Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

# From Words to Sentences and Back: Characterizing Context-dependent Meaning Representations in the Brain

A dissertation presented by


**Nora Elsa Aguirre Sampayo**

(a.k.a. Nora E. Aguirre-Celis and Nora E. Celis)


Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

in

Artificial Intelligence

Monterrey, Nuevo León, December 2nd, 2021

Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey
School of Engineering and Sciences

The committee members, hereby, certify that have read the dissertation presented by Nora Elsa Aguirre Sampayo and it is fully adequate in scope and quality as a partial requirement for the degree of Doctor of Philosophy in Artificial Intelligence.

Dr. Risto Miikkulainen
The University of Texas in Austin
Advisor

Dr. Manuel Valenzuela-Rendón.
Tecnológico de Monterrey
Co-advisor

Dr. Hugo Terashima
Tecnológico de Monterrey
Examiner

Dr. Francisco Cantu
Tecnológico de Monterrey
Examiner

Dr. Uli Grasemann
The University of Texas in Austin
External Examiner

Dr. Hugo Terashima Marín
Dean of Graduate Studies
School of Engineering and Sciences

Monterrey, Nuevo León, December 2nd, 2021

To my boys:

Nacho, Nachito, Milo & Aaron

# Acknowledgements

I am forever grateful to my advisor, Risto Miikkulainen for his time, patience, and energy vested in me. His commitment to guiding and accompanying me into this journey is exceptional and unique. He has always led me to the right path and then encourage me to find my own solutions. He provided me with the opportunities, freedom, knowledge, guidance, and support to complete my research. He is an amazing inspiration and role model to me both in science and in life.

Thanks to Manuel Valenzuela for his instrumental support. Thanks to Jeff Binder from Medical College of Wisconsin, Rajeev Raizada and Andrew Anderson from University of Rochester, and Mario Aguilar and Patrick Connolly from Teledyne Scientific Co. for providing the data and insight for this research. They deserve the most sincere and deep acknowledgment for sharing with me their expertise and work, for teaching me everything I know about fMRI and CAR theory while working together during our time in the IARPA project.

Special thanks (gracias) to all the volunteers that participated in my experiments. They are part of this research. Thank you for evaluating my results. I would like to thank all of my friends, those I met in the NNR group and those I have met across the places I have lived. They provided the emotional support, companionship, and encouragement needed to persist throughout this journey.

My mom and dad, although they are no longer with us, I want to thank them for their infinite love and sacrifices, for empowering me, and for teaching me, among many things, how to be persistent! Thank you for the best brother and sister that I could ever asked, their unconditional love and care kept me going. To my mother in-law, even though she is no longer with us, she has been always an inspiration to me. To my beloved extended family that I have always received love, support, and encouragement throughout my life.

I am in debt for life to my husband Nacho, and each of my boys Nachito, Milo and Aaron. They are a blessing to me! They have always provided everything I could ever dream of, and I owe each of them every moment of joy in my life. Thanks for believing in me, for their never-ending support, for listening, for working as a team by helping me and each other to make the best of our lives together. I will be forever grateful to my husband for constantly supporting me, for enjoying side by side a beautiful life, and for sponsoring those wonderful trips we spend together to present my work. You are my everything!

This work was supported in part by IARPA-FA8650-14-C-7357 and by NIH 1U01DC014922.

From Words to Sentences and Back: Characterizing
Context-dependent Meaning Representations in the Brain

by
Nora Elsa Aguirre Sampayo

# Abstract

How do people understand concepts such as *olive oil*, *baby oil, lamp oil,* or *oil paint*? Embodied approaches to knowledge representation suggest that words are represented as a set of features that are the basic components of meaning. In particular, Binder et al. (2009) grounded this idea by mapping semantic features (attributes) to different brain systems in their Concept Attribute Representations (CAR) theory. Their fMRI experiments demonstrated that when humans listen or read sentences, they use different brain systems to simulate seeing the scenes and performing the actions that are described. An intriguing challenge to this theory is that concepts are dynamic, i.e., word meaning depends on context. This dissertation addresses this challenge through the Context-dEpendent meaning REpresentations in the BRAin (CEREBRA) neural network model. Based on changes in the fMRI patterns, CEREBRA quantifies how word meanings change in the context of a sentence. CEREBRA was evaluated through three different computational experiments and through behavioral analysis. The experiments demonstrated that words in different contexts have different representations, that the changes observed in the concept attributes encode unique conceptual combinations, and that the new representations are more similar to the other words in the sentence than to the original representations. The behavioral analysis confirmed that the changes produced by CEREBRA are actionable knowledge that can be used to predict human responses.

Together, these experiments constitute a comprehensive evaluation of CEREBRA's context-based representations as well as the soundness of CAR theory. Thus, CEREBRA is a useful tool for understanding how semantic knowledge is represented in the brain, and for providing a human-like context-based representations for NLP applications.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

How is human knowledge organized, stored, and used? Visual, motor, somatosensory, emotional, social and other brain systems enable humans to acquire, store, and integrate knowledge, to make sense, structure the endless stream of information, and navigate in the external world. As a result, humans not only create an internal representation of what is experienced by seeing, hearing, touching, smelling or tasting, they mentally represent complex instantiations of both the external and internal world. They create representations of everything they experience (e.g., objects, actions), and things they have never encountered (e.g., unicorns, dragons, dinosaurs). Furthermore, most of this knowledge relies on their ability to form semantic representations by combining existing concepts, to achieve an unlimited number of meanings. The goal of this dissertation is to develop a computational framework to understand how this knowledge is represented in the brain.

## 1.1 MOTIVATION

A word meaning is more than an entry in a dictionary. It involves a vast amount of knowledge relating the scenes and experiences people encounter (i.e., a rich encyclopedic knowledge), a set of referents to which the word properly applies (i.e., *the boy was angry* vs. *the chair was angry*), combination of other words, and grammatical constructions in which the word occurs. The meaning of the word varies from situation to situation and across contexts of use. For example, the word *small* means something different when used to describe a mosquito, a whale, or a planet. The properties associated with *small* vary in context-dependent ways: It is necessary to know what the word means, but also the context in which is used, and how the words combine in order to construct the word meaning.

While humans have a remarkable ability to form new meanings by combining existing concepts, modeling this process is challenging (Hampton, 1997; Janetzko 2001; Middleton et al., 2011; Murphy, 1998; Sag et al., 2002, Wisniewski, 1997, 1998). The same concept can be combined to produce different meanings: *corn oil* means oil made of corn, *baby oil* means oil rubbed on babies, and *lamp oil* means oil for lighting lamps (Wisniewski, 1997). Since *lamp* is an object, oil is likely to be a member of the inanimate category. However, *corn* and *baby* are living things, which suggest otherwise. How do language users determine the membership structure of such combinations of concepts, and how do they deduce the interpretation? As this example illustrates, there is no simple rule for how *oil* combines with other concepts. Computational models of such phenomena could potentially shed light into human cognition and advance AI applications that interact with humans via natural language. Such applications need to be able to understand and to form by themselves novel combination of concepts. Consider for example virtual assistants such as Siri, OK Google, or Alexa. These applications are built to answer questions posed by humans in natural language. All of them have natural language processing software to recognize speech and to give a response. However, whereas humans process language at many levels, machines process linguistic data with no inherent meaning. Given the ambiguity and flexibility of human language, modeling human conceptual representations is essential in building AI systems that interact effectively with humans.

Although early efforts were restricted to behavioral observations, experimental methods to map neural structures have made possible to study the neural mechanisms underlying the semantic memory system. For instance, neuroimaging (fMRI) technology provides a way to measure brain activity during word and sentence comprehension. When humans listen or read sentences, they are using different brain systems to simulate seeing the scenes and performing the actions that are described. As a result, parts of the brain that control these actions “light up” during the fMRI experiments. Hence, semantic models have become a popular tool for prediction and interpretation of brain activity using fMRI data.

Recently, Machine Learning systems in vision and language processing have been proposed based on single-word vector spaces. They are able to extract low-level features in order to recognize concepts (e.g., cat), but such representations are still shallow and fall short from symbol grounding (meaning). In general, these models build semantic representations from text corpora, where words that appear in the same context are likely to have similar meanings (Burgess, 1998; J. Devlin et al., 2018; Harris, 1970; Landauer & Dumais, 1997; Mikolov et al., 2013; Peters et al., 2018). However, such representations lack intrinsic meaning, and therefore sometimes even different concepts may appear similar. This problem has driven researchers to develop new componential approaches, where concepts are represented by a set of basic features, integrating textual and visual inputs. (Anderson et al., 2019; Bruni et al., 2014; Silberer & Lapata, 2014; Vinyals et al., 2015). Still, even with these multimodal embedding spaces, such vector representations fall short of symbol grounding; a truly multimodal representation should account for the full array of human senses (Bruni et al., 2014).

On the other hand, embodiment theories of knowledge representation (Barsalou, 1987; Binder et al., 2009; Landau et al., 1998; Regier, 1996) provide a direct analysis in terms of sensory, motor, spatial, temporal, affective, and social experience. Further, these theories can be mapped to brain systems. Recent fMRI studies helped identify a distributed large-scale network of sensory association, multimodal and cognitive regulatory systems linked to the storage and retrieval of conceptual knowledge (Binder et al., 2009). This network was then used as a basis for Concept Attribute Representation (CAR) theory, a semantic approach that represents concepts as a set of features that are the basic components of meaning, and grounds them in brain systems (Binder et al., 2009, 2011, 2016).

