al. 2009, Hagoort 2017) by several studies demonstrating the fast integration of non-syntactic cues early during online processing (Spivey-Knowlton and Sedivy 1995, Altmann and Steedman 1988, Taraban and McClelland 1988, Traxler and Tooley 2007, Mohamed and Clifton 2011). For the processing of ambiguous PPs, it has been shown that facilitated processing of verb-attachments is modulated by referential information imposed by the context (Altmann and Steedman 1988) as well as semantic content of the preceding verb (Spivey-Knowlton and Sedivy 1995). More concretely, Spivey-Knowlton et al. have shown that action verbs bias expectations towards verb-attachment while verbs referring to mental states (e.g. the spy hoped for ..) or perception can bias towards noun-attachment (Spivey-Knowlton and Sedivy 1995). The authors explain this by different types of verbs being associated with certain thematic roles to different degrees (e.g. action verbs occur with an instrument more often than perception verbs). As a consequence, reading time differences that have originally been interpreted to be a direct consequence of hierarchical structure building, could be reflecting predictions about upcoming semantic content instead.

In a more recent study, Boudewyn and colleagues argued against this alternative hypothesis of PP reading differences being caused by varying semantic predictions. They investigated the neural activity evoked by verb- and noun-attached prepositional phrases through event-related potentials (ERPs). In addition to the classically observed delay in reading times, their noun-attached stimuli evoked larger positive potentials around 600 ms (P600) (as compared to their verb-attached versions). Importantly, they showed that the amplitude of this P600 was reduced when noun-attached targets followed noun-attached primes

(Boudewyn et al. 2014). Boudewyn and colleagues are not the first ones to report structural priming effects. In fact, syntactic priming has been reported already some 35 years ago, showing that speakers are more likely to repeat a given syntactic structure in their utterances than to switch between two conceptually equal alternatives (Bock 1986). To evoke priming of hierarchical structure, researchers explicitly vary lexical information while keeping syntactic structure stable. More recent investigations indicate, however, that event structure (i.e. thematic roles) as well as lexical information can to a large degree account for many priming results and hence priming solely on the structural level has not been definitively proven yet (Ziegler et al. 2019). Other confounding factors that can evoke priming and are often contrasted along side syntactic structures are information structure, syntax-animacy mapping and rhythmic priming. Boudewyn et al. argue for their priming effect to be structural in nature based on the timing of their observed ERP effect. Differences in ERPs have been generally interpreted as neural markers for a difference in processing (for example more or less engagement of the underlying neuronal population). The P600, specifically, has been reported most often in the context of syntactic violations or anomalies. Hence, the authors interpret this priming effect to reflect facilitated structural processing of an originally dis-preferred structure. Still, ERP effects need to be interpreted with caution, since their relationship to underlying cognitive mechanisms is unclear. For example, recent computational cognitive models of language processing illustrate that ERP markers can be modelled as reflecting general update or error signals, without restricting them to any specific linguistic operation (Rabovsky et al. 2018, Fitz and Chang 2018).

In addition, most ERP research so far reflects only a one-sided measure of the neural code. Namely, the dominant analysis approach has been to treat ERPs as unidimensional point-estimates. Computing signal amplitude separately for a given channel and time point and averaged over trials, subjects and eventually space and time. As a consequence, such analyses can only detect univariate effects and are highly sensitive to subject-level variability. With the recent increase in computing power and developments of multi-variate pattern analysis (MVPA) we can now capture richer multidimensional information encoded across several channels or source points (Guggenmos et al. 2018, Norman et al. 2006). Through MVPA, researchers have been able to uncover additional task-relevant brain regions (Jimura and Poldrack 2012) and characterise the specific computations

needed for ambiguity resolution in more detail (Tyler et al. 2013). Furthermore, MVPA has the potential to be sensitive to distributed neural representations of the content whereas univariate methods have been thought to be most sensitive to the engagement of basic processing operations (Raizada et al. 2010, Mur et al. 2009, Okada et al. 2010). Although not every effect revealed through MVPA is necessarily indicative of an underlying distributed neural code (Davis et al. 2014), the technique has nonetheless been successfully used to reveal higher-level structure in the neural signal for domains other than language (e.g. for hierarchical motor sequences Yokoi and Diedrichsen 2019). MVPA might hence be better suited to target hierarchical structure building during language processing than previous univariate methods.

In this study, we revisit processing of structurally ambiguous PPs with the approach of MVPA in order to more directly tap into representations of hierarchical structure underlying language comprehension. In contrast to early psycholinguistic approaches we do not assume that noun or verb-attached prepositional phrases are processed differently from each other in the sense of one structure being more preferred over another. Rather we ask, whether it is possible to find a neural correlate of the hierarchical phrase structure of a sentence (i.e. neural patterns that distinguish between verb- and noun-attached PPs), given completely ambiguous syntactic cues.

## 3.2 Methods

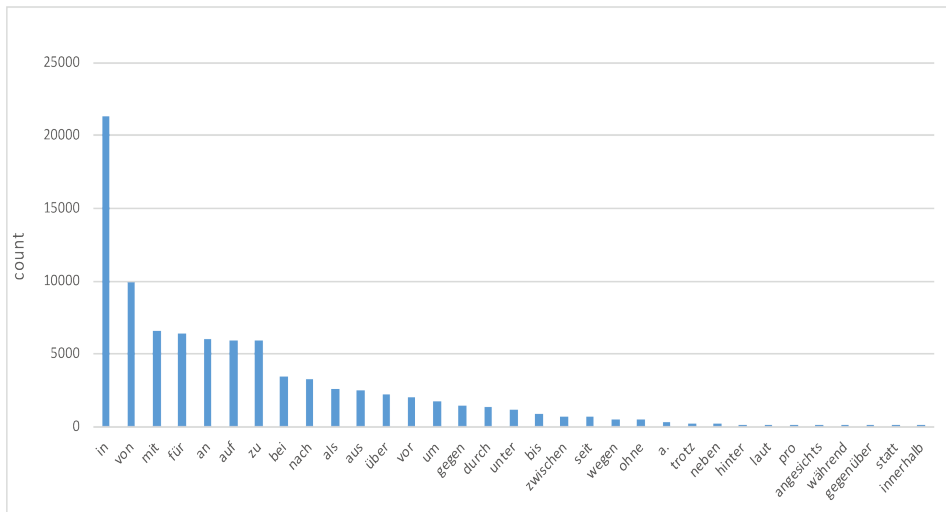
### Stimulus Material

#### Corpus Analysis

All stimuli were created in German. Since most of the previous literature had looked at prepositional phrases in English, we first conducted a corpus analysis to determine which German preposition will most likely be ambiguous with respect to structural attachment of the prepositional phrase.

For our corpus analysis we used the TIGER corpus, a manually annotated corpus of 40,000 German sentences (Brants et al. 2004). The corpus is available at [www.ims.uni-stuttgart.de](http://www.ims.uni-stuttgart.de) in both xml as well as conll09 format. We used the xml version for queries with the TIGERSearch Tool as well as the conll09 version for quick extraction of frequency statistics using the bash shell command `awk`. We





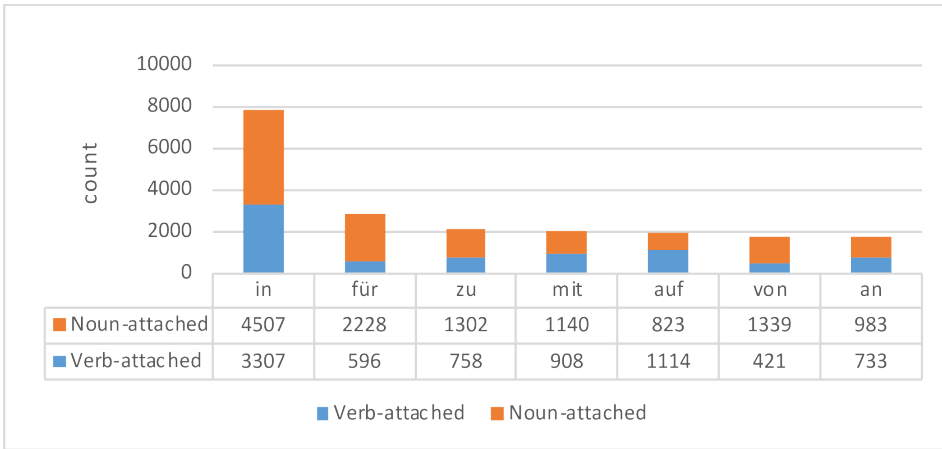
**Figure 3.1.: Tiger corpus frequencies per preposition.**

Total number of occurrence for the 33 most frequent prepositions based on the German "Tiger" corpus.

extracted separate frequency information per preposition and structure (noun-attached and verb-attached prepositional phrases) through the TIGERSearch software (see appendix for details on the TIGERSearch queries).

### Stimuli

Based on the corpus search, we selected the preposition "mit" (engl.: with) because it occurs with high frequency (Figure 3.1) and equally often within both noun- and verb attached phrases (Figure 3.2). We created a stimulus set of 100 sentence pairs in German. All sentences consisted of nine words each, a subject-verb-object structure in the main clause followed by a four word prepositional phrase including the preposition and a determiner-adjective-noun phrase. This sentence structure was syntactically ambiguous with respect to the attachment site of the prepositional phrase. Within a given pair, the same prepositional phrase was presented while the sentence context leading up to it was manipulated. Based on the combined semantic information of the sentence context and the prepositional phrase, the interpretation of the most plausible attachment could be disambiguated. To steer the preferred attachment interpretation, we manipulated the sentence context in two ways. In half of the sentence pairs we varied the main verb,



**Figure 3.2.: Tiger corpus attachment proportions per preposition.**

Frequency of verb- and noun-attached phrase constructions (not restricted to sentence final PP) for the seven most frequent prepositions in the corpus.

we call this the verb condition (examples 3.4 & 3.5). Sentence pairs in the verb condition were constructed such that the noun in object position could potentially be modified by the PP but did not have a particularly strong semantic association with the PP internal noun. By presenting these sentences with a verb for which modification through the PP internal noun was either allowed or forbidden (or at least unlikely), a verb-attached interpretation could either be encouraged or prevented respectively. In the other half of the sentence pairs, we exchanged agent and patient identity across the two sentences. In the following, I will refer to this as the role condition (examples 3.6 & 3.7). For sentence pairs in the role condition the two nouns preceding the PP had a varying degree of semantic association to the PP internal noun while the verb was held stable with a mild semantic association to the PP internal noun and optional modification through a PP. This lead to a noun-attached interpretation if the more strongly associated noun occurred in object position (the noun immediately preceding the PP) but to a verb-attached interpretation when it occurred in subject position. In both the role and the verb condition, each verb was repeated exactly two times across all sentences. We explore difference between verb and role conditions in the behavioral data but collapse across both conditions when analysing the neural data. Finally 100 filler sentences with varying syntactic structure were created.

### **Verb condition**

3.4 Die Partei besitzt eine Untergruppe mit einigen Argumenten

*engl.: The party has a subgroup with questionable arguments.*

3.5 Die Partei überzeugt eine Untergruppe mit einigen Argumenten *engl.: The party convinces a subgroup with questionable arguments.*

### **Role condition**

3.6 Das Kind verängstigt das Insekt mit dem giftigen Stachel

*engl: the child frightens the insect with the poisonous sting*

3.7 Das Insekt verängstigt das Kind mit dem giftigen Stachel

*engl: the insect frightens the child with the poisonous sting*

### **Pre-test**

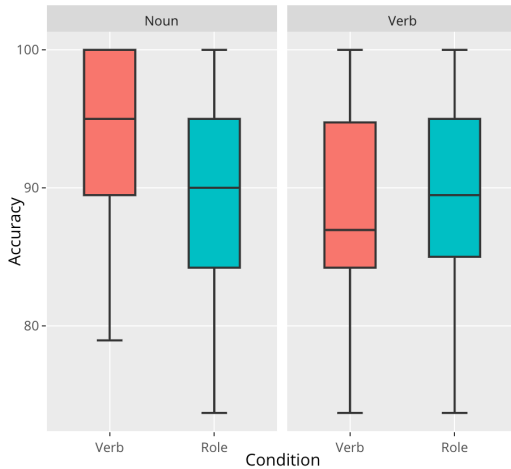
For the majority of the sentences, the overall semantics licensed both PP attachments, even if they were constructed such that one attachment should be perceived as more plausible. To verify that our manipulation evoked the intended sentence interpretation we pre-tested all stimuli via an online questionnaire, created with the survey tool Limesurvey (Limesurvey GmbH 2012). During this online questionnaire, 62 native German speakers with a mean age of 25 (range 19-33) judged for each stimulus-sentence whether it contained a verb- or noun-attached prepositional phrase and how plausible they found the sentence (on a scale from 1 to 5). All subjects gave informed consent prior to filling in the survey and received financial reimbursement. Based on the answers we selected 200 sentences out of a larger set of 469 sentences according to criteria described in detail below (see Table 3.2 and 3.3 in Appendix for the final selection of sentences as used in the MEG experiment).

First, subjects were instructed about the difference in attachments. This was done using unambiguous stimuli and a non-formal intuitive explanation like “In the verb-attached case the prepositional phrase says something about the verb”. Subjects were then asked to formulate the rule to distinguish the two attachments in their own words and were presented with four unambiguous practice items. Finally, they would read 80 to 100 sentences one by one and for each sentence decide between verb- or noun attachment. Ten seconds after a sentence appeared

on screen a pop-up window encouraged subjects to answer faster. This time limit was chosen to force subjects to answer intuitively. However, many subjects would need more time on certain trials. After selecting their answer they could continue with the next item at their own pace. Half way through the questionnaire subjects were encouraged to take a longer break if needed. The stimulus list was split up into three parts to keep the duration of each survey to about 30 minutes. Each subject saw one of the possible lists in a pseudo-random order, so that sentences from the same pair were at least four items apart. Three subjects were excluded either based on poor performance on the practice items (less than three correct), because their average reaction time diverged extremely from the average (greater than 1.5 times the interquartile range) or because they had less than 60% correct answers to those sentences that were semantically completely unambiguous.

The survey results were analyzed using R version 3.6.3 and the lme4 package for linear mixed-effects models (Bates et al. 2015). Pairs of sentences were selected if both received at least 74% of answers consistent with the intended attachment. With more than 74% of answers being consistent with the intended attachment we can exclude the alternative hypothesis of random behavior at an alpha level of 0.05 given a binomial distribution and 20 data samples per item. The selection was made so that every verb was repeated exactly two times and there were equal amounts of sentences in both verb and role condition.

On pre-test results for the final selection of sentences, we used a generalised linear mixed effects model (GLMM) with a logit link function fit by maximum likelihood to examine the relationship between accuracy (i.e. percentage of answers in line with our expectations), reaction time, plausibility ratings (on a scale of 1 to 5), context manipulation (verb condition or role condition) and attachment type (verb- or noun-attached). A mixed logit model appropriately accounts for binomial response variables (Jaeger 2008), in our case hits or misses (correctly identifying an attachment according to intended sentence meaning or not). The model thus allowed us to test whether there were systematic differences in processing noun- or verb-attached sentences, as well as systematic differences between our different context manipulation conditions while controlling for between-subject variance. We specified accuracy (hit or miss) as the dependent variable and reaction time, plausibility rating, and context condition as fixed effects. Additionally, the model included random-effect terms for items (intercept only) and subject (intercept and slope). The model was fully saturated with all two-way interaction effects.

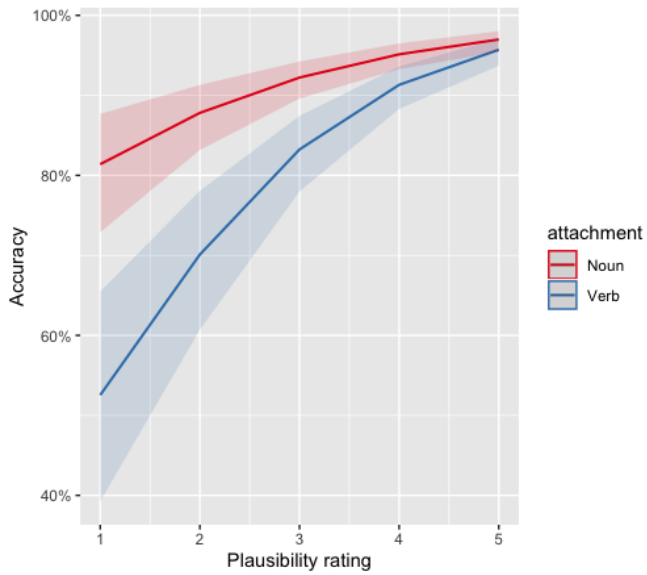


**Figure 3.3: Pre-test proportion of correct responses averaged across all subjects.**

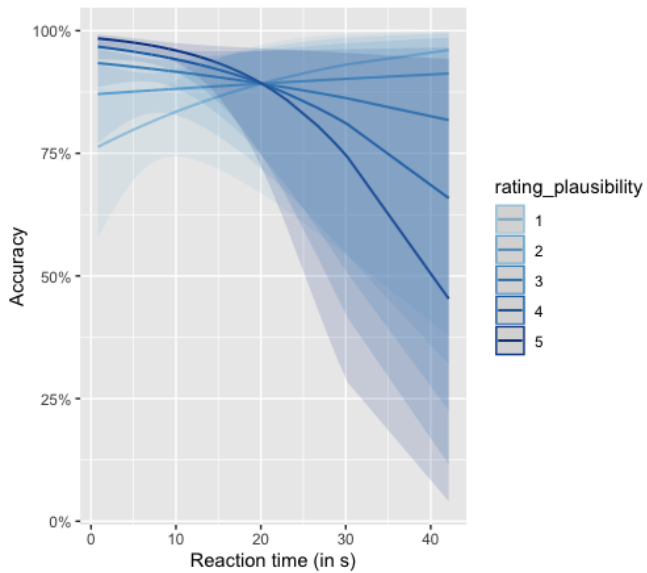
Accuracies are plotted separately for verb condition (red), role condition (blue) and for noun-attached sentences (left-most graphs) and verb-attached sentences (rightmost graphs).

GLMM results indicate a significant effect of attachment type and plausibility, with factor level contrasts revealing that subjects were more often correct for noun-attached items (see Figure 3.3) and high plausibility ratings led to high accuracy. There was a significant Attachment type  $\times$  Plausibility interaction. Factor level contrasts revealed that the effect of high plausibility leading to high accuracy was stronger for verb-attached sentences than noun-attached sentences (see Figure 3.4). The context manipulation effect was not significant and only the interaction Context Manipulation  $\times$  Attachment was significant, indicating that only for noun-attached sentences were items more often correctly interpreted in the verb condition compared to the role condition (see Figure 3.3). Finally, the interaction of Reaction Time  $\times$  Plausibility was significant. As illustrated in Figure 3.5, high plausibility ratings only lead to higher accuracy if reaction times were fast. In summary, whether sentences were constructed to fit the verb or the role condition did not lead to large differences in accuracies, although sentences in the verb condition were slightly biased towards a noun-attached interpretation. Most of the items used in the experiment received a plausibility rating of higher than 3 on average with only four items with an average rating below 3 and verb-attached sentences receiving on average slightly higher plausibility ratings.

**Figure 3.4:** Interaction between plausibility ratings and attachment type. Mean Accuracy per plausibility rating is plotted for noun-attached (red) and verb-attached (blue) items.



**Figure 3.5:** Interaction between reaction times and plausibility ratings. Mean Accuracy per reaction time is plotted for different plausibility ratings. The higher the plausibility the darker the color.



## Experiment

10 Native German speakers (mean age = 22 years, 3 male) were seated in a magnetically shielded room and read sentences word-by-word while their neural activity was recorded using Magnetoencephalography (MEG). All subjects gave informed consent prior to filling in the survey and received financial reimbursement or credits. All stimuli were presented using the Presentation software (Version 16.0, Neurobehavioral Systems, Inc). Sentences were presented in pseudo-random order and word-by-word in four blocks with self-paced pauses in between blocks. In 25% of all trials a comprehension question would follow the sentence. Comprehension questions were either directed at identifying the agent or patient of the sentence (“Who has the bucket” or “Who is being carried”) or they would target the semantic dependency of the prepositional attachment (example question following 3.5: “Who has the questionable arguments”). The question was presented together with two answers, one on the left and one on the right side of the screen. Subjects indicated which answer they chose by pressing a button with their index finger corresponding to the position of the answer on the screen. The comprehension questions were meant to ensure that subjects were engaged and attentive during the task and that they fully parsed the presented sentences on both a semantic as well as structural level. Prior to the main experiment subjects received four practice trials to familiarise themselves with the pace of the presentation. Words were presented sequentially on a back-projection screen, placed in front of them (vertical refresh rate of 60 Hz) at the centre of the screen, in a white font, on a black background. Each word was separated by an empty screen for 200 ms and the final word of each sentence was followed by a 2000 ms blank screen. Duration of each word on screen was 392 ms on average and varied with word length with a minimum duration of 300 ms and maximum duration of 500 ms (formula:  $300 \text{ ms} + \text{number of letters} * 1000/60$ ). The inter-sentence interval was jittered between 500 and 1000 ms. Within two weeks after the MEG experiment, subjects filled out a questionnaire rating each stimulus sentence as either noun- or verb attached and as plausible on a scale from 1 to 5. This questionnaire was the same as the one used for the pre-test but contained only those stimuli that had been used during the MEG experiment.

MEG data were collected with a 275 axial gradiometer system (CTF). The signals were analog low-pass-filtered at 300 Hz and digitized at a sampling frequency of 1,200 Hz. The position of the subject’s head was registered to the MEG-sensor

array using three coils attached to the subject's head (nasion, and left and right ear canals). Throughout the measurement, the head position was continuously monitored using custom software (Stolk et al. 2013). During breaks the subject was instructed to reposition to the original position if needed. Subjects were able to maintain a head position within 5 mm of their original position. Three bipolar Ag/AgCl electrode pairs were used to measure the horizontal and vertical electrooculogram and the electrocardiogram. In addition to the brain signal, we acquired T1-weighted magnetic resonance (MR) images of each subject's brain using 3 Tesla Siemens PrismaFit and Skyra scanners. All scans covered the entire brain and had a voxel size of  $1 \times 1 \times 1 \text{ mm}^3$ . Finally, we recorded the subject's head shape with the Polhemus for better co-registration of MEG and anatomical scans.

## Preprocessing & Source reconstruction

Data were pre-processed using the Fieldtrip toolbox in MATLAB (Oostenveld et al. 2011). For the decoding analysis the Donders machine learning toolbox (Van Gerven et al. 2013) was used in combination with custom-made MATLAB scripts. The data were segmented into epochs around word onset with a 200 ms pre-stimulus period. To detect muscle artifacts, data was bandpass filtered between 110 Hz and 140 Hz and the trials with large variance were excluded upon inspection (less than 4% of all critical trials). Data was filtered between 0.1 Hz and 40 Hz. Independent component analysis (ICA) was used to remove artifacts stemming from the cardiac signal and eye blinks. For each subject, the time course of the independent components was correlated with the horizontal and vertical EOG signals as well as the ECG signal to identify and subsequently remove contaminating components.

We used linearly constrained minimum variance beamforming (LCMV) (Van Veen et al. 1997) to reconstruct activity onto a parcellated cortically constrained source model. For this, we computed the covariance matrix between all MEG-sensor pairs as the average covariance matrix across the cleaned single trial covariance estimates. This covariance matrix was used in combination with the forward model, defined on a set of 7842 source locations per hemisphere on the subject-specific reconstruction of the cortical sheet to generate a set of spatial filters, one filter per dipole location. Individual cortical sheets were generated with the Freesurfer package (Dale et al. 1999, version 5.1) ([surfer.nmr.mgh.harvard.edu](http://surfer.nmr.mgh.harvard.edu)). The forward model was computed using FieldTrip's singleshell method (Nolte



2003), where the required brain/skull boundary was obtained from the subject-specific T1-weighted anatomical images. We further reduced the dimensionality of the data, by grouping source points into 374 parcels, using a refined version of the Conte69 atlas. These parcels were used as searchlights in the subsequent analyses.

## Multivariate decoding analysis

### Gaussian Naïve Bayes

We trained a Gaussian Naïve Bayes classifier (GNB) (Mitchell and Others 1997) to identify cognitive states associated with underlying sentence structure from the pattern of brain activity evoked by reading the final word of a prepositional phrase. The GNB is a generative classifier that models the conditional probability  $P(x_j|Y_i)$  of signal amplitude  $x$  (at a given sensor/voxel  $j$ ) given that the stimulus is of a class  $Y_i$  (noun- or verb-attached prepositional phrase) using a univariate Gaussian and assuming class conditional independence. The mean and variance of this distribution is estimated on a subset of the trials (training set). The remaining data (test set) is then classified as the class  $Y_i$  whose posterior probability  $P(Y_i|x)$  is maximal among all classes. The corresponding classification rule is:

$$Y \leftarrow \underset{y_j}{\operatorname{argmax}} P(Y = y_j) \prod_j P(X_j|Y = y_j) \quad (3.1)$$

Classification results were evaluated using 20-fold cross-validation, so that accuracy was always based on test data that were disjoint from the training set. 20 folds were chosen for a good balance between amount of training data per fold and computational speed. Accuracy was estimated as the percentage of correctly classified trials across all folds. Classifiers were trained using a sliding time-window approach, where for each time-point, MEG data from all sensors and all time-points  $\pm 50$ ms were concatenated into a single vector (length = vertices  $\times$  time-points). We also trained the same classifier on source-reconstructed data using a spatial searchlight approach in addition to the sliding time-window. The searchlight procedure followed the parcellation of the cortical sheet. For each parcel and time-point a classifier was trained on source data of all vertices within

that parcel, while concatenating across all time-points within a sliding window of width 100 ms.

All parameters chosen for the classification analysis were manually optimised based on accuracy of an orthogonal classification task, namely to distinguish neural patterns evoked by either reading the main verb or the second noun (object noun) of the sentences. Decoding which of these different word classes was being presented robustly resulted in accuracies significantly higher than chance performance. Within our stimulus design, word class was confounded with ordinal word position in the sentences. Therefore, we conducted a control analysis on the same ordinal word positions within only filler items (where sentence structure varied and therefore nouns and verbs did not always occur at the same sentence position). This control analysis did not yield comparably high decoding accuracies. We compared the performance of the verb-noun classifier given different sliding time window widths (50 ms, 100 ms or 200 ms) and feature transformations (concatenating vs averaging over time dimension, feature selection, orthogonalisation and feature reduction through principal component analysis (PCA), gaussianisation).

PCA transforms the data into linearly uncorrelated components, ordered by the amount of variance explained by each component. Using these uncorrelated components as features can improve the decoding performance of classifiers such as GNB, which assume no feature covariance (Grootswagers et al. 2017). Furthermore, PCA allowed our feature selection to be based on a data-driven approach by keeping only a subset of components that explain highest variance. We observed that both orthogonalising of features (sensor-time points) using PCA and feature reduction by restricting training to the first 60 components only, boosted classification accuracy. Further feature selection based on signal strength (selecting features based on largest difference in means between classes) did not improve accuracy beyond the effects of feature reduction based on PCA. Gaussianisation of the sensor-level data prior to classification analysis or broadening the training time window did not yield large differences in performance. Based on these comparisons we then continued to train the classifier on the noun- vs. verb-attachments with the optimal parameters.

## Representational similarity analysis

Prepositional phrase attachment is interpreted based on the semantic information given the context preceding the phrase. We therefore predicted that there might be reactivation of this semantic information (i.e. those semantic features that most strongly influence the attachment) after the disambiguating sentence-final word. We tested this hypothesis through representational similarity analysis (RSA) (Kriegeskorte et al. 2008), representing semantic content by means of a high-dimensional word-embedding vector (semantic vectors). For the word-embeddings we relied on pre-trained models published by facebookresearch<sup>1</sup> which had been trained on German Wikipedia using fastText (Bojanowski et al. 2017; Grave et al. 2018).

First, we ensured that the semantic information captured by the word-embeddings is also encoded in the neural signal. We extracted all segments of neural data time-locked to each word presented and further restricted the selection to either content words only for this analysis or sentence-final words (as described in detail below). We then generated pairwise similarity measures between those words by computing the euclidean distance between their corresponding word-embedding vectors (semantic similarity model). Repeated presentations of the same word were treated as separate words (i.e. not averaged across). In the same way, we computed pairwise similarity measures for the corresponding segments in the neural signal, i.e. the pairwise neural similarity during reading of the same words. Words that were not present in the vocabulary of the pre-trained embeddings were excluded from both semantic model and neural data, which left 387 trials in total. Neural similarity was computed based on a moving searchlight by concatenating all samples within a 100 ms time-window and across source locations within a given parcel, and this was repeated for all parcels and shifting time-windows (between word onset and 800 ms post onset) with an 80% overlap in time. Finally, semantic similarity and neural similarity were correlated (Spearman correlation) at each searchlight position. This resulted in a map indicating when and where neural activity reflected semantic information about the perceived words.

Crucially, we then generalised this RSA to the post-sentence phase, when subjects were reading the final, disambiguating word. For this, we re-computed the neural similarity, this time based on neural activity evoked by the final word. For

---

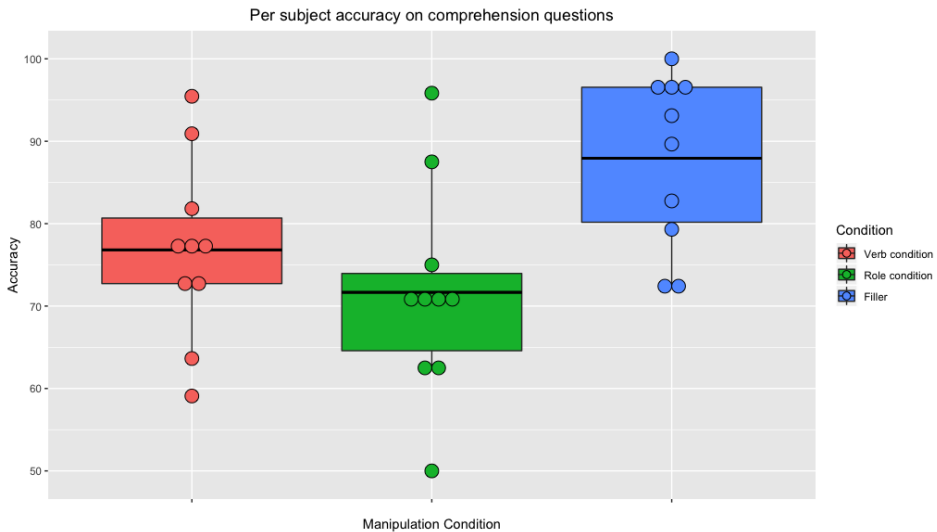
<sup>1</sup><https://ai.facebook.com/tools/fasttext>

each Verb-attached and each Noun-attached PP instead of the word-embedding of the final noun we assign the word-embedding vector of the preceding verb or noun respectively (i.e. of the most plausible attachment points). We then recomputed the euclidean distance between word-embedding vectors for all trial pairs, which now expresses for each sentence pair the semantics similarity with respect to the disambiguated attachment sites. Any significant correlations between the neural similarity and the attachment site semantic similarity indicate when and where neural patterns evoked by reading the final noun are also encoding (i.e. reactivate) information about the preceding verb or noun respectively.