An intriguing challenge to semantic modeling is that concepts are dynamic, i.e., word meaning depends on context and recent experiences (Pecher, Zeelenberg, & Barsalou, 2004). For example, a pianist would invoke different aspects of the word *piano* depending on whether he will be playing in a concert or moving the *piano*. When thinking about a coming performance, the emphasis will be on the piano's function, including sound and fine hand movements. When moving the piano, the emphasis will be on shape, size, weight and other larger limb movements (Barclay et al., 1974). This dissertation addresses this challenge based on the CAR theory. The assumption is that words in different contexts have different representations. Therefore, different features in CARs should be weighted differently depending on context, that is, according to the combination of concepts that occur in the sentence.

### 1.2 Problem Statement

The main focus of this dissertation is to understand how word meanings change in the context of a sentence. There are three challenges that need to be addressed:

1. How are concepts represented in the brain? Componential theories of lexical semantics assume that they consist of a set of features that constitute the basic

components of meaning. CAR theory represents such features in terms of known brain systems, relating semantic content to systematic modulation in neuroimaging activity.

2. How do word meanings change in the context of a sentence? A word is broken down into various features that can become active at different rates in different situations. According to the CAR theory, the weights given to different feature dimensions are modulated by context.
3. What tools and approaches can be used to quantify such changes in meaning? CAR theory assumes that context modifies the baseline meaning of a concept. A computational model can test this assumption by using sentence fMRI patterns and the CAR semantic feature model to characterize how word meanings are modulated within the context of a sentence.

The first two points constitute the foundation for the third, which is the main technical challenge of this dissertation. Conceptual representations are dynamic, changing not only in response to context and the combination of words occurring in the sentence, but also in response to the knowledge an individual brings with them, i.e., experience and culture. All these dynamic effects grant words their transformation powers adjusted by a semantic system that is (1) experience-based in that it allows representations to change over time, and (2) distributed so that various features can become active at different rates in different situations (Yee & Thompson-Schill, 2016). The next section outlines an approach to these challenges.

In general, a "concept" is defined as the mental representation of a semantic object and a "word" is the symbolic name of it. Sometimes multiple words are required to identify a concept, such as *lamp oil*. "Word" and "concept" are used interchangeably in this dissertation when there is no risk of confusion.

## 1.3 APPROACH

The first two challenges are addressed using the CAR theory. The approach to the third challenge is to develop a model called CEREBRA, or Context-dependent mEaning REpresentation in the BRAin, a neural network model based on CAR theory.

The main point throughout is that conceptual knowledge is built from experience. Particularly, humans learn concepts from birth on through their senses and mental states and they are encoded according to the way they are experienced (e.g., seeing a dog is a visual experience). Since each person's experiences involve different times, locations, cultures, and people, concepts are not static but change throughout lifetime. The CAR theory provides a direct correspondence between conceptual content and neural representations. It suggests that concepts can be represented through a number of semantic dimensions that correspond to different brain systems, such as sensory, motor, visual, spatial, temporal, and affective, and based on the way concepts are acquired. Word meanings are represented as a set of weighted such dimensions or attributes, modulated by context. This dissertation integrates a subject-generated instantiation of the CAR theory into the model design. These attribute representations of word meaning provide a novel and powerful tool for investigating individual differences in conceptual organization. Also, the attributes capture fundamental elements of conceptual knowledge, which allows for a highly interpretable analysis of where systematic differences in conceptual content occur, and how they associate to each brain system.

CAR theory suggests that different properties of word meaning are activated in different contexts, and it is possible to capture these changes in fMRI images. Together, the CAR theory, the fMRI images, and the CEREBRA model form the groundwork for the last challenge: a tool to quantify how word meanings change in the context of a sentence. The main idea is to train a neural network to predict what the sentence fMRI should be, based on the CARs, and then use the FGREP mechanism (Forming Global Representations with Extended Backpropagation; Miikkulainen et al., 1988) to adjust the CARs so that the prediction becomes correct. As a result, the modified CARs indicate how the meaning changed in context.

This dissertation evaluates CEREBRA's context-dependent representations through three computational experiments and a behavioral analysis. The first experiment analyzes interesting context effects for different shades of meaning for a number of words and subjects. The results showed that different meanings for words were activated in different contexts. The second experiment extends the analysis by using combinations of concepts across subjects. It focuses on different types of conceptual combinations and their effect on word meanings by analyzing statistically significant changes for individual sentences across multiple fMRI subjects. The results showed that CEREBRA identifies several types of combinations such as property, thematic, hybrid and centrality. The third experiment demonstrates that similar sentences have a similar effect, and this effect is consistent across all words in the sentence. It analyzes the differences and correlates these changes to the CARs of the other words in the sentence. In other words, it statistically quantifies the conceptual combination effect across sentences and subjects. This analysis was performed across the entire corpus of sentences, for all subjects, and the results demonstrated that the new CARs were more similar to the other words in the sentence than to the original CARs, indicating how features of the context were transferred to each word in the sentence. The fourth experiment, a behavioral analysis, evaluates CEREBRA's context-based representations via human judgements. First, CEREBRA is used in a number of sentences to determine how the generic meaning of a word would have to change in order to account for the context. Then, the survey characterizes the changes in human subjects over the same sentences, to demonstrate that the changes produced by the model agree with human responses. The results confirmed that the context-dependent changes produced by CEREBRA are actionable knowledge that can be used to predict human judgements. Collectively, these experiments constitute a comprehensive approach to evaluate the CEREBRA's context-based representations as well as to assess the soundness of CAR theory as a semantic model of the brain.

CEREBRA opens the door for cognitive scientists to achieve better understanding and form new hypotheses about how semantic knowledge is represented in the brain. Additionally, the context-based representations produced by the model could be used for a broad range of artificial natural language processing systems, where grounding concepts as well as novel combinations of concepts is essential.

### 1.4 Overview of the Dissertation

In the following chapters, a computational framework developed to characterize context-dependent representations in the brain is first described. Then a series of experiments are presented to test the hypothesis that different properties of word meaning are activated in different contexts, and that it is possible to see those changes in the corresponding fMRI images using the computational model. The experiments start by characterizing individual cases and gradually incorporating more general analyses, eventually showing that the effects are robust and emerge across the entire collection of data and subjects. The human subject study completes the investigation by showing that the effects are real and understandable to the subjects. The experiments thus provide a comprehensive analysis towards understanding how the brain constructs sentence meanings from word features.

More specifically, Chapter 2 presents the foundations on how the semantic knowledge is grounded in the brain, includes different semantic models, and the tools and approaches used in this dissertation.

Chapter 3 introduces the framework to account for context effects in the brain. The chapter describes the computational model that is based in the brain-based theory, is implemented using neural networks, and is designed to interpret fMRI patterns to reflect the semantic space of the brain.

Chapter 4 demonstrates that context-based changes are likely to be nonlinear: A linear mapping (regression) and a nonlinear mapping (neural network model) are compared in these representation cases, focusing on a verb, a noun and an adjective.

Chapter 5 tests the context effect systematically across subjects and across examples. It studies the effects of similar context on different words and of different contexts. It also characterizes differences in context in terms of the centrality of meaning.

Chapter 6 expands the analysis through an aggregation study to demonstrate that the effect is robust and general across the entire corpus of sentences and case roles. Similar sentences have a similar effect, and this effect is consistent across all words in the sentence.

Chapter 7 provides evidence that the context-induced changes are real and actionable: The context-dependent changes are consistent with human judgements.

Chapter 8 reviews the model limitations and discusses the soundness of the brain-based theory. In addition, it outlines an NLP application that utilizes the context-based word representations, and proposes future work by integrating linguistic knowledge to the CEREBRA model.

Chapter 9 summarizes the contributions of this dissertation and concludes that the computational model is a useful tool for interpreting fMRI patterns to reflect the semantic space of the brain. Its functionality can further assist researchers to understand how the brain's semantic knowledge is organized, and it provides a human-like context-based representations toward AI applications of natural language understanding.

## 1.5 CONTRIBUTIONS

Overall, this research contributes to the development of a unified theory of concepts and the organization of the semantic space. CEREBRA model extends the CARs static representations into context-modified representations. As a result, CEREBRA advances grounded word representations by encompassing the full range of human experience as a fully multimodal semantic model. Further, it provides a mechanism for adapting representations to context so that robots can behave more robustly. Finally, provides a mechanism for mapping fMRI to interpretable representations so that insights can be obtained.

# Chapter 2

# Background and Related Work

Concept representation and word meaning are the main focus of this dissertation. Thus, the key issues are: How are concepts represented in the brain? How is word meaning represented? How do word meanings change during concept combination or under the context of a sentence? What tools and approaches serve to quantify such meaning representation changes? This chapter presents an overview of the foundations necessary to answer these questions.

## 2.1 MOTIVATION

Over the last few years, many empirical findings have shown the connection between perception (information collected from the environment) and action (information emitted to the environment) with language comprehension (Barsalou, 2008, Binder & Desai, 2011; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012). Listening to a sentence such as *The child threw the book* regulates the activation of hand-related areas of the motor cortex, even if individuals are not performing any hand-related action. These findings suggest that linguistic meaning is grounded in such systems (i.e., the motor system), challenging the traditional idea that word meaning can be explained on amodal and abstract symbols proposed by classic cognitivists (Fodor, 1983; Pylyshyn, 1984).

Meaning comes across in different ways: the dark clouds in the sky mean rain, a yellow bus means school bus, a mountain lion nearby means danger. For humans, linguistic meaning makes it possible to build complex ideas (i.e., create an unlimited set of meanings from different combination of concepts), and thus connect the internal knowledge with the external world.

Just like language, this internal knowledge is acquired from infancy. Large amounts of knowledge are learned from interacting with the environment: manipulating tools, preparing, smelling, and tasting food, watching animals move or make sounds, listening to natural sounds, music, and noise. From these perceptions, people develop internal representations (knowledge) to discuss and exchange ideas about the world. This knowledge is called semantic knowledge, and the memory involved in its representation and processing is called the semantic memory (Binder et al., 2009, Binder & Desai, 2011; Cree & McRae, 2003; McRae & Jones, 2013).

Neuroimaging (fMRI) technology has recently become a major tool to study semantic knowledge. It measures human brain activity non-invasively during word and sentence comprehension. Several fMRI studies on healthy volunteers and brain-impaired patients support the claim that these knowledge representations are at least partly "embodied" in different neural systems that play an essential role during conceptual acquisition and recall (Barsalou, 2008, Binder & Desai, 2011; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012).

Contemporary theories of semantic knowledge representation suggest that concepts are represented as sets of features that are the basic components of meaning. In the embodied approach, the meaning of a concept is not a set of verbal features that people associate with the concept, but rather a set of neural processing modalities that are involved while experience instances of a concept (Barsalou, 2008, Binder et al., 2009, Cree & McRae, 2003). For example, while experiencing a *dog* for the first time, individuals gather their understanding by the sensory input associated with them, i.e., *a dog is loud and furry*. The neural processing modalities involved in this experience are

Visual, Auditory and Somatosensory. Every time they experience the concept *dog* in the future, it will stimulate these same neural processing modalities (Barsalou, 2008, Binder et al., 2009, 2011, 2016a, and 2016b; Damasio, 1989). Thus, concepts are instrumental to cognition because they eliminate the need to re-learn an object's features every time it is experienced (Murphy, 2002).

Studying the neurobiology of the semantic memory, Binder et al. (2009, 2011, 2016a, 2016b) identified a distributed large-scale brain network of semantic processing systems that addresses many of the unresolved issues on concept representation (i.e., symbol vs. feature-based representation, feature selection, amodal vs. multimodal representation). Subsequently, they defined a model of semantic representations based entirely on such set of neural processing systems. This approach provides a direct correspondence between semantic content and neural representations (i.e., concept grounding; Harnad, 1990), and suggests that concepts can be represented through a number of weighted semantic features corresponding to different brain areas (Binder et al., 2009, 2011, 2016a, 2016b).

Binder's brain-based semantic feature model provides the background of the theoretical part of this dissertation. Likewise, a neural network model together with a collection of fMRI data constitute the foundation for the experimental portion. Combined, they form the basis for the framework developed in Chapter 3, to identify the changes in word meaning under different contexts.

The following sections review the theory of concept representations and word meaning, including: (1) fMRI technology and grounded cognition, (2) semantic organization in the brain, (3) text-based and multimodal computational semantic models, (4) the neural-based semantic model, and (5) the FGREP neural network model for learning representations. Together, they provide the basis for the neural-based semantic model developed in this dissertation to address the questions posed in the introductory paragraph.

## 2.2 fMRI Technology and Grounded Cognition

Are concepts grounded in perceptual and motor systems? Neuroimaging tools have become suitable tools to study the relationship of symbols, meaning and grounding. During concept understanding, this technology has shown significant correlation between the activity of a certain brain areas and properties of the task. This section briefly reviews brain imaging technology, followed by an analysis of the different theories of perception and cognition based on it.

### 2.2.1 The fMRI Technology

The availability of brain imaging tools has enhanced interest in how concepts are represented in human neural systems. Functional Magnetic Resonance Imaging (fMRI) is a technique for measuring brain activity (H. Devlin, 2018). It works by detecting the changes in blood oxygenation and flow that occur in response to neural activity. When a brain area is more active it consumes more oxygen, thus the blood flow increases to the active area. MRI uses magnetic resonance to form images of the anatomy of the brain. MRI becomes functional by adding a blood oxygen level-dependent (BOLD) signal, to reveal the anatomy of the brain and to show where blood oxygen is being heavily used during certain tasks. Thus, fMRI is an extension of MRI to capture functional changes in the brain caused by neuronal activity.

As a brain imaging technique, fMRI has the following significant advantages over other technologies such as EEG (Electroencephalography) and TMS (Transcranial Magnetic Stimulation): (1) It is non-invasive and does not involve radiation, therefore is safe for the subject, (2) It is possible to read the brain activations while a subject is performing cognitive tasks, and (3) It achieves high spatial resolution. On the other hand, it is limited of temporal resolution because the hemodynamic response imposes a major constraint on the time-precision of the measurement.

The two different states of hemoglobin, i.e., oxygen-rich oxyhemoglobin vs. oxygen-poor deoxyhemoglobin, differ in their magnetic properties. The large fMRI magnets are sensitive to changes in the concentration of deoxyhemoglobin. As neural activity increases, blood flow to the vasculature of the brain increases, altering this concentration. Consequently, fMRI is used to produce activation maps showing which parts of the brain are involved in a specific mental process.

The cylindrical tube of an MRI scanner houses a powerful electro-magnet. A typical research scanner has a field strength of 3 Teslas, about 50,000 times greater than the Earth's magnetic field; in some cases, scans up to 7 Teslas have been used (Morris et al., 2019; Yacoub et al., 2001).

Significant progress has been achieved over recent decades in understanding how conceptual knowledge is represented in the human brain. In particular, fMRI technology has encouraged cognitive scientists to explore the hypothesis regarding the knowledge organization and content of the brain. This technology allows researchers to observe regions of interest of brain activity while participants perform cognitive tasks such as reading sentences.

### 2.2.2 Grounded Cognition

The interaction and connection between body and mind, i.e., sensory-motor experience, and conceptual processing, has been studied for a long time. The history of philosophy on human cognition includes philosophers divided among empiricism, idealism, and rationalism (Borghesani, 2017). According to the proponents of the first view (e.g., Thomas Hobbes, John Locke, David Hume), at birth human's mind is a tabula rasa, i.e., a clean slate ready to be filled with knowledge acquired through sensory-motor experiences. Idealists (e.g., Plato, Kant) believe that humans are born with innate ideas, basic conceptual knowledge that does not require learning processes. Rationalists (e.g., Descartes, Spinoza, Leibniz) reject the connection of knowledge with perception, and claim that knowledge can be derived from reasoning independently of any sensory data.

Until recently, Descartes' mind-body dualism theory (i.e., a separation between mental and physical properties) was the norm, however, contemporary cognitive scientists revived the empiricist view that argues that sensory and motor experiences form the basis of conceptual knowledge. Under the label of Embodied Cognition, this theory is based on the hypothesis that all kinds of cognitive processes are rooted in the perceptual and action systems (Andrews et al., 2009, Binder et al., 2011).

Within the Embodied Cognition principles, the degree of embodiment varies across theories from strongly embodied to completely disembodied representations. Disembodied models propose a complete separation in which conceptual processing is based entirely on amodal, symbolic representations (Fodor, 1983; Pylyshyn, 1984). Other theories like Grounding by Interaction propose that conceptual and perceptual representations are different but interact closely, so that amodal symbols can derive content from the perceptual knowledge (Damasio, 1989; Patterson et al., 2007). Contrary to both of these theories, Strong Embodiment theory states that conceptual representation is grounded in the sensorimotor system. This approach proposes a direct connection between experiences and the semantic memory (Barsalou, 2008; Martin, 2007). In contrast to all of these theories, neuroanatomical evidence for multiple modality-specific systems converging on a common semantic network suggests a process of Embodied

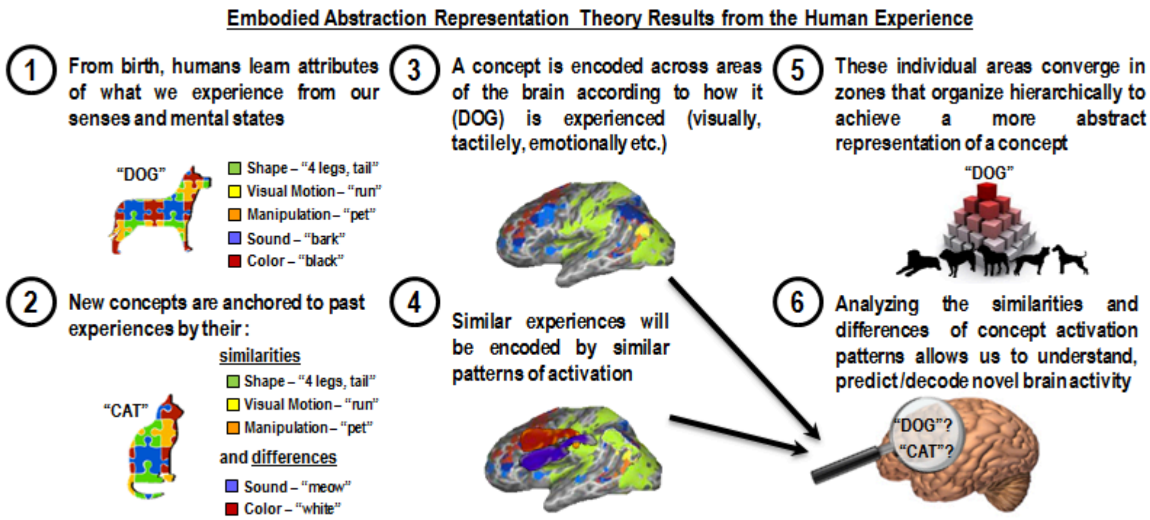



Figure 2.1: Embodied Abstraction. The central idea is that concept knowledge is built from experience. Therefore, knowledge representation is not static but changes with experience. The attribute representation captures the embodied abstraction of the concept as experienced by humans. (Reproduced with permission from Binder et al., 2011).

Abstraction, in which conceptual representations are formed in multiple levels of abstraction from sensory, motor and affective input (Binder et al., 2009, 2011, 2016a, 2016b).