### Significance testing of decoding accuracy

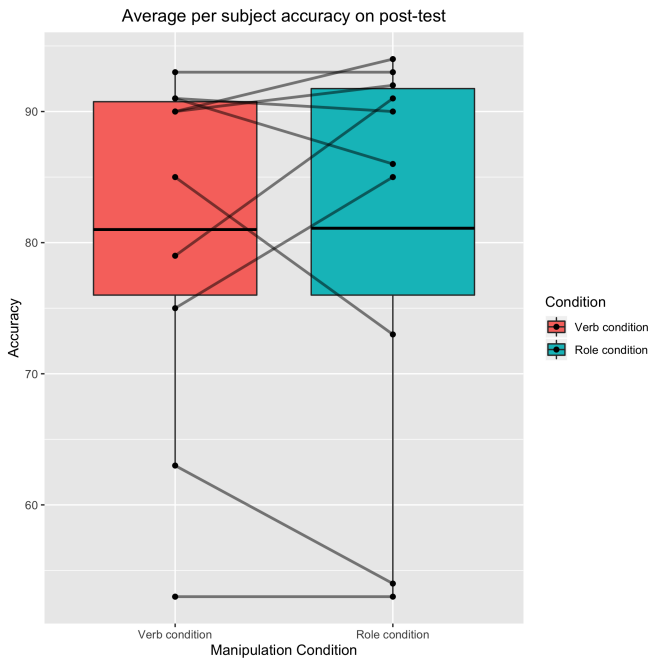
When evaluating significance of group-level accuracy differences between two classifiers (GNB vs. logistic regression; part-of-speech classifier vs. word position control) we relied on non-parametric permutation testing (Maris and Oostenveld 2007), randomly swapping observed accuracy between classifiers. For statistical evaluation of the GNB classifier against chance level we relied on information prevalence inference (Allefeld et al. 2016) based on subsampling of single-subject permutations. Prevalence inference tests the significance of above-chance accuracy in the majority of subjects given the permutation distribution at an alpha level of 0.05. Permutation tests are preferred over traditional tests against theoretical chance level, given that the small amount of trials (typical for neuroimaging studies) will lead to larger cross-validation errors (Varoquaux 2017). Therefore, we computed null-distributions on randomly re-labeled data for the GNB classification task. For the binary classification task we randomly selected half of the items per category (either attachment type of part of speech) and switched their labels in order to maintain an equal amount of items per class. For analyses conducted on the source-reconstructed data we used one fixed set of permutations of the observations for each searchlight to preserve spatial correlations. The procedure of generating a permutation and subsequent classification/prediction using permuted labels/semantic vectors was repeated 100 times per subject.

To evaluate statistical significance of the correlation values resulting from the RSA analysis, we used nonparametric permutation tests against a baseline of zero, including cluster-based correction for multiple comparisons across time and space.



**Figure 3.6.:** Accuracy of comprehension questions for each subject per manipulation condition.

Accuracy across subjects depicted separately for each manipulation condition: Verb condition (left, red), role condition (middle, green) and filler items (right, blue). Individual subject accuracies are plotted as dots.



**Figure 3.7:** Accuracy of attachment rating for each subject per manipulation condition.

Average accuracy is plotted separately for verb condition (left, red) and role condition (right, green). Individual subject accuracies (percentage of items correctly classified) are plotted as black dots.

## 3.3 Results

### Behavioral

In the MEG experiment, all subjects had higher than chance level performance on answering the comprehension questions. On average they gave 77% correct answers on sentences from the verb condition, 72% correct answers for the role condition and 88% correct answers on filler sentences. While performance on the filler items was above chance for all subjects, some subjects performed at chance for questions from the verb and role conditions (see Figure 3.6). Since correct answers to target items depended on the interpretation of the prepositional phrase attachment, this suggests, that some subject's attachment interpretations differed from the norm (as determined by the pre-test). Within a week after the MEG experiment, each subject had filled in an online post-test, explicitly rating all stimulus sentences as either noun or verb attached (following the methods from the pre-test). Average accuracy across subjects on this post-test did not differ between conditions (verb and role condition both 81% correct) and subjects interpreted the sentences mostly as intended. Except for two subjects, who performed close to chance, subjects had a minimum accuracy of 79% (see Figure 3.7).

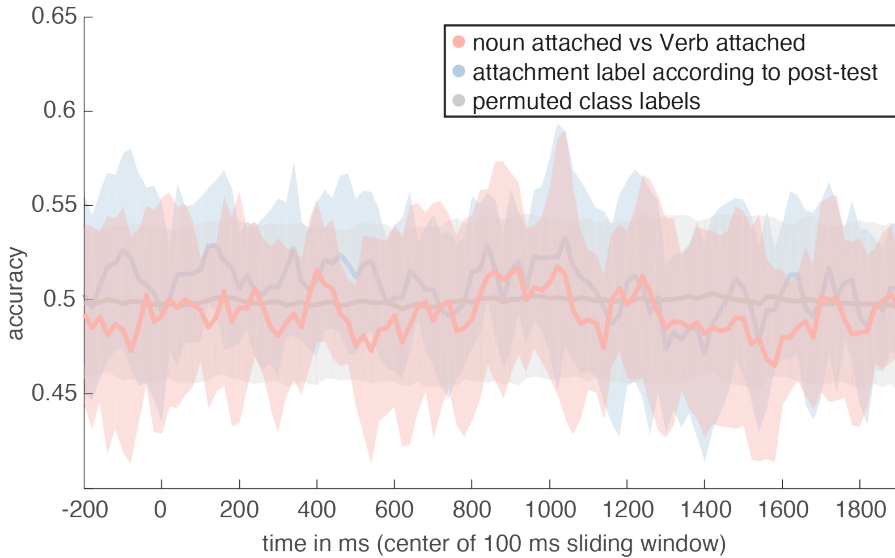
### Multivariate pattern analysis

#### 2-way classification Noun-attached vs Verb-attached

Our main analysis of interest, the 2-way classification of different phrase structure (Noun attachment vs. Verb attachment) did not reach above chance-level accuracy at any time window up to 2 seconds after onset of the final word of a sentence. We observed this null-finding both, when items were labeled according to the general pre-tested attachments, but also when items were labeled according to subject-specific post-tests (see red and blue graphs respectively in Figure 3.8).

#### 2-way classification Noun vs Verb

The 2-way classification on whether the currently seen stimulus was a verb or a noun based on sensor-level MEG data reached a maximal average accuracy (across subjects) of 67% at 160 ms after word onset and was significantly more accurate as compared to the word position classifier ( $p=0$ , cluster-corrected permutation



**Figure 3.8.: Attachment classification in sensor space.**

Accuracy of a Gaussian Naive Bayes classifier is plotted for two-way classification of attachment type (noun-attached vs verb-attached). Accuracy is shown for both, a classifier trained on items labeled according to coherent interpretations of sentences during pre-test (red) and a classifier trained on items labeled according to subject-specific post-test interpretations (blue). Observed accuracy was tested against a baseline performance estimate generated by repeatedly classifying data after permuting labels (grey).

tests) up until 460 ms after word onset (see Figure 3.9). Note that classification accuracy is already significantly above chance before the onset of the noun/verb. This is due to the fact that nouns were always preceded by a determiner and verbs by a noun, effectively turning the baseline period into a determiner vs. noun classification sample. PCA transformation of the data led to higher classification accuracy as compared to training on the raw features. Additional feature selection based on class means did not lead to further increases in accuracy (see Figure 3.10). Training the classifier on moving windows of length 100 ms not only was more efficient in terms of computation time but also led to higher classification accuracies as compared to training the classifier per time point (see Figure 3.11). Concatenating sensors of all time points mostly lead to slightly higher accuracies as compared to averaging over time points before training.

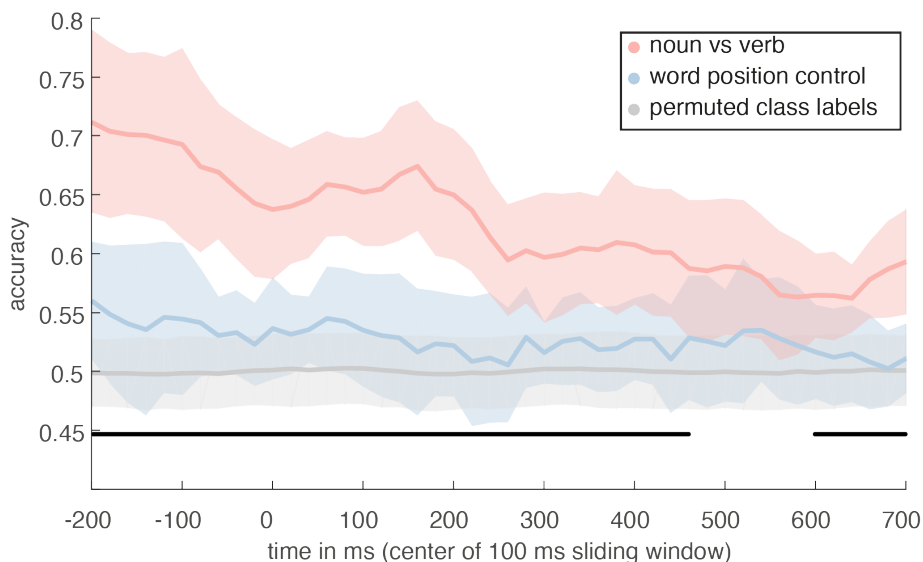
Besides Naive Bayes, we also tested different classification algorithms, i.e. support vector machines and logistic regression. None of these resulted in higher classification accuracies for the classification of nouns vs verbs (see Figure 3.12) as compared to Naive Bayes. Logistic regression performed better than Naive Bayes for the classification of determiner vs noun.

Given that nouns and verbs have some systematic orthographical differences in German, we wanted to know whether classification success was mostly driven by low-level visual cortex. To investigate this, we source-reconstructed the MEG data and trained several classifiers on different regions across the cortex (searchlight approach). While classification accuracies were overall lower than those observed based on the sensor-level data, they were highest in occipital areas (see Figure 3.13). However, classification was also significantly above chance in more anterior cortical areas. With increasing time since word onset, classification accuracy increased as well in more anterior, bilateral occipito-temporal areas (see Figure 3.13 middle panel for Brodmann area 37). Between 340 ms and 540 ms, higher level areas like left inferior central and inferior frontal areas contain information about the noun-verb distinction (see Figure 3.13 lower panel for Brodmann area 43).

### Generalization over time

Concerning the hypothesis that combinatorial processes involve a reanalysis of the to be combined parts, we tested whether after the onset of the final word of the sentence (the word which disambiguated the structural attachment of the prepositional phrase) the encoded information of the preceding noun or verb would be reactivated in the presence of either a noun- or verb-attachment respectively. We first investigated whether there was a reactivation of morphosyntactic information (part of speech) by generalising the 2-way classification trained on brain data measured during reading of noun and verbs preceding the prepositional phrase to the period following the final word of the sentence. Even though the final word was always a noun we hypothesised that only verb-attached prepositional phrases would in addition lead to verb-like activity patterns following the final word. However, contrary to our hypothesis the classifier trained on nouns and verbs in the context did not accurately classify the post-sentence period of verb-attached prepositional phrases as more verb-like (see Figure 3.14).





**Figure 3.9.: Part-of-speech 2-way classification in sensor space.**

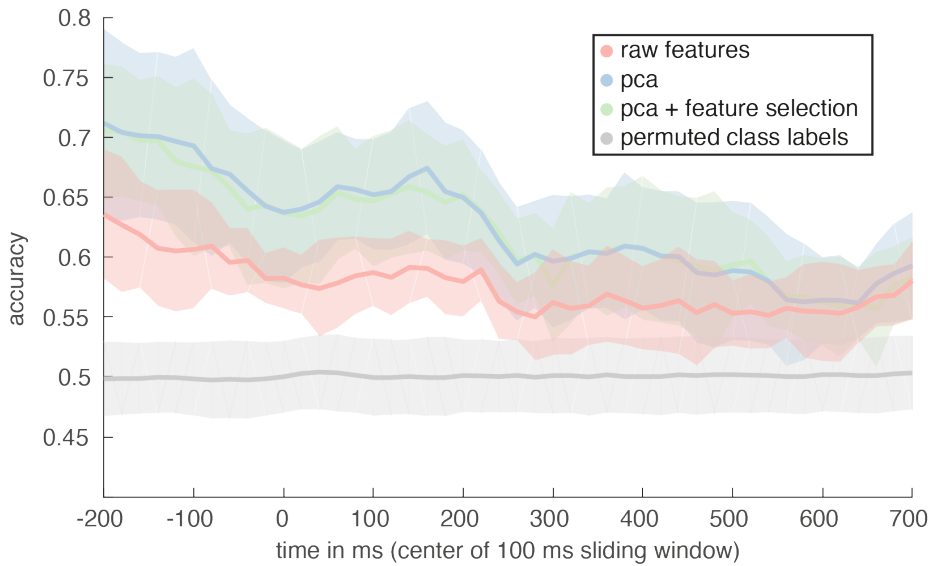
Accuracy is plotted for part-of-speech classification (nouns vs verbs) using Gaussian Naive Bayes (red) and for classification of word position in filler sentences (blue, varying part-of-speech categories). Black lines indicate when part-of-speech classification is significantly higher as compared to classification on filler items. In addition, a chance performance distribution generated by repeatedly classifying data after permuting labels is depicted in grey.

## RSA

For our stimuli, the interpretation of a prepositional phrase attachment was purely driven by semantic content. Therefore, we might also expect any reactivation to occur in the form of semantic information. We therefore tested whether at the time of disambiguation, any of the semantic information of preceding context would be reactivated. Specifically, we expected the semantics of the verb to be more strongly activated at the end of a verb-attached prepositional phrase and the semantics of the noun to be more strongly activated at the end of a noun-attached prepositional phrase.

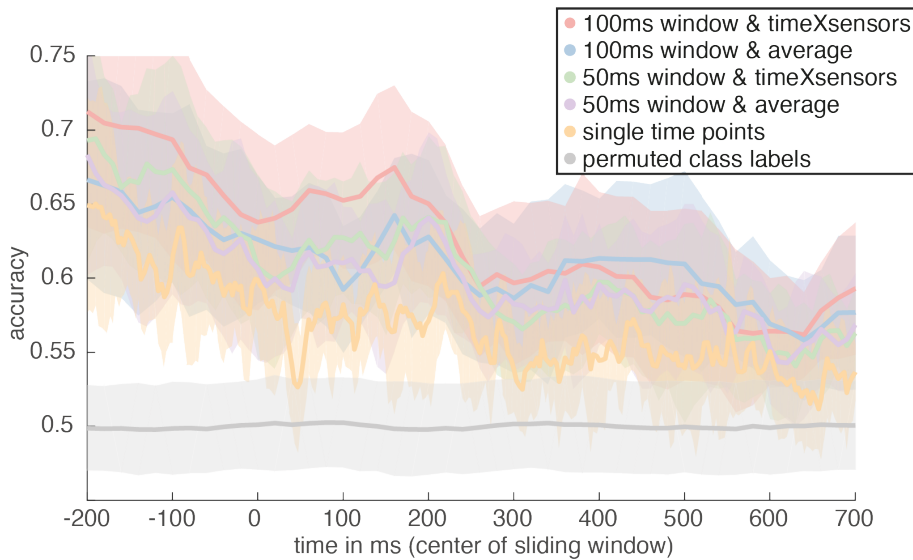
Our RSA revealed significant correlations between a model of the trial-by-trial similarity derived from word embeddings and the pairwise similarity derived from neural data evoked by the corresponding words (see Figure 3.15). Activity patterns that correlated with semantic similarity first emerged in a window from

380 ms to 480 ms in superior parietal cortex. Between 440 and 600 ms after word onset, semantic information was represented more extensively across parietal, temporal and occipital regions. Areas in which activity patterns significantly correlated with semantic similarity included posterior parietal cortex, somatosensory cortex, angular gyrus, fusiform gyrus, auditory cortex and posterior parts of the superior temporal gyrus. Late after onset, from 560ms to 720ms only areas in the ventral occipital lobe remained significantly correlated. When we generalised the RSA to the final word of the sentence, however, there was no significant correlation with semantic similarity in any brain area and hence no evidence for semantic reactivation.



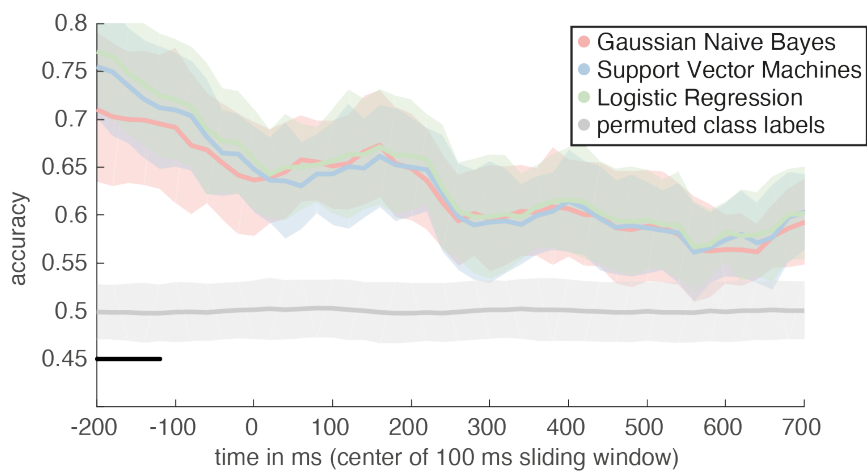
**Figure 3.10.: Feature transformation for 2-way classification.**

Accuracy is plotted for part-of-speech classification (nouns vs verbs) using Gaussian Naive Bayes and different feature reduction choices. Line plots represent the mean accuracy across all subjects and shaded areas represent its standard deviation. We first select evoked neural data from a 100ms (moving) time window and concatenate across all sensors and time points within that window, such that each sensor  $\times$  time point equals one feature. We compare performance of a classifier trained on either the original features (red), on a dimensionality reduced sensor space after selecting only the first 60 components using principal component analysis (PCA, blue) or on a reduced feature space using PCA as well as further only selecting the 150 sensor  $\times$  timepoints with the largest difference in class means (green). A baseline performance estimate was generated by repeatedly classifying data after permuting labels (grey). While feature space reduction through PCA improved classification accuracy, feature selection based on class means did not yield further improvements.



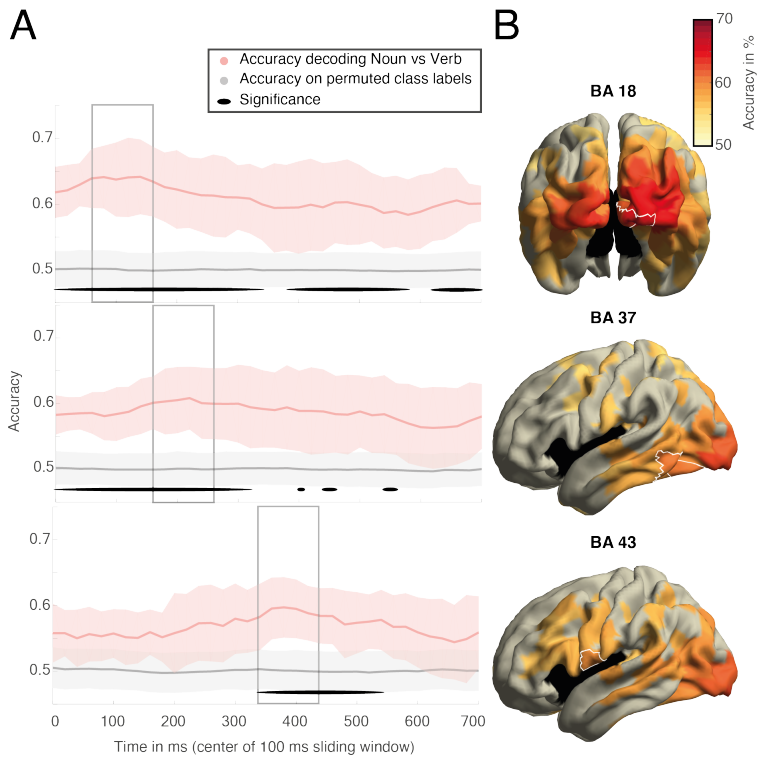
**Figure 3.11.: Time dimension for 2-way classification.**

Accuracy is plotted for part-of-speech classification (nouns vs verbs) using Gaussian Naive Bayes and different options for how to treat time. Line plots represent the mean accuracy across all subjects and shaded areas represent its standard deviation. A baseline performance estimate was generated by repeatedly classifying data after permuting labels (grey). Our moving window approach with window width of 100ms (red & blue) is most efficient in terms of computational time needed. On top of that, reducing the width of the window to 50 ms (green & purple) or even computing a separate model per time point (yellow) did not yield better classification performance. Further, for a window width of 100ms averaging over time points before training the classifier (blue) yielded lower accuracy as compared to concatenating across sensors and time points (red).



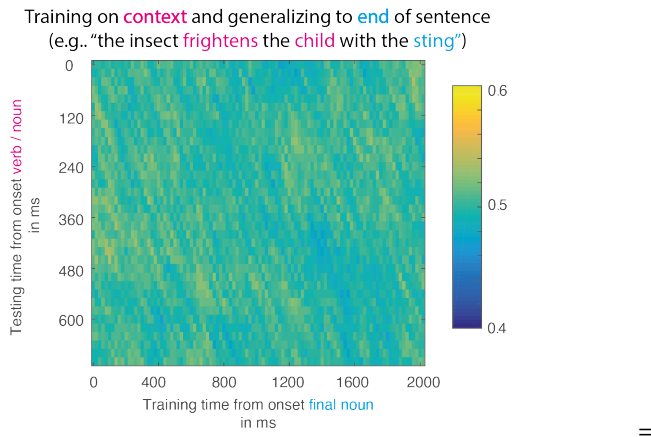
**Figure 3.12.: Comparison of different classification algorithms.**

Accuracy is plotted for part-of-speech classification (nouns vs verbs) using three different linear classifier: Gaussian Naive Bayes (red), support vector machines (blue) and logistic regression (green). Line plots represent the mean accuracy across all subjects and shaded areas represent its standard deviation. A baseline performance estimate was generated by repeatedly classifying data after permuting labels (grey). Significant differences in accuracy between different classifiers is indicated by a black bar.



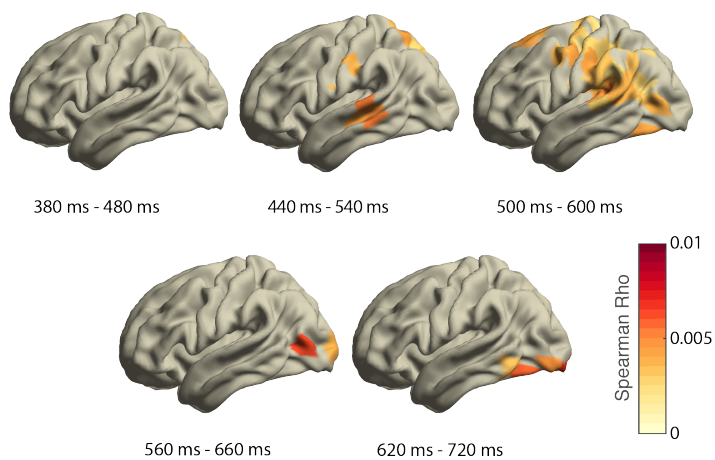
**Figure 3.13.: Part-of-speech 2-way classification in source space.**

Panel A: Accuracy is plotted over time for part-of-speech classification (nouns vs verbs) using Gaussian Naive Bayes (red). Observed accuracy was tested for significance (prevalence statistics, significant time points marked with black line) against a baseline performance estimate generated by repeatedly classifying data after permuting labels (grey). The upper, middle and lower panel display the mean accuracy over time for right occipital parcels (BA 18), left occipitotemporal parcels (BA 37) and left sub-central parcel (BA 43) respectively. Panel B: Cortical maps show the spatial patterns of classification accuracy, masked for significance. White contours outline the parcels for which time-courses are plotted in panel A respectively. Cortical maps contain averaged accuracies over the time-windows defined by the grey boxes.



**Figure 3.14.: Generalised classification accuracy for part-of-speech.**

We trained a classifier to distinguish between nouns and verbs based on the neural data evoked by reading one or the other. While training this classifier on a moving time window starting at onset of the noun/verb, we then tested whether the learned weights would generalise to data recorded while reading the end of the corresponding sentence. To illustrate this on a specific stimulus example, on a sentence like 3.7 “The insect frightens the child with the poisonous sting”, we would train the classifier on distinguishing activity evoked by “frightens” from activity evoked by “child” but we would test the classifier on activity evoked by “sting”. Given that this sentence contains a verb-attached preposition, the correct label for the classifier to identify would be “verb”, regardless of the final word always being a noun. Color codes for classification accuracy at any given training-by-testing time tile. Generalised classification accuracy is not significantly above chance-level at any time point.



**Figure 3.15.:** Searchlight RSA analysis on semantic information as measured by word embeddings.

Cortical maps show the spatial patterns of correlations with the semantic similarity model (masked for significance) averaged across several time windows. Colour codes strength of correlation.



## 3.4 Discussion

In this study we applied MVPA to probe the neural signal for hierarchical structure building during online reading of structurally ambiguous sentences. Subjects read sentences containing verb-attached and noun-attached prepositional phrases ambiguous with respect to their attachment. We successfully applied a Naive Bayes classifier to classify part-of-speech information of the current stimulus from the multidimensional evoked neural activity. We also successfully extracted neural patterns encoding semantic information of content words as subjects were reading them, through modelling the pairwise semantic similarity structure of all word pairs (RSA) with corpus-extracted word-embeddings. However, none of these measures revealed encoding of different underlying hierarchical phrase structure for verb- vs noun-attached sentences at the end of the sentence, when attachment information was disambiguated through combined semantic information. That is, we did not find traces of stronger reactivation of either verb or noun in verb- or noun-attached sentences respectively; not in terms of their part-of-speech identity nor in terms of their semantic content. Nor were we able to directly train a classifier to distinguish between verb- and noun-attached PPs across varying lexical material. In the following, we will discuss several potential explanations for the absence of an effect.

### Signal-to-noise ratio

Could it be that our analyses were simply not sensitive enough to reveal effects of high-level processes such as phrase structure building? Previous literature relying on MVPA to capture higher-level language processing does not necessarily suggest high-level effects to be smaller as compared to more perception related effects. For example, Tyler et al. used an RSA approach to investigate the temporally unfolding syntactic computations during listening of temporarily ambiguous sentences (Tyler et al. 2013). While their more perceptual word identity model correlated robustly with neural activity ( $\rho > 0.015$ ), when probing more abstract syntactic processing they found both small and large effects. Specifically, their model quantifying verb sub-categorization information was only marginally significant and correlations were much weaker ( $\rho \approx 0.005$ ) and only occurred on the word following the verb ( $n+1$ ). Their model distinguishing ambiguous from unambiguous sentences, however, correlated even more strongly ( $\rho > 0.020$ )

with neural activity at late time points. Unfortunately, it is not straightforward to compare these effect sizes to our study. Our approach is novel in that we tried to directly probe neural representations of hierarchical phrase structure rather than its consequence on ongoing processing demands (e.g. memory requirements [Nelson et al. 2017](#) or processing effort due to ambiguity [Tyler et al. 2013](#)). Therefore, it is not immediately clear from those prior studies whether an MVPA approach is powerful enough to reveal representations of phrase structure directly.

Through additional analyses, targeting orthogonal syntactic information such as part-of-speech we tried to somewhat assess the sensitivity of our approach. Our Naive Bayes classifier reached a maximum average accuracy of 67% when trained to distinguish nouns from verbs. Above chance level performance was observed robustly across all subjects. Part-of-speech although not directly indicative of hierarchical structure, is a higher-level syntactic feature and hence our classifier captured information beyond perceptual signals. It is important to note, that within our design, the part-of-speech contrast is partly confounded by physical attributes of the stimulus. Specifically, nouns and verbs differ in their form as well as their syntactic function (e.g. the majority of verbs ended in the same inflexional syllable -t signalling third person singular). We must assume that any decoding success is partly due to stimulus form. Still, our observations that part-of-speech information can be decoded from anterior brain regions in addition to occipital cortex suggests that information was not solely based on the wordform differences. Hence, while the part-of-speech classifier provides some indication to the utility of the data with respect to higher-level features, it does not necessarily ensure the success of decoding more higher-level phenomena such as hierarchical structure.

Furthermore, we also set out to find semantic and syntactic reactivation of structurally relevant context as a direct consequence of phrase structure building. Brain data and semantic models correlated with a maximum correlation coefficient smaller than 0.01. This coefficient describes the correlation with data evoked by stimuli on screen and correlations can be expected to be substantially smaller when looking at the reactivation period. It is plausible to assume that reactivated neural patterns are harder to detect, as they are not directly evoked by a stimulus. In the present analyses, we focused on the time window following the onset of the final word. Content of the final word, however, was orthogonal to the supposedly reactivated information. For example, the last word of the sentence was always a

noun and the same nouns (same semantic information) were presented in both verb- and noun-attached version. Nonetheless, in half of the trials (namely the verb-attached phrases), we would expect reactivation to reflect semantic and syntactic information of the preceding verb. The question is, whether MVPA is sensitive to internally generated, behaviourally relevant information, even with interfering material driving the neural response. While decoding of semantic category membership has been shown in the absence of a stimulus on screen (Simanova et al. 2015), this was only shown for single words. To our knowledge there are no language studies explicitly probing reactivation in sentence context through MVPA. Within vision research, however, it has been shown that during a visual working memory tasks, information about stimulus orientation could be decoded from EEG during the retention period only through perturbation using an impulse stimulus (so called ‘ping’) but would otherwise be undetected (Wolff et al. 2017). The authors argue that relevant information is not encoded explicitly in a persistent activity state but through an item-specific neural response profile that needs to be probed in order to affect ongoing neural activity. This might also explain why previous effects of prepositional phrase attachment ambiguity were found not directly following the disambiguating word but on subsequent words (Taraban and McClelland 1988, Boudewyn et al. 2014). Since we did not have a sentence continuation after the disambiguating noun, we may have been less sensitive to alterations in response profile caused by attachment structure.

Finally, it is possible, that our sensitivity was reduced by temporal variability in processing of the ambiguous sentences. It can be observed in the literature, that decoding accuracies are usually largest soon after stimulus onset and then decrease with increasing time (Cichy et al. 2014, van Es et al. 2020). We observe a similar pattern for our part-of-speech classification performance, which peaks very early after word onset (160 ms) but then decreases sharply until 250 ms after onset and continues to decrease thereafter. Thus, most information seems to be already encoded in the onset-potential or at least the neural signal might become more salient due to onset-related synchronisation of postsynaptic potentials. Effects of hierarchical structure building however may be less strictly time-locked events. Specifically, the varying difficulty in resolving structural ambiguities in our stimuli might have caused the signal to be jittered in time such that any reactivation might be less consistently synchronised across trials and subjects. Generally, each stimulus evokes a cascade of brain processes (both bottom-up and top-down)

which all can vary slightly in their duration depending on context and individual and may therefore lead to more substantial variation in later, high-level brain processing as compared to initial bottom-up processing. Such temporal variability might have led to lower sensitivity for finding our effect as well. Future analyses should take temporal variability explicitly into account to not encounter the same issue. To achieve this, probabilistic frameworks for data-driven estimation of brain states could be used to align processing and overcome temporal variability. For example, Vidaurre et al. have developed an analysis that not only defines multiple representational states that dynamically encode the stimulus but also specifies which of these states is active when in time (Vidaurre et al. 2019).

### Shallow processing

Assuming that our signal to noise ratio in principle allows to capture neural representations of hierarchical structure, we will now turn to some more cognitive explanations for our failure to decode such structural representations. It is possible that readers do not compute phrase structure by default and at all times. Specifically, our experiment may have discouraged any detailed syntactic processing and subjects may have been engaged in “shallow” processing instead, similar to what has been reported before for garden-path sentences under the term “good-enough processing” (Ferreira and Patson 2007, Ferreira and Lowder 2016, Traxler 2014). The idea of good-enough processing is that readers often arrive at a semantic proposition when interpreting a sentence without conducting a full syntactic (re)analysis. The recently established link between shallow processing and information structure (Ferreira and Lowder 2016) further increases the plausibility of prepositional phrases falling victim to this strategy as well. Specifically, Ferreira & Lowder suggest that processing effort is usually directed towards parts of a sentence that constitute new rather than given information. The motivation for such a strategy is twofold. Firstly, it would maximise the success of integration of newly received information. And secondly, since given information links to prior discourse it is also more likely to be redundant and therefore more likely to survive “shallow” processing. It might not be obvious why our experiment should be affected by such shallow processing, given that we presented subjects with unrelated sentences without any larger discourse context to drive information structure. PPs are, however, making up the subordinate clause of the sentence, which is standardly viewed as communicating previously known information

(Hornby 1974) rather than new. Hence, it is possible that structurally inherent information structure in sentences with PPs causes readers to allocate less processing resources onto the structural disambiguation of the attachment. This would also be in line with processing accounts where hierarchical operations are not assumed as the default (Frank et al. 2012). It is assumed that such processing strategies can be overwritten by strong task demands. For example, previous research has shown that syntactic task demands can reveal a P600 when there was none evoked by a purely semantic task (Mongelli 2020). Indeed, many previous studies probing syntactic processing make use of syntactic tasks such as grammaticality judgments (Tyler et al. 2013). In our study, however, subjects had to respond in only 25% of the trials and even on those trials, comprehension questions were not always probing knowledge about the PP region. The absence of a task and the fact that thematic role assignment could only be based on semantic cues in the first place may have discouraged a deep analysis of phrase structure.

The good-enough processing hypothesis further implies that hierarchical structure need not be computed at all in order to assign thematic roles. Instead, the semantic implications of the assigned thematic roles would be the sole outcome of successful sentence processing. Semantics of thematic roles are more complex and numerous than their possible corresponding phrase structures. Through adopting a strictly binary distinction of verb- and noun attachments we have intentionally ignored this semantic variation to target only the structural differences. However, as mentioned before, phrase structure and thematic roles are somewhat related and hence can easily become confounded. In fact, the relationship between thematic roles and syntactic structure is somewhat asymmetric to begin with. While any given thematic role is always bound to a certain syntactic structure<sup>2</sup>, this is not a bidirectional relationship. For example an instrument role will always be expressed in a verb-attached PP, but not every verb-attached phrase structure is necessarily carrying information about instruments (see sentences 3.8 & 3.9 for alternative role example).

3.8 The girl cuts the apple with a knife. (instrument role)

3.9 The girl cuts the apple with vigour. (manner role)

Taraban et al. have shown that previously reported reading time effects of PPs can be explained largely by expectations about thematic role. Specifically, they

---

<sup>2</sup>Assuming that the thematic role is explicitly expressed and does not result from coercion

showed that unexpected structural attachment (verb- or noun attachment) do not delay reading times beyond the effect of thematic role expectations (Taraban and McClelland 1988). The P600 effects reported by Boudewyn et al. could have also been driven by the semantics of the associated thematic roles rather than structure per se. In their stimulus set all verb-attached stimuli contained PPs expressing an instrument and all noun-attached PPs expressed an attribute. Moreover, most of their sentences contained action verbs (which bias towards expectations for instrument roles to begin with). Their P600 could therefore just as well be a marker for surprisal due to the unexpected thematic role in noun-attached sentences. In our study, we had more varying verb types (almost a third of all verbs were perception verbs) and more varying thematic roles (see table 3.1). However, the definition of thematic roles can be murky and the less common ones are usually poorly defined. With the exception of agent and patient role, the psychological reality of certain thematic roles (even as prominent as the instrument role) can be debated (Rissman and Majid 2019). It is therefore difficult to systematically manipulate this dimension. Nonetheless, through using more varied thematic roles and verbs we have created a more naturalistic stimulus set as compared to previous studies, potentially weakening effects of thematic role expectations, that likely have been driving previous findings of divergent neural activity between noun- and verb-attached PPs.

VA	action	I M G	The painter paints the wall with the fresh paint. The student writes the exam with few errors. The state supplies households with a power grid.
	perception	I/M G	The customer angers the waitress with her rude manners.
NA	action	AT AC	The politician pays the taxi driver with the annoying manners. The intern wraps the bread with the organic butter.
	perception	AT AC	the chef likes the salad with the local herbs. The paramedic spots the sick person with a furry teddybear.

**Table 3.1.:** *Example sentences. For each verb-attached (VA) or noun-attached (NA) PP several thematic roles could occur within the stimuli. Possible roles are instrument role (I), manner role (M), goal role (G), attribute role (AT) accompanying role (AC). Categorisation of thematic roles following those in Taraban and McClelland 1988*

In conclusion, with this study we could not identify a neural representation of hierarchical structure using MVPA. We did show, however, that our MVPA approach was in principle sensitive to both syntactic and semantic information

encoded in the neural signal. Further, we did not find any differences between processing verb- or noun-attached prepositional phrases unlike previous studies have suggested. We speculate that this was partly due to our well controlled and semantically varied sentence material. In the future, a more fine-grained characterisation of the semantic dimensions driving attachment decisions and the systematical manipulation of thematic roles may help to establish any differences in processing PPs at a purely structural level.

## 3.5 Appendix

### TIGERSearch queries

We defined the number of ambiguous prepositional phrases (PPs) as those phrases that dominate a preposition and directly follow a noun:

```
(10) [pos="NN"].#pp:[cat="PP"]& #pp > #prep:["APPR" | pos="APPRART"]$
```

We extracted frequency counts for all postnominal modifiers (noun-attached) within the ambiguous PPs, excluding those cases where the PP is topicalized (sentence-initial and therefore not ambiguous):

```
(2) #noun:[pos="NN"].#pp:[cat="PP"]& #phrase > #noun &  
#pp > #prep:[pos="APPR" | pos="APPRART"]&  
#n >MNR #pp & #phrase > @1 #x & [cat="VROOT"] !>@1 #x
```

Similarly, we extracted frequency counts for all verb modifiers (verb-attached) within the ambiguous PPs:

```
(3) #noun:[pos="NN"].#pp:[cat="PP"]&  
#pp > #prep:[pos="APPR" | pos="APPRART"] & #n > MO #pp
```



## Stimulus Material - Verb condition

**Table 3.2.:** Stimulus Material - Verb condition

Sentence (Attachment)
Das Amt belohnt einen Arbeiter mit einer höheren Position. (VA)
Das Amt empfiehlt einen Arbeiter mit einer höheren Position. (NA)
Der Beirat besetzt die Ämter mit den besten Arbeitern. (VA)
Der Beirat sucht die Ämter mit den besten Arbeitern. (NA)
Der Camper mag die Suppe mit der frischen Petersilie. (NA)
Der Camper würzt die Suppe mit der frischen Petersilie. (VA)
Die Chefin meidet den Mitarbeiter mit der faltbaren Karte. (NA)
Die Cousine erneuert die Reifen mit dem feinen Flickzeug. (VA)
Die Cousine verschenkt die Reifen mit dem feinen Flickzeug. (NA)
Die Diebin beneidet ihren Komplizen mit der einzigen Pistole. (NA)
Die Diebin rettet ihren Komplizen mit der einzigen Pistole. (VA)
Der Förster befördern das Holz mit der roten Markierung. (NA)
Der Förster markiert das Holz mit der roten Markierung. (VA)
Die Fotografen benötigen eine Kamera mit dem wertigen Objektiv. (NA)
Die Fotografen erweitern eine Kamera mit dem wertigen Objektiv. (VA)
Die Gärtnerin beschenkt die Dame mit den weißen Rosen. (VA)
Die Gärtnerin kennt die Dame mit den weißen Rosen. (NA)
Der Gast beschriftet die Serviette mit einer mobilen Handynummer. (VA)
Der Gast findet die Serviette mit einer mobilen Handynummer. (NA)
Der Großvater backt die Brezel mit dem groben Salz. (NA)
Der Großvater bestreut die Brezel mit dem groben Salz. (VA)
Der Ingenieur beschmiert die Kette mit dem klebrigen Öl. (VA)
Der Ingenieur verpackt die Kette mit dem klebrigen Öl. (NA)
Die Investoren besetzen die Betriebe mit einigen fleißigen Tagelöhnern. (VA)
Die Investoren suchen die Betriebe mit einigen fleißigen Tagelöhnern. (NA)
Der Junge beneidet seinen Bruder mit dem dicken Seil. (NA)
Der Junge rettet seinen Bruder mit dem dicken Seil. (VA)
Der Kellner füllt die Tasse mit dem heißen Kaffee. (VA)
Der Kellner hält die Tasse mit dem heißen Kaffee. (NA)

Continued on next page

**Table 3.2 – continued from previous page**

<b>Sentence (Attachment)</b>
Der Koch mag den Salat mit den lokalen Kräutern. (NA)
Der Koch würzt den Salat mit den lokalen Kräutern. (VA)
Der Konditor backt den Kuchen mit den bunten Streuseln. (NA)
Der Konditor bestreut den Kuchen mit den bunten Streuseln. (VA)
Der Küchenchef füllt den Topf mit der gestrigen Suppe. (VA)
Der Küchenchef hält den Topf mit der gestrigen Suppe. (NA)
Der Kunde benötigt einen Computer mit einer modernen Tastatur. (NA)
Der Kunde erweitert einen Computer mit einer modernen Tastatur. (VA)
Die Kundin bezahlt die Kellnerin mit den unhöflichen Manieren. (NA)
Die Kundin verärgert die Kellnerin mit den unhöflichen Manieren. (VA)
Die Landwirte sperren die Wiesen mit den stacheligen Zäunen. (VA)
Die Landwirte umfahren die Wiesen mit den stacheligen Zäunen. (NA)
Die Nichte meidet die Patentante mit der riesigen Torte. (NA)
Die Partei überzeugt eine Untergruppe mit einigen fraglichen Argumenten. (VA)
Die Partei besitzt eine Untergruppe mit einigen fraglichen Argumenten. (NA)
Die Pflegerin beschenkt eine Seniorin mit ganz viel Liebe. (VA)
Die Pflegerin kennt eine Seniorin mit ganz viel Liebe. (NA)
Die Politikerin bezahlt den Taxifahrer mit der dreisten Art. (NA)
Die Politikerin verärgert den Taxifahrer mit der dreisten Art. (VA)
Der Polizist braucht seinen Kollegen mit dem anonymen Telefon. (NA)
Der Polizist verständigt seinen Kollegen mit dem anonymen Telefon. (VA)
Der Praktikant beschmiert das Brot mit der organischen Butter. (VA)
Die Praktikant verpackt das Brot mit der organischen Butter. (NA)
Der Prüfer sperrt die Zone mit dem rot-weißen Absperrband. (VA)
Der Prüfer umfährt die Zone mit dem rot-weißen Absperrband. (NA)
Die Reiterin belohnt ein Pferd mit einem neuen Sattel. (VA)
Die Reiterin empfiehlt ein Pferd mit einem neuen Sattel. (NA)
Die Schülerin schreibt die Klausur mit nur wenigen Fehlern. (VA)
Die Schülerin zeigt die Klausur mit nur wenigen Fehlern. (NA)
Der Sekretär schreibt das Protokoll mit der schönen Handschrift. (VA)
Der Sekretär zeigt das Protokoll mit der schönen Handschrift. (NA)
Der Spion beschriftet das Notizbuch mit einer wertvollen Information. (VA)

Continued on next page

**Table 3.2 – continued from previous page**

<b>Sentence (Attachment)</b>
Der Spion findet das Notizbuch mit einer wertvollen Information. (NA)
Der Staat beliefert die Haushalte mit einem robusten Stromnetz. (VA)
Der Staat zählt die Haushalte mit einem robusten Stromnetz. (NA)
Die Trainerin schlägt den Hund mit einem langen Stock. (VA)
Die Trainerin sieht den Hund mit einem langen Stock. (NA)
Der Verbrecher besänftigt den Anwalt mit den cleveren Ausreden. (VA)
Der Verbrecher bevorzugt den Anwalt mit den cleveren Ausreden. (NA)
Der Verein überzeugt ein Komitee mit einer dynamischen Rhetorik. (VA)
Der Verein besitzt ein Komitee mit einer dynamischen Rhetorik. (NA)
Die Zentrale braucht das Flugzeug mit dem digitalen Funkgerät. (NA)
Die Zentrale verständigt das Flugzeug mit dem digitalen Funkgerät. (VA)
Die Züchterin schlägt das Tier mit der kurzen Leine. (VA)
Die Züchterin sieht das Tier mit der kurzen Leine. (NA)
Der Produzent beliefert die Fabriken mit den seltenen Teilen. (VA)
Der Produzent zählt die Fabriken mit den seltenen Teilen. (NA)
Der Unternehmer besänftigt den Geldanleger mit den klugen Sprüchen. (VA)
Der Unternehmer bevorzugt den Geldanleger mit den klugen Sprüchen. (NA)
Die Cousine erneuert den Raumduft mit einem handlichen Nachfüller. (VA)
Die Cousine verschenkt den Raumduft mit einem handlichen Nachfüller. (NA)
Der Bote befördert die Kisten mit dem gelben Etikett. (NA)
Der Bote markiert die Kisten mit dem gelben Etikett. (VA)
Die Chefin gratuliert dem Mitarbeiter mit der faltbaren Karte. (VA)
Die Nichte gratuliert der Patentante mit der riesigen Torte. (VA)
Der Maler begutachtet die Wand mit der frischen Farbe. (NA)
Der Maler bemalt die Wand mit der frischen Farbe. (VA)
Der Schamane begutachtet die Maske mit der braunen Kreide. (NA)
Der Schamane bemalt die Maske mit der braunen Kreide. (VA)
Der Arzt entdeckt den Säugling mit einem flauschigen Teddy. (NA)
Der Arzt ermuntert den Säugling mit einem flauschigen Teddy. (VA)
Der Sanitäter entdeckt den Kranken mit einem kuscheligen Bären. (NA)
Der Sanitäter ermuntert den Kranken mit einem kuscheligen Bären. (VA)
Die Blinde ertastet das Wesen mit den zarten Fingern. (VA)

Continued on next page

**Table 3.2 – continued from previous page**

<b>Sentence (Attachment)</b>
Die Blinde verehrt das Wesen mit den zarten Fingern. (NA)
Die Kaiserin ertastet das Geschöpf mit den feinen Händen. (VA)
Die Kaiserin verehrt das Geschöpf mit den feinen Händen. (NA)
Der Junggeselle erfreut die Angebetete mit einem hübschen Kleid. (VA)
Der Junggeselle wählt die Angebetete mit einem hübschen Kleid. (NA)
Der Kandidat erfreut die Kandidatin mit einem strahlenden Lächeln. (VA)
Der Kandidat wählt die Kandidatin mit einem strahlenden Lächeln. (NA)

## Stimulus Material - Role condition

**Table 3.3.:** Stimulus Material - Role condition

<b>Sentence (Attachment)</b>
Der Wärter streichelt den Elefant mit dem grauen Rüssel. (NA)
Der Elefant streichelt den Wärter mit dem grauen Rüssel. (VA)
Der Wanderer schubst den Bock mit dem gekrümmten Horn. (NA)
Der Bock schubst den Wanderer mit dem gekrümmten Horn. (VA)
Der Hirsch trifft den Krieger mit dem klobigen Gewehr. (NA)
Der Krieger trifft den Hirsch mit dem klobigen Gewehr. (VA)
Die Robbe bespritzt die Animateurin mit dem vollen Eimer. (NA)
Die Animateurin bespritzt die Robbe mit dem vollen Eimer. (VA)
Der Papagei ärgert den Pilger mit dem spitzen Schnabel. (VA)
Der Pilger ärgert den Papagei mit dem spitzen Schnabel. (NA)
Der Schüler kitzelt den Kater mit dem weißen Schnurrhaar. (NA)
Der Kater kitzelt den Schüler mit dem weißen Schnurrhaar. (VA)
Der Doktor grüßt den Patient mit dem brandneuen Stethoskop. (VA)
Der Patient grüßt den Doktor mit dem brandneuen Stethoskop. (NA)
Der Mieter erwartet den Klempner mit der dreckigen Rohrzange. (NA)
Der Klempner erwartet den Mieter mit der dreckigen Rohrzange. (VA)
Die Zahnfee überrascht die Tochter mit dem wackeligen Zahn. (NA)
Die Tochter überrascht die Zahnfee mit dem wackeligen Zahn. (VA)
Das Maskottchen umarmt das Mädchen mit den pelzigen Armen. (VA)
Das Mädchen umarmt das Maskottchen mit den pelzigen Armen. (NA)
Der Sänger winkt dem Fan mit der akustischen Gitarre. (VA)
Der Fan winkt dem Sänger mit der akustischen Gitarre. (NA)
Der Dirigent folgt dem Musiker mit der lieblichen Geige. (NA)
Der Musiker folgt dem Dirigent mit der lieblichen Geige. (VA)
Die Betreuer geleiten die Senioren mit den klapprigen Rollatoren. (NA)
Die Senioren geleiten die Betreuer mit den klapprigen Rollatoren. (VA)
Der Sanitäter holt den Urlauber mit der faltbaren Trage. (VA)
Der Urlauber holt den Sanitäter mit der faltbaren Trage. (NA)

Continued on next page

**Table 3.3 – continued from previous page**

Sentence (Attachment)
Der Reiter überholt den Biker mit dem schweren Motorrad. (NA)
Der Biker überholt den Reiter mit dem schweren Motorrad. (VA)
Die Mütter bedrängen die Obsthändler mit den sperrigen Kinderwägen. (VA)
Die Obsthändler bedrängen die Mütter mit den sperrigen Kinderwägen. (NA)
Das Kleinkind berührt das Pony mit der weichen Schnauze. (NA)
Das Pony berührt das Kleinkind mit der weichen Schnauze. (VA)
Der Kaiser erheitert den Hofnarr mit der bunten Perücke. (NA)
Der Hofnarr erheitert den Kaiser mit der bunten Perücke. (VA)
Die Erzählerin lauscht der Greisin mit dem piepsenden Hörgerät. (NA)
Die Greisin lauscht der Erzählerin mit dem piepsenden Hörgerät. (VA)
Der Milliardär begegnet dem Bauarbeiter mit dem teuren Cabrio. (VA)
Der Bauarbeiter begegnet dem Milliardär mit dem teuren Cabrio. (NA)
Der Fußballer nervt den Schiri mit der schwarzen Pfeife. (NA)
Der Schiri nervt den Fußballer mit der schwarzen Pfeife. (VA)
Der Adler verfolgt den Jäger mit der rostigen Flinte. (NA)
Der Jäger verfolgt den Adler mit der rostigen Flinte. (VA)
Der Kassierer erreicht den Käufer mit dem vollen Wagen. (NA)
Der Käufer erreicht den Kassierer mit dem vollen Wagen. (VA)
Der Specht lockt den Käfer mit dem glänzenden Panzer. (NA)
Der Käfer lockt den Specht mit dem glänzenden Panzer. (VA)
Die Soldaten bekriegen die Indianer mit den vergifteten Pfeilen. (NA)
Die Indianer bekriegen die Soldaten mit den vergifteten Pfeilen. (VA)
Der Büffel bekämpft den Tiger mit den breiten Tatzen. (NA)
Der Tiger bekämpft den Büffel mit den breiten Tatzen. (VA)
Der Hausmeister erschreckt den Greis mit dem klappernden Gebiss. (NA)
Der Greis erschreckt den Hausmeister mit dem klappernden Gebiss. (VA)
Der Kurier ohrfeigt den Butler mit dem silbernen Tablett. (NA)
Der Butler ohrfeigt den Kurier mit dem silbernen Tablett. (VA)
Das Kind verängstigt das Insekt mit dem giftigen Stachel. (NA)
Das Insekt verängstigt das Kind mit dem giftigen Stachel. (VA)
Die Kuh bedroht die Wilde mit der brennenden Fackel. (NA)

Continued on next page

**Table 3.3 – continued from previous page**

<b>Sentence (Attachment)</b>
Die Wilde bedroht die Kuh mit der brennenden Fackel. (VA)
Das Rind attackiert das Publikum mit den spitzen Hörnern. (VA)
Das Publikum attackiert das Rind mit den spitzen Hörnern. (NA)
Das Einhorn beschützt das Fräulein mit dem leuchtenden Horn. (VA)
Das Fräulein beschützt das Einhorn mit dem leuchtenden Horn. (NA)
Der Radler behindert den Bauer mit dem dreckigen Trecker. (NA)
Der Bauer behindert den Radler mit dem dreckigen Trecker. (VA)
Der Chor animiert den Pensionär mit seinem alten Krückstock. (NA)
Der Pensionär animiert den Chor mit seinem alten Krückstock. (VA)
Der Sänger begleitet den Violinist mit seiner kostbaren Violine. (NA)
Der Violinist begleitet den Sänger mit seiner kostbaren Violine. (VA)
Der Knecht empfängt den König mit seinem prächtigen Zepter. (NA)
Der König empfängt den Knecht mit seinem prächtigen Zepter. (VA)
Der Ninja schützt den Meister mit den uralten Weisheiten. (NA)
Der Meister schützt den Ninja mit den uralten Weisheiten. (VA)
Die Schwangere verblüfft die Hebamme mit ihrer jahrelangen Erfahrung. (NA)
Die Hebamme verblüfft die Schwangere mit ihrer jahrelangen Erfahrung. (VA)
Der Fuchs verletzt den Igel mit den kleinen Stacheln. (NA)
Der Igel verletzt den Fuchs mit den kleinen Stacheln. (VA)
Das Volk vertreibt das Militär mit den grässlichen Waffen. (NA)
Das Militär vertreibt das Volk mit den grässlichen Waffen. (VA)
Die Beute reizt die Krake mit den flinken Tentakeln. (NA)
Die Krake reizt die Beute mit den flinken Tentakeln. (VA)
Der Elch rammt den Wolf mit seinem enormen Geweih. (VA)
Der Wolf rammt den Elch mit seinem enormen Geweih. (NA)
Der Samurai verwundet den Alligator mit dem antiken Schwert. (VA)
Der Alligator verwundet den Samurai mit dem antiken Schwert. (NA)
Die Muschel bezwingt die Möwe mit ihrer harten Schale. (VA)
Die Möwe bezwingt die Muschel mit ihrer harten Schale. (NA)
Die Wühlmaus befühlt die Schnecke mit den wendigen Fühlern. (NA)
Die Schnecke befühlt die Wühlmaus mit den wendigen Fühlern. (VA)
Die Bäuerin liebkost die Miezekatte mit den rosa Pfoten. (NA)

Continued on next page

**Table 3.3 – continued from previous page**

<b>Sentence (Attachment)</b>
Die Miezekatze liebkost die Bäuerin mit den rosa Pfoten. (VA)
Der Badegast schikaniert den Delphin mit den kräftigen Flossen. (NA)
Der Delphin schikaniert den Badegast mit den kräftigen Flossen. (VA)
Der Eigentümer erzürnt den Mechaniker mit dem schmutzigen Werkzeug. (NA)
Der Mechaniker erzürnt den Eigentümer mit dem schmutzigen Werkzeug. (VA)
Die Mücke quält die Urlauberin mit dem aggressiven Mückenspray. (NA)
Die Urlauberin quält die Mücke mit dem aggressiven Mückenspray. (VA)
Die Fliege plagt die Hündin mit dem wedelnden Schwanz. (NA)
Die Hündin plagt die Fliege mit dem wedelnden Schwanz. (VA)







# Measuring compositional sentence meaning through behaviour

A sentence is more than just a collection of words. In fact, the meaning of a sentence is defined not only by the words it contains but also by their structured relations. Therefore, a sentence can be likened to other high-dimensional stimuli such as visual scenes. Having quantitative models of such complex, high-dimensional stimuli can help evaluate cognitive models of language processing. We tested, whether the online measurements of perceived similarity can capture both semantic as well as structural dimensions of simple transitive sentences. Specifically, we collected similarity judgments of 200 subjects through an online multiple arrangement task. We find that group-level averages of perceived similarity reveal a strong bias towards the main verb of the sentence. Furthermore, we show how non-negative matrix factorisation of similarity judgment data can reveal multiple underlying dimensions reflecting not only semantic but also structural information. Finally, we provide an example of how similarity judgments on sentence stimuli can serve as benchmarks for artificial neural networks models by comparing our behavioural data against sentence similarity extracted from three state-of-the-art models.

## 4.1 Introduction

Human language use is special due to its unbounded creativity. We can flexibly produce and understand never before encountered combinations of words. Underlying this ability, it has been theorised, must be the compositional nature of language, i.e. that any sentence can be described as a function of its parts (i.e. words) and the rules to combine them (i.e. grammar) (Fodor and Pylyshyn 1988). In other words, compositional language is high-dimensional in terms of not only semantic features but also structure. For example, a bag of words (e.g. [biting, old, dog, lady]) can be regarded as high-dimensional in meaning, to the extent that it contains multiple elements, each defined through a set of semantic features. In contrast to a bag of words, a compositional sentence additionally encodes relational roles (i.e. thematic roles such as “dog as agent”) that add structural information to the elements (e.g. “the dog bit the old lady”). Such relational roles are a form of abstract knowledge that is extremely important in language use as it provides the basis for the systematicity that allows us to flexibly assigning the role of the agent to different exemplars (e.g. “the old lady bit the dog”).

There is an ongoing debate in cognitive neuroscience, as to how our brains might learn and represent such relational roles (Rabovsky and McClelland 2020; Puebla et al. 2021). Modelling the output of this unknown neural function for compositional meaning formation, i.e. the resulting mental representations, could help us better understand what neural mechanisms are at play. Representational models have already advanced our understanding of the neurobiology of semantics. For example, through explicitly modelling word semantics, researchers could not only confirm a crucial role for the anterior temporal lobe (in line with previous patient data) but also identify additional frontal and parietal brain areas to be sensitive to modality-independent semantic representations (Bruffaerts et al. 2019). Yet, when it comes to meaning beyond the single word level, it is not trivial to quantitatively describe what makes up a compositional representation in the first place.

In the current study, we investigated whether a behavioural measure of perceived sentence similarity can serve as a quantitative representation of high-dimensional compositional meaning.

## Similarity as a window into mental representations

Similarity judgments can be used as a proxy for representational content. Specifically, we can ask experimental subjects to make explicit similarity judgments in the lab under controlled conditions and those judgments presumably reflect people's mental representations of the judged items at that instance. Similarity judgments therefore allow us to capture mental representations without having to specify their exact content. For example, the underlying rules that determine meaning composition within any given sentence might be underspecified but we can nonetheless determine how that sentence's meaning compares to others. Perceived similarity across multiple sentences may then approximate representations and can be further related to the representational geometry in the brain using multivariate analysis techniques (Kriegeskorte et al. 2008). For example, past research has shown that similarity judgments of images can capture perceptual representations (Hebart et al. 2020) that correlate with multivariate neural representations and can be used to study the temporal dynamics of object recognition in the brain Cichy et al., 2019.

Here we study, whether similarity judgments can approximate the compositional nature of linguistic representations as well as they have been shown to capture high-dimensional mental representations of images.

## Similarity of compositional meanings

The notion of similarity, that we are after, should capture the high-dimensional nature of compositional representations according to at least two aspects. First, it needs to account for similarity in semantic features. Hence, it should reflect that the sentence "John loves Mary" is similar to "John likes Mary" but both are dissimilar to "John hates Mary". Second, given identical semantic features it should reflect role assignment across sentences. For example, two sentences with complete role reversal (John loves Mary vs. Mary loves John) should be recognised as dissimilar. Previous studies on visual scene perception indicate that relational information might indeed influence perceived similarity. For example, participant's judgment of scene similarity varied more strongly when the diverging feature was structurally aligned across scenes (a change in an existing element can be big or small) as compared to when it was not structurally aligned (adding a

new element creates a difference independent of its exact features)([Markman and Gentner 1996](#)).

Finally, both semantic and relational features can interact in complex ways in compositional sentences. For example, if sentences contain bidirectional verbs (e.g. John greeted Mary vs. Mary greeted John), their perceived similarity under role reversal might decrease less as opposed to sentences describing more unidirectional actions. Furthermore, roles are not rigidly defined in terms of syntactic arguments only (e.g. subject, object) but carry semantic content ([Holyoak 2005](#)). For verbs that convey a mental state, such as “to surprise” or “to notice”, the roles of the agent and patient can be defined as the causal element of the experience (the stimulus) and the undergoer of the experience (the experiencer) respectively. Under this semantic definition, the mapping between syntactic subject and syntactic object on the one hand and agent/stimulus and patient/experiencer roles on the other will depend on the specific verb semantics. For example, the cat (subject) in [example 4.1](#) maps better onto woman (object) than deer (subject) in [example 4.2](#) since they are both causal for the events described ([Frankland and Greene 2020](#)).

4.1 The cat surprised the man.

4.2 The deer noticed the woman.

For the current study, we are less interested in semantic modulations of event roles but rather aim to isolate the general effects of role-filler assignment. Therefore, we removed semantic constraint as much as possible from our stimuli. To reduce semantic constraint of the verb, we selected only unidirectional action verbs with consistent agent and patient roles (see details about stimuli in [methods](#)) and relatively arbitrary verb-noun combinations (e.g. “the electrician encourages the guitarist”, “the guitarist pushes the athlete”). As a result, the compositional meaning of these sentences can be simply modelled through the conjunctive contribution of the words’ semantics and their roles ([Goldstone and Son 2005](#)). Note, however, that it is impossible to completely remove any effects of combinatorial semantics whatsoever. For example, while “guitarist” may refer to roughly the same concept (e.g. person, on a stage) independent of the role the word occurs in, some additional features such as “aggressive” may be activated, when assigned the agent as compared to the patient role of the action “to push”. In that sense, the very strict theoretical account of compositionality we have described so far is somewhat at odds with the high degree of flexibility and idiosyncrasy observed in

human language comprehension (Rabovsky and McClelland 2020). Nonetheless, for the current study, an idealised assumption of compositionality is sufficient as a theoretical model for studying the systematicity of relational roles.