These levels of abstraction are not accessed or activated under all conditions. Instead, access depends on the type of task, i.e., familiar context and rapid processing needs (Binder & Desai, 2011). The top level contains schematic representations that are highly abstracted from specific representations in the primary perceptual-motor systems. In novel contexts or when the task requires deeper processing, sensorimotor-affective systems make a major contribution to supplement the representations. This hierarchical representation is known as an heteromodal approach (Binder & Desai, 2011; Thompson-Schill, 2003).

Figure 2.1 shows the central idea of this approach as concept representation proposed by Binder et al. (2009, 2011, 2016a, 2016b). The main point is that conceptual knowledge is built from experience: (1) humans learn concepts (list of features) from birth throughout senses and mental states; (2) when learning new concepts, humans use past experiences to find similarities and differences; (3) concepts are encoded according to the way they were experienced; (4) similar experiences are encoded with similar neural patterns; (5) the encoding is done by a distributed neural network that includes a convergence zone to generalize the concept representation (i.e., classification); and (6) by analyzing similarities and differences of concept activation patterns, it is possible to understand novel brain activity. A comprehensive description of his theory will be presented in Section 2.5.

## 2.3 SEMANTIC ORGANIZATION IN THE BRAIN

Neuroimaging techniques have increased interest in neural and semantic organization of the brain. This section reviews the Semantic Memory function, followed by relevant issues regarding concepts, word meanings, conceptual combinations, and sentence context and contextual modulation (i.e., how sentence context modulates the meaning of a word).

### 2.3.1 Semantic Memory

Although philosophers have been wondering for centuries over the nature of concepts, semantic memory became a topic of formal study in cognitive science only recently (Binder & Desai, 2011). Tulving (1972) proposed that long-term memory is subdivided into declarative (facts) and procedural (skills) components. Declarative memory is further divided into two different but interacting systems:

- Semantic Memory is vital for understanding the meanings of words as well as knowing the facts about the world. It includes general knowledge independent of personal experiences with no connection to specific time or place in which it was acquired.

- Episodic Memory contains information about individual's own personal experiences. Episodic information is stored with information about when and where it was learned.

Therefore, semantic memory is the aspect of human memory that corresponds to general knowledge of objects, word meanings, facts and people, without connection to any particular time or place (Patterson et al., 2007). For example, knowing that you have been confined to your house since early March of year 2020 due to the COVID-19 pandemic, is stored in episodic memory. Knowing that Coronavirus refers to a highly contagious disease, that primarily spread between people during close contact, via small droplets such as coughing, sneezing, and talking, and common symptoms including cough, fever, fatigue, shortness of breath, and loss of smell, are all forms of semantic memory.

Consequently, semantic memory is linked to language because it includes word meanings, which are mostly shared across individuals in a given culture but can differ between cultures. For instance, English and Italian speakers both have different words for the body parts *foot* (It. *piede*) and *leg* (It. *gamba*), while Japanese speakers have one word *ashi* that refers to both *foot* and *leg* (Borghesani, 2017; Vigliocco et al., 2007). Thus, if word meanings map into people's mental representation of the world (e.g., objects, actions, events), are concepts and word meaning representations the same? Conceptual knowledge is believed to be universal (i.e., consistent) in its core features across cultures, contrary to the cross-linguistic variability of word meanings, like the example above (Vigliocco et al., 2007). Next section will discuss such issues referring to concepts, word meanings, their representations, and how they are related.

### 2.3.2 Concepts and Word Meaning

Categories of concepts include people, events, objects or ideas. Concepts in turn can be concrete or abstract. Concrete concepts like fruits, animals, buildings, and man-made tools have shape, dimensions, something that can be seen, heard or touched. Abstract concepts such as happiness, beauty, energy, idea, or holiday are not possible to perceive with the senses. These concepts are often difficult to understand specially by young children but as they grow, they develop strategies to master them. Usually, abstract concepts include a variety of experiential domains (spatial, temporal, social, affective, cognitive etc.) and the general agreement is that many abstract concepts are learned by experience with complex situations and events. (Barsalou, 1999; Binder, 2016a; Borghi et al., 2011; Vigliocco et al., 2009; Wiemer-Hastings & Xu, 2005).

Many researchers assume that concepts and word meanings are the same, or at least are linked on a one-to-one mapping (Humphreys et al., 1999). However, they cannot be the same simply because speakers of any language have far many more concepts than words (Vigliocco et al., 2007). According to Barsalou (1987), meanings are generated when a word is recognized in interaction with its context. Thus, word meanings use concepts but are not equal to concepts (Barsalou et al., 1993). Thus, how is the meaning of a word represented?

The earliest theory of concept representation may be traced to Aristotle (Cohen & Murphy, 1984). The **Classical** view is based on the idea that concepts are defined by a list of characteristics (features) necessary for the object or instance to be a member of that category. Each concept has a definition characterizing its "essence" and providing the necessary and sufficient conditions for concept membership. For example, the concept *bachelor* includes the characteristics *unmarried*, *adult* and *male*. Objects that do not match all the features cannot accurately be named by that word. An object that is missing the feature *unmarried* might be called a *man*, if the *married* feature is present, a *husband*.

This view of concepts was dominant until 1950s. It was then challenged by studies done by Rosch (Rosch & Mervis, 1975) demonstrating that people do not hold such lists of attributes to decide a category membership. For example, the concept *bachelor* does not match the Pope, although he definitely satisfies *bachelor's* definition (Cohen & Murphy, 1984). She found that individuals keep a mental picture of what made up an example of a member of a class, giving rise to the prototype theory (Rosch & Mervis, 1975).

The **Prototype** theory builds on the typicality effect, i.e., that concepts are organized around examples. A prototype is an object or item that is the most typical of that concept. The prototype could be a real example that has been experienced or an amalgamation of various examples of the concept (i.e., an abstraction). To determine category membership for novel items, the number of components that the two concepts have in common is compared. For example, the reason *apples* are considered to be more typical than *plums* is because the concept *apple* shares many features with the concept *fruit*. Although, strong evidence supported this theory, some problems started to emerge for cases where participants judged items to be more typical of a category in some situations than in others. For example, *cottage cheese* is more typical of *pudding* than *cheese*, but it is categorized as *cheese*.

The **Exemplar** theory challenged the other theories by proposing that specific examples of concepts are learned, and stand as the representation, instead of a generalized prototype, or a list of specific mandatory characteristics. Not all the experienced examples are stored in memory, but many are retained. This theory faced challenges too, as in the number of exemplars that should be stored take larger space than a single prototype would need (Brooks, 1978).

Barsalou (1993) proposed the **Experiential** view based on the idea that conceptual representations are learned from experience and are built from smaller components. The representations are more abstract and schematic, excluding many details present during perception (e.g., *chair* representation might include "seat", "back" and "legs", while omitting color and texture). These representations do not appear from the basic sensory-motor modalities; instead, perceptual representations emerge from any aspect of experience, including proprioception (sense of position, location, and orientation), introspection (self-examination of representation states), emotions, and so on.

In this theory, the representation of meaning is considered in terms of a list of features, that is, the properties collectively express the meaning of a word. However, some approaches differ in the type of conceptual features used, with some concepts relying on sensory-related properties, and others on motor-related properties (e.g., Barsalou, 1993; Cree & McRae, 2003; Damasio, 1989; Vigliocco et al., 2004). Consequently, some studies have attempted to gain insight into those dimensions of meaning by asking participants to provide a list of the features (semantic feature norms) that they believe to be important in describing and defining the meaning of a given word (Cree & McRae, 2003; McRae et al., 1997; Vigliocco et al., 2004; Vinson & Vigliocco, 2002). A common characteristic of these models is the fact that the semantic features are chosen a priori by the investigators and may not reflect the full range of properties of meaning that may be relevant to the representations of the words (Vigliocco et al., 2007). In addition, the features are problematic in the sense that these features are complex concepts themselves, i.e., “seat’, ‘back’, and “legs”, are components of a larger entity such as *chair*, but they are no more basic that the entity they define.

A major departure from verbal feature analysis is an abstract embodiment theory of knowledge representation that provides a straightforward analysis of conceptual content in terms of sensory, motor, affective, and other experiential phenomena and their corresponding modality-specific neural representations. The neurobiologically defined experiential attributes provide a set of primitive features for the analysis of conceptual content, while simultaneously grounding concepts in experience (Binder et al., 2011).

The main idea in this brain-based theory is that people weigh concept features differently based on context, i.e., they construct meaning dynamically according to the combination of concepts that occur in the sentence. Thus, this model allows conceptual representations to be dynamically sensitive to context (Binder et al., 2011; Yee & Thompson-Schill, 2016). More details about this brain-based concept attribute representation theory are given in Section 2.5.

Finally, returning to the questions of how the meaning of a word is represented and how concepts and word meanings are related, there seems to be a clear distinction between conceptual and semantic levels of representation. One way in which this distinction is achieved is by assuming that concepts consist of grounded componential feature representations, and word meanings (lexical semantics) bind these representations with the goal of using them in language. Simply put, word meanings are represented as lists of features that together express the meaning of the word (Damasio et al., 2004; Vigliocco et al., 2004, 2007). Next two sections discuss how do word meanings change during concept combination and under the context of a sentence.

### 2.3.3 Conceptual Combination

Conceptual combination is the process where complex concepts are constructed from simpler constituents (e.g., *olive oil* from *olive* and *oil*). Generating such combinations is an intriguing example of a high level cognitive process that humans perform very quickly. People are likely to create novel noun-noun phrases in their conversations i.e., *dancer game*, and listeners are capable at understanding them, i.e., playing games by dancing (Murphy, 1988).

In language, complex concepts are expressed by noun phrases of the form adjective-noun (i.e., *yellow car*), noun-noun (i.e., *trash can*), noun-verb (i.e., *duck slept*), and verb-noun (i.e., *drank tea*). Wisniewski (2000) distinguished three types of conceptual combination interpretations: (1) property (or attribute) combination involves one or a few properties of one word applied to the combination (e.g., *red apple* is an apple that can be red), (2) relational combination usually requires a thematic relation such as an association or world knowledge to explain the combination (e.g., *apple basket* is a basket that holds apples), and (3) hybrid combination normally applies to a conjunction of the constituents or a cross between them (e.g., *apple pie* is a pie made of apples), (Wisniewski, 1997).

For example, listeners must realize that *red apple* could mean just a fruit having a certain color by selecting salient features that dominate in the combination. The noun *apple* is defined by color, size, shape, taste etc. and one or more of those dimensions will be modified during the attribute combination. In relational combination, the modifier features have nothing to do with the combination. For example, *apple basket* or *apple pie* contain a variety of relations that often do not include *apple's* features as in *apple baskets* are not edible, red or a fruit. To help understand that *apple pie* is made of apples, but *apple baskets* are not, a thematic relation needs to be built based on world knowledge about plausible combinations. In the hybrid combination, interactions during conceptual combination are likely to take the form of additional processing within the neural systems where attribute representations overlap. This additional processing is necessary to compute the new attribute representations resulting from conceptual combination, and these new representations tend to have added salience. Usually, the more communalities the constituents have, the more likely can be interpreted as a hybrid combination. For example, *apple pie*, both constituents share features such as taste, smell, and edible.

All three types of combinations contribute significantly to the construction of new or complex concepts (Gagné & Shoben, 1997; Murphy 1990; Pecher, Zeelenberg, & Barsalou, 2004). Similarly, a mechanism proposed by (Medin & Shoben, 1988) known as the centrality effect, plays a useful role in the process of conceptual combination not being considered in Wisniewski's (1997) models. Centrality expresses the idea that some attributes are true to many different concepts, but they are more important to some concepts than others (e.g., it is more important for basketball than for cantaloupes to be round).

The Centrality effect cannot be used to combine concepts by adding or changing a single feature. Attributes are not independent of each other, and the interpretations relies on the relationship among attributes and on the multidimensional effects that an attribute has on a noun. For example, *small pen* vs. *small car*, to interpret such combinations, world knowledge needs to be used to obtain the correct relationship.

### 2.3.4 Sentence Context/Contextual Modulation

While comprehending a sentence, sensory stimuli must be mapped into meanings. This process not only involves determining the meaning of individual words but also the meaning of the combination of words that appear in a sentence (Humphries et al., 2008).

Context provides the background for the real action of the main events. More importantly, as a consequence of learning and semantic processing, context often helps select appropriate behaviors and determine the explicit and implicit content of human thought (Rudy, 2009). For example, the nouns *boat* and *basketball* each have their own meanings, however, when the words appear in the context of a sentence such as *Chris used a basketball as a life preserver when the boat sank* (Barsalou, 1982), the context brings up unusual features to mind such as "basketball floats". Thus, when words share features, those aspects of the word representation that are relevant to the context are strengthened (Hampton, 1997; Kiefer & Pulvermüller, 2012; Medin & Shoben, 1988; Mitchell & Lapata, 2010; Murphy, 1990; Wisniewsky, 1998).

Along those lines, conceptual representations are dynamic, changing not only in response to context as it relates to stimulus modality and task, but also in response to the context an individual brings with them, i.e., recent or long-term experience. Putting all dynamic effects together creates a (moving) picture adjusted by a semantic system that is (1) experience-based such that it allows representations to change over time, and (2) distributed such that various features can become active at different rates, in different situations (Yee & Thompson-Schill, 2016).

Thus far, the fMRI technology and the main topics of semantic organization have been reviewed. Next section will present different computational models of conceptual representation. Such models have the potential to explain the organization of the semantic memory and thus provide the groundwork for this dissertation.

## 2.4 Computational Models of Semantic Representation

Current computational models of semantic representation address how such representations may develop in the first place. (Yee et al., 2018). While there are many models in the literature, most of them fall into two general classes: theories based on relations among words, i.e., those in which a word's meaning is represented by way of its relation to other words (Bruni et al., 2014; Kintsch & Mangalath, 2011; Mitchell, 2008; Mitchell & Lapata, 2010), and feature-based, i.e., those in which a word's meaning is represented as basic components of meaning, which together make the meaning of the word. Feature-based models further differ in the way the features are defined, i.e., whether they are abstract or embodied (Cree & McRae, 2003; McRae et al., 1997; Vigliocco et al., 2004; Vinson & Vigliocco, 2002).

The following models are briefly reviewed with emphasis on aspects relevant to this dissertation. First, a currently dominant model which constructs semantic representations based on relations between words (a.k.a. distributional semantics), is described. This model uses large corpora of texts in order to compute aspects of a word's meaning based on those other words found in the same linguistic contexts, i.e., *dog* is related to *leash, bone,* and *collar* (Mitchell et al., 2008). Then, an alternative view, where models derive semantic representations using multimodal information i.e., textual and feature-based visual inputs, is discussed. Last, Section 2.5 describes a third kind of feature-based model proposed by (Binder et al., 2009, 2011, 2016a, 2016b) that uses "experiential attributes" to represent word meaning. This third approach is the basis for the semantic representations in this dissertation.

### 2.4.1 Text-based Semantic Representation

A text-based computational model designed to predict neural activation patterns to understand how conceptual knowledge is represented in the brain was first proposed by Mitchell et al. (2008). The model uses word co-occurrence statistics derived from a large text corpus to model semantic content. The central idea is that the neural basis of the semantic representation of concrete nouns is related to the distributional properties of those words in a large scale corpus of language.

A collection of fMRI patterns for 60 different word-picture pairs that include five items from each of 12 semantic categories (animals, body parts, buildings, building parts, clothing, furniture, insects, kitchen items, tools, vegetables, vehicles, and other man-made items) were recorded. A set of 25 verbs that reflect the basic sensory and motor activities, actions performed on objects, and actions involving changes to spatial relationships was then selected: *see*, *hear*, *listen*, *taste*, *smell*, *eat*, *touch*, *rub*, *lift*, *manipulate*, *run*, *push*, *fill*, *move*, *ride*, *say*, *fear*, *open*, *approach*, *near*, *enter*, *drive*, *wear*, *break*, and *clean*. These verbs were used as semantic features in two ways: by counting word co-occurrence with the object stimulus word, and through fMRI patterns (for each verb).

The model predicts the neural activation for any given stimulus word *w* using a two-step process (Fig. 2.2). The first step encodes the meaning of stimulus word *w* as a 25-dimensional vector, with each value corresponding to the normalized sentence-wide co-occurrence of stimulus word *w* with one of 25 selected verbs using a large text corpus (Brants & Franz, 2006). For example, one intermediate semantic feature might be the frequency with which *celery* co-occurs with the verb *taste*. The second step predicts the neural fMRI activation at every voxel location in the brain as a weighted sum of neural activations contributed by each of the intermediate semantic features.

The results identify a direct, predictive relationship between the statistics of word co-occurrence in text and the neural patterns associated with thinking about the word meaning. Also, the computational model trained to make these predictions shed light into

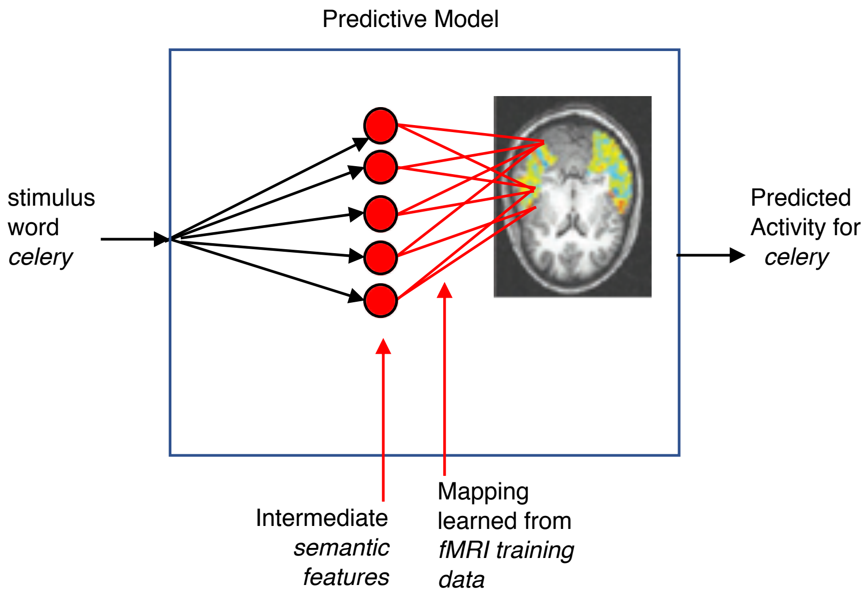



Figure 2.2: Predicting fMRI activation for nouns. First, the model encodes the meaning of the input word via intermediate semantic features (extracted from a large corpus of text). Then, predicts the fMRI image as a linear combination on the fMRI signatures associated with each of these intermediate semantic features. Predicting fMRI activation using text-based semantic representations. (Mitchell et al., 2008).

how the neural activity can be associated with different semantic components of the objects. This model focuses on encodings of abstract semantic concepts denoted by words and predicts brain-wide fMRI activations based on text corpus features that capture semantic aspects of the stimulus word, compared to visual or auditory features that capture perceptual aspects.