## Measuring similarity through behaviour

Perceived similarity can be measured by means of different behavioural tasks. Earlier used methods include asking people to freely sort a set of items into piles (free sorting), to make speeded same/different judgments (implicit measure, inter-item confusability), to determine the odd one out of three items (triad test), to rate the similarity of two items on a scale (pairwise judgments) or to indicate similarity between items by placing them either close by (similar) or far apart (dissimilar)(geometrical tasks). In the current study, we implement a geometrical task because it provides several advantages over other methods. All of these methods can in principle be used to collect continuous similarity measures. Whenever binary similarity judgments are acquired, e.g in free sorting or the triad test, continuous values can be obtained by combining data across multiple participants or repeated presentations of the same pair. Both pairwise judgments and geometrical tasks have the advantage of probing continuous valued similarity at the single participant level. Beyond that, geometrical tasks additionally allow for the most time-efficient sampling. This is because in a geometrical arrangement task, participants are asked to arrange several items, randomly scattered across a screen, within a circular 2D space, such that the distance between items is proportional to each pair's similarity. Spatial adjustment (via drag and drop) of each individual item hence communicates multiple similarity judgments at once. The time to acquire pairwise similarity judgments, in contrast, grows quadratically as a function of total set size, since  $n(n-1)/2$  judgments are necessary for a set of  $n$  items. In practice, the pairwise similarity judgements method has been shown to last 5 - 6 times longer as compared to a geometrical task on the same stimuli (Hout et al. 2013). Furthermore, the similarity ratings attained through geometrical tasks have been shown to correlate highly with pairwise similarity ratings. Therefore, geometrical tasks are to be preferred over other methods when sampling similarity judgments for larger sets of items as well as for more complex items such as sentences, which by their nature will require longer processing times than pictures or single words.

## High-dimensional representations

Since sentences contain high-dimensional meaning, an appropriate tool for quantifying sentence meaning should be able to capture multiple semantic dimensions simultaneously. Despite their 2D nature, geometrical tasks capture higher-dimensional representations already through a single arrangement of a set of items, as has been shown for both visual objects and single words (Richie et al. 2020; Hout et al. 2013). Recent extensions of the geometric arrangement task, that go beyond the single arrangement, have made it even more sensitive to high-dimensional representations. Kriegeskorte and colleagues suggest to have subjects perform multiple arrangements of subsets of items that are adaptively designed for optimal measurement efficiency (Kriegeskorte and Mur 2012). The final representational similarity matrix (RSM) is then computed by combining evidence across all subset arrangements.

Past studies have successfully applied the multi-arrangement task to quantify high-dimensional mental representations of naturalistic images (King et al. 2019) and visual scenes (Groen et al. 2017). To our knowledge, it has not been shown that the multi-arrangement task is equally suitable for sentence stimuli, which just like visual scenes include complex structural information, but unlike scenes may be perceived as more fragmented due to separation between words.


In this study, we tested the feasibility of using a multiple arrangement tasks to collect similarity judgments for sentences, that specify simple relational roles on semantically varying agents and patients. We report online-acquired similarity judgments on (1) isolated nouns and on (2) sentences containing those nouns and evaluate the sensitivity of perceived similarity to relational roles. Additionally, we provide an example use case of sentence similarity judgments, namely, as a benchmark for computational models of human sentence processing.

## 4.2 Methods

### Stimuli creation

We created a set of German sentences ( $n = 48$ ) describing simple transitive events, such that the similarity between events could be captured by a small number of meaning dimensions. For this, we selected 36 words that belonged to six thematic categories, i.e. 24 nouns & 12 verbs from four profession themes and two





Nouns				Verbs	
medicine	manual labor	sport	music	communication	physical
Sanitäter <i>paramedic</i>	Elektriker <i>electrician</i>	Boxer <i>boxer</i>	Bassist <i>bassist</i>	bestärken <i>encourage</i>	stoßen <i>push</i>
Internist <i>internist</i>	Handwerker <i>artisan</i>	Läufer <i>runner</i>	Geiger <i>violin player</i>	ermuntern <i>encourage</i>	schubsen <i>push</i>
Pfleger <i>nurse</i>	Klempner <i>plumber</i>	Sprinter <i>sprinter</i>	Musiker <i>musician</i>	bejubeln <i>cheer</i>	schlagen <i>beat</i>
Radiologe <i>radiologist</i>	Zimmermann <i>carpenter</i>	Athlet <i>athlete</i>	Pianist <i>pianist</i>	loben <i>praise</i>	verscheuchen <i>chase</i>
Therapeut <i>therapist</i>	Mechaniker <i>mechanic</i>	Fußballer <i>soccer player</i>	Sänger <i>singer</i>	ermutigen <i>encourage</i>	verprügeln <i>beat</i>
Chirurg <i>surgeon</i>	Tischler <i>carpenter</i>	Sportler <i>sportsman</i>	Gitarrist <i>guitarist</i>	trösten <i>comfort</i>	schütteln <i>shake</i>

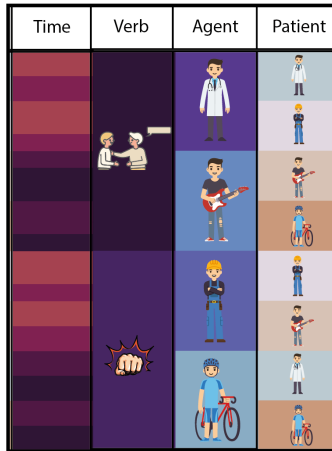
Figure 4.1.: Stimulus vocabulary

action themes respectively (see figure 4.1). Across thematic categories, words were matched (nouns and verbs separately) according to number of letters, number of syllables and frequency. In addition, we took care that noun categories did not systematically differ from each other in their suffixes. Importantly we chose words that could be combined more or less arbitrarily without imposing strong constraints with respect to meaning amongst each other. For example, whether a surgeon is pushing or praising does not change the properties/semantic features associated with the concept of the surgeon itself. Hence, we assume the compositional meaning of each sentence to be a straightforward conjunction of the individual words and their respective thematic roles within the event.

From this vocabulary of 36 words, we formed sentences by pseudorandomly combining nouns and verbs from this vocabulary (e.g. “This morning the paramedic praised the electrician.”, see full stimulus set in Appendix). The randomisation was generated according to six constraints: Nouns were combined, such that each noun in agent position would be (1) paired equally often with a patient from either the same semantic category or one of the other semantic categories (e.g. “the paramedic praised the electrician.” & “the paramedic encouraged

**Figure 4.2: Stimulus randomization.**

The diagram illustrates for all sentences (rows) which elements make up any given sentence. Elements are colour coded to indicate the identity of the temporal adverbs (1st column), the semantic category of the verb (2nd column), the semantic category of the agent role (3rd column) and the semantic category of the patient role (4th column). Given any sequence of adverb, verb and agent, the patient semantic category is uncertain, since there are always to categories that occur with equal probabilities.



the nurse.”) Also, each noun (2) appeared in both object and subject position (e.g. “the paramedic praised the electrician” & “the boxer hit the paramedic”) and (3) each noun category appeared equally often. Subsequently, noun pairs and verbs were combined, such that (5) each noun category would occur equally often with verbs from both action categories. Finally, in the beginning of each sentence, we added a temporal adverb (e.g. “This morning”, “Yesterday” etc.) and then constructed the sentence according to VSO word order. The temporal adverbs were distributed such that (6) each adverb could precede two of the four noun categories and either of the verb categories (see figure 4.2).

We quantified the semantic similarity between each possible pair of sentences by counting how many words from the same semantic category occur in the same thematic roles across two sentences (see Table 4.1 for examples).

### Online Multi-arrangement Task

200 native German speakers rated the perceived similarity of our stimuli. During the multiple arrangement task, participants were asked to arrange the sentences on a computer screen inside a white circular arena by using computer mouse drag and drop operations. The distance of the placed sentences indicates the perceived similarity. Usually, participants would see the full stimulus set on the first trial and subsequent trials consist of a subset of those stimuli. The subset selection is based on an adaptive procedure aimed at 1) minimising uncertainty

1	This morning the paramedic praised the electrician. <i>Heute Morgen lobte der Sanitäter den Elektriker</i>
2	Today the surgeon encouraged the carpenter. <i>Heute ermutigte der Chirurg einen Tischler.</i>
3	Earlier the carpenter chased the electrician away. <i>Vorhin verscheuchte der Zimmermann den Elektriker.</i>
4	Yesterday the soccer player hit the boxer. <i>Gestern schlug der Fußballer einen Boxer.</i>

**Table 4.1.: Example sentences.** Sentences are listed in English translation (original German stimuli in italics below). 1 and 2 are considered most similar because thematic categories of agent, patient and verb are shared. 1 and 3 are more dissimilar in comparison, since verbs are of different thematic category, regardless of final noun identity. 1 and 4 are most dissimilar since they do not share any semantic category.

for all possible pairs of sentences (e.g. items that initially are placed very close to each other) and 2) better approximate the high-dimensional perceptual representational space (Kriegeskorte and Mur 2012). Often these subsequent trials included items that were placed close together in the initial trial. As subsequent trials included fewer items, this allowed participants to refine their judgements with distinctions that are more difficult to carry in the context of the whole set and the limited arena space. As a consequence, each trial will only provide a subset of pairwise distances (similarity judgments), however. To extract one single estimate of all pairwise similarities, the overlapping subsets are first scaled and then combined as weighted averages (see Kriegeskorte and Mur 2012 for details). Due to the iterative procedure, the task is very efficient at obtaining reliable high-dimensional similarity judgments for our 48 individual sentences within 60 minutes per participant (or 24 nouns within 30 minutes). The behavioral data was collected using the Meadows web-based platform for psychophysical experiments (<http://meadows-research.com>). Online participants were recruited from the Prolific online participant pool (<http://www.prolific.co>).

Half of the participants were presented with the nouns that were used to generate sentences and the other half rated the full sentences. Subjects were instructed to place all nouns/sentences inside the white area in a manner that reflects the similarity between the described people/events. The instructions did not specifically mention that people could be categorised into professions or that verbs could be categorised into positive and negative actions. Nonetheless, the majority of subjects mentioned those dimensions in their debrief. In addition,

those subjects seeing the events were instructed to not place items only according to a single word in the sentence but rather pay attention to all elements. Prior to the main task, an example was shown for how an arrangement could look like, using nouns/sentences that were not part of the stimulus set.

### Noun similarity

For the noun similarity task, we collected arrangement data from 100 subjects (mean age = 31, std = 9), of which 41 were female. Subjects arranged all 24 unique nouns on the first trial and each subsequent trial contained subsets of those nouns. Therefore, there were no unseen item pairs and we did not exclude any of the subjects. On average, subjects completed 37 trials (std = 8) before they either reached a minimum evidence level of 0.5 or 30 minutes had passed.

### Event similarity

To our knowledge we are the first to apply this task to full sentences instead of individual words or pictures. In order to present the sentence stimuli in a format suitable for the task, we split each sentence into 3-4 lines such that it could be presented within a square box. In order to minimise the influence of the verb, sentences were broken up, such that the verb never appeared on a line on its own. In the sentence similarity judgment task, an additional practice trial preceded the main task. The practice was based on three sentences, of which two were semantically synonymous and the third describing a completely different event. The main task was complete once a subjects had reached a minimum evidence level of 0.5 for each item pair or 60 minutes had passed. On average subjects completed 112 trials (std = 41). Due to space limitations, participants were not presented with all sentences in the beginning. Instead they saw only 10 sentences during the first trial, and at least 3 items or up to a maximum of 10 on each subsequent trial. Seven subjects were excluded because they executed too few trials within the 60 minutes, i.e. they rated less than 50 item pairings. Two subjects were excluded because they did not perform as expected on the practice trial. The remaining 87 subjects (mean age = 30, std = 9; 39 female) were all German native speakers.

## Matrix factorisation

We used non-negative matrix factorisation (NMF) [Lee and Seung, 1999](#) to investigate which underlying dimensions played a role in the sentence similarity ratings. For this we first concatenate all individual subject RSMs into a large Matrix  $D$  of dimensions number of subjects  $\times$  number of item pairs. NMF allows us to decompose  $D$  into two non-negative matrices, which gives a lower rank approximation for  $D$ . Basically, we decompose this matrix into two matrices, such that  $D = W \times H$ , where  $W$  is the  $|n| \times k$  mixing matrix that contains the weights for constructing  $N$  observed subject similarity judgments from the  $k$  components, and  $H$  is a  $k \times |t|$  factorisation matrix that contains the  $k$  latent components capturing underlying pattern of pairwise item similarity. Note that, by definition, all  $H_{i,j} \geq 0$ . We applied the NMF implementation of scikit-learn [Pedregosa et al., 2011](#), which finds the optimal decomposition by iteratively optimising the distance between  $D$  and the matrix product  $WH$  using the squared Frobenius norm as the distance function.

The NMF algorithm requires to specify the number of latent components  $k$  to be extracted. In order to get an estimate of what the optimal  $k$  would be, we computed the NMF repeatedly ( $n = 1000$ ) with different random initialisations, each time limiting the factorisation to an increasing amount of components ( $1 < k < 20$ ). For each of the 1000 random initialisation we checked for each additional component for how many subjects it would receive maximal mixing weights. Although each additional component will further optimise the fit to the data, we only regard it as informative, if it captures general judgment patterns, i.e. receive maximal weights for multiple subjects, rather than individual solutions. On average it took 7 components to capture all patterns in the data, that generalised across at least 2 subjects. We then fixed the number of components to 7 and again computed 1000 factorisations using different random initialisations. Based on the resulting 7000 components, we ran agglomerative hierarchical cluster analysis to determine which underlying components are reliably found throughout repeated factorisations. Based on visual inspection of the within- and between-cluster similarity we decided on a distance threshold of 0.9, such that we could define 5 clusters of components, that would reliably emerge across multiple factorisations (at least 990 times out of 1000) and were for the most part interpretable in terms of the underlying similarity patterns (see [Figure 4.4](#)). From each cluster, we computed one final component, by taking the average across all cluster exemplars (centroid).

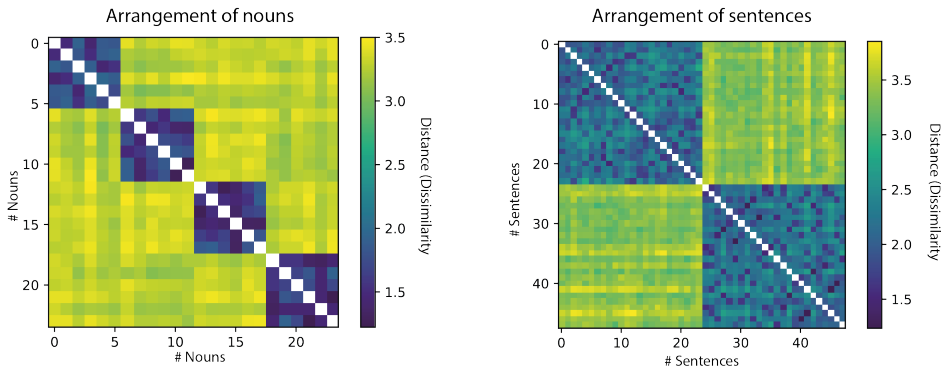
Based on the resulting 5 components we again computed the unmixing matrix using the same optimisation algorithm. In order to qualitatively assess the final factorisation, we computed the Spearman rank-ordered correlation between each of the components and the upper triangular vectors of our binary model matrices for the verb category, the agent category and the patient category respectively.

## Sentence similarity based on ANN generated embeddings

We extracted sentence embeddings from three pre-trained ANN models, GPT2 (Radford et al. 2019), BERT (Devlin et al. 2018) and SBERT (Reimers and Gurevych 2020b), and compared their pairwise similarity to our behavioural similarity judgments. For the BERT and GPT2 architectures, embeddings were extracted from models trained on German texts and implemented in PyTorch with the Huggingface module. Specifically, we used the bert-base-german-cased model and german-gpt2 (<https://huggingface.co/>). For BERT embeddings, we extracted activation based on units from layer 12 and special token “[SEP]”, which marks the end of a sentence. For the GPT2 embeddings, we extracted activation based on units from layer 12 and the final word token, ignoring punctuation. For the SBERT architecture, we used a pre-trained model architecture implemented in PyTorch with the Sentence-transformers module <https://www.sbert.net/>. SBERT is not available in a German-only version, so we used the multilingual model distiluse-base-multilingual-cased, which supports a range of languages including German Reimers and Gurevych 2020a.

## 4.3 Results

The multi-arrangement task provides pairwise distances (dissimilarities) for all item pairs (e.g. pairs of nouns or pairs of sentences). These pairwise distances for  $n$  items can be visualised as a so-called representational similarity matrix (RSM)  $H = n \times n$ , such that each entry in the matrix  $H_{i,j}$  contains the dissimilarity between item  $i$  and item  $j$ . For both groups of 100 subjects separately, we extracted one continuous matrix by first normalising the individual subject RSMs by their standard deviation, and then averaging over all subjects. Figure 4.3 depicts the resulting group averages.



**Figure 4.3.: Mean dissimilarity for nouns and sentences.**

Left: Nouns are sorted (along rows and columns) according to thematic categories (order: medicine, manual labor, sports and music) and within each category the same order as depicted in figure 4.1 is maintained. Right: Sentences are sorted according to the full stimulus list (see appendix or figure 4.2), i.e. all sentences containing communicative verbs and both agent and patient from the category "medicine" first, followed by all sentences containing communicative verbs and agent from category "medicine" plus patient from the category "manual labor" and so on and so forth.

Results for the noun similarity task reflect that subjects easily picked up on the thematic categories and arranged nouns according to their professions. The average dissimilarity within any given category was lower (mean = 1.6, std = 0.08) as compared to the average dissimilarity across categories (mean = 3.3, std = 0.04) and the correlation with a binary, theoretical model of noun similarity, encoding all within-category pairs with a distance of 0 and all across-category pairs with a distance of 1, was high ( $\rho = 0.71$ ). In the event similarity task, the average dissimilarities are most highly correlated with a binary theoretical model of verb category ( $\rho = 0.86$ ). Based on the average it is therefore not clear whether participants took into account all semantic dimensions of the event or rather arranged sentences only based on the verb semantics.

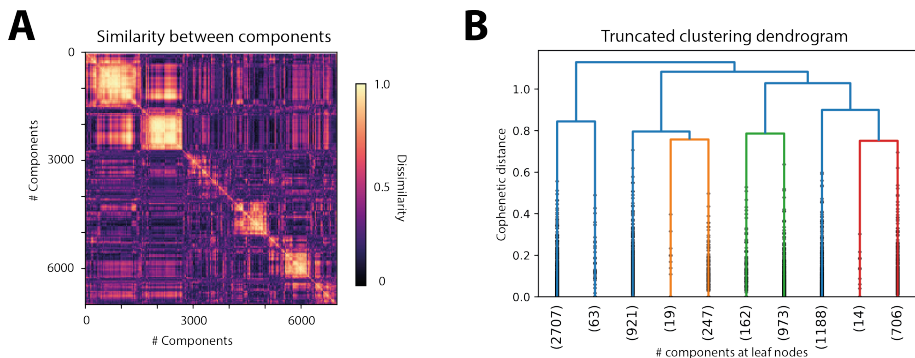
## Factorising high-dimensional representations

Even though it seems as if event similarity ratings are mostly driven by the semantics of the verb, subjects reported multiple strategies for solving the task. Indeed a data-driven factorisation revealed underlying components that influenced similarity judgments beyond verb semantics. We identified five components (see figure 4.5 A), which robustly resulted across 1000 factorisations with random initial weights and together explain more than 95% of variance in the data. The first component reflected the verb similarity (correlation with a binary verb category model was 0.87). The second component did not reflect any of our categorical dimensions, instead after inspection we found that it reflected sorting according to verb identity (see figure 4.5 B left). Component three reflected sorting according to the temporal adverb of each sentence (see figure 4.5 B right). Finally, component five reflected sorting of sentences according to semantic similarity of both the agent ( $\rho = 0.58$ ) and the patient ( $\rho = 0.29$ ) of the sentence. Component four remained elusive but was negatively correlated with the model for verb semantics ( $\rho = -0.36$ ). Based on the mixing matrix (see figure 4.5 C), we observed that similarity according to verb semantic category was weighted highest for most subjects. Nonetheless, the majority of subjects took into account additional semantic dimensions when arranging the sentences, namely the specific verb identity, the temporal information, the semantics of the agent role and the semantics of the patient role.

## Sentence similarity based on ANNs

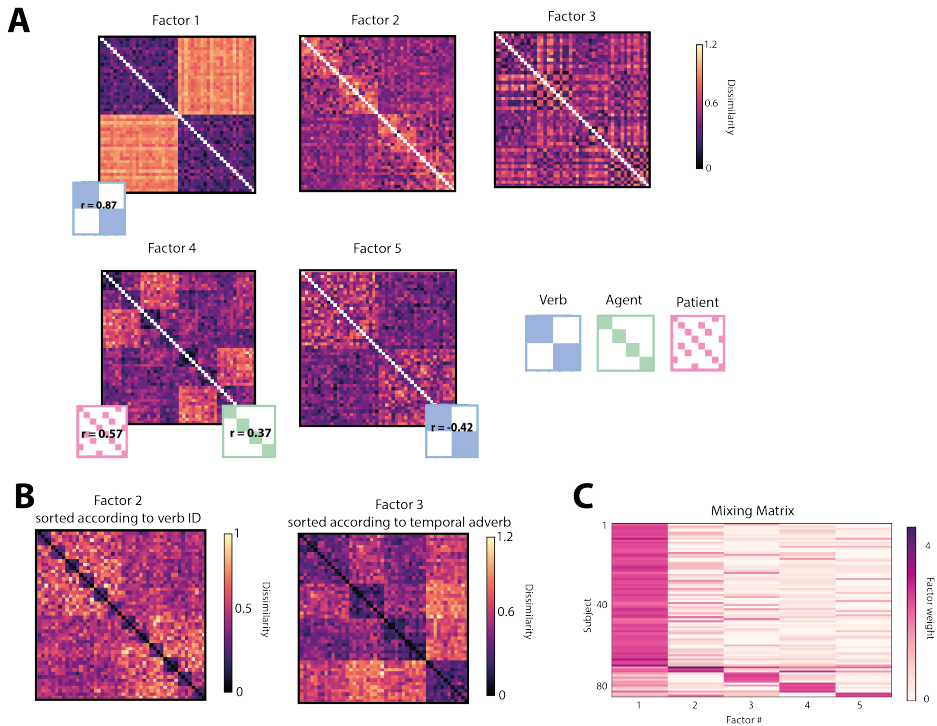
We evaluated the pairwise similarity based on sentence embeddings generated by three ANN models. All models produced embeddings that captured verb, agent and patient categories to some extent (pairwise cosine similarity between sentence pairs was on average higher for items that shared categories than items with different categories across all three dimensions). Nonetheless, we observed differences in the strength with which each model captured different dimensions of event meaning. For example, While the GPT2 model had overall a low fit to our event model, it captured each dimension more or less equally. In contrast, both BERT and SBERT produced embeddings that loaded more strongly on certain dimensions. While BERT embeddings most strongly encoded the agent and the verb dimensions, the embeddings produced by SBERT contain less information





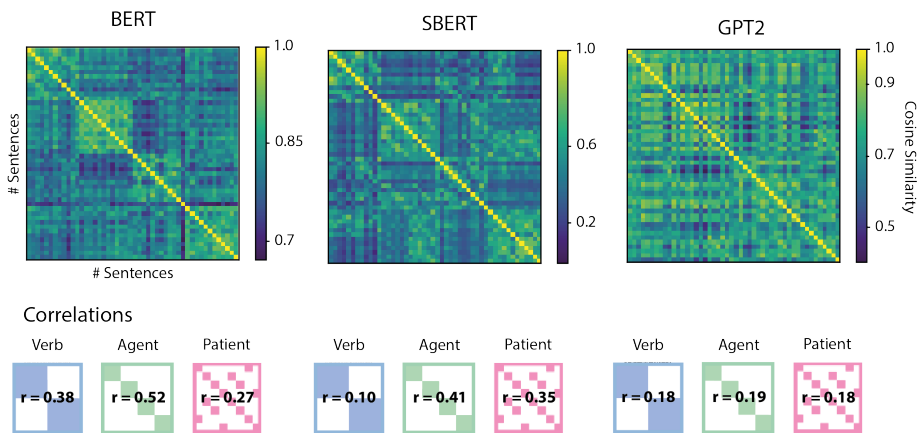
**Figure 4.4.: Clustering of latent factors driving event similarity task results**

A: Pairwise similarity between all components (1000 repetitions of factorisation into 7 components each yields 7000 components in total) is depicted. Lighter colour codes for similarity (absolute Pearson correlation). Components are sorted according to order determined by hierarchical clustering algorithm. B: Truncated dendrogram showing only the last 10 merges across all components. On the x axis, for each cluster the label indicates the amount of leaf nodes (components) belonging to the depicted cluster. Horizontal lines indicate a merge of leaves into a new cluster. The height of the horizontal lines indicates the distance between the merged sub-clusters. As can be seen some sub-clusters merge only relatively few components (e.g. 24 in left branch of orange cluster). The threshold of distance 0.9 for determining clusters was in part motivated to result in roughly equally sized clusters (panel B) that seem most coherent based on their inter- & intra-cluster similarity (panel A).



**Figure 4.5.: Latent non-negative factors.**

Panel A: Each factor is visualised as a sentence-by-sentence matrix ( $n = 48$ ) with sentences in the same order as listed in the full stimulus list (see appendix or figure 4.2), i.e. all sentences containing communicative verbs and both agent and patient from the category "medicine" first, followed by all sentences containing communicative verbs and agent from category "medicine" plus patient from the category "manual labor" and so on and so forth. For each factor its Spearman correlation with the theoretical semantic category models for verb (blue), agent role (green) and patient role (pink) is shown for those semantic dimensions best capturing the similarity pattern expressed by the factor. Both factors 2 and 3 were poorly correlated with any of the semantic category models. Their patterns are visualised by re-ordering sentences according to verb ID (panel B, left) and temporal adverb identity (panel B, right) respectively. The order of the temporal adverbs was the following: "Today"/"This morning", "Earlier", "In the morning" or "Yesterday"/"Yesterday evening". Panel C: Mixing weights are depicted per subject (rows) and factor (columns). Darker colour indicates higher weights. Subjects are sorted according to their maximal factor weight.



**Figure 4.6.:** Representational similarity matrix for all stimulus sentences based on ANN generated embeddings

Pairwise cosine similarity is plotted for each sentence pair  $n_{i,j}$  ( $i = 1,2,\dots,48$  and  $j = 1,2,\dots,48$ ) of the stimulus set and three different ANN architectures. Sentences are sorted according to the full stimulus list (see appendix or figure 4.2), i.e. all sentences containing communicative verbs and both agent and patient from the category "medicine" first, followed by all sentences containing communicative verbs and agent from category "medicine" plus patient from the category "manual labor" and so on and so forth. Pearson correlation coefficients for theoretical models of either verb (blue), agent (green) or patient (pink) semantic category are depicted below each model.

about verb semantics but instead strongly encode the patient role filler. Interestingly, while SBERT is the only model optimised specifically for pairwise sentence similarity judgments (the training goal most resembling the multiple arrangement task), it is also producing sentence embeddings with the worst fit to our observed human behavioural data (see figure 4.6).

## 4.4 Discussion

We applied a geometrical multiple arrangement task to acquire similarity judgments for 24 profession nouns and 48 compositional sentences describing simple transitive events. Similarity judgments revealed a sensitivity to the thematic category of professions when arranging nouns. Although sentences contained those same nouns, the thematic category of profession was less prominent in the sentence-by-sentence similarity judgments. Instead, average sentence similarity

judgments seemed to be highly sensitive to the semantics of the main verb. Matrix factorisation, however, revealed that subjects additionally took into account relational role information when arranging the sentences. Overall, the multi-arrangement task is suited for efficient sampling of similarity judgments even for complex linguistic stimuli. Importantly, it allowed us to quantify high-dimensional mental representations of compositional meaning.

## Context & salience affect similarity judgments

We found the semantics of the verb to be the dominant dimension according to which subjects arranged the sentences in our study. There are multiple explanations for this finding. First of all, the importance of the verb may not come as a surprise given its special linguistic status in the sentence. It has been argued in the past, that subtle features of verb semantics, such as subcategorisation information, immediately affect online comprehension and can even be exploited to predict sentence structure (Hare et al. 2004; McRae et al. 1997). Furthermore, verbs are thought to be linked to so-called “event templates” (Tenny 1994; Jackendoff 1992; McKoon and Macfarland 2002). Event templates formally conceptualise an event by establishing which primitive “event kind” it belongs to and by specifying the syntactic argument positions of its entities in a sentence. In our stimulus material, the selected verbs can be said to instantiate the semantic primitive  $ACT(x,y)$ . This means, that the event involves entity  $x$  acting upon entity  $y$ , where  $x$  and  $y$  map onto syntactic subject and syntactic object respectively. The exact meaning of the verbs will further specify the event, e.g. entity  $x$  is acting upon entity  $y$  through positive, communicative gestures for a verb like “to encourage” or through negative, physical impact for a verb like “to hit”. The verb “to break”, on the other hand, would instantiate a very different event template, i.e.  $CAUSE(a,BECOME\ in\ STATE(x))$ , where entity  $x$  undergoes a change of state (intact to broken) through external force of  $a$ . The two verbs “to break” and “to hit” hence differ in the amount of sub-events that it takes to characterise them. It has been experimentally demonstrated that such event templates are implicitly taken into account as we process sentences. For example, words that are presented in the same template across sentences can prime each other later on (McKoon and Ratcliff 2008) and more complex event templates will slow down reaction times during lexical decisions (McKoon and Macfarland 2002). In the present study, both semantic categories of verbs instantiate more or less similar event templates (one causative

event involving two entities). Therefore it is unlikely that the perceived similarity was affected by differences in event template complexity. Nonetheless, people may have been naturally biased towards attending the verb of a sentence (rather than the nouns) given that it carries crucial information about the event template.

Unfortunately, within our study we cannot distinguish an a priori verb bias from other factors such as valence or context-specific effects such as salience. We specifically chose contrasting verbs that could either express a positive, communicative event or a negative, physical contact event. This difference in valence might have made the verb semantic information more salient as compared to that of nouns. Additionally, in the context of the larger stimulus set, verbs could be broadly divided into only two categories, whereas nouns were more varied, stemming from four distinct thematic categories. The fact that there were fewer semantic categories for verbs may have added to their overall salience.

Alternative models of similarity have been developed to address the effects of salience as well as other supposed shortcoming of the geometric approach such as the assumption of symmetry ( $\text{distance}(A,B) = \text{distance}(B,A)$ ) which has been disconfirmed in empirical data (Tversky 1977). As a solution, Tversky suggested a set-theoretic model, within which similarity between two items is a function of the set of their shared features and the two respective sets of their distinctive features. This approach allows context to modulate how strongly certain features are activated and as a consequence influence similarity, accounting for salience effects. For example, when adding a single item to a set of items, which varies in a specific feature, that so far had been shared by all items. The addition of the new variance in that feature will increase perceived similarity of all original items.

It is correct that geometric models per definition impose certain assumptions such as symmetry onto similarity relations. Similarity judgments collected throughout multiple arrangements, however, are not completely incompatible with observations of asymmetry and salience effects. In fact, the repeated sampling of subsets of the total stimulus set assumes the existence of a multitude of conceptual spaces, some more and some less salient, under which similarity can be defined. Under this assumption, distance from A to B might differ from distance from B to A, if the order of comparison evokes different similarity spaces (Decock and Douven 2011). When combining similarity judgments across subset arrangements those subtle differences get lost and the principle of symmetry will be enforced. Although the final similarity matrix cannot explicitly speak to the

multiple underlying dimensions anymore, we have shown that they can nonetheless be extracted through data-driven factorisation. The same holds true for more or less salient features within a stimulus set.