This distributional, or text-based, model of semantic representations helps understand how concepts are represented in the brain. Such models became popular because they make good predictions about the structure of the semantic space in the brain. Also, they are fast and easy to construct on a very large scale to model the way humans acquire and represent knowledge. In contrast to this alternative, semantic representations based on multimodal information are reviewed next.

### 2.4.2 Multimodal Semantic Representations

Many experimental studies in language acquisition suggest that word meaning arises not only from exposure to the linguistic environment but also from our interaction with the physical world (Andrews et al., 2009; Bruni et al., 2014; Feng & Lapata, 2010; Lenci, 2008). There is therefore an increasing interest on grounding semantic models in sensory modalities (Andrews et al., 2009; Bruni et al., 2014; McRae et al., 2005).

Early studies demonstrated that combining text-based distributional information with man-made conceptual attributes is an alternative for perceptual experience, providing a good approximation to human-like semantic representations (Andrews et al., 2009). However, in most cases both components were derived from linguistic information. Recent models of brain analysis have used multi-modal inputs in which the features are assumed to be grounded (Anderson et al., 2013; Feng & Lapata, 2010; Silberer et al., 2017).

Along these lines, Feng & Lapata (2010) proposed the first multimodal distributional semantic model that combines visual and linguistic representations of word meaning. The model learns from natural language corpora paired with images, assuming the images describe some of the document's content. It extracts the semantic representation from a large collection of BBC News articles and their associated images without human involvement. Words are represented by their distribution over a set of latent multimodal dimensions, or topics derived from the textual and visual features. The idea is to build a common latent space by merging the two sources of information. The main limitations of this approach are that textual and visual data must be extracted from the same corpus, constraining the choice of corpora to be used, and the approach does not allow much flexibility in how the two information channels are combined.

Similarly, Anderson et al. (2013) tested whether image-based models capture the semantic patterns that emerge from fMRI recordings of neural signal. They used the fMRI collection of Mitchell (2008), focusing on a subset of 51 words out of the original 60 words. They collected text co-occurrence statistics from ukWaC corpora and the English Wikipedia's full content (2009) combined, by selecting 20K, 5K, and 5K most frequent nouns, adjectives and verbs respectively.

The visual model was based on that of Bruni et al. (2014), which extracts features from images separately from the object and its surrounding context, leading to a better performance. Particularly when evaluating inter-object similarity, the context in which the object is located can otherwise contribute significantly to semantic representation. The collection for the visual data was extracted from ImageNet (Deng et al., 2009), a large image database organized on top of the WordNet hierarchy, known for high quality images with concept annotations.

Once the two modalities were built, they were combined by concatenation, and the relationship between the distributional models and the brain data was evaluated by using representational similarity analysis (i.e., correlation analysis). They found that image-based distributional semantic largely correlate with fMRI-based neural similarity patterns mainly for categories of concrete concepts. Correlations at the conceptual level were low, which suggests the need to develop better distributional models and/or reduce the noise inherent in neural data. Moreover, image-based models complement a state-of-the-art text-based model, with the best performance achieved when the two modalities are combined.

A novel approach developed by Silberer et al. (2014) utilized multimodal representations by using stacked autoencoders (deep networks) to learn higher level embedding from text and visual input (Silberer et al., 2017). Existing models (including those reported above) present words as vectors resulting from a combination of representations with sometimes different statistical properties that do not necessarily have a natural equivalence (e.g., text and images). The Silberer et al. model computes meaning representations at a finer level of granularity for individual words and is unique in its use of attributes as a means of representing the textual and visual modalities. Since humans acquire a large amount of their semantic knowledge from perceptual input (Anderson et al., 2013; Bruni et al., 2014; Silberer and Lapata, 2017), the goal was to capture such intrinsic meaning.

Earlier work by Silberer et al. (2013) had shown that automatically predicted visual attributes can be excellent substitutes for feature norms (i.e., features obtained from human participants to describe the meaning of a word). The Silberer et al. (2013) model then learned multimodal representations from attributes that were automatically inferred from text and images. Textual attributes were extracted from a 2009 download of the English Wikipedia using Strudel, a fully automatic approach for extracting weighted word attribute pairs (e.g., swan-bird:n) from a lemmatized and POS-tagged corpus. This step returned a total of 2,362 dimensions for the text-based vectors. Visual vectors were obtained by using an SVM-based attribute classifier that predicts visual attributes for images. The dataset was a taxonomy of 636 visual attributes (e.g., has_wings, made_of_wood) and nearly 700K images from ImageNet, and was used to form 414-dimensional vectors for each noun. A stacked autoencoder was then used to project these linguistic and visual vectors onto a unified representation that fuses the two modalities together. The goal was to learn multiple levels of representations through a hierarchy of network architectures, where higher-level representations are expected to help define higher-level concepts.

Experimental results in two tasks, i.e., word similarity and word categorization, showed that their model outperforms other models trained on the same attribute-based input. Also, the evaluation revealed that the bimodal model (combined representation) is superior to the unimodal, and that higher-level unimodal representations are better than the original input.

These multimodal studies suggest that semantic features provide useful insight into several cognitive process regarding concept representation, categorization, and semantic memory. They are useful in different tasks such as testing hypothesis, constructing experimental stimuli, and designing experiments. In addition, semantic features are a convenient approach because: (1) they are a natural (human) way to express salient properties of word meanings (McRae et al., 2005), (2) they allow for easy integration of different modalities, and (3) they adequately describe visual components, e.g., objects, scenes (Silberer and Lapata, 2012).

The computational models presented in this section claimed to have grounded the semantic representations by using features from the visual modality along with text. However, truly multimodal representations should account for the full array of human senses (Bruni et al., 2014). Next section presents such an approach: An abstract embodiment theory of knowledge representation that uses a set of neural features for the analysis of conceptual content, while simultaneously grounding concepts in experience (Binder et al., 2011).

### 2.5 The Brain-based Semantic Representations

The vector representations reviewed in the previous section are based on word co-occurrence and other automatically or manually generated feature modalities which do not provide precise information about the experienced features of the concept itself

(Anderson et al., 2016). They are missing intrinsic knowledge (i.e., they are ungrounded). Moreover, the abstract description of such representations significantly restricts the possibility of capturing important aspects of conceptual knowledge representation.

One alternative proposed by Binder et al. (2009, 2011, 2016a, 2016b) is to model concept representations based on known brain systems. Most importantly, such representations are not limited to the classical sensory-motor dimensions associated to the strong embodied theories (Section 2.2.2). The following sections provide an overview of Binder's proposed semantic brain systems, the heteromodal semantic processing model, and the theory of Concept Attribute Representation (CAR).

### 2.5.1 Neural Brain Systems

The CAR theory (a.k.a. The Experiential attribute representation model) is supported by substantial evidence on how humans acquire and learn concepts through sensory-motor, affective, social, and cognitive interactions with the world (Binder et al., 2009, 2011, 2016a, 2016b). The central axiom of this theory is that concept knowledge is built from experience, as a result, knowledge representation in the brain is not static. This process starts from birth: babies learn about objects through sensory input (e.g., *parrot* is *loud*). As they develop, the dimensions to such concepts expand through more modalities, including visual, somatosensory, and auditory (e.g., *parrot* is *green*, has *feathers*, and is *musical*). Later in life, humans connect previous concepts to new ones (e.g., *parrots* are similar to *penguins*), while also learning how to differentiate between the concepts.

The theory suggests that conceptual knowledge can be decomposed into a set of features that are mapped to individual brain systems. It is based on these assumptions: (1) recalling a concept stimulates the features that were active when the concept was first experienced; (2) concepts with similar features produce similar neural patterns; and (3) context modifies the baseline meaning of a concept. This last assumption is the focus of this dissertation.

Table 2.1 lists the set of features defined in CAR theory. The first column shows the brain systems: Vision, Somatosensory, Auditory, Gustatory, Olfactory, Motor, Spatial, Event, Cognitive, Evaluation, Emotion, Drive, and Attention. The second column shows the list of 66 features embodying the concept representations. The third gives a high-level description of each feature.

These features were selected based on physiological evidence with two assumptions: (1) All aspects of mental experience can contribute to concept acquisition and, consequently, to concept composition; (2) experiential phenomena are grounded in neural processors representing a particular kind of experience. These features are all visible in the brain at the macroscopic scale of in vivo imaging, and they are represented as a continuous one-dimensional variable (i.e., a scalar quantity). Next, each brain system will be outlined with the proposed group of features and how they apply to a number of concepts.