## Quantifying similarity through vector space models

An alternative approach to quantify semantic representations is based on distributional semantics. Distributional semantics rely on the idea that associated words cooccur in similar contexts. Consequently, given a large variety of contexts, we can find a function that maps each word onto a high-dimensional numerical vector (embedding) capturing a word's association to all other words in the vocabulary. Recently, artificial neural networks (ANNs) have become prominent for generating word embeddings as a byproduct of unsupervised learning tasks. ANNs are usually trained to predict a word based on its preceding or surrounding context and require large linguistic corpus data for training. The most recent generation of ANNs (e.g. [Vaswani et al. 2017](#), [Devlin et al. 2018](#), [Radford et al. 2019](#)) is able to capture contextualised word meanings, taking into account local sentence context. For example, they will assign distinct vector representations to the word "bank" if preceded by either "river" or "money". This newest generation of algorithms excels at multiple natural language processing tasks such as text generation, translation, question answering and cloze tasks ([Brown et al. 2020](#)).

Due to their broad success at language tasks, ANNs have caught the interest of language scientists not only as tools for natural language processing but also as mechanistic models for the brain's language processing. Indeed, several research groups have shown that internal representations in ANNs, emerging during training, predict brain activity above chance ([Pereira et al. 2018](#), [Mitchell et al. 2008](#), [Abnar et al. 2019](#), [Schrimpf et al. 2020](#), [Toneva et al. 2020](#)). Specifically, those models implementing attention mechanisms ([Vaswani et al. 2017](#)), like the GPT2 model, seem to outperform other architectures ([Schrimpf et al. 2020](#)). At the same time, several researchers have raised the criticism that ANNs cannot capture structural relational meaning ([Gershman and Tenenbaum 2015](#), [Puebla et al. 2021](#)). We need focused test-sets, that explicitly probe those semantic dimensions impacting human behaviour in order to evaluate what representations these models are learning and how similar they are to neural representations in humans.

We propose that similarity judgments sensitive to event structure might provide a meaningful benchmark for evaluating ANNs as models for human language processing. Previously, ANNs' have been evaluated not only in terms of their predictability for brain signals, but also with respect to human linguistic behaviour such as for example reading times and gaze-duration (Schrimpf et al. 2020, Van Schijndel and Linzen 2018, Merx and Frank 2020). Measures of perceived similarity can provide an additional benchmark of human behaviour against which computational models could be evaluated. This approach of evaluating computational models based on human similarity ratings has already proven to deliver insights with respect to ANNs trained on visual object recognition (Jozwik et al. 2017, Peterson et al. 2017). For example, researchers are starting to identify which parameters of the model architecture (e.g. layer depth) are crucial for learning human-like representations (Jozwik et al. 2017). We have shown that similarity judgments collected through the multiple arrangement task capture high-dimensional event meaning to some degree. These similarity judgments could therefore serve as a benchmark for human perception of event structure. As an example, we compared contextualised embeddings of our stimuli from three state-of-the-art ANNs: GPT2, BERT and its extension, SBERT (Reimers and Gurevych 2020b). While all ANN models seemed to capture information about event roles to some degree, none of them reproduced the verb bias we observed in our behavioural data. This could suggest that current ANNs might not exploit information in the same way humans do. Whether ANNs are unable to capture human biases needs to be further tested, however, with a larger behavioural dataset and carefully controlled stimuli to exclude valency effects of local context. Finally, a comparison with human behavioural data cannot speak to the predictive performance and utility of ANNs as engineering solutions for language tasks. Instead, similarity judgments provide a benchmark to evaluate the ANNs' utility as mechanistic models of human sentence processing irrespective of their highly successful application as chatbots or for machine translation.

## Conclusion and outlook

In conclusion, we evaluated the multiple arrangement task as a suitable tool for quantifying complex semantic representations. The similarity judgments presented here captured multiple dimensions of event meaning while also being sensitive to biases in human sentence comprehension. Recently, it has been shown

that the multiple arrangement task can be scaled up when combined with wide online distribution (Hebart et al. 2020). In the future, the multiple arrangement task could be used to collect perceived sentence similarity on a larger scale. Such a database could provide an additional benchmark when evaluating ANNs as models for human language processing.



## 4.5 Appendix

### Stimuli

**Table 4.2.:** Full set of sentences

Sentence
Heute Morgen bestärkte der Therapeut einen Sanitäter.
Heute ermunterte der Sanitäter den Pfleger.
Heute Morgen bejubelte der Radiologe den Internisten.
Vorhin ermutigte der Pfleger einen Chirurgen.
Vorhin tröstete der Chirurg den Radiologen.
Vorhin lobte ein Internist den Therapeuten.
Heute Morgen ermunterte der Therapeut den Handwerker.
Heute Morgen lobte der Sanitäter den Mechaniker.
Heute tröstete der Chirurg einen Tischler.
Heute Morgen ermutigte der Internist einen Zimmermann.
Vorhin bestärkte der Pfleger den Klempner.
Vorhin bejubelte der Radiologe den Elektriker.
Am Vormittag bejubelte der Bassist den Pianisten.
Am Vormittag ermutigte der Sänger den Gitarristen.
Gestern Abend ermunterte der Pianist einen Geiger.
Gestern tröstete der Geiger den Sänger.
Gestern bestärkte der Gitarrist einen Musiker.
Gestern Abend lobte der Musiker den Bassisten.
Am Vormittag bestärkte der Bassist einen Läufer.
Am Vormittag lobte der Pianist den Sportler.
Am Vormittag tröstete der Sänger einen Athleten.
Am Vormittag ermunterte der Gitarrist den Fußballer.
Gestern Abend ermutigte der Geiger den Sprinter.
Gestern Abend bejubelte der Musiker einen Boxer.
Heute Morgen schlug der Zimmermann einen Handwerker.
Heute verprügelte der Mechaniker einen Klempner.
Continued on next page

**Table 4.2 – continued from previous page**

Sentence
Heute stieß der Klempner einen Tischler.
Heute Morgen verscheuchte der Tischler einen Elektriker.
Vorhin schubste der Handwerker den Mechaniker.
Vorhin schüttelte der Elektriker den Zimmermann.
Heute Morgen schubste der Klempner einen Sänger.
Heute Morgen verscheuchte der Mechaniker einen Gitarristen.
Heute stieß der Elektriker den Bassisten.
Vorhin schüttelte der Zimmermann den Pianisten.
Vorhin schlug ein Handwerker den Geiger.
Vorhin verprügelte der Tischler den Musiker.
Am Vormittag schüttelte der Sportler den Radiologen.
Am Vormittag schlug der Läufer den Chirurgen.
Am Vormittag verscheuchte der Athlet einen Internisten.
Gestern Abend schubste der Boxer einen Sanitäter.
Gestern stieß ein Sprinter einen Pfleger.
Gestern Abend verprügelte ein Fußballer einen Therapeuten.
Am Vormittag verscheuchte der Boxer einen Läufer.
Am Vormittag schubste ein Sprinter den Athleten.
Am Vormittag schlug der Fußballer einen Boxer.
Gestern Abend verprügelte der Läufer den Sportler.
Gestern stieß der Sportler den Sprinter.
Gestern Abend schüttelte der Athlet den Fußballer.





## Neural dynamics of combinatorial sentence processing

Successful sentence comprehension requires not only the processing of individual words but also their combination into higher-level meaning based on syntactic rules. Without this combinatorial processing we would not be able to extract relational event information such as knowing who did what to whom. Recent cognitive computational models of sentence processing make contradicting predictions about the timing of combinatorial processing in the brain. For example, it is debated whether incoming input can update the overall event representation immediately (within 400 ms after onset) or only late (600 ms after onset) after word-specific processing has been completed. We recorded magnetoencephalography while subjects read sentences describing simple transitive events. Using representational similarity analysis, we tracked multivariate neural correlates of relational event information during reading. Comparing a model for relational event information with a word-specific semantic models, we found that only event information strongly modulated neural activation patterns in inferior frontal, anterior temporal and posterior parietal brain regions. Importantly, we found those areas to encode the event most strongly within time windows as early as 250 ms to 350 ms after sentence-final word onset. Our study provides a detailed description of the spatio-temporal neural dynamics related to processing combinatorial event meaning. The results support language processing accounts that assume contin-

ously unfolding neural event representations to be immediately modulated by any incoming information.

## 5.1 Introduction

To comprehend a sentence, we need to not only understand each word's meaning but also integrate it into a broader event representation. Events in the world are usually composed of entities and actions. For example, when reading a newspaper article about an assault, full comprehension of the event consists in an understanding of who did what to whom. Extracting this information requires combinatorial processing of multiple aspects. One needs to activate not only the action encoded in the verb (e.g. to beat) but also the general roles or arguments this action requires (e.g. an agent and a patient). These general event roles need to be assigned to the specific event participants (e.g. burglar and grandmother) usually based on syntax. Correct role assignment is often crucial for comprehension and, in this case, can make the difference between a tragedy (burglar beats up grandmother) or a hero story (grandmother beats up burglar). In the current study, we investigate the neural basis of combinatorial processing during sentence reading using Magnetoencephalography (MEG). Specifically, we model sentence meaning in terms of semantics and role assignments to capture the temporal dynamics of the brain signal when processing event structure.

Past neuroscience research has identified several brain areas that underly combinatorial processing, including angular gyrus, posterior & anterior temporal lobe as well as left inferior frontal gyrus and ventro-medial prefrontal gyrus (Pylkkänen 2019). Widely used, but broad, experimental contrasts (e.g. sentences vs word lists or sentences vs phrases) have provided conflicting evidence by implicating varying combinations of areas to be actively involved during combinatorial processing (e.g. Matchin et al. 2019, Hultén et al. 2019). The reason why such an extensive and variable network of areas seems to be involved might be due to the loose definition of the term “combinatorial processing”. It implies multiple and potentially overlapping processing steps. Namely, combinatorial processing may refer to verb argument structure activation and role assignment as described above, but may also encompass simple semantic combination (e.g. red apple), syntactic structure building or even more controlled processes such as revision and plausibility evaluation.

Event-related potentials (ERPs) are powerful neural markers that can provide fine-grained information about combinatorial processing, because their temporal response profile depends on subtle stimulus manipulations. Therefore, many

hypotheses have been raised, trying to link temporally specific ERP signatures to cognitive processes (Bornkessel-Schlesewsky and Schlesewsky 2008, Kaan et al. 2000, Baggio and Hagoort 2011, Sassenhagen et al. 2014, Fitz and Chang 2018). While much knowledge has been gained about which linguistic features modulate ERPs, it is still debated in how far they reflect aspects of combinatorial processing. In fact, ERP effects can sometimes be consistent with several alternative underlying cognitive mechanisms. For example, the N400, a robust negative ERP peaking at around 400 ms after word onset, has been reported to be stronger after words that are incongruent with preceding context as compared to congruent continuations (e.g. she likes to drink her coffee with *dog* vs *cream*). However, several decades worth of reports on the N400 have not led to a consensus whether the effect is a marker of pre-activation of the expected continuation (lexical level) or integration difficulties following the incongruent word (sentence level). This is best illustrated in the contrasting predictions stemming from two current computational models of the N400 effect, one by Brouwer et al. (Brouwer et al. 2012, Brouwer et al. 2017), the other by Rabovsky et al. (Rabovsky et al. 2018, Rabovsky and McClelland 2020). Both model the N400 as the update within one of the hidden layers in an artificial neural network model. Brouwer et al. distinguish between two processing steps, early non-combinatorial retrieval and subsequent late integration, with N400 effects being linked strictly to the former. Rabovsky et al, on the other hand, link the N400 to the update of a full probabilistic event representation and hence implicitly assume that combinatorial information, including relational roles, is already available within the N400 processing time window. Both computational models successfully simulate several empirically observed N400 effects, while holding different underlying assumptions about when in time certain combinatorial sub-processes such as role assignment come into play. Therefore, there is a need to complement the currently available ERP data in order to distinguish between these alternative hypotheses.

Multi-variate analyses techniques allow us to capture rich multidimensional information encoded across several channels or source points (Guggenmos et al. 2018) and can thus provide us with additional insights into neural processing above and beyond univariate ERP analysis. One example for multivariate analyses providing new insights into combinatorial processing is a study by Lyu et al. (Lyu et al. 2019). The authors explicitly modelled both general noun semantics as well as context-dependent noun semantics (restricted through the preceding verb). Only



the context-dependent semantics were reflected in the neural patterns and these appeared as early as 240 ms after noun onset. This is evidence for early combinatorial processing and complements univariate analyses of brain data with respect to conceptual combinations (Pylkkänen 2020). Lyu et al.'s findings, however, were limited to semantic combination, as they specifically modulated semantic constraint but kept syntactic structure constant. We aim to extend those findings by investigating combinatorial processing at the syntax-semantics interface. Specifically, we will rely on more arbitrary verb-noun combinations but additionally vary role-filler assignments across sentences. Based on fMRI data, researchers have put forward potential neural correlates of role assignment (Bornkessel et al. 2005) and even achieved to localise fine-grained relational role information through multivariate analysis (Frankland and Greene 2015). Nonetheless, there is little data on the time-course of sentence comprehension including role assignment in the brain. Again, some researchers have evaluated the sensitivity to relational information expressed in ERPs, with mixed evidence for a modulation of the N400 (Frisch and Schlesewsky 2001, Paczynski and Kuperberg 2011).

In this study, we model the semantic content of sentences via item-by-item similarity, taking into account relational roles and use MEG to track neural representations with high temporal resolution. The temporal dynamics of encoded event structure representations can provide evidence for or against alternative computational hypotheses of sentence processing, which can in turn inform the interpretation of ERP components and hence improve our understanding of combinatorial processing.

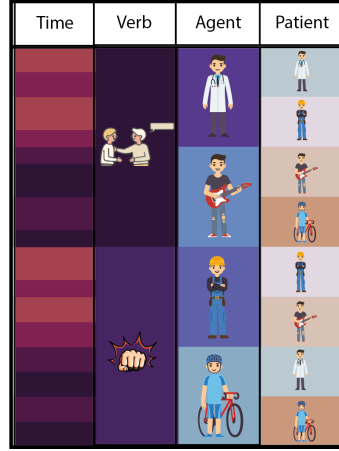
## 5.2 Methods

### Subjects

18 German native speakers participated in the study. Data from two participants was discarded due to below-average accuracy on the behavioral task or extensive muscle activity (see details below). The remaining 16 subjects (8 female) had an average age of 24 years (std = 4 years), were mostly right-handed (3 left-handed), had normal or corrected-to-normal vision and reported no history of neurological, developmental, or language deficits. The study was approved by the local ethics committee (Central Committee on Research involving Human Subjects

**Figure 5.1: Stimulus randomization.**

The diagram illustrates for all sentences (rows) which elements make up any given sentence. Elements are colour coded to indicate the identity of the temporal adverbs (1st column), the semantic category of the verb (2nd column), the semantic category of the agent role (3rd column) and the semantic category of the patient role (4th column). Given any sequence of adverb, verb and agent, the patient semantic category is uncertain, since there are always to categories that occur with equal probabilities.



the local “Committee on Research Involving Human Participants” in the Arnhem–Nijmegen region) and followed the guidelines of the Helsinki declaration. All subjects gave written informed consent before participation and received monetary compensation for their participation.

## Sentence material

### Stimuli creation

For stimuli, we used 48 German sentences, each describing a simple transitive event between one agent and one patient. The stimulus creation is described in detail in chapter 4. During the MEG experiment each unique event was presented in both active and passive voice versions. Each unique sentence was presented four times in total, two times per MEG session (384 trials in total). Importantly, agent and patient identity was balanced across sentences, such that the agent of one sentence would appear as the patient in another sentence. Furthermore, the pairing of agents and patients was pseudo-randomised, such that in the active sentences the semantic category of the final noun (the patient) could not be predicted based on the preceding context (see figure 5.1; more details on the randomisation in chapter 4).

We collected similarity judgments for both the 48 sentences (active voice) as well as the 24 nouns that appeared as agent or patients within the sentences. 200 participants completed a multiple arrangement task, half of them rating the

sentences and the other half the nouns (see details on the task in Chapter 4). In addition, each person participating in the present MEG experiment completed the multiple arrangement task on the sentences within one week after the final MEG session.

## Experimental design

All subjects participated in two MEG sessions. In each session, participants silently read sentences and responded to comprehension questions while their brain activity was recorded. In addition, within one week after the second MEG session, they completed a multiple arrangement task on the same sentences. During the MEG sessions, sentences were presented word-by-word in a rapid sequential manner. Words were presented on a back-projection screen placed in front of the participant (vertical refresh rate of 60 Hz) in a black monospaced font on a grey background. The 96 unique sentences were presented in pseudo-random order and repeated once in each MEG session (total number of trials per participant  $n = 384$ ). The order of sentence presentation was constrained, such that sentence repetitions would be at least five trials apart and any given noun would not be repeated for at least three consecutive sentences. The set of sentences presented in the second MEG session was exactly the same as in the first but in a different pseudo-random order. Due to technical difficulties, one subject saw only 75% of sentences during the first MEG session. Those missing 25% of sentences were added to the second MEG session instead.

Each trial started with a fixation cross. The fixation cross stayed on screen for 1 to 1.5 seconds and participants were instructed to blink during that period. Then, each word appeared for 350 ms with a 550 ms blank period in between words. The final word of each sentence was followed by a blank screen for 2 seconds. Following each sentence, a question was presented. The comprehension question asked to specify either the agent or the patient of the previously read sentence (e.g. Who was encouraged? or Who encourages?). Three possible responses were presented below the question and participants chose one of the options by pressing either the first, second or third button corresponding to the first, second and third response on the screen from left to right. Two of the response options consisted of the agent and patient presented in the previous sentence and the third option was always “Niemand” (engl. No-one). In 50% of trials, the verb included in the question was not the verb actually presented in the previous sentence, hence

“Niemand” was the correct answer. In these trials, when participants correctly selected “Niemand”, a new question with the expected verb would appear and participants could respond again. After every 48 trials, participants could take a self-paced break before advancing to the next block.

## MEG data

We recorded brain activity using Magnetoencephalography (MEG) with a 275 axial gradiometer system (CTF) while participants read the sentences. The signals were analog low-pass filtered at 300 Hz and digitised at a sampling frequency of 1200 Hz. The subject’s head was registered to the MEG sensor array using three coils attached to the subject’s head (nasion, and left and right ear canals). For the second MEG session, subjects were instructed to find a comfortable position in the scanner, that was aligned with the first MEG session head position as closely as possible. Throughout the measurements, the head position was continuously monitored using custom software (Stolk et al. 2013). During breaks, the subject was allowed to reposition to the original position if needed. Subject’s gaze direction and pupil size were continuously recorded using an SR Research Eyelink 1000 eye-tracking device (RRID: SCR\_009602). We acquired T1-weighted magnetic resonance (MR) images of each subject’s brain using 3 Tesla Siemens PrismaFit and Skyra scanners. All scans covered the entire brain and had a voxel size of  $1 \times 1 \times 1 \text{mm}^3$ . A vitamin E capsule was placed as a fiducial marker behind the right ear to allow a visual identification of left–right consistency. Finally, we recorded the subject’s head shape with the Polhemus for better co-registration of MEG and anatomical scans.

Data were pre-processed using the Fieldtrip toolbox in MATLAB (Oostenveld et al. 2011). Each MEG session was preprocessed and source-reconstructed separately. The data was filtered with a DFT filter to remove 50 Hz line noise and its harmonics at 100 Hz and 150 Hz, with a data padding of 10 s. Subsequently, we epoched the data based on word onset and downsampled to 600 Hz. Independent component analysis (ICA) was used to remove artifacts stemming from the cardiac signal and eye blinks. For each subject, the time course of the independent components was correlated with the horizontal and vertical movement as well as blinks as recorded by the eye-tracker. In addition, samples contaminated by muscular activity (20 - 280 Hz) and superconducting quantum interference device jumps were replaced by “Not a Number” before further analysis. Finally, each

trial was de-meant based on a 200 ms baseline window and low-pass filtered at 40 Hz.

We used minimum norm estimation (MNE, [Hämäläinen and Ilmoniemi 1994](#)) to reconstruct activity onto a parcellated cortically constrained source model. Individual cortical sheets were generated with the *Freesurfer* package ([Dale et al. 1999](#), version 5.1) ([surfer.nmr.mgh.harvard.edu](http://surfer.nmr.mgh.harvard.edu)). The forward model was computed using *FieldTrip*'s singleshell method ([Nolte 2003](#)), where the required brain/skull boundary was obtained from the subject-specific T1-weighted anatomical images. Each cortical sheet defined 7842 dipole locations per hemisphere. We computed minimum norm estimates of source activity for all dipole orientations. The current density was estimated using depth-weighting and regularisation with the noise covariance estimated on all presented words ([Dale et al. 2000](#)). Finally, further reduced the dimensionality of the data, by grouping dipole locations into 374 parcels, using a refined version of the *Conte69* atlas. These parcels were used as searchlights in the subsequent analyses.

## Representational similarity analysis

We computed sentence-by-sentence dissimilarity estimates based on both neural dissimilarity (neural RDM) and cognitive models (model RDMs) for both sentences as well as sentence-final nouns. By comparing the neural RDM with each model RDM respectively, we probed when and where in the brain the neural activity reflected the semantic category of the noun or the event model of the entire sentence as participants read the sentence-final noun.

For the cognitive model of noun semantics we coded sentence pairs ending in a noun from the same thematic category with 0 and all other pairs as 1. Note that, although we refer to this model as the noun semantics model, the model is rather coarse and only captures semantic differences related to broad profession categories. For the cognitive model of the full event meaning (event model), we coded a pair of sentences as maximally similar (0) if both sentences shared their verb, their agent and their patient role fillers from the same semantic categories respectively. Further, we coded sentence pairs as increasingly dissimilar (1-3) depending on how many of the fillers belonged to distinct categories across two sentences (see [Table 5.1](#)). In addition, we created a model of lexical identity, encoding words as identical in wordform (0) or not (1). The lexical identity model served as a control for low-level features and was partialled out when

computing correlations between models and neural RDM. Finally, we extracted model RDMs based on the behavioural results from the multiple arrangement task. These RDMs modelled the distance of a sentence pair as either (1) the average perceived similarity between their final nouns (averaged across responses from 100 participants not taking part in the MEG experiment), (2) the average perceived similarity of their underlying events (averaged across responses from 100 participants not taking part in the MEG experiment) or (3) the similarity of the underlying events as indicated by the individual MEG participant's responses.

1	This morning the paramedic praised the electrician. <i>Heute Morgen lobte der Sanitäter den Elektrikern</i>
2	Today the surgeon encouraged the carpenter. <i>Heute ermutigte der Chirurg einen Tischler.</i>
3	Earlier the carpenter chased the electrician away. <i>Vorhin verscheuchte der Zimmermann den Elektriker.</i>
4	Yesterday the soccer player hit the boxer. <i>Gestern schlug der Fußballer einen Boxer.</i>

**Table 5.1.:** **Example sentences.** Sentences are listed in English translation (original German stimuli in italics below). 1 and 2 are considered most similar because thematic categories of agent, patient and verb are shared. 1 and 3 are more dissimilar in comparison, since verbs are of different thematic category, regardless of final noun identity. 1 and 4 are most dissimilar since they do not share any semantic category.

The neural RDM was constructed using a spatio-temporal searchlight approach. We extracted source activity time-locked to the final word of each sentence for all vertices within a parcel and a 100 ms sliding time window (25 ms overlap). The source activity time-courses were averaged over repetitions of identical sentences (4 repetitions per item) and concatenated along both vertex- and time dimensions before calculating the pairwise euclidean distance among all possible sentence pairings. This resulted in a 48 x 48 neural RDM centered at each time point and parcel which was compared against the cognitive model RDMs using Pearson correlation. The output of the comparison consisted of a time course of model fit for each parcel. For the noun semantics model we computed partial correlations, partialling out lexical identity. For the event similarity model we computed two different correlations coefficients: First we correlated the event similarity model while partialling out lexical identity. This partial event model was still somewhat correlated with noun semantics (hence referring to it as event&noun model going forward). Second we correlated the event similarity model with

noun semantics RDM partialled out (event model). A 1-tailed 1-sample t test was conducted at each searchlight with the fits of all participants for a given model RDM to test whether the mean model fit is larger than 0. We performed cluster permutation tests for multiple comparison correction, randomly flipping the sign of individual participant model fits 1000 times (Maris and Oostenveld 2007) and applying a parcel-wise and cluster-wise alpha threshold of  $p < 0.05$ .

## 5.3 Results

### Behavioral

All 16 participants were able to respond to the comprehension questions correctly. The average accuracy across sessions was 92 percent (std = 0.05) and performance did not differ between active or passive sentences (mean difference of 0.008 percent, t-test  $p > 0.05$ ) but improved slightly for the second session (mean difference of 0.04 percent, t-test  $p < 0.05$ ).

### RSA analysis

To reveal the neural dynamics of combinatorial processing, we correlated the dissimilarity captured within cognitive models with the dissimilarity based on the corresponding brain activity. In addition to our model of interest, the event model, we also tested a model for lexical identity, targeting purely bottom-up processing as well as a noun semantics model, targeting word-specific semantic processing. A large, bilaterally distributed network of areas was activated by the sentence-final noun in an item-specific manner (Figure 5.2 A, Lexical identity; see also figure 5.6 for right hemisphere results and 5.8 and 5.9 for medial plots). When evaluating the noun semantics model, we controlled for item-specific activation by partialling out the lexical identity of the stimulus.

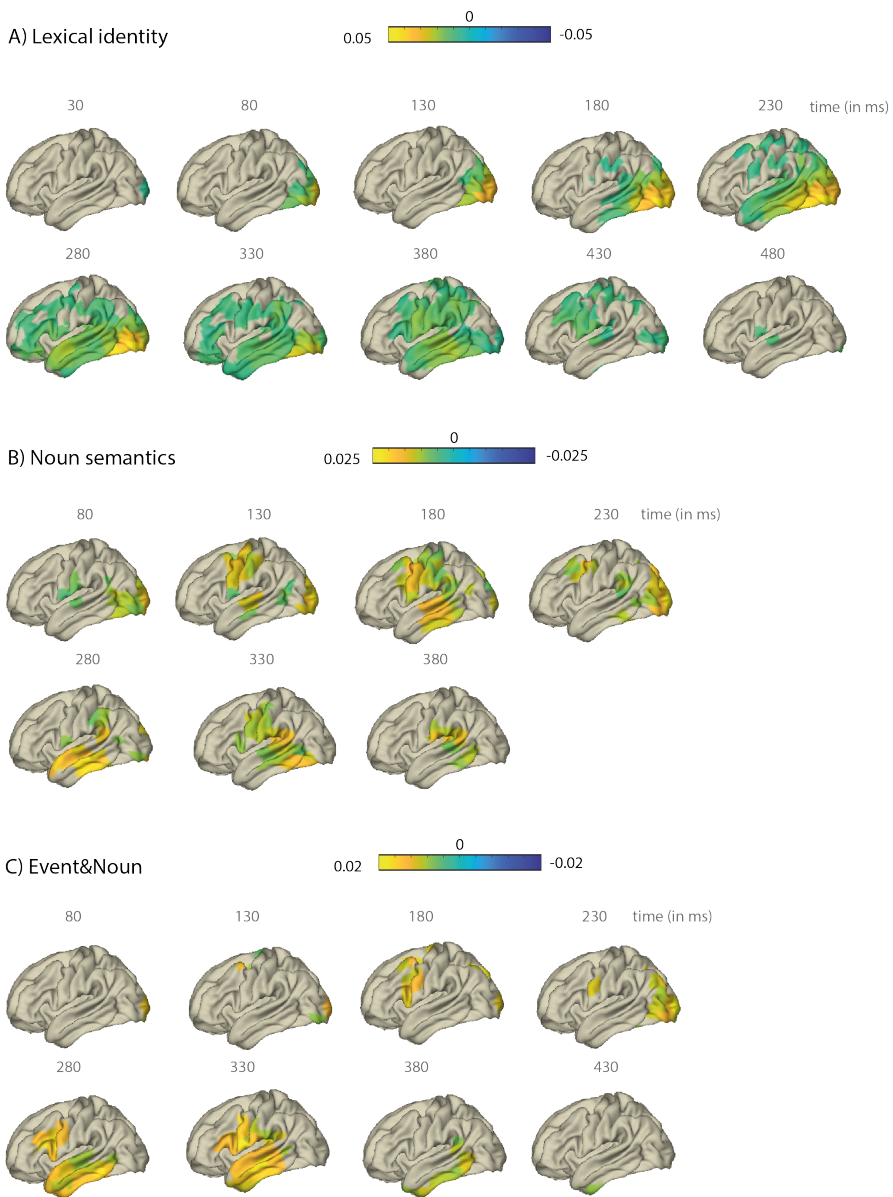
The correlations between the neural similarity patterns and the noun semantics model were significantly higher than chance (cluster-corrected non-parametric permutation test,  $p = 0.002$ ). The earliest timepoints at which brain activity was modulated by noun semantics were between 50 ms and 150 ms after final-noun onset in both lateral and medial occipital cortex as well as superior temporal cortex (Figure 5.2 B Noun semantics). From 130 ms after word onset onwards

noun semantics were additionally encoded in several parietal areas, including motor and somatosensory cortex as well as posterior parts of superior, middle and inferior temporal gyri. Around 280 ms after onset, brain activity in angular gyrus, triangular inferior frontal gyrus and more anterior parts of the superior temporal cortex was modulated by noun semantics.

The event&noun model, with lexical identity partialled out, was also significantly correlated with neural similarity (cluster-corrected non-parametric permutation test,  $p = 0.007$ ). Event similarity modulated neural activity in many of the same regions as observed for the noun semantics model (Figure 5.2 C, see figure 5.9 for right hemisphere results). One caveat is that the event&noun model is correlated with the model for final noun category ( $\rho = 0.54$ ). Since model fits of the event&noun model are overall weaker as compared to the noun semantics model, it might be that noun semantic similarity only is driving the observed effect. We note, however, that the model including the event similarity modulated brain activity in a couple of additional areas, not observed for the encoding of noun semantics. Specifically, in addition to visual and superior temporal areas, event similarity strongly modulated the brain signal extending throughout the entire anterior temporal lobe (ATL), into posterior regions of angular gyrus and superior parietal regions as well as more anterior regions of the inferior frontal cortex (Brodmann area 45). In figure 5.3, we have summarised which areas belong to the supra-threshold cluster with highest correlations for both model RDMs separately and indicated areas for which effects differ.

In order to account for the collinearity between the noun semantics model and the event&noun model we additionally computed partial correlations between event model and neural RDM, partialling out noun semantics. This partial correlation was not significant at any point in space or time in our whole-brain analysis, after correcting for multiple comparisons (see figure 5.4 & 5.7 for the uncorrected results,  $p < 0.05$ ). Nonetheless, we inspected the peak latency of the correlations for some indication of the time-course of combinatorial processing. We observed the highest correlations after sentence-final noun onset in a parcel in middle temporal cortex, with two prominent peaks, one around 325 ms ( $\rho = 0.015$ ) and another around 550 ms ( $\rho = 0.02$ , see figure 5.5). Furthermore, We were interested in peak correlations of those areas, which we identified earlier as being modulated by event similarity but not necessarily noun semantics in the previous comparisons controlling for lexical identity only. These areas included

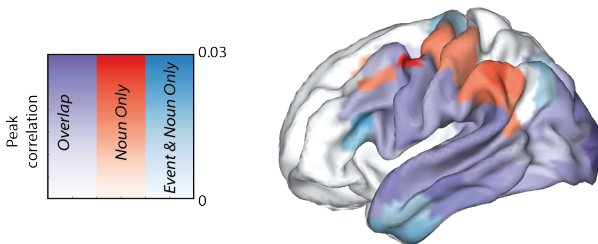




**Figure 5.2.: Spatio-temporal maps of model correlations.**

Figure caption on next page.

**Figure 5.2.:** Pearson correlation coefficients are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Time labels indicate the center point of the temporal searchlight window, which was 100 ms wide and had 25 ms overlap with neighbouring time windows. (A) The first row shows correlation of neural RDM with the RDM for lexical identity (i.e. wordform identity). (B) The second row shows partial correlation of neural RDM with Noun semantics RDM, partialling out lexical identity RDM. (C) The third row shows partial correlation of neural RDM with event similarity RDM, partialling out lexical identity RDM. All spatio-temporal maps are masked for significance (non-parametric permutation test with cluster-based correction,  $p < 0.05$ ).

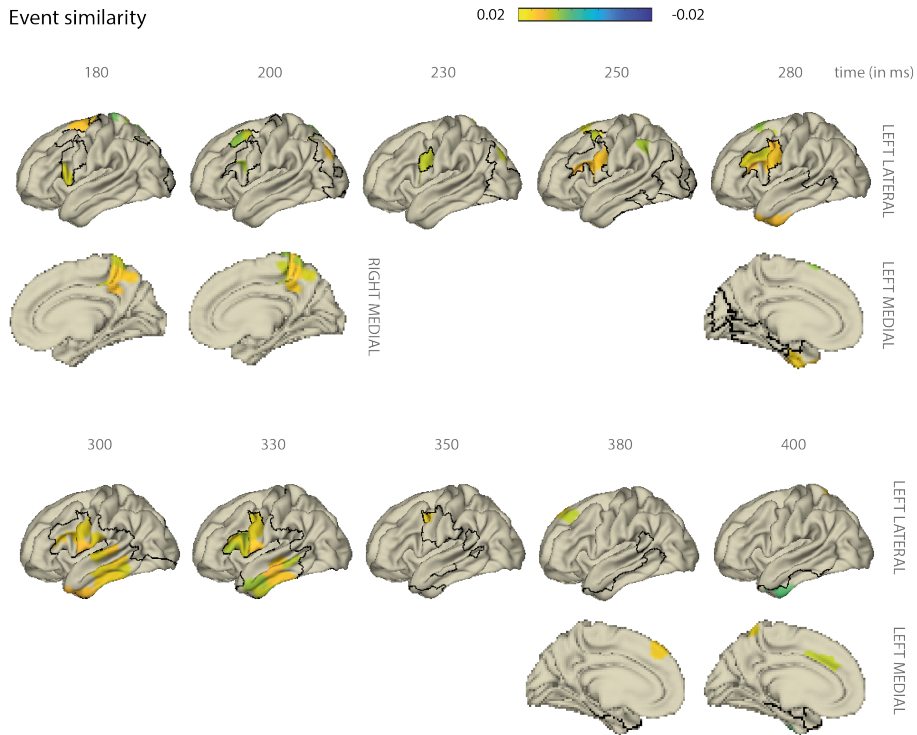


**Figure 5.3.:** Spatial overlap between noun semantics model and event&noun model.

Peak correlations are plotted per parcel, strength of correlation is indicated by color intensity. In a given parcel, the color indicates, whether correlation strength was significantly higher than 0 for the noun semantics model only (red), for the event&noun model only (blue) or for both (purple). In red, peak correlations for the noun semantic model are plotted. In blue, peak correlations for the event&noun model are plotted. In purple, the maximal correlation value taken from either models is plotted.

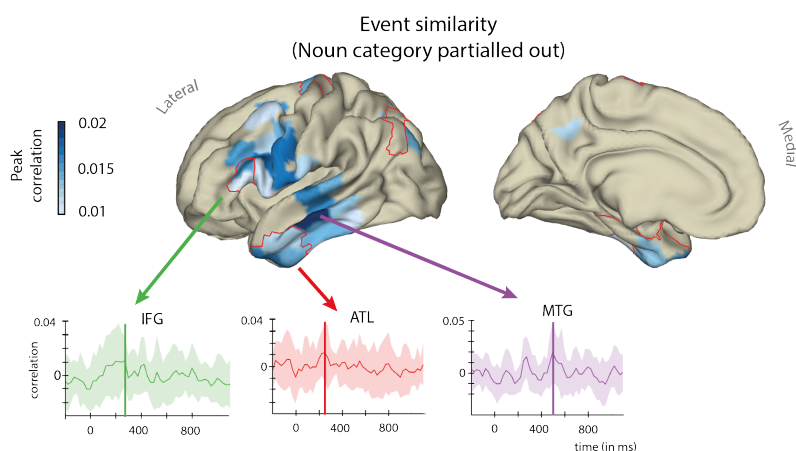
the angular gyrus, inferior frontal gyrus and ATL. When controlling for noun semantics, correlations in angular gyrus were no longer noteworthy. Both ATL as well as inferior frontal gyrus, however, were still modulated by event similarity, peaking around 300 ms ( $\rho = 0.01$ ) and 325 ms ( $\rho = 0.01$ ) after onset respectively ( $p < 0.05$ , uncorrected).

Finally, the behavioral RDM, modelling perceived similarity of noun semantics based on the multiple arrangement task was significantly correlated with the neural data. The spatio-temporal dynamics of the correlations were largely overlapping with the results from the theoretical model for noun semantics (see figure 5.10 in appendix). This was to be expected, given the high correlation between the theoretical and the behavioral model for noun semantics. Correlation strength, however, was slightly stronger as compared to the correlation with the theoretical model for noun semantics. The other behavioural RDMs, modelling



**Figure 5.4.:** Spatio-temporal partial correlation maps for the event model.

Pearson partial correlation coefficients between neural RDM and event model are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Results are based on partial correlation, partialling out the noun semantics RDM. Time labels indicate the center point of the temporal searchlight window, which was 100 ms wide and had 25 ms overlap with neighbouring time windows. Spatio-temporal maps are shown only for those time points for which the event&noun model was significant and masked for  $p < 0.05$  (non-parametric permutation test uncorrected). Black outlines indicate the spatial extent of the supra-threshold cluster for the event&noun model effects.



**Figure 5.5.: Overview peak correlations for event model.**

Parcel-wise peak correlations are shown for the event model. Colour codes for strength of correlation. Blue outlines indicate areas which exhibited an effect of correlations with the event&noun model but not the noun semantics model. For three areas of interest, middle temporal gyrus (mTG), triangular part of the inferior frontal gyrus (IFG) and anterior temporal lobe (ATL), the timecourse of the average correlation coefficients are plotted in blue, green and red respectively. Shaded areas indicated the standard deviation at a given time point. The point of maximum correlation (marked by vertical lines) appeared at 300 ms (ATL), 325 ms (IFG) and 550ms (mTG) after sentence-final noun onset.

perceived similarity of event semantics were not statistically significant after correcting for multiple comparisons. When inspecting the uncorrected results, peak activation was more spatially scattered and more fleeting as compared to the theoretical model results (see figure 5.11 A & B in appendix), making an interpretation difficult.

## 5.4 Discussion

In this study, we investigated the spatio-temporal dynamics of brain activity related to combinatorial processing. Specifically, we targeted comprehension of event structure, which requires assignment of entities into common event roles such as agent and patient. To this end, we modelled sentence similarity, taking into account both semantic and syntactic cues. We then probed where in the brain and how fast after the sentence-final noun this sentence similarity was reflected

in MEG-recorded brain signals. Since we were interested in neural processing of sentence meaning beyond the single word level, we explicitly compared the results of the event similarity against a model of only the final noun semantics. In direct comparison, both models have their maximal fit to neural data within several parcels across visual cortex, inferior frontal cortex and temporal cortex. In addition, we observed a significant effect of the event model in several areas for which the noun semantics model was not consistently correlated, namely, parts of the inferior IFG, ATL and angular gyrus. This finding is consistent with prior research implicating those three areas in the processing of combinatorial language. When controlling for noun semantics, the event model fit was still consistently higher than zero in middle and superior temporal gyri as well as IFG and ATL. Finally, when inspecting the time-course within those areas, we observed maximum model fits within the first 400 ms after sentence-final noun onset. This observation, conflicts with serial accounts of sentence processing, that maintain a strictly non-compositional nature of early processing up to 400 ms after word onset (Brouwer et al. 2012, Brouwer et al. 2017). Instead, it seems that combinatorial processes occur early and presumably simultaneously with emerging word specific activation patterns. This pattern of simultaneous activation of both word-level and sentence-level representations is compatible with neural processing accounts assuming continuously unfolding activity patterns that are immediately modulated by any incoming information (Baggio and Hagoort 2011, Rabovsky et al. 2018).

Our findings are also partly in line with a recent concurrent fMRI and MEG study on semantic composition (Lyu et al. 2019). This study provided evidence for early effects of semantic combination as well as early directed connectivity from left IFG to left middle temporal gyrus (LMT). Lyu et al. presented their participants with spoken sentences, each with a verb phrase containing a direct-object noun. They report that a model for verb and noun semantic interaction was significantly correlated to activity patterns evoked by the direct object noun, first in ventral IFG (BA 47) and at later time point (360 ms post onset) in lateral IFG (BA 45). In our data, the event model was highly correlated with both anterior (BA 45) and posterior (BA 44) parts of the lateral IFG. In contrast to Lyu et al., however, we do not find effects in the most ventral parts of the IFG (BA 47), neither in our model for noun semantics nor for our event model. The inferior frontal gyrus has previously been implicated in syntactic processing, specifically during integration

of words into a sentence structure (Snijders et al. 2009, Hagoort 2017) and in the presence of long-distance dependencies (Leiken et al. 2015) but also for effortful or flexible semantic processing (Binder and Desai 2011). Lyu et al. proposed activation in BA 45 to reflect control processes related to selection of contextually relevant semantic properties. This interpretation is also supported by a large meta-analysis of more than 80 studies, which has identified a dissociation between BA 45 and BA 44 for semantically and syntactically demanding stimuli respectively (Hagoort and Indefrey 2014). Based on previous paradigms it was not obvious whether either of these areas would also be sensitive to event structure, which results from an interaction of syntactic and semantic cues. Our data indicate that processing of event structure involves both areas, with the more anterior BA 45 reflecting event structure but not necessarily word-level semantics. The fact that, unlike Lyu et al, we find effects of combinatorial processing in BA 44 may be related to additional syntactic processing demands in our experiment. That is, we varied active and passive voice across sentences such that participants had to pay attention to syntactic cues in order to reach full comprehension of the underlying event. Fine-grained spatial differences in activation within IFG need to be interpreted with caution, however, given the limited spatial resolution of MEG as well as individual variation in underlying cytoarchitectonic profiles.

There were several other areas encoding the event, including bilateral angular gyrus and left ATL, which have previously been implicated in combinatorial processing (Pylkkänen 2019). The angular gyrus is thought to be part of a semantic memory system. Specifically, it seems to be involved in representing event concepts (Binder and Desai 2011) with a causal role in the integration of lexical-semantic information (Price et al. 2016). In our data, we observe the peak model correlations in left angular gyrus to be weaker in comparison to other left-hemispheric areas and only the effect in the right hemisphere to survive when controlling for noun semantics. Although evidence for a hemispheric dissociation for angular gyrus regarding its role in combinatorial processing is mixed (Graves et al. 2010, Williams et al. 2017), prior studies have mainly relied on minimal phrases for stimuli. Our MEG data, evoked by full sentences, supports the claim that event-relational information is mainly processed within right angular gyrus, while lexical semantics primarily engage the left-hemisphere homolog. Concerning the role of the ATL, most evidence for its role in combinatorial processing stems from phrase-level combinations. Pylkkänen et al. have argued, based on a series

of experiments, that the ATL is sensitive to conceptual specificity of lexical combinations rather than syntactic combination (e.g. “Indian food” being more specific than “Asian food”, [Pylkkänen 2019](#), [Zhang and Pylkkänen 2015](#)). Furthermore, it seems that ATL is most sensitive to highly associated phrase combinations, while the LMT tends to respond more to combinations with low associations ([Li and Pylkkänen 2020](#)). This might explain, why our arbitrary verb-noun combinations were best encoded in the middle portion of the temporal gyrus. In addition, we show an effect of event structure in left ATL with a similar temporal profile (peak around 300 ms) as previously reported effects. Therefore, we suggest that the ATL’s role in combinatorial processing should not be seen as restricted to basic phrasal combination.

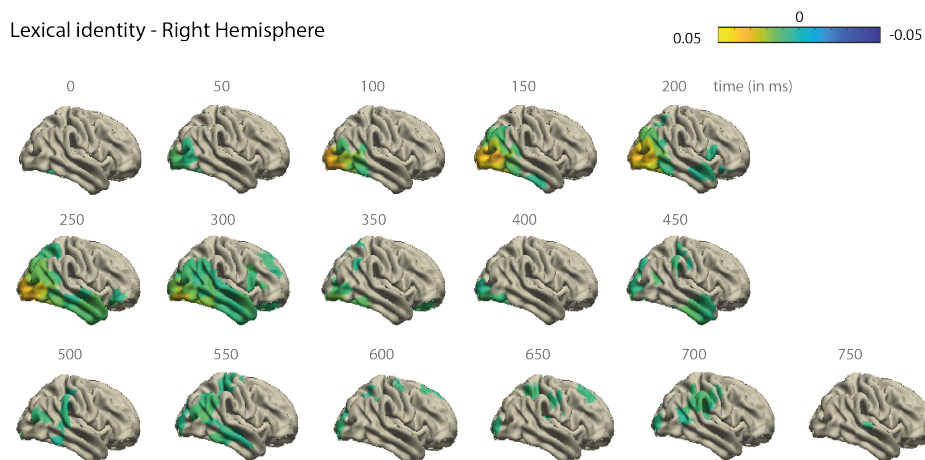
One potential restriction of our study, is the limited vocabulary size used for generating our stimuli. The size was restricted by a combination of considerations: We wanted to sample unique items repeatedly to be able to extract meaningful underlying activity patterns. At the same time, our goal was to reduce the semantic predictability of role-fillers. Although predictive processes play an important role in sentence processing, the ability to combine arbitrary words into novel, unexpected combinations, i.e. combinatorial processing, is often seen as a uniquely human and powerful capability. Therefore, we ensured that fillers could plausibly occur with any of the verbs and would appear equally often in both agent or patient roles. We recorded two sessions for each subject to accommodate both the repetitions of items as well as the fully crossed pairing of roles and fillers while also ensuring the comfort of our subjects sitting in the scanner. Nevertheless, these constraints limited the number of unique sentences we could present and may have affected the temporal dynamics of neural activity in our experiment. For example, recognition of words on the screen may have been accelerated due to fewer competing candidates, once the participants had recognised the four dominant thematic categories. However, during natural language processing, context often provides a wealth of information that can equally pre-select a restricted pool of potential themes. It remains to be seen, whether the here observed temporal dynamics generalise to more naturalistic environments.

In conclusion, we show that combinatorial representations, including information about event roles, can be encoded early in the brain signal, within the first 400 ms after final word onset. We find activity related to event structure within

several left-lateralised brain areas, including the middle & superior temporal gyrus, lateral inferior frontal gyrus, the anterior temporal lobe and potentially angular gyrus. This study complements previous univariate approaches to probe the neural dynamics of combinatorial processing and provides new insights into the time-course of processing fine-grained event information. Importantly, it provides evidence against a strict temporal separation of word-specific and integrated sentence-level processing.

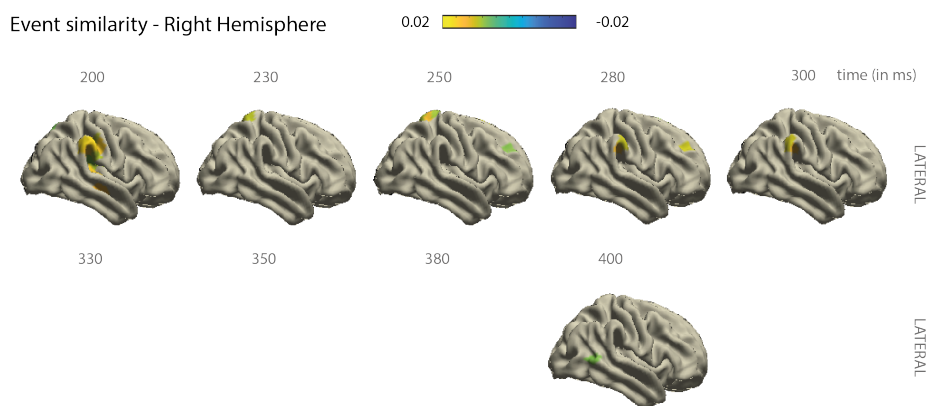


## 5.5 Appendix



**Figure 5.6.:** Right hemisphere correlation with lexical identity RDM.

Pearson correlation coefficients are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Time labels indicate the center point of the 100ms wide temporal searchlight window. All spatio-temporal maps are masked for significance (non-parametric permutation test with cluster-based correction,  $p < 0.05$ ).

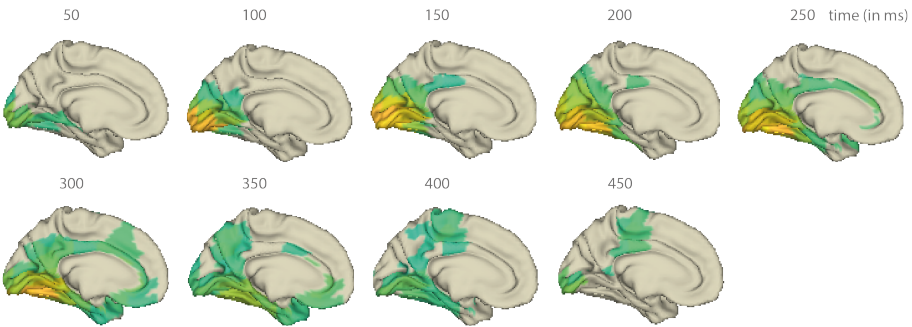
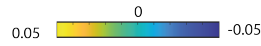


**Figure 5.7.:** Right hemisphere correlations with event similarity RDM.

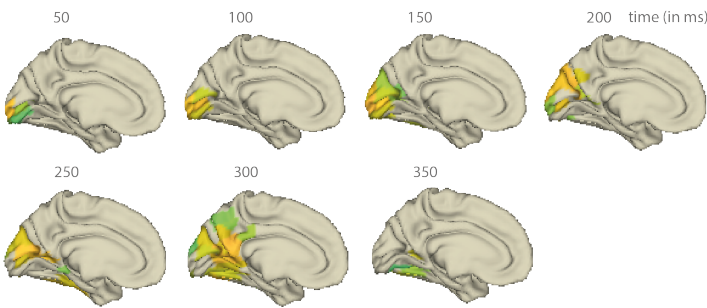
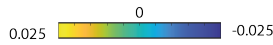
Pearson correlation coefficients are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Time labels indicate the center point of the 100ms wide temporal searchlight window. All spatio-temporal maps are masked for significance (non-parametric permutation test uncorrected,  $p < 0.05$ ).

Medial View - Left Hemisphere

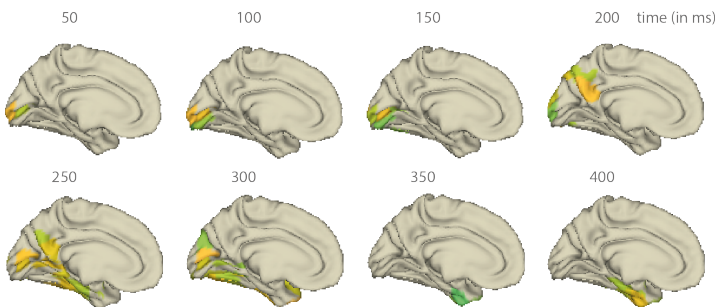
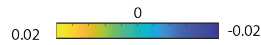
A) Lexical identity



B) Noun semantics



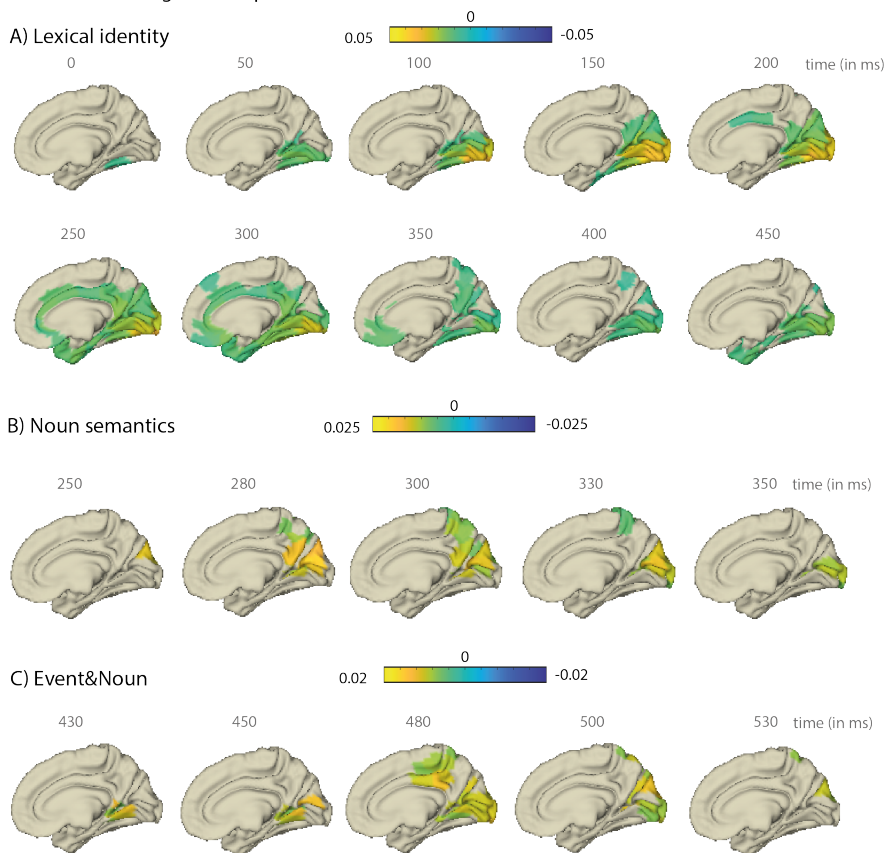
C) Event&Noun



**Figure 5.8.:** Spatio-temporal maps of model correlations - left medial brain view.

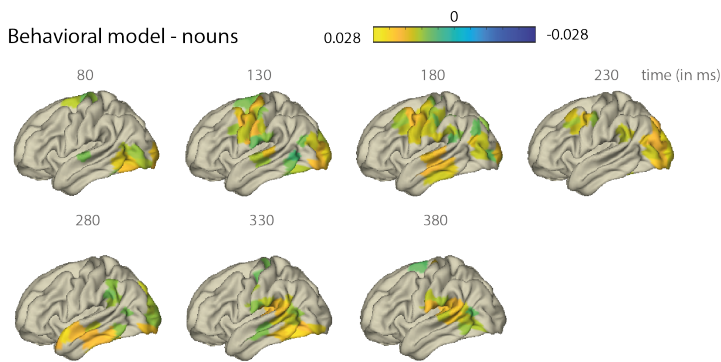
Pearson correlation coefficients are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Time labels indicate the center point of the 100ms wide temporal searchlight window. Correlations are shown for three different model RDMs, namely lexical identity RDM (A), Noun semantics RDM with lexical identity partialled out (B) and event+ RDM, with lexical identity partialled out (C). All spatio-temporal maps are masked for significance (non-parametric permutation test uncorrected,  $p < 0.05$ ).

Medial View - Right Hemisphere



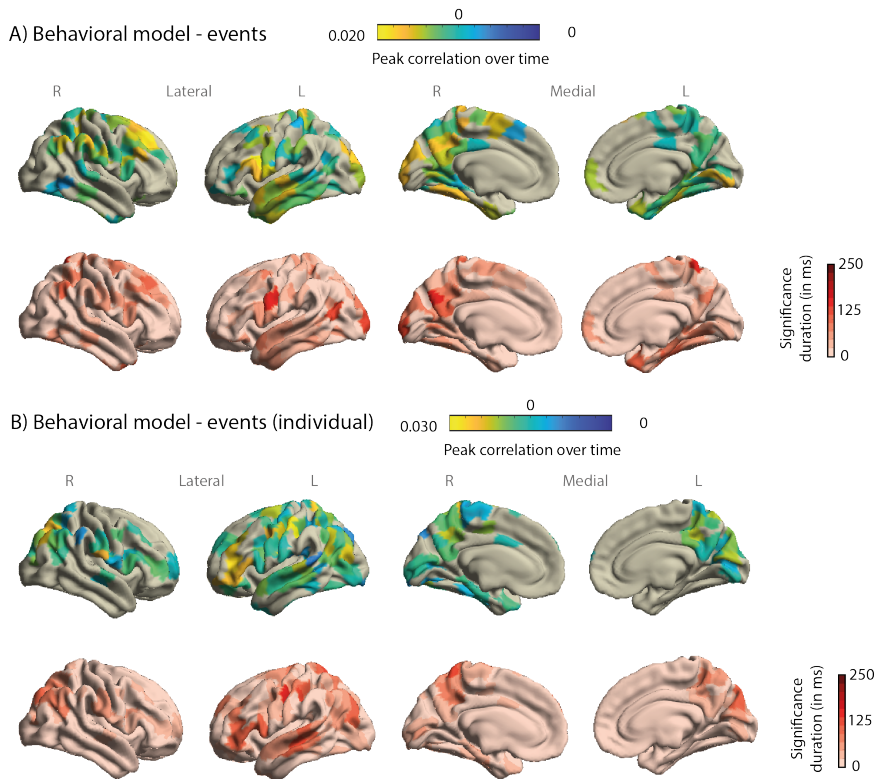
**Figure 5.9.:** Spatio-temporal maps of model correlations - right medial brain view.

Pearson correlation coefficients are plotted for each spatio-temporal searchlight. Color codes for strength of correlation. Time labels indicate the center point of the 100ms wide temporal searchlight window. Correlations are shown for three different model RDMs, namely lexical identity RDM (A), Noun semantics RDM with lexical identity partialled out (B) and event+ RDM, with lexical identity partialled out (C). All spatio-temporal maps are masked for significance (non-parametric permutation test uncorrected,  $p < 0.05$ ).



**Figure 5.10.:** Spatio-temporal maps of behavioral model fit for noun semantics.

Average perceived similarity between sentence-final nouns (non-MEG participants): Spatio-temporal map of Pearson correlation coefficients are masked for significance (non-parametric permutation test with cluster-based correction,  $p < 0.05$ ). Time labels indicate the center point of the temporal searchlight window, which was 100 ms wide and had 25 ms overlap with neighbouring time windows.



**Figure 5.11.:** Spatio-temporal maps of behavioral model fit for event semantics.

Results are shown for the two behavioral models. (A) Average perceived similarity of the event semantics (non-MEG participants) and (B) individual perceived similarity of events (MEG participants): Behavioral RDMs for event semantics were not significantly correlated. Summary statistics are plotted in the form of peak correlation across time (rows 3 & 5, masked for significance uncorrected  $p < 0.05$ ) and duration of uncorrected significance (rows 4 & 6, red).







## Discussion

The goal of this thesis was to investigate the neural correlates of abstract mental representations of language during sentence comprehension. By combining several MVPA techniques, such as multiset canonical correlation, classification and representational similarity analysis with MEG-recorded and source-reconstructed neural data, I investigated both the spatial as well as the temporal dynamics of modality-independent sentence processing in general and the processing of sentence and event structure more specifically.

In **Chapter 2**, I investigated the spatiotemporal dynamics of sensory modality-independent processing by comparing neural signals of subjects either reading or listening to sentences. I effectively applied multiset canonical correlation analysis to align brain signals across multiple subjects and thereby boosted aspects in the signal that were common to all. Specifically, I demonstrated that such spatial alignment across subjects makes it possible to capture subtle word-by-word fluctuations in the neural signal. I found that modality-independent processing is supported by a widely distributed network within the left hemisphere including classical language areas such as left prefrontal, superior and middle temporal areas, and anterior temporal lobe, but also parts of the control network as well as subcentral and more posterior temporal-parietal areas. Modality-independent sentence processing started in temporal areas but rapidly spread to the other regions involved.

In **Chapter 3**, I asked whether the brain activity evoked during reading reflects the structural interpretation of a sentence. Specifically, subjects read sentences that were syntactically ambiguous with respect to the structural attachment of a prepositional phrase. In a rating study, we confirmed that people were in-

deed able to disambiguate the prepositional phrase attachment of these sentences based on semantic information. Based on MEG recordings, however, we were not able to find distinct neural activation patterns for distinct structural attachment. We demonstrated that the MEG data carried sufficient information about word-specific features to classify both their syntactic (part-of-speech) as well as semantic content. This suggests that subjects may not fully parse the structure of sentences when reading only for comprehension.

In **Chapter 4**, I tested in how far behavioural measures of similarity can serve as quantitative models of combinatorial sentence meaning. To this end, I collected similarity judgments for simple transitive sentences by means of a geometry multiple arrangement task. I showed that the similarity judgments captured multiple dimensions of event meaning but also reflected verb biases in human event perception. I further demonstrated how similarity-based representational models of sentence meaning can serve as a benchmark for current artificial neural network models. Based on the example of three state-of-the art ANNs, I exemplify that such models might fall short in fully capturing task-induced biases observed in humans.

In **Chapter 5**, I continue to use similarity based models of simple transitive sentences to investigate the neural correlates of combinatorial processing. Using RSA on source-reconstructed MEG data, I show that both word-specific as well as event meaning are represented in distributed, left-lateralised brain networks, that overlap in middle superior temporal cortex as well as some frontal and inferior parietal areas. Furthermore, I find that event meaning is additionally represented in brain areas that don't strongly code for individual noun semantics, namely, anterior parts of the IFG, the anterior temporal lobe and the angular gyrus. Within those areas, combinatorial event meaning is encoded early in the brain signal, within the first 400 ms after sentence-final word onset. This pattern of simultaneous early activation of both word-level and sentence-level representations is compatible with neural processing accounts assuming continuously unfolding activity patterns that are immediately modulated by any incoming information.

## 6.1 Temporal variability of brain states

Our results in Chapters 2 & 5 illustrate that the neural activity in response to invariant stimulus features is highly dynamic. Abstract neural representations

seem to be encoded in complex spatiotemporal dynamics rather than by single fixed transient activation pattern that is either “on or off”. This observation is in line with dynamic coding frameworks in which information is encoded in dynamic transitions between activation “states” (Stokes et al. 2015). Furthermore, it has recently become possible to specifically test the temporal robustness of neural activation patterns by generalising decoding of stimulus features over time (“temporal generalisation method”, King and Dehaene 2014). Studies relying on this method, confirm that abstract stimulus features are encoded through a cascade of transient neural processes, hierarchically unfolding over time (Gwilliams and King 2020). Therefore, taking into account the temporal dynamics of the neural signal should increase the decodability of abstract stimulus features. This has been nicely illustrated, for example, in the study of the olfactory system, where odours have been shown to be best decoded from spiking neural activity when the time-course of the response was taken into account (Mazor and Laurent 2005). Similarly, in Chapter 3 we show that syntactic information is better decoded when classifiers are trained on concatenated rather than averaged time points. Throughout my thesis, I leveraged the high temporal resolution of MEG recordings and took both temporal and spatial dimensions into consideration when probing multivariate neuronal patterns. To this end, I applied MVPA on shifting time windows of MEG data.