**Visual System** Visual features include luminance, size, color, texture, shape, motion, and biological motion. Within the shape-perception system, there are separate subsystems that primarily process faces, human body parts, and three-dimensional spaces. The following list outlines how each Visual feature applies to different concepts:

- Bright or Dark: Apply to concepts that are brilliant or obscure, e.g., sun, light, shine, ink, night, dark, and black.
- Color: Applies to color concepts, e.g., green, red.
- Motion, Fast or Slow: Apply to concepts that involve movement, e.g., run, dash, zoom, and walk, but also to nouns such as bullet, jet, hare, snail, and tortoise, and

Table 2.1: The twelve brain systems and the 66 features used as the basis for the CAR theory. The first column lists the brain systems. The second column includes the list of features as basic components of meaning. The third column presents a description of each feature. List of attributes representing the semantic system proposed by Binder et al., (2009, 2011).

| BRAIN SYSTEMS | FEATURES | EXPLANATION | BRAIN SYSTEMS | FEATURES | EXPLANATION |
|---|---|---|---|---|---|
| | Vision | something that you can easily see | S | Landmark | having a fixed location, as on a map |
| | Bright | visually light or bright | P | Path | showing changes in location along a particular direction to path |
| V | Dark | visually dark | A | Scene | bringing to mind a particular setting or physical location |
| I | Color | having a characteristic or defining color | T | Near | often physically near to you (within easy reach) in everyday life |
| S | Pattern | having or defining visual texture or surface pattern | | Toward | associated with movement toward or into you |
| I | Large | large in size | | Away | associated with movement away from or out of you |
| O | Small | small in size | | Number | associated with a specific number or amount |
| N | Motion | showing a lot of visually observable movement | | Time | an event that occurs at a typical or predictable time |
| | Biomotion | showing movement like that of a living thing | E | Duration | an event that has a predictable duration, whether short or long |
| | Fast | showing visible movement that is fast | V | Long | an event that lasts a long period of time |
| | Slow | showing visible movement that is slow | E | Short | an event that lasts a short period of time |
| | Shape | having a characteristic or defining visual shape or form | N | Caused | caused by some clear preceding event, action, or situation |
| | Complexity | visually complex | T | Consequential | likely to have consequences (cause other things to happen) |
| | Face | having a human or human-like face | | Social | an activity or event that involves an interaction between people |
| | Body | having a human or human-like body parts | C | Human | having human or human-like intentions, plans, or goals |
| S | Touch | something that you could easily recognize by touch | O | Communication | a thing or action that people use to communicate |
| O | Temperature | hot or cold to the touch | G | Self | related to your own view of yourself, part of YOUR self-image |
| M | Texture | having a smooth or rough texture to the touch | | Cognition | a form of mental activity or a function of the mind |
| S | Weight | light or heavy in weight | E | Benefit | someone or something that could help or benefit you or others |
| | Pain | associated with pain of physical discomfort | V | Harm | someone or something that could cause harm to you or others |
| | Audition | something that you can easily hear | A | Pleasant | someone or something that you find pleasant |
| A | Loud | making a loud sound | L | Unpleasant | someone or something that you find unpleasant |
| U | Low | having a low-pitched sound | | Happy | someone or something that makes you feel happy |
| D | High | having a high-pitched sound | E | Sad | someone or something that makes you feel sad |
| I | Sound | having a characteristic or recognizable sound or sounds | M | Angry | someone or something that makes you feel angry |
| T | Music | making a musical sound | O | Disgusted | someone or something that makes you feel disgusted |
| | Speech | someone or something that talks | T | Fearful | someone or something that makes you feel afraid |
| G | Taste | having a characteristic or defining taste | | Surprised | someone or something that makes you feel surprised |
| S | Smell | having a characteristic or defining smell or smells | DR | Drive | someone or something that motivates you to do something |
| M | Head | associated with actions using the face, mouth or tongue | | Needs | someone or something that would be hard to live without |
| O | UpperLimb | associated with actions using the arm, hand or fingers | ATT | Attention | someone or something that grabs your attention |
| T | LowerLimb | associated with actions using the leg or foot | | Arousal | someone or something that makes you feel alert or excited (+/-) |
| O | Manipulation | a physical object you have personal experience using | | | |
| R | Object | a physical object | | | |

to verbs that describe patterns of movement such as turn, roll, bounce, float, spin, and twist.

- Biological motion: Applies to concepts that are key for animate entities related to face and body part actions, e.g., boy, girl, woman, and man.
- Shape: Applies to mass concepts, e.g., butter, coffee, rice. In contrast, Shape does not apply to substance concepts, e.g., water, metal, plastic.

- Complexity: Applies to concepts related to animals, plants, and tools, e.g., car, plane, computer, dog, chicken, bird, hammer, cloud. Animal concepts tend to have more complex shapes.
- Face & Body: Apply to many human concepts including Face and Body themselves, but also to general human types and roles, e.g., boy, man, nurse. The Face feature may have a high value for specific individuals such as brother, father, and Einstein.

**Somatosensory System** Somatosensory features represent body location (somatotopic), body position and joint force (proprioception), pain, and surface characteristics of objects (texture and temperature). Additionally, these features are multimodal because a body location touched by an object is the same as body location involved in manipulative motor actions of such an object. These features encode the action as the salient characteristic of the object rather than the tactile features (which is represented through the Motor System). The following list outlines how each Somatosensory feature applies to different concepts:

- Touch, Temperature, Texture, Weight & Pain: Apply to all living things, e.g., boy, mouse, duck, bird, tree; to events and places, e.g., summer, beach; to buildings and natural disasters, e.g., church, hospital, hurricane; to human roles, e.g., criminal, teacher, lawyer; to verbs, e.g., fly, kick, walk; and to artifacts, e.g., car, plane, newspaper.

**Auditory System** The auditory cortex includes low-level areas tuned to frequency, amplitude, and spatial location, as well as higher levels that specialize in auditory objects such as environmental sounds and speech sounds. The following list outlines how each Auditory feature applies to different concepts:

- Low pitch, High pitch and Loud: Apply to concepts with auditory characteristic, e.g., mouse noise, bell chime, piano sound, explosion sound.
- Music: Applies to concepts that refer to musical features, e.g., sing, song, melody, harmony, rhythm; and to entities, e.g., instruments, artist, theater.
- Speech: Applies to concepts that refer to speech-like sound, e.g., reporter, farmer, party, duck, church, embassy.

**Gustatory & Olfactory Systems** Both features refer to concepts with Taste and/or Smell experiences such as food, drinks, places, events, animals, humans, plants and actions, e.g., bread, tea, coffee, restaurant, beach, forest, party, dog, boy, tree, dead.

**Motor System** These features capture the degree to which a concept is associated with actions involving specific body parts. The following list outlines how each Motor feature applies to different concepts:

- Head: Applies to concepts related to face, mouth, and tongue movements, e.g., spoke, dinner, reporter, school, interview.
- Upper limb: Applies to concepts related to arm, hand, or fingers movement, e.g., restaurant, injured, write, pencil, football, piano, play, pilot, throw.
- Lower limb: Applies to concepts associated to leg or foot movement, e.g., kick, walk, football, steal, beach, cross.
- Object & Manipulation: Apply to concepts that refer to actions and personal experience performed with an object or instrument, e.g., steal, kick, walk, damage, flower, boat, television, pilot, author.

**Spatial System** Spatial features generally involve multimodal inputs. For instance, perception of three-dimensional space is experienced via movement of the body through space, and often include the Visual and Auditory systems. Prepositions (i.e., about, from, in, far, behind) express most of the spatial content that occurs in English language, but this semantic representation focus instead on verbs, nouns and adjectives. The following list outlines how each Spatial feature applies to different concepts:

- Landmark: Applies to concepts referring to large entities that have a fixed location, like mountains, buildings, and parks, e.g., church, hospital, library, lake, forest.
- Path: Applies to concepts denoting direction of motion and a type of path, e.g., hurricane, ascend, climb, rise, fall, jump, circle, run, walk, swim, mouse, horse, pilot.
- Scene: Applies to concepts describing spaces and buildings, e.g., kitchen, library, hospital, church; and to concepts that evoke scenes by thematic relations such as oven evokes kitchen.
- Near: Applies to concepts related to paths from an observer, e.g., arrive, depart, leave, return; and to entities associated with a path, e.g., bird, butterfly, bicycle, plane.
- Toward & Away: Apply to concept objects that are within reach, to actions performed on objects, and to objects that move toward and away to the self, e.g., pencil, book, give, punch, tell, throw, acquire, catch, receive, eat, car, hurricane, football, plane.

**Temporal & Causal Systems** Temporal features refer to duration and temporal order of events. Causal features pertain to cause-and-effect relationship between concepts. The following list outlines how Temporal and Causal features apply to different concepts:

- Number: Applies to concepts referring to recurrent events, e.g., shower, breakfast, lunch, weekend, summer, school, payday.

- Duration: Applies to concepts associated with a typical duration, e.g., breakfast, lecture, movie, shower; and to concepts that specifically refer to duration, e.g., minute, hour.
- Long & Short: Apply to concepts associated with a long or short duration, e.g., childhood, college, life, blink, flash, sneeze.
- Time: Applies to concepts associated with a particular or predictable point in time like recurring daily events, e.g., shower, breakfast, commute, lunch, dinner; to annual events, e.g., holidays, seasonal phenomena; and to phases of life, e.g., infancy, childhood, retirement.
- Causal: Applies to concepts associated with a clear preceding cause such as an event or situation, e.g., infection, kill, laugh, spill; and to those that occur without an apparent cause like natural phenomena, e.g., tornado, flood.
- Consequence: Applies to concepts referring to events and actions that have probable consequences, e.g., dead mouse, accident, summer, party, steal, injured, protest, break, hurricane.

**Social & Cognitive Systems** Social features capture intentionality. They include events with social interactions as well as the degree to which a thing or an action reflects human or human-like intentions, plans, or goals. These features usually refer to people's roles and actions. Social and Cognitive concepts involve partial simulation of the Cognitive system and a partial activation of the Somatosensory and Motor systems. The following list outlines how Social and Cognitive features apply to different concepts:

- Self: Applies to concepts (nouns and verbs) related to one's view of oneself, and to concepts that describe an activity that is typical of oneself, e.g., party, school, dinner, teacher, scientist, write, walk.

- Communication: Applies to verbal and nonverbal concepts associated with social interaction and communication, e.g., speech, book, meeting, speak, read, meet, give, editor.
- Human & Cognition: Apply to concepts that evoke human activities and thought, e.g., teacher, scientist, steal, write, voter, party, watch, school.