While the high temporal resolution of MEG in principle allows to capture information encoded in dynamic transitions, trial-by-trial variability along the temporal dimension can also pose a challenge. Specifically, by aligning shifting time windows of activation across trials, I have assumed the neural correlates of abstract representations to be somewhat synchronous. While onset profiles of neural activity usually demonstrate high synchronisation for low-level sensory information, modality-independent and higher-level processing, might be less synchronised by external stimulation. Indeed, recent studies, that have estimated functional brain states using data-driven approaches, report high temporal variability of those states at the range of hundreds of milliseconds (Vidaurre et al. 2019). Similarly, it has been shown that multivariate approaches to subject alignment improve sensitivity to trial-specific fluctuations even more when taking into account temporal in addition to spatial variability of the MEG signal (Huizeling et al. 2020).

Trial-by-trial variability could potentially explain our reduced sensitivity for event semantic effects at the group-level in Chapter 5. There, we observed clear peaks in the model fit of individual subjects. Nonetheless, because the latencies of those peaks were slightly jittered in time from subject to subject, effects on the group level were rather weak. It is possible, that trial-by-trial variability of brain states at the individual trial level, has caused those inter-individual differences. In the future, MVPA analyses of linguistic representations will profit from estimating functionally relevant brain states on a trial-by-trial basis and extracting neural activity from state-aligned time windows.

## 6.2 Task modulation on language processing

Given how quickly children learn a language without formal instruction, it is tempting to assume that the regularities of language are extracted effortless and automatic. This, however, doesn't necessarily seem to be the case. During learning, people often seem to require explicit instruction in order to efficiently generalise rules beyond a known context (Detterman 1993). Indeed, once explicit instructions are provided, humans have been shown to outperform artificial neural networks in generalising abstract rules given only few exemplars (Lake and Baroni 2018, Lake et al. 2019). Even then, however, reaction times on a single trial can take up to a full minute (Johnson et al. 2021), a time scale that doesn't match the rapid process of language comprehension. Similarly, evidence from artificial grammar learning paradigms (AGL) suggest that extraction of abstract sequential structure might actually be somewhat hard. AGL paradigms usually eliminate any language-specific lexical information by presenting rule-generated sequences of non-words. Although several studies have reported successful learning of underlying grammatical rules, it has also been argued that the more complex structures are not necessarily learned automatically nor at an abstract level (Poletiek 2002, Lai and Poletiek 2011, Wilson et al. 2020, Conway 2020). During natural language comprehension specifically, it has been questioned whether people always automatically extract hierarchical phrase structure when processing sentences (Frank et al. 2012).

Redundant information in the language input may enable comprehension without explicit representation of abstract structure. Instead of computing a full structural parse of a sentence, people have been shown to sometimes rely on an un-

derspecified, or “shallow” interpretation for comprehension (Ferreira and Patson 2007). Shallow processing may often be sufficient due to the high redundancy of information in the language input. For example, often semantic and syntactic cues enforce sentential structure equally. When reading “the dog bit the man”, both the word order as well as the semantic world knowledge (i.e. dogs being more likely to bite people than vice versa) reinforce the interpretation of the sentence. This redundancy of information is also the reason why patients suffering from Broca’s aphasia have seemingly intact comprehension abilities while being strongly impaired in their production of grammatical sentences (Caramazza and Zurif 1976). By leveraging semantic information, they can somewhat compensate deficits in syntactic processing. Beyond mere redundancy, recent research extending the AGL paradigm, suggest that semantic information may be crucial for recognition of categories and grounding thereof in world knowledge, which in turn allows more generalised learning of statistical dependencies (Poletiek et al. 2021).

In light of the above observations, it seems that language processing can occur at varying levels of depth and effort depending on the current internal goal of the receiver. For example, when engaging in smalltalk with my neighbour on the hallway, extracting the “gist” of what is being said, i.e. shallow processing, might be sufficient. When completing an assessment test for a job application, however, I will analyse each sentence more carefully. Therefore, we need to take into account the processing goal of our subjects when investigating their neural correlates of sentence structure in the lab. Passive comprehension of sentences may not provoke a deep enough processing of the stimuli. Rather, in order to elicit neural representations of sentence structure it might be necessary to apply a behavioural task that explicitly probes structural knowledge. In the current work, we had more success revealing sentence-level representations when applying a behavioural tasks which emphasised the dimensionality of interest. Specifically, in the experiment presented in Chapter 3 only 25% of trials were followed by a general comprehension question and we did not find any evidence for neural representations of hierarchical phrase structure. In contrast, in Chapter 5, we found evidence for neural encoding of event semantic structure when subjects were probed for their comprehension of event roles after every single trial. Importantly, perceived event similarity as measured through a different behavioural task did not model the neural code for event structure as well as a theoretical model giving equal weights to all elements of the sentence. This was likely because the similarity

judgment task resulted in strong focus towards the main verb, whereas the task performed in the MEG scanner in addition emphasised the event roles. Taken together, the experiments in this thesis thus highlight the non-uniform nature of language comprehension. In the future it will be important to investigate in more detail which sentence comprehension tasks can reliably elicit abstract structural representations in the brain signal.

### 6.3 How domain general are abstract neural representations?

The high-level abstract mental representations I investigated in this thesis, although evoked through language stimuli, need not necessarily be language-specific. For example, many earlier studies have compared hierarchical structure of language and music (Zuidema et al. 2018) and thematic roles such as agent and patient have long been assumed to map onto similar non-linguistic event representations (Jackendoff 1990).

The current evidence for domain-general encoding of abstract representations can appear somewhat contradictory. On the one hand, thematic roles, such as those targeted in Chapter 4 and 5, are often considered domain-general notions, that form part of the core knowledge of general - and potentially even pre-verbal - cognition (Rissman and Majid 2019). This assumption is supported by several behavioural findings demonstrating susceptibility to agent and patient roles in young children. Similarly, in adults, abstract role concepts have been shown to influence behaviour across domains. For example, an implicit hierarchy of thematic roles determines their ordering during language production as well as their overall salience in visual change detection tasks (Ünal et al. 2021 for review). Furthermore, neural correlates of agent and patient roles have been consistently found in similar brain areas across studies using either language or visual stimuli (Frankland and Greene 2015, Wang et al. 2016). On the other hand, recent data from global aphasia patients suggests a dissociation between linguistic and nonverbal role assignment (Ivanova et al. 2021). A similar discrepancy between neuroimaging and neuropsychological data had previously been reported when comparing structural processing across the domains of language and music (Patel 2003).

One explanation addressing the discrepancy is that abstract representations might be encoded in distributed working memory traces that include both domain-specific as well as domain-general codes. Having a somewhat redundant coding scheme for abstract stimulus features at different levels of domain-specificity could potentially help avoid interference between sequentially presented input and lead to more robustness after lesions (Christophel et al. 2017). Indeed, in Chapter 5 we observe encoding of event structure in multiple cortical areas including left inferior frontal gyrus, anterior and superior temporal cortex and right angular gyrus. Simultaneous activation of prefrontal and temporal areas is often reported during high-level language processing and the anatomical and function connectivity profile between frontal and temporoparietal cortex are well documented (Hagoort 2013). Specifically, directed information flow has been shown to occur most prominently and early from temporal regions to left IFG (Schoffelen et al. 2017, Lyu et al. 2019). Moreover, the inferior prefrontal cortex has been shown to encode domain-general information in cognitive domains other than language (Stokes et al. 2013). Therefore, findings of distributed encoding of abstract representations might reflect a “representational loop”, including both language-specific as well as domain-general neural representations, the latter being most likely hosted in prefrontal areas.

MVPA is a suitable technique not just for identifying neural correlates of higher-level representational content but also comparing these representations across different cognitive domains. For example, classifiers can be trained on one stimulus set and tested on another (cross-decoding) to check whether the information they pick up on, generalises across tasks or paradigms. RSA is especially suited to allow comparisons even if the dimensionality of estimation parameters is highly variable across domains (e.g. different amounts of time-points or when comparing MEG sensors with MRI voxels), since it maps the multivariate brain signal onto the abstract space of representational content. In the future, more MVPA studies should explicitly test the cross-domain generalisability of neural correlates underlying abstract representations during sentence comprehension. Knowing whether such representations are encoded by the brain in a domain-general manner is crucial for developmental theories. For example, this knowledge might lend support for a role of domain-general generalisation principles in language learning, eliminating the need for language-specific innate biases (Perfors et al. 2011).

## 6.4 Conclusion

In this thesis, I have identified neural representations of abstract language properties during sentence comprehension and described their spatiotemporal dynamics. I have demonstrated that abstract language processing is supported by a number of widely distributed and mostly left-lateralised cortical areas, potentially implying a redundant neural code for abstract properties of language. Throughout this distributed brain network, I have provided evidence for early encoding of both word-specific but also full sentence-level meaning. This supports the view that incoming words are immediately processed and combined into one unified sentence interpretation. Moreover, I have shown that task demands have important consequences for the neural representations of language. Specifically, I have shown that certain behavioural tasks might in fact bias perception towards individual words within a sentence and I have argued that a lack of task goal might eliminate the need to neurally encode abstract sentence structure altogether. This likely reflects different “modes” of language processing, with varying degrees of depth and effort. Overall, by taking into account the multivariate nature of the brain signal along the dimensions of space, time and even subjects, I have revealed the neural dynamics of language processing in greater detail and begun to explore the dimensionality of the underlying representational content.



# Bibliography

- Abnar, S., Beinborn, L., Choenni, R., and Zuidema, W. (2019). Blackbox meets blackbox: Representational similarity and stability analysis of neural language models and brains. *arXiv preprint arXiv:1906.01539*.
- Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of child language*, 26(2):339–356.
- Allefeld, C., Görden, K., and Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, 141:378–392.
- Altmann, G. and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3):191–238.
- Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366.
- Baggio, G. and Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26(9):1338–1367.
- Baker, B., Lansdell, B., and Kording, K. (2021). A Philosophical Understanding of Representation for Neuroscience. *arXiv preprint arXiv:2102.06592*.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Haubo Bojesen Christensen, R., and Others (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.
- Bellmund, J. L., Gärdenfors, P., Moser, E. I., and Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415).

- Bemis, D. K. and Pykkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8):1859–1873.
- Ben-Yakov, A., Honey, C. J., Lerner, Y., and Hasson, U. (2012). Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage*, 63(1):501 – 506.
- Berl, M. M., Duke, E. S., Mayo, J., Rosenberger, L. R., Moore, E. N., VanMeter, J., Ratner, N. B., Vaidya, C. J., and Gaillard, W. D. (2010). Functional anatomy of listening and reading comprehension during development. *Brain and Language*, 114(2):115–125.
- Binder, J. R. and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11):527–36.
- Bock, K. (1986). Syntactic persistence in language processing. *Cognitive Psychology*, 18:355–387.
- Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002). Modality independence of word comprehension. *Human Brain Mapping*, 16(4):251–261.
- Bornkessel, I., Zysset, S., Friederici, A. D., Von Cramon, D. Y., and Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage*, 26(1):221–233.
- Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59(1):55–73.
- Boudewyn, M. A., Zirnstein, M., Swaab, T. Y., and Traxler, M. J. (2014). Priming prepositional phrase attachment: evidence from eye-tracking and event-related potentials. *Quarterly journal of experimental psychology*, 67(3):424–54.
- Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirra, S., König, E., Lezius, W., Rohrer, C., Smith, G., and Uszkoreit, H. (2004). TIGER: Linguistic Interpretation of a German Corpus. *Research on Language and Computation*, 2(4):597–620.

- Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Todd Constable, R., Fulbright, R. K., Magnuson, J. S., Van Dyke, J. A., and Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, 47(4):416–431.
- Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2):163–173.
- Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences*, 75012(May).
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. J. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, 41:1318–1352.
- Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about Semantic Illusions : Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446:127–143.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., and Others (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Bruffaerts, R., De Deyne, S., Meersmans, K., Liuzzi, A. G., Storms, G., and Vandenberghe, R. (2019). Redefining the resolution of semantic knowledge in the brain: Advances made by the introduction of models of semantics in neuroimaging. *Neuroscience and Biobehavioral Reviews*, 103(March):3–13.
- Caramazza, A. and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4):572–582.
- Carpentier, A., Pugh, K. R., Westerveld, M., Studholme, C., Skrinjar, O., Thompson, J. L., Spencer, D. D., and Constable, R. T. (2001). Functional MRI of language processing: Dependence on input modality and temporal lobe epilepsy. *Epilepsia*, 42(10):1241–1254.
- Cattinelli, I., Borghese, N. A., Gallucci, M., and Paulesu, E. (2013). Reading the reading brain: A new meta-analysis of functional imaging data on reading. *Journal of Neurolinguistics*, 26(1):214–238.

- Chan, A. M., Halgren, E., Marinkovic, K., and Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, 54(4):3028–3039.
- Chee, M. W. L., O’Craven, K. M., Bergida, R., Rosen, B. R., and Savoy, R. L. (1999). Auditory and visual word processing studied with fMRI. *Human Brain Mapping*, 7(1):15–28.
- Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, 21(2):111–124.
- Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J., and Charest, I. (2019). The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. *NeuroImage*, 194(January):12–24.
- Cichy, R. M., Pantazis, D., and Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3):455–462.
- Cichy, R. M., Ramirez, F. M., and Pantazis, D. (2015). Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *NeuroImage*, 121:193–204.
- Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., and Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *NeuroImage*, 22(1):11–21.
- Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, 112(August 2019):279–299.
- Coopmans, C. W., Hoop, H. D., Kaushik, K., Hagoort, P., and Martin, A. E. (2021). Structure-(in)dependent Interpretation of Phrases in Humans and LSTMs. *Proceedings of the Society for Computation in Linguistics*, 4(58).
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., and Halgren, E. (2000). Dynamic statistical parametric mapping: Combining

- fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67.
- Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., and Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage*, 97:271–283.
- de Cheveigné, A., Di Liberto, G. M., Arzounian, D., Wong, D. D., Hjortkjær, J., Fuglsang, S., and Parra, L. C. (2018). Multiway canonical correlation analysis of brain data. *NeuroImage*, 186:728–740.
- Decock, L. and Douven, I. (2011). Similarity After Goodman. *Review of Philosophy and Psychology*, 2(1):61–75.
- Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., and Gallant, J. L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience*, 39(39):7722–7736.
- Detterman, D. (1993). The case for prosecution: Transfer as an epiphenomenon. *Transfer on Trial Intelligence Cognition and Instruction*, (January 1993).
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dinga, R., Schmaal, L., Penninx, B. W., van Tol, M. J., Veltman, D. J., van Velzen, L., Mennes, M., van der Wee, N. J., and Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017). *NeuroImage: Clinical*, 22:101796.
- Fahrenfort, J. J., Grubert, A., Olivers, C. N., and Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports*, 7(1):1–15.
- Fedorenko, E., Mineroff, Z., Siegelman, M., and Blank, I. (2018). Word meanings and sentence structure recruit the same set of fronto-temporal regions during comprehension. *bioRxiv preprint*.

- Fedorenko, E., Nieto-Castañón, A., and Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, 50(4):499–513.
- Ferreira, F. and Lowder, M. W. (2016). *Prediction, Information Structure, and Good-Enough Language Processing*, volume 65. Elsevier Ltd.
- Ferreira, F. and Patson, N. D. (2007). The ‘Good Enough’ Approach to Language Comprehension. *Language and Linguistics Compass*, 1(1-2):71–83.
- Fitz, H. and Chang, F. (2018). Sentence-level ERP effects as error propagation: A neurocomputational model. *Cognitive Psychology*, pages 1–46.
- Fitz, H., Uhlmann, M., Van Den Broek, D., Duarte, R., Hagoort, P., and Petersson, K. M. (2020). Neuronal spike-rate adaptation supports working memory in language processing. *Proceedings of the National Academy of Sciences of the United States of America*, 117(34):20881–20889.
- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3/71.
- Fox, N. P., Leonard, M. K., Sjerps, M. J., and Chang, E. F. (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *eLife*, 9:1–43.
- Frank, S. L. and Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22(6):829–834.
- Frank, S. L., Bod, R., and Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, 279(1747):4522–4531.
- Frankland, S. M. and Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(37):11732–11737.
- Frankland, S. M. and Greene, J. D. (2020). Two ways to build a thought: distinct forms of compositional semantic representation across brain regions. *Cerebral Cortex*, 30(6):3838–3855.
- Frazier, L. Y. N. and Rayner, K. (1982). Making and Correcting Errors during Sentence Comprehension : Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive psychology*, (14):178–210.

- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological reviews*, 91(4):1357–92.
- Frisch, S. and Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport*, 12(15):3391–3394.
- Gershman, S. J. and Tenenbaum, J. B. (2015). Phrase similarity in humans and machines. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 776–781.
- Geschwind, N. (1979). Specializations of the human brain. *Scientific American*, 2(241):180–201.
- Gilead, M., Trope, Y., and Liberman, N. (2020). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.
- Goldstone, R. L. and Son, J. (2005). Cambridge Handbook of Thinking and Reasoning, chapter Similarity.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M. S. (2014). Mne software for processing meg and eeg data. *NeuroImage*, 86:446 – 460.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.
- Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., and Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *NeuroImage*, 53(2):638–646.
- Groen, I. I., Greene, M. R., Baldassano, C., Fei-Fei, L., Beck, D. M., and Baker, C. I. (2017). Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *bioRxiv*, pages 1–26.
- Grootswagers, T., Cichy, R. M., and Carlson, T. A. (2018). Finding decodable information that can be read out in behaviour. *NeuroImage*, 179(June):252–262.

- Groetswagers, T., Wardle, S. G., and Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of cognitive neuroscience*, 29(4):677–697.
- Guggenmos, M., Sterzer, P., and Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, 173(August 2017):434–447.
- Gwilliams, L. and King, J. R. (2020). Recurrent processes support a cascade of hierarchical decisions. *eLife*, 9:1–20.
- Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, 4(July):1–13.
- Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience and Biobehavioral Reviews*, 81:194–204.
- Hagoort, P. and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600 to syntactic visual presentation. *Neuropsychologia*, 38:1531–1549.
- Hagoort, P. and Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, 37(1):347–362.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497.
- Hämäläinen, M. S. and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & biological engineering & computing*, 32(1):35–42.
- Hare, M., McRae, K., and Elman, J. L. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, 19(2):181–224.
- Haxby, J. V., Guntupalli, J. S., Nastase, S. A., and Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *Elife*, 9:e56601.



- Hebart, M. N. and Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, 180(July 2017):4–18.
- Hebart, M. N., Zheng, C. Y., Pereira, F., and Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, 4(11):1173–1185.
- Hickok, G. and Poeppel, D. (2007). The Cortical Organisation fo Speech Processing. *Nature*, 8(5):393–402.
- Hillebrand, A. and Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, 16(3A):638–650.
- Holyoak, K. (2005). *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press Cambridge.
- Homae, F., Hashimoto, R., Nakajima, K., Miyashita, Y., and Sakai, K. L. (2002). From perception to sentence comprehension: The convergence of auditory and visual information of language in the left inferior frontal cortex. *NeuroImage*, 16(4):883–900.
- Hornby, P. A. (1974). Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior*, 13(5):530–538.
- Hout, M. C., Goldinger, S. D., and Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, 142(1):256–281.
- Huizeling, E., Arana, S., Hagoort, P., and Schoffelen, J. M. (2020). Lexical frequency and sentence context influence the brain’s response to single words. *bioRxiv*.
- Hultén, A., Schoffelen, J.-M., Uddén, J., Lam, N. H., and Hagoort, P. (2019). How the brain makes sense beyond the processing of single words – An MEG study. *NeuroImage*, 186:586–594.
- Ivanova, A. A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R., and Fedorenko, E. (2021). The language network is recruited but not required for non-verbal semantic processing. *Neurobiology of Language*, 2(2).
- Jackendoff, R. (1992). *Semantic structures*, volume 18. MIT press.

- Jackendoff, R. (2003). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- Jackendoff, R. S. (1990). Semantic structures current studies in linguistics series 18.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4):434–446.
- Jimura, K. and Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, 50(4):544–552.
- Jobard, G., Vigneau, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34(2):784–800.
- Johnson, A., Vong, W. K., Lake, B. M., and Gureckis, T. M. (2021). Fast and flexible: Human program induction in abstract reasoning tasks. *arXiv preprint arXiv:2103.05823*.
- Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., and Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in Psychology*, 8(OCT):1726.
- Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2):159–201.
- Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5):679–685.
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451.
- King, A. J., Teki, S., and Willmore, B. D. (2018). Recent advances in understanding the auditory cortex [version 1; peer review: 2 approved]. *F1000Research*, 7(0).
- King, J.-r. and Dehaene, S. (2014). Characterizing the dynamics of mental representations : the temporal generalization method. *Trends in Cognitive Sciences*, 18(4):203–210.

- King, M. L., Groen, I. I., Steel, A., Kravitz, D. J., and Baker, C. I. (2019). Similarity judgments and cortical visual responses reflect different properties of object and scene categories in naturalistic images. *NeuroImage*, 197(October 2018):368–382.
- Kriegeskorte, N. and Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, 42:407–432.
- Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868.
- Kriegeskorte, N. and Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, 3:1–13.
- Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4.
- Lai, J. and Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, 118(2):265–273.
- Lake, B. and Baroni, M. (2018). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. *35th International Conference on Machine Learning, ICML 2018*, 7:4487–4499.
- Lake, B. M., Linzen, T., and Baroni, M. (2019). Human few-shot learning of compositional instructions. *arXiv*.
- Lam, N. H., Hultén, A., Hagoort, P., and Schoffelen, J. M. (2018). Robust neuronal oscillatory entrainment to speech displays individual variation in lateralisation. *Language, Cognition and Neuroscience*, 33(8):943–954.
- Lam, N. H., Schoffelen, J.-M., Uddén, J., Hultén, A., and Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *NeuroImage*, 142:43–54.
- Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

- Leiken, K., McElree, B., and Pylkkänen, L. (2015). Filling predictable and unpredictable gaps, with and without similarity-based interference: Evidence for LIFG effects of dependency processing. *Frontiers in Psychology*, 6(NOV):1–16.
- Li, J. and Pylkkänen, L. (2020). Disentangling semantic composition and semantic association in the left temporal lobe. *bioRxiv preprint*.
- Limesurvey GmbH (2012). LimeSurvey: An Open Source survey tool.
- Lindenberg, R. and Scheef, L. (2007). Supramodal language comprehension: Role of the left temporal lobe for listening and reading. *Neuropsychologia*, 45(10):2407–2415.
- Liuzzi, A. G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., Storms, G., Dupont, P., and Vandenberghe, R. (2017). Cross-modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage*, 150:292–307.
- Luyckx, F., Nili, H., Spitzer, B., and Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning. *eLife*, 8:1–19.
- Lyu, B., Choi, H. S., Marslen-Wilson, W. D., Clarke, A., Randall, B., and Tyler, L. K. (2019). Neural dynamics of semantic composition. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42):21318–21327.
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., and Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3):487–497.
- Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.
- Markman, A. B. and Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, 24(2):235–249.
- Matchin, W., Brodbeck, C., Hammerly, C., and Lau, E. (2019). The temporal dynamics of structure and content in sentence comprehension: Evidence from fMRI-constrained MEG. *Human Brain Mapping*, 40(2):663–678.
- Maye, J., Weiss, D. J., and Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1):122–134.

- Mazor, O. and Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, 48(4):661–673.
- McClelland, J. L. and Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In *Parallel distributed processing: explorations in the microstructure of cognition, vol. 2: psychological and biological models*, pages 272–325. MIT press.
- McKoon, G. and Macfarland, T. (2002). Event templates in the lexical representations of verbs. *Cognitive Psychology*, 45(1):1–44.
- McKoon, G. and Ratcliff, R. (2008). Meanings, propositions, and verbs. *Psychonomic Bulletin & Review*, 15(3):592–597.
- McRae, K., Ferretti, T. R., and Amyote, L. (1997). Thematic Roles as Verb-specific Concepts. *Language and Cognitive Processes*, 12(2-3):137–176.
- Menenti, L., Petersson, K. M., Scheeringa, R., and Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience*, 21(12):2358–2368. PMID: 19016600.
- Merkx, D. and Frank, S. L. (2020). Comparing Transformers and RNNs on predicting human sentence processing data. *arXiv preprint arXiv:2005.09471*.
- Michael, E., Keller, T. A., Carpenter, P., and Just, M. A. (2001). fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, 13(4):239–252.
- Mitchell, T. M. and Others (1997). Machine learning.
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M. M., Malave, V. L., Mason, R. A., and Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–5.
- Mohamed, M. T. and Clifton, C. J. (2011). Processing temporary syntactic ambiguity: The effect of contextual bias. *Quarterly journal of experimental psychology (2006)*, 64(9):1797–1820.

- Mongelli, V. (2020). *The role of neural feedback in language unification: How awareness affects combinatorial processing*. PhD thesis, Radboud University Nijmegen Nijmegen.
- Mur, M., Bandettini, P. A., and Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI - An introductory guide. *Social Cognitive and Affective Neuroscience*, 4(1):101–109.
- Murphy, G. (2004). *The big book of concepts*. MIT press.
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). ‘Primitive intelligence’ in the auditory cortex. *Trends in Neurosciences*, 24(5).
- Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S. S., Naccache, L., Hale, J. T., Pallier, C., et al. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114(18):E3669–E3678.
- Noble, C. H., Rowland, C. F., and Pine, J. M. (2011). Comprehension of Argument Structure and Semantic Roles: Evidence from English-Learning Children and the Forced-Choice Pointing Paradigm. *Cognitive Science*, 35(5):963–982.
- Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, 48(22):3637–3652.
- Noppeney, U. and Price, C. J. (2004). An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience*, 16(4):702–713.
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9):424–430.
- Obleser, J. and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, 13(1):14–19.
- 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.

- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:1–9.
- Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage*, 49(3):1943–1948.
- Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, 88(2):109–123.
- Paczynski, M. and Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language and Cognitive Processes*, 26(9):1402–1456.
- Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6):2522–7.
- Papanicolaou, A. C., Kilintari, M., Rezaie, R., Narayana, S., and Babajani-Feremi, A. (2017). The Role of the Primary Sensory Cortices in Early Language Processing. *Journal of Cognitive Neuroscience*, 29(10).
- Parra, L. C. (2018). Multi-set Canonical Correlation Analysis simply explained. *arXiv:1802.03759*, (11 February 2018).
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7):674–681.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., and Others (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830.
- Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., Botvinick, M., and Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, 9(1):963.
- Perfors, A., Tenenbaum, J. B., and Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3):306–338.

- Peterson, J. C., Abbott, J. T., and Griffiths, T. L. (2017). Adapting Deep Network Features to Capture Psychological Representations: An Abridged Report. In *IJCAI*, pages 4934–4938.
- Poletiek, F. H. (2002). Implicit learning of a recursive rule in an artificial grammar. *Acta Psychologica*, 111(3):323–335.
- Poletiek, F. H., Monaghan, P., van de Velde, M., and Bocanegra, B. R. (2021). The Semantics - Syntax Interface: Learning Grammatical Categories and Hierarchical Syntactic Structure through Semantics. *bioRxiv preprint*, page 6.
- Popov, V., Ostarek, M., and Tenison, C. (2018). Practices and pitfalls in inferring neural representations. *NeuroImage*, 174(March):340–351.
- Price, A. R., Peelle, J. E., Bonner, M. F., Grossman, M., and Hamilton, R. H. (2016). Causal evidence for a mechanism of semantic integration in the angular Gyrus as revealed by high-definition transcranial direct current stimulation. *Journal of Neuroscience*, 36(13):3829–3838.
- Puebla, G., Martin, A. E., and Dumas, L. A. (2021). The relational processing limits of classic and contemporary neural network models of language processing. *Language, Cognition and Neuroscience*, 36(2):240–254.
- Pulvermüller, F., Shtyrov, Y., and Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2):81–94.
- Pylkkänen, L. (2019). The neural basis of combinatorial syntax and semantics. *Science*, 366(6461):62–66.
- Pylkkänen, L. (2020). Neural basis of basic composition: What we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791).
- Rabovsky, M., Hansen, S. S., and McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, 2(9):693–705.
- Rabovsky, M. and McClelland, J. L. (2020). Quasi-compositional mapping from form to meaning: A neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791).



- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Raizada, R. D., Tsao, F. M., Liu, H. M., and Kuhl, P. K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: Prediction of individual differences. *Cerebral Cortex*, 20(1):1–12.
- Rayner, K., Carlson, M., and Frazier, L. (1983). The Interaction of Syntax and Semantics During Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(3):358–374.
- 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.
- Reimers, N. and Gurevych, I. (2020a). Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Reimers, N. and Gurevych, I. (2020b). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pages 3982–3992.
- Reverberi, C., G6rger, K., and Haynes, J. D. (2012). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex*, 22(6):1237–1246.
- Richie, R., White, B., Bhatia, S., and Hout, M. C. (2020). The spatial arrangement method of measuring similarity can capture high-dimensional semantic structures. *Behavior Research Methods*, 52(5):1906–1928.
- Rissman, L. and Majid, A. (2019). Thematic roles : Core knowledge or linguistic construct ? *Psychonomic Bulletin & Review*.
- Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137:29–39.

- Schoffelen, J.-M., Hultén, A., Lam, N., Marquand, A. F., Uddén, J., and Hagoort, P. (2017). Frequency-specific directed interactions in the human brain network for language. *Proceedings of the National Academy of Sciences*, 114(30):8083–8088.
- Schoffelen, J.-M., Oostenveld, R., Lam, N. H. L., Uddén, J., Hultén, A., and Hagoort, P. (2019). A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific Data*, 6(1):1–17.
- Schrimpf, M., Blank, I., Tuckute, G., Kauf, C., and Hosseini, E. A. (2020). Artificial Neural Networks Accurately Predict Language Processing in the Brain. *bioRxiv preprint*.
- Schwettmann, S. E., Tenenbaum, J. B., and Kanwisher, N. (2019). Invariant representations of mass in the human brain. *eLife*, 8:1–14.
- Segaert, K., Kempen, G., Petersson, K. M., and Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: An fMRI study. *Brain and Language*, 124(2):174–183.
- Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., and Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, 109(7):1214–1226.e8.
- Simanova, I., Hagoort, P., Oostenveld, R., and Van Gerven, M. A. J. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex*, 24(2):426–434.
- Simanova, I., van Gerven, M., Oostenveld, R., and Hagoort, P. (2010). Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS ONE*, 5(12).
- Simanova, I., van Gerven, M. A. J., Oostenveld, R., and Hagoort, P. (2015). Predicting the Semantic Category of Internally Generated Words from Neuromagnetic Recordings. *Journal of Cognitive Neuroscience*, 27(1):35–45.
- Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., and Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology*, 3:1–10.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., and Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence

- comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, 19(7):1493–1503.
- Sona, D., Veeramachaneni, S., Olivetti, E., and Avesani, P. (2007). Inferring cognition from fmri brain images. In *International Conference on Artificial Neural Networks*, pages 869–878. Springer.
- 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.
- Spivey-Knowlton, M. and Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55(3):227–267.
- Stevens, J. P. (2012). Canonical Correlation. In *Applied multivariate statistics for the social sciences*, pages 395–412. Routledge.
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375.
- Stokes, M. G., Wolff, M. J., and Spaak, E. (2015). Decoding Rich Spatial Information with High Temporal Resolution. *Trends in Cognitive Sciences*, 19(11):636–638.
- Stolk, A., Todorovic, A., Schoffelen, J. M., and Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *NeuroImage*, 68:39–48.
- Su, L., Fonteneau, E., Marslen-Wilson, W., and Kriegeskorte, N. (2012). Spatiotemporal searchlight representational similarity analysis in EMEG source space. *Proceedings - 2012 2nd International Workshop on Pattern Recognition in NeuroImaging, PRNI 2012*, pages 97–100.
- Taraban, R. and McClelland, J. L. (1988). Constituent Attachment and Thematic Role Assignment in Sentence Processing: Influences of content-based expectations. *Journal of Memory and Language*, 27:597–632.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285.
- Tenny, C. (1994). *Aspectual roles and the syntax-semantics interface*, volume 52. Springer.

- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3):209–253.
- Toneva, M., Mitchell, T. M., and Wehbe, L. (2020). The meaning that emerges from combining words is robustly localizable in space but not in time. *bioRxiv*.
- Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, 18(11):605–611.
- Traxler, M. J. and Tooley, K. M. (2007). Lexical mediation and context effects in sentence processing. *Brain Research*, 1146(1):59–74.
- Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., and Lingnau, A. (2015). MEG Multivariate Analysis Reveals Early Abstract Action Representations in the Lateral Occipitotemporal Cortex. *Journal of Neuroscience*, 35(49):16034–16045.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.
- Tyler, L. K., Cheung, T. P., Devereux, B. J., and Clarke, A. (2013). Syntactic computations in the language network: Characterizing dynamic network properties using representational similarity analysis. *Frontiers in Psychology*, 4(MAY):1–19.
- Uddén, J., Hultén, A., Schoffelen, J.-M., Lam, N., Harbusch, K., van den Bosch, A., Kempen, G., Petersson, K. M., and Hagoort, P. (2019). Supramodal sentence processing in the human brain: fmri evidence for the influence of syntactic complexity in more than 200 participants. *BioRxiv*, page 576769.
- Ünal, E., Ji, Y., and Papafragou, A. (2021). From Event Representation to Linguistic Meaning. *Topics in Cognitive Science*, 13(1):224–242.
- Van Den Brink, D. and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPS. *Journal of Cognitive Neuroscience*, 16(6):1068–1084.
- van Es, M. W., Marshall, T. R., Spaak, E., Jensen, O., and Schoffelen, J.-M. (2020). Phasic modulation of visual representations during sustained attention. *European Journal of Neuroscience*.
- Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001). An Integrated Software Suite for Surface-based Analyses of Cerebral Cortex. *Journal of the American Medical Informatics Association*, 8(5):443–459.

- Van Gerven, M., Bahramisharif, A., Farquhar, J., and Heskes, T. (2013). Donders Machine Learning Toolbox (DMLT). <https://github.com/distrep/DMLT>, version, 26(06):2013.
- Van Schijndel, M. and Linzen, T. (2018). Modeling garden path effects without explicit hierarchical syntax. In *CogSci*.
- Van Veen, B. D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880.
- Varoquaux, G. (2017). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, (April):1–10.
- 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.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.
- Vidaurre, D., Myers, N. E., Stokes, M., Nobre, A. C., and Woolrich, M. W. (2019). Temporally Unconstrained Decoding Reveals Consistent but Time-Varying Stages of Stimulus Processing. *Cerebral Cortex*, 29(2):863–874.
- Vigneau, M., Beaucousin, V., Mazoyer, B., Vigneau, M., Tzourio-Mazoyer, N., Petit, L., Jobard, G., Hervé, P.-Y., Crivello, F., Zago, L., and Mellet, E. (2010). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? *NeuroImage*, 54(1):577–593.
- Wang, J., Cherkassky, V. L., Yang, Y., Chang, K.-m. K., Vargas, R., Diana, N., and Just, M. A. (2016). Identifying thematic roles from fMRI-measured neural representations. *Cognitive Neuropsychology*, 3294(November):1–8.
- Wasmuht, D. F., Spaak, E., Buschman, T. J., Miller, E. K., and Stokes, M. G. (2018). Intrinsic neuronal dynamics predict distinct functional roles during working memory. *Nature Communications*, 9(1).
- Weber, K., Lau, E. F., Stillerman, B., and Kuperberg, G. R. (2016). The Yin and the Yang of prediction: An fMRI study of semantic predictive processing. *PLoS ONE*, 11(3):1–25.

- Williams, A., Reddigari, S., and Pylkkänen, L. (2017). Early sensitivity of left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia*, 100(April):131–143.
- Wilson, B., Spierings, M., Ravnani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., and Rey, A. (2020). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*, 12(3):843–858.
- Wolff, M. J., Jochim, J., Akyürek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, 20(6):864–871.
- Yokoi, A. and Diedrichsen, J. (2019). Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex. *Neuron*, 103(6):1178–1190.e7.
- Zhang, L. and Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage*, 111:228–240.
- Ziegler, J., Bencini, G., Goldberg, A., and Snedeker, J. (2019). How abstract is syntax? Evidence from structural priming. *Cognition*, 193(December 2018):104045.
- Zuidema, W., Hupkes, D., Wiggins, G., Scharff, C., and Rohrmeier, M. (2018). Formal models of structure building in music, language and animal song. *The origins of musicality*, page 253.

# Appendix



A

English summary

Nederlandse samenvatting

Acknowledgements

Biography

## A.1 English summary

Language is both highly variable but also structured and full of regularities. We are experts at extracting any regularities and abstracting across variation when interpreting language - even if not always aware of it. For example, we are able to recognise words fairly well, irrespective of who is speaking them, whether the person is whispering or yelling, speaking with an accent, speaking slow or fast. We abstract across varying physical instantiations of a word every day often without being aware. Furthermore, the order that words appear in, is somewhat regular within a given language. In English we may say “John loves Mary” but not “Kissed John Mary”. At a young age, we pick up on morphosyntactic rules, such as inflections for past tense (I walk-ed) and apply them flexibly to new words. This skill is often only noticed, when it fails, i.e. when children overgeneralise, applying a rule where it is not suitable (\*I fall-ed). In this thesis, I set out to find patterns in brain activity, occurring as we perceive language, that might correspond to this abstract knowledge of linguistic regularities. Today we are in a better position than ever to find abstract representations in patterns of brain activity due to increased computational power and advanced techniques such as multivariate pattern analysis (MVPA). Those techniques allow us to draw information from activity patterns across populations of neurons, potentially capturing spatial codes. The brain data I analysed in this work was recorded using Magnetoencephalography (MEG), which is a non-invasive neuroimaging technique, that records magnetic fields produced by electrical currents of large neuronal populations. Due to its high temporal resolution, this neuroimaging technique is ideally suited to study fast and highly dynamic brain processes such as language comprehension.

I start out, in **chapter 2**, by characterising brain activity patterns that are occurring across different physical instantiations of a word. When driving a car, you can be alerted about an upcoming crossroad by either reading the word “stop” on a traffic sign or by other passengers yelling “stop”. Therefore, there must be common brain activations independent of the input modality, even though the primary sensory brain areas initially activated by either visual or auditory stimulation are located in different parts of the cortex. To characterise the dynamics of modality-independent brain activity, I analysed a large-scale MEG dataset of 200 human subjects either listening to or reading varying sentences. Because word-



specific neural signals can vary a lot from one individual to another, I first applied a transformation to each subject's neural data using multiset canonical correlation, such that they became more comparable. It turns out that a large network of left-lateralised cortical brain areas respond to words in a modality-independent manner. This modality-independent processing was significant in temporal areas starting around 325 ms after word onset and rapidly spread to frontal and parietal regions. Importantly the modality-independent brain activity occurred in both primary auditory and general language areas and was only visible when adjusting for idiosyncrasies of individual subject data.

In **chapter 3**, I turn to the question whether brain activity reflects abstract sentence structure. In a sentence, words are grouped together into phrases and phrases in turn can be nested within other phrases (e.g. [the woman [who owns a dog] chases the cat]). This nested relationship constrains which words are combined in meaning (e.g. here "a dog chases" is not part of the meaning of the sentence). We seem to pick up on structural configurations of words and even repeat them during conversation. For example, having just heard a passive sentence, we are more likely to express a response using a passive as well. Do we form neural representation of sentence structure that are abstract and hence independent of specific word meaning? To answer this question, I recorded neural activity of people reading structurally ambiguous but semantically obvious prepositional phrases ("the woman sees the dog with binoculars"). Prepositional phrases (e.g. "with binoculars") can provide subordinate information about entities ("dog with binoculars") or actions ("seeing with binoculars"). Through subtle changes in the sentences, I manipulated this information structure towards one or the other. While participants were able to disambiguate the prepositional phrases when asked about each sentence explicitly, there was no indication of an abstract hierarchical representation given their neural activity during reading. The neural data did carry information about semantic content of each individual word of a sentence, however, as well as information about the type of word (e.g. verb or noun) being read at any given moment.

Finally, in **chapters 4 & 5** I turn my focus onto abstract relational properties, that emerge from the combined semantic and syntactic information of a sentence. For example, Given the order of words in "John loves Mary", we identify John as the one who loves or the agent of the event and Mary as the one being loved or the patient of the event. The notions of agent and patient can be applied to either John

or Mary or any person and occur across many semantic contexts (the one who loves, the one who eats, ...). Agent and patient are therefore considered abstract relational concepts. Again, we are sensitive to this systematicity in language even before formal education. For example, young children are able to infer agent and patient roles correctly from unknown verbs (e.g. “the bunny is blicking the frog”).

In **chapter 4**, I investigated, whether asking people to judge how similar sentences are to each other, could reveal the multitude of meaning dimensions that make up a sentence’s meaning - such as individual word meanings or abstract relations like agent and patient. I collected similarity judgments of several transitive sentences (e.g. “the doctor encouraged the athlete”) using a geometric task, in which similar sentences needed to be placed close to each other on a screen. As had been shown with visual scenes before, similarity judgments of sentences captured multiple dimensions of event meaning including relational information such as who is the agent and who is the patient of the described event. However, those dimensions were overshadowed by the strong verb bias, where the semantic similarity of the verbs between two events had the strongest influence on their perceived similarity. A dimensionality reduction of the group-results using non-negative matrix factorisation allowed to recover the additional relational meaning dimensions. Finally, I compared the similarity judgments to sentence similarity according to current artificial computer models of language. Although they captured some of the same dimensions, they did not all capture the human bias for verb semantics.

I then continue to use the similarity structure across those transitive sentences in **chapter 5**, where I investigate the neural correlates evoked while reading such sentences. I collected MEG data from people reading the sentences for comprehension and answering simple questions such as “who was doing X” or “Who was being X’ed”? I find that across readers both semantic and relational information are encoded in distributed, left-lateralised brain networks. These networks included middle superior temporal cortex as well as frontal and inferior parietal cortical areas. In addition, only relational information but not semantic information seemed to be encoded in anterior parts of the IFG, the anterior temporal lobe and the angular gyrus. Importantly, both types of sentence meaning are activated simultaneously and early, within the first 400 ms after sentence-final word onset. This suggests, that our brains integrate newly read words immediately into the

full sentence context, instead of sequentially processing a word's meaning first and then its relation to other words in the sentence.

The results of this thesis demonstrate how multivariate patterns of brain activity - as measured through MEG - can reveal neural representations of abstract knowledge during sentence comprehension. Throughout the chapters, I have provided several examples for how MVPA can be applied to neural data, not only to quantify different dimensions of neural representations, but also to factor out individual subject variability. With the help of these techniques, I have characterised the dynamics of abstract neural representations in time and space. I have focused on different levels of abstraction, namely abstraction across physical differences in presentation as well as abstraction of higher-level relational concepts. Both seem to be supported by a distributed network of several left-lateralised cortical areas, potentially implying a redundant neural code. In addition, abstract properties of words and their relations are processed early on when read in sentence context. I did not find evidence, however, for an abstract representation of phrase structure, possibly due to the lack of task demands. Comparing tasks across studies, we observe clear biases in perception of sentence meaning when participants were making explicit judgments, whereas when reading without any goal they seem to form somewhat impoverished neural representations. By tracking neural representations of sentence meaning over time, we can gain valuable insights into the neural dynamics of sentence comprehension. This work provides a first exploration of such neural dynamics and illustrates the benefits of MVPA, when combined with suitable language tasks.

## A.2 Nederlandse samenvatting

Taal is zowel ontzettend variabel als gestructureerd en vol regelmatigheden. Door ons vermogen om taal te interpreteren, zijn we experts in het afleiden van allerlei regelmatigheden en het abstraheren over veel variatie in die taal - zelfs als we er niet eens bewust van zijn. We zijn bijvoorbeeld in staat om woorden vrij goed te herkennen, onafhankelijk van wie ze uitspreekt, of de persoon fluistert of roept, met een accent praat, of snel of langzaam praat. Elke dag abstraheren we taal over verschillende fysieke instanties van een woord, vaak zonder dat we het doorhebben. Ook is de volgorde waarin woorden voorkomen enigszins regelmatig binnen een bepaalde taal. In het Nederlands kun je zeggen “Jan houdt van Marieke” maar niet “Kust Jan Marieke”. Vanaf een jonge leeftijd beginnen we morfosyntactische regels te herkennen, zoals inflectie in de verleden tijd (“ik fiets-te”), en passen we ze flexibel toe op nieuwe woorden. Deze vaardigheid wordt vaak pas opgemerkt als dit misgaat, zoals wanneer kinderen een regel overgeneraliseren en toepassen als deze niet passend is (“ik val-de\*”). In dit proefschrift is mijn doel om patronen te vinden in hersenactiviteit die ontstaan wanneer we taal waarnemen en die kunnen overeenkomen met deze abstracte kennis van taalkundige regelmatigheden. Tegenwoordig bevinden we ons in een betere positie dan ooit om abstracte representaties in patronen van hersenactiviteit te vinden door toegenomen computer- en rekenkracht, en geavanceerde technieken zoals multivariate pattern analysis (MVPA). Die technieken stellen ons in staat om informatie te halen uit patronen van activiteit over neuronpopulaties heen, die mogelijk ruimtelijke codes bevatten. De data afkomstig van hersenen die ik in dit werk heb geanalyseerd, was gemeten met behulp van Magnetoencephalography (MEG). Dit is een niet-invasieve neuroimagingtechniek die magnetische velden opvangt, geproduceerd door elektrische stromingen van grote neuronpopulaties. Door haar hoge temporele resolutie is deze neuroimagingtechniek bij uitstek geschikt voor het bestuderen van snelle en zeer dynamische hersenprocessen zoals taalbegrip.

Ik begin, in **hoofdstuk 2**, met het karakteriseren van patronen van hersenactiviteit die voorkomen over verschillende fysieke instanties van een woord. Als je aan het autorijden bent, kun je gewaarschuwd worden voor een kruispunt door het woord ‘stop’ te lezen op een verkeersbord of door andere passagiers die ‘stop’ roepen. Daarom moeten er gemeenschappelijke hersenactivaties zijn die

onafhankelijk zijn van de modaliteit van de input, zelfs als de primaire somatosensorische hersengebieden die aanvankelijk geactiveerd worden door visuele of auditieve stimulatie zich bevinden in verschillende delen van de hersenschors. Om de dynamieken van modaliteitsonafhankelijke hersenactiviteit te karakteriseren heb ik een MEG-dataset van grote schaal geanalyseerd, namelijk van 200 proefpersonen die naar gevarieerde zinnen luisteren of deze lezen. Omdat neurale signalen die woordspecifiek zijn veel kunnen variëren tussen individuen heb ik eerst een transformatie toegepast op de neurale data van elke proefpersoon met behulp van multiset canonical correlation om ze meer vergelijkbaar te maken. Dit resulteerde in het gegeven dat een groot netwerk aan linksgelateraliseerde gebieden in de hersenschors reageerden op woorden op een modaliteitsafhankelijke manier. Deze modaliteitsafhankelijke manier van verwerking was significant in temporele hersengebieden vanaf 325 ms na het horen van het begin van een woord, en dit verspreidde snel naar frontale en parietale gebieden. Belangrijk is dat de modaliteitsonafhankelijke hersenactiviteit plaatsvond in zowel primaire auditieve hersengebieden als algemene hersengebieden voor taal. Ook was dit alleen zichtbaar wanneer aangepast werd voor de idiosyncratische eigenschappen in de data van individuele proefpersonen.

In **hoofdstuk 3** focus ik me op de vraag of hersenactiviteit abstracte zinsstructuur weerspiegelt. In een zin worden woorden samen in een woordgroep gezet en woordgroepen kunnen vervolgens ook weer in andere woordgroepen ingebed worden (bijv.: [de vrouw [die een hond heeft] achtervolgt de kat]). Deze inbedding beperkt welke woorden in betekenis gecombineerd worden. We lijken structurele configuraties te herkennen en zelfs te herhalen in conversatie. Bijvoorbeeld, wanneer we net iemand een passieve zin hebben horen uitspreken, zijn we meer geneigd om ook te antwoorden met een passieve zin. Vormen we neurale representaties van zinsstructuur die abstract zijn en daarom onafhankelijk van de specifieke woordbetekenis? Om deze vraag te beantwoorden heb ik hersenactiviteit gemeten van mensen die structureel ambigue maar semantisch duidelijke voorzetselvoorwerpen lezen (“de vrouw ziet de hond met een verrekijker”). Voorzetselvoorwerpen (zoals “met een verrekijker”) kunnen onderliggende informatie geven over entiteiten (“hond met een verrekijker”) of acties (“zien met een verrekijker”). Met behulp van subtiele veranderingen in de zinnen heb ik deze informatiestructuur de ene of de andere kant op beïnvloed. Hoewel de proefpersonen de betekenis van de voorzetselvoorwerpen konden afleiden toen er expliciet

naar elke zin gevraagd werd, was er geen indicatie van een abstracte hiërarchische representatie gerelateerd aan hun hersenactiviteit tijdens het lezen. De hersendata bevatte wel informatie over de semantische inhoud van elk individueel woord van een zin naast ook informatie over het type woord (zoals een werkwoord of zelfstandig voornaamwoord) die iemand op een gegeven moment las.

Tot slot, in **hoofdstuk 4 en 5**, leg ik me toe op abstracte relationele eigenschappen die ontstaan uit gecombineerde semantische en syntactische informatie. Bijvoorbeeld, bij de volgorde van de woorden “Jan houdt van Marieke”, identificeren we Jan als degene die houdt van of de agens van de gebeurtenis en Marieke als diegene van wie gehouden wordt of de patiëns van de gebeurtenis. De noties van agens en patiëns kunnen toegepast worden op zowel Jan als Marieke als elk ander persoon en komen over veel verschillende semantische contexten voor (degene die houdt van, degene die eet, ...). Agens en patiëns worden daarom gezien als abstracte relationele concepten. Ook voor deze systematiek in taal zijn we gevoelig zelfs voordat we formeel onderwijs hebben gehad. Jonge kinderen zijn bijvoorbeeld al in staat om agens- en patiënsrollen correct af te leiden van werkwoorden die ze niet kennen (bijv. “het konijn glakt de kikker”).

In **hoofdstuk 4** bestudeerde ik of het mogelijk is om, door mensen te vragen de gelijkenis van twee zinnen te beoordelen, de vele dimensies in betekenis die de betekenis van een zin bepalen, te onthullen, zoals individuele woordbetekenissen of abstracte relaties zoals agens en patiëns. Ik heb beoordelingen van gelijkenis verzameld van meerdere transitieve zinnen (bijv.: “de dokter moedigt de atleet aan”) door middel van een geometrische taak, waarin zinnen die op elkaar leken dichtbij elkaar geplaatst moesten worden op een scherm. Zoals eerder met visuele scenes is aangetoond, vangen beoordelingen van gelijkenis van zinnen meerdere dimensies van de betekenis van een gebeurtenis, inclusief relationele informatie zoals wie de agens en wie de patiëns is van de gegeven gebeurtenis. Deze dimensies werden echter overschaduwed door de sterke voorkeur voor werkwoorden, waarbij de semantische gelijkenis van de werkwoorden tussen twee gebeurtenissen een sterkte invloed had op de ervaren gelijkenis. Een dimensionaliteitsreductie van de groepsresultaten door middel van non-negative matrix factorisation zorgde ervoor dat de additionele relationele betekenisdimensies werden behouden. Ten slotte vergeleek ik de gelijkenisbeoordelingen met gelijkenis van de zin bepaald door huidige kunstmatige computermodellen van taal. Ook al vingen deze modellen sommige van dezelfde dimensies, vingen ze niet alle menselijke voorkeur

voor werkwoordsbetekenis. Ik vervolg het gebruik van de gelijkende structuur over deze transitieve zinnen in **hoofdstuk 5**. Hierin onderzoek ik de neurale processen die worden veroorzaakt en die samenhangen met deze zinnen. Ik heb MEG-data verzameld van mensen die zinnen lazen, probeerden te begrijpen en vervolgens simpele vragen beantwoordden zoals “wie deed X?” of “wie werd er geXd?”. Ik vond dat voor alle lezers zowel semantische als relationele informatie gecodeerd werd in verspreide, linksgelateraliseerde hersennetwerken. Onder deze netwerken vielen zowel middel superieure temporale schors als frontale en inferieur parietale gebieden in de hersenschors. Daarnaast leek er alleen voor relationele informatie maar niet voor semantische informatie gecodeerd te worden in de anterieure delen van de IFG, de anterieure temporale kwab en de angulaire gyrus. Belangrijk is dat beide typen van zinsbetekenis tegelijk en vroeg geactiveerd worden, binnen de eerste 400 ms na de start van het laatste woord in een zin. Dit suggereert dat onze hersenen pas gelezen woorden meteen integreren in de context van de hele zin in plaats van achtereenvolgend. De resultaten van dit proefschrift demonstreren hoe multivariabele patronen in hersenactiviteit - gemeten met MEG - neurale representaties aan abstracte kennis gedurende zinsbegrip kunnen onthullen. Door de hoofdstukken heen heb ik verschillende voorbeelden geleverd voor hoe MVPA toegepast kan worden op neurale data, niet alleen om verschillende dimensies aan neurale representaties te kwantificeren maar ook om individuele subjectvariatie weg te nemen. Met behulp van deze technieken heb ik de dynamieken van abstracte neurale representaties in tijd en ruimte gekarakteriseerd. Ik heb gefocust op verschillende niveaus van abstractie, namelijk abstractie over fysieke verschillen in presentatie en abstractie van relationele concepten van een hoger niveau. Beide lijken ondersteund te worden door een verspreid netwerk van meerdere linksgelateraliseerde gebieden in de hersenschors, wat mogelijk een overbodige neurale code impliceert. Bovendien, abstracte eigenschappen van woorden en hun relaties worden vroeg verwerkt tijdens het lezen van zinscontext. Ik vond echter geen bewijs voor een abstracte representatie voor woordgroepstructuur, mogelijk door het gebrek aan de vereisten van de taak. Bij het vergelijken van taken van verschillende studies observeren we duidelijke voorkeuren in de perceptie van zinsbetekenis wanneer proefpersonen expliciete beoordelingen maakten, terwijl ze enigszins zwakke neurale representaties leken te vormen bij het lezen zonder een bepaald doel. Door het tracken van neurale representaties van zinsbetekenis over tijd kunnen we waardevolle inzichten verkri-

ngen in de neurale dynamiek van zinsbegrip. Dit werk biedt een eerste verkenning van zulke neurale dynamieken en illustreert de voordelen van MVPA wanneer dit gecombineerd wordt met geschikte taaltaken.



## A.3 Acknowledgements

**Jan-Mathijs**, I remember the first months of us working together very well. Me showing you some puzzling data points, you asking “Can I quickly try something?” and next I knew you were sitting in front of my computer, typing vigorously, producing plot after plot of increasingly detailed aspects of the data, that I hadn’t even known was possible. These moments in which I was frantically trying to follow what was happening on the screen, have taught me many things: First and foremost, how to code in MATLAB. Second, that there is always a new way to look at old data, and most importantly the joy and thrill of data analysis! Thank you so much for your support during the past years.

**Peter**, Thank you for the opportunity to join the NBL group and for giving me the freedom to pursue my interests. Throughout my PhD you provided me with what I needed most at the times when I needed it most urgently, whether this was critique or reassurance, urgency or patience. Your passion for science is inspiring and I appreciate the many fundamental discussions we’ve had about sentence processing. Thank you for always taking the time to engage with my questions, even on your back-to-back meeting days.

**Milena**, ich bin sehr dankbar, dass ich zwei meiner PhD Projekte gemeinsam mit dir entwickeln und umsetzen durfte. Deine Arbeit und die Zeit die ich in deinem Lab in Potsdam verbracht habe, haben mir einen neuen Blick auf meine Forschungsfragen ermöglicht. Vor allem aber hat mich deine positive, authentische Art aus einem tiefen Motivationsloch gerettet. Du bist für mich ein Vorbild für female leadership.

**Anne**, you are the other powerhouse of a female scientist, who has had an impact on me. Not only have you guided me through the scientific aspects of the Master’s internship, you haven often provided me with valuable advice and strengthened my confidence in my own skills. You are that person that can crack a joke and then say something deeply wise within the same breath. Je suis super contente de t’avoir rencontré!

I had the privilege of being part of not only one but two academic communities, the Donders Institute and the Max Planck Institute for Psycholinguistics. The many wonderful individuals at both sites have made the back-and-forth from lab meeting at one to office space at the other a true pleasure rather than a nuisance. Thank you to all lab managers, admins, TGs, pizza dough wizards and rest

of the staff at both Institutes for all their help along the way and plenty of nice conversations.

A special thanks to the **FieldTrip team** and to **Robert & Jan-Mathijs** for allowing me to contribute my tiny part to it. Teaching and tutoring during the toolkits has been an invaluable experience for me (Thank you **Diego** for inviting me to Madrid!). I have a lot of respect for the dedication with which you develop and maintain not only the software itself but also the community around it. **Robert**, your ability to create educational spaces that are based on inclusivity and mutual support is unparalleled and I benefitted a lot from having access to them.

At the Trigon, I shared my office with a bunch of wonderful people that made the daily grind more enjoyable. **Claudia**, thank you for many welcome distracting conversations about all the things that matter in life (especially dogs & kids). **Anke**, thank you for much needed hugs and for going through all the highs and lows together. **Mats & Kris**, thanks for many wonderful moments “growing up” in science together throughout Masters and PhD. Also, thank you Mats for your warm welcome in Oxford! At the MPI, I was part of the IMPRS graduate school and lucky to work with a group of amazing, dedicated IMPRSers putting together the first virtual edition of the IMPRS conference. **Julia & Merel**, I had a blast working with you and the others on this project and making it all happen. **Anne**, unsere gemeinsame Mittagspausen haben mir immer so viel Energie gegeben, Danke für die vielen Gespräche. **Kevin**, thank you for all your hard work on improving the IMPRS and for truly caring about the well-being and successes of “your” PhD students. I always felt like you had my back.

The NBL family has a very special place in my heart. Not only is the group full of extremely talented and interesting personalities, but the kindhearted spirit of all interactions and the genuine interest to learn from each other, made it extraordinary. Thank you to everyone for your valuable feedback over the years. **Yingying**, thank you for your funny spirit and for letting me win every single squash game we ever played, I’m sure it was all a fake to teach me some life lesson. **Ellie**, thank you for letting me come along when you tackled the MOUS dataset, for Victoria sponge cake and many good conversations about food. Thank you to my fellow PhD students, for filling the PhD meetings with your interesting ideas and sharing your journeys every week. **René**, you coached me in a variety of ways, from my IMPRS interview presentation to my first steps on a longboard.

You clearly are a jack of all trades, thank you for all of it. Thank you **Ksenija & Laura** for always being ready to buddy in the MRI lab as I acquired my last anatomicals. **Rowan**, MEG session were much more enjoyable when accompanied by your philosophical musings. Thank you **Teun** for sharing your tips about data viz with me. Thank you **Nienke** for some very pieciful forest walks. Thank you **Cas** for never giving up on the “green ideas sleep furiously” example. Thank you **Rowan, Cas, Micha & Alessio** for always being up for a discussion about deep learning and language processing even during summer break. **Margot**, I was very glad when you joined our trigon-based NBL club. I feel like covid robbed us of some lunch breaks but I am glad for the time we did get to hang out, a thousand thank yous for your wonderful translation job on my thesis summary.

**Annie**, wer hätte gedacht dass aus einer flüchtigen Begegnung im Hol­ländisch Kurs so eine schöne Freundschaft erwächst. Wir haben über Kaffee, Matcha und etlichen Gläsern Wein Businessspläne geschmiedet und nebenbei unsere Leben rauf und runter analysiert. Ich bin sehr froh dass es dich gibt. Gerade im letzten PhD Jahr, als wir zu zweit inmitten von Covid dem Ende entgegen fieberten, waren unsere gemeinsamen Spaziergänge oft wie eine vertraute Insel, die mich immer wieder in die Realität zurückgebracht haben.

**Julia**, meine Zeit in Nijmegen und der PhD sind untrennbar verflochten mit unserer gemeinsamen Zeit. Vom Häuschen in Bottendaal, über gemeinsame Roadtrips und Tatoos, das Fiebern auf die PhD Positionen und später die gemeinsamen NBL lab meetings inklusive geheimer Blicke (That’s flirting!). Vielen Dank, dass du in den letzten Jahren mit mir durch dick und dünn gegangen bist. Wenn unsere Freundschaft ein Kuchen wäre, dann wäre es eine mehrstöckige (vegane) Schokotorte. Eine Torte die lange wehrt, die man noch drei Tage später aus dem Kühlschrank holt und sich daran erfreut.

None of the things I have achieved during my PhD would have been possible if it weren’t for my **mother**. She has sacrificed much to provide me with opportunities that she didn’t have and she has never let me doubt my abilities in achieving anything. Danke Mamita, für deine lebenslange Unterstützung und dein Vertrauen in jede meiner Entscheidungen.

**Matthias**, I feel incredibly lucky to have found a partner who is also my best friend and who understands me so deeply as you do. On top of that, how lucky are we? Having found in each other that other person to endlessly and sometimes furiously discuss with and to have a shared passion, science. Throughout working

on this thesis I have more than once doubted myself and I cannot describe how much your support and your faith in me has helped me to continue nonetheless. Thank you for making me the best version of myself. You light my fire baby.

## A.4 Biography

Sophie Arana was born on April 30th, 1990, in Berlin, Germany. She completed a Bachelor's degree in German language and literature, with a minor in computer science at Freie Universität Berlin and a Master's degree (cum laude) in Cognitive Neuroscience at the Radboud University in Nijmegen. During her Master's studies, she worked on neural entrainment to speech under the supervision of Anne Kösem. In 2017, she started as a PhD candidate in the Neurobiology of Language lab at the Max Planck Institute for Psycholinguistics. Under the supervision of Peter Hagoort, Jan-Mathijs Schoffelen and later Milena Rabovsky she investigated different levels of abstract neural representations during sentence comprehension. She is currently working as a postdoctoral researcher in the Human Information Processing lab at Oxford University.