**Emotion & Evaluation Systems** Emotion features include the degree of association of a target concept with each of Ekman's basic emotions: anger, disgust, sadness, happiness, and surprise (Ekman, 1992). Also, the features include affective states that arise from a combination of more fundamental dimensions known as valence and arousal, leading to a separate features for pleasantness and unpleasantness. The following list outlines how Affect and Evaluation features apply to different concepts:

- Happy, Sad, Angry, Disgusted, Fearful, and Surprised: Apply to concepts related to an entity, object, event or place that evoke happiness, e.g., party, summer, beach; sadness, e.g., hurricane; accident, injured; anger, e.g., politician, protest, break; disgust, e.g., steal, terrorist, mouse; fear, e.g., policeman, flood, hospital; and surprise, e.g., interviewed, jury, party.
- Benefit, Harm, Pleasant, and Unpleasant: Apply to concepts associated to an entity, object, event or place that cause benefit, e.g., school, minister, book; harm, e.g., storm, steal, criminal; pleasantness, e.g., forest, theater, vacation; unpleasantness, e.g., commander, trial, storm.

**Drive System** These features represent the degree of general motivation associated with the target concept, and to which the concept itself refers as a basic need. The following list outlines how each feature applies to different concepts:

- Drives: Applies to concepts that motivates to do something. They are closely related to emotions, which are produced when needs are fulfilled or frustrated, e.g., school, teacher, steal, write, read, voter, run, vacation.
- Needs: Applies to concepts that help reach homeostasis (equilibrium) or enable growth-like physiological needs, e.g., food, rest, sleep, sex; also, security, social contact, and approval, e.g., walk, food, water, hospital, happy, friendly, bed.

**Attention System** These features represent the degree to which a concept is associated to an entity, object, event or place that elicits attentiveness or alertness. Some concepts (e.g., scream) are so strongly associated with attention that the attentional event becomes part of the conceptual representation. The following list outlines how Attention and Arousal features apply to different concepts:

- Attention: Applies to concepts that prompt mental focus, e.g., flower, mouse, theater, dog, pilot, listen, television, protest, storm.
- Arousal: Applies to concepts that are associated to being awake or reactive, e.g., school, books, morning, criminal, hospital, party, football, happy, negotiate.

Putting all these together, the following section reviews the neuroanatomical model of the semantic memory proposed by Binder et al. (2009, 2011, 2016a, 2016b) by recognizing the regions underpinning the brain systems described here.

### 2.5.2 Heteromodal Semantic Processing Model

As explained in Section 2.2.2, the heteromodal semantic system proposed by Binder et al. (2009, 2011, 2016a, 2016b) is based on the idea that semantic knowledge is built from multiple levels of abstraction of sensory, motor, and affective input. Hubs or convergence zones then process semantic information (Damasio, 1989; Patterson, 2007). These hubs receive and combine input from multiple modality-specific areas. They bind together features and transform their input such that they represent similarity among concepts that cannot be captured based on individual sensory or motor modalities (Patterson et al., 2007).

Analogously to the distributional models of text (Landauer & Dumais, 1997; Lund & Burgess, 1996), heteromodal representations capture conceptual similarity by combining weighted inputs from multiple neural channels. In addition, these abstract representations capture thematic relations between concepts (as explained in Section 2.3.3) arising from spatial, temporal or functional proximity rather than from feature similarity. Most importantly, they provide a direct mapping to symbolic representations allowing for a rapid response in language interactions and to support communication tasks that do not require access to details (i.e., small talk).

The proposed neural architecture is shown in Figure 2.3. Concepts are abstracted from perceptual, motor, and affective experiences that engage strongly modal, low-level brain systems, indicated in green. This abstraction process produces modality-specific conceptual representations in modal convergence zones, indicated in yellow. Multimodal generalization and language promote gradual development of supramodal concept representations in the areas indicated in red. They include a temporal lobe convergence zone emphasizing object knowledge, an inferior parietal convergence zone emphasizing event knowledge, and a posterior cingulate/precuneus convergence zone emphasizing the encoding of meaningful events into episodic memory (functioning as an interface between the semantic neural systems and the hippocampal system). This hierarchical abstraction scheme is most obvious in the ventral visual object recognition pathway on

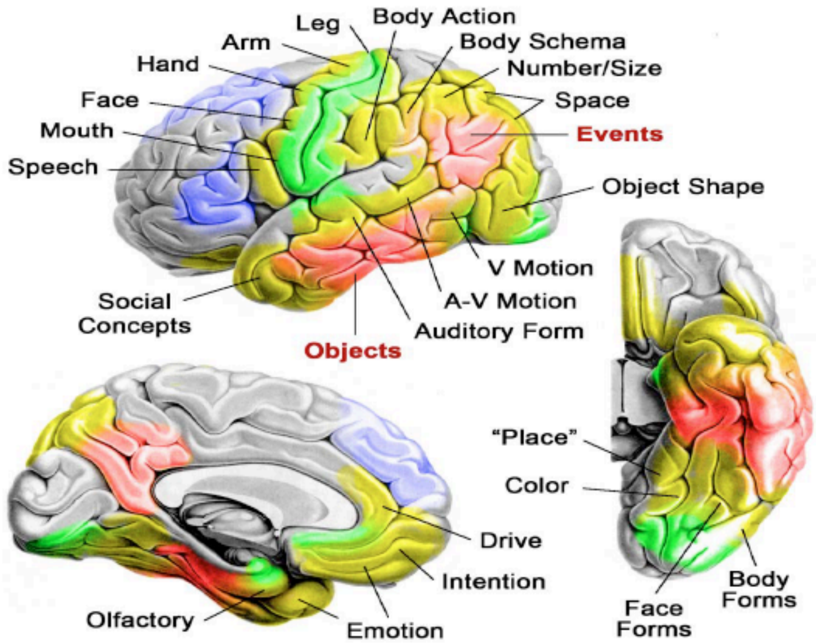



Figure 2.3: Neural architecture for semantic processing. Modality-specific sensory, action, and emotion systems (green regions) provide experiential input to high-level temporal and inferior parietal convergence zones (yellow regions). Multimodal generalization and language promote gradual development of abstract representations of entity/object and event knowledge in the high-level temporal and inferior parietal convergence zones (red regions). Dorsomedial and inferior prefrontal cortices (blue regions) control the goal-directed activation and selection of the information stored in temporoparietal cortices. Neural architecture of conceptual representations assumed to be hierarchical and convergent over multiple levels (Reproduced with permission Binder et al., 2011).

the right side of the figure, but can also be observed in the auditory object recognition pathway running from superior to lateral temporal lobe, the pathway for action concepts running from sensorimotor cortex to inferior parietal lobe, and the pathway for emotion representation running from subgenual cingulate and amygdala to orbital frontal cortex and temporal pole. Blue indicates prefrontal regions linked specifically with semantic retrieval and selection processes.

Binder et al. (2009) asserted that the semantic neural systems are widespread and take a large part of the cortex in the human brain. The areas implicated in these processes can be grouped into three broad categories: posterior heteromodal association cortex (AG, MTG, and fusiform gyrus), specific subregions of heteromodal prefrontal cortex (dorsal, ventromedial, and inferior prefrontal cortex), and medial paralimbic regions with strong connections to the hippocampal formation (parahippocampus and posterior cingulate gyrus). These neural systems support conceptual processing such that humans can use language, make plans, solve problems, and be creative (Binder et al., 2009). Appendix A includes a Glossary that lists the functions for each of these brain regions. Next section introduces the CAR theory, the semantic vectors proposed to map concept representations (i.e., set of features) to brain systems as a way to represent the semantic space of the brain.

### 2.5.3 Concept Attribute Representation (CAR) Theory

In CAR theory, neurobiologically defined "experiential attributes" form a set of primitive features for semantic representations. This set of features capture aspects of experience that are central to the acquisition of event and object concepts, both abstract and concrete. The features correspond to the brain systems described in Section 2.5.1. This approach establishes direct correspondence between conceptual content and neural representations, that is conceptual grounding (Section 2.2.2). By defining conceptual content in terms of brain systems results in a closed and relatively small set of basic features and offers a powerful solution to the problem of feature selection.

The features are weighted according to statistical regularities. The semantic content of a given concept is estimated from ratings provided by human participants. For example, concepts referring to things that make sounds (e.g., *explosion*, *thunder*) receive high ratings on a feature representing auditory experience, relative to things that do not make a sound (e.g., *milk*, *flower*).

Each word is modeled as a collection of 66 features that captures the strength of association between each neural attribute and word meaning. Specifically, the degree of activation of each attribute associated with the concept can be modified depending on the linguistic context, or combination of words in which the concept occurs. Thus, people weigh concept features differently to construct a representation specific to the combination of concepts in the sentence.

Figure 2.4 shows the weighted CARs for the concepts *bicycle* and *table*. The weight values represent average human ratings for each feature. Given that both concepts are objects, they get low weighting on animate attributes such as Face, Body, Speech, Human, Communication, and emotions such as Sad, Angry, Disgust and Fear, and high weighting on attributes like Vision, Shape, Touch, and Manipulation. However, they also differ in expected ways, including stronger weightings for *bicycle* on Motion, Biomotion, Fast Motion, Lower Limb and Path, and stronger weightings for *table* on Large, Smell, Head, Scene, Near, and Needs.

In contrast to concrete concepts, abstract concepts refer directly to cognitive events (such as adventure, marriage, future, death), states (such as decide, judge, recall, think), mental "products" of cognition (such as idea, memory, opinion, thought), social cognition (such as justice, liar, promise, trust), and affective states (such as anger, fear, sad, happy, disgust). These concepts are learned in large part by generalization across these cognitive experiences in exactly the same way as concrete concepts are learned through generalization across perceptual and motor experiences.

One key aspect of CAR theory, compared to the strong embodied approaches, is that these cognitive and affective mental experiences count as much as sensory-motor experiences. Experiencing an affective event is like experiencing a sensory-motor event except that the perception is internal rather than external. The 66 features on this model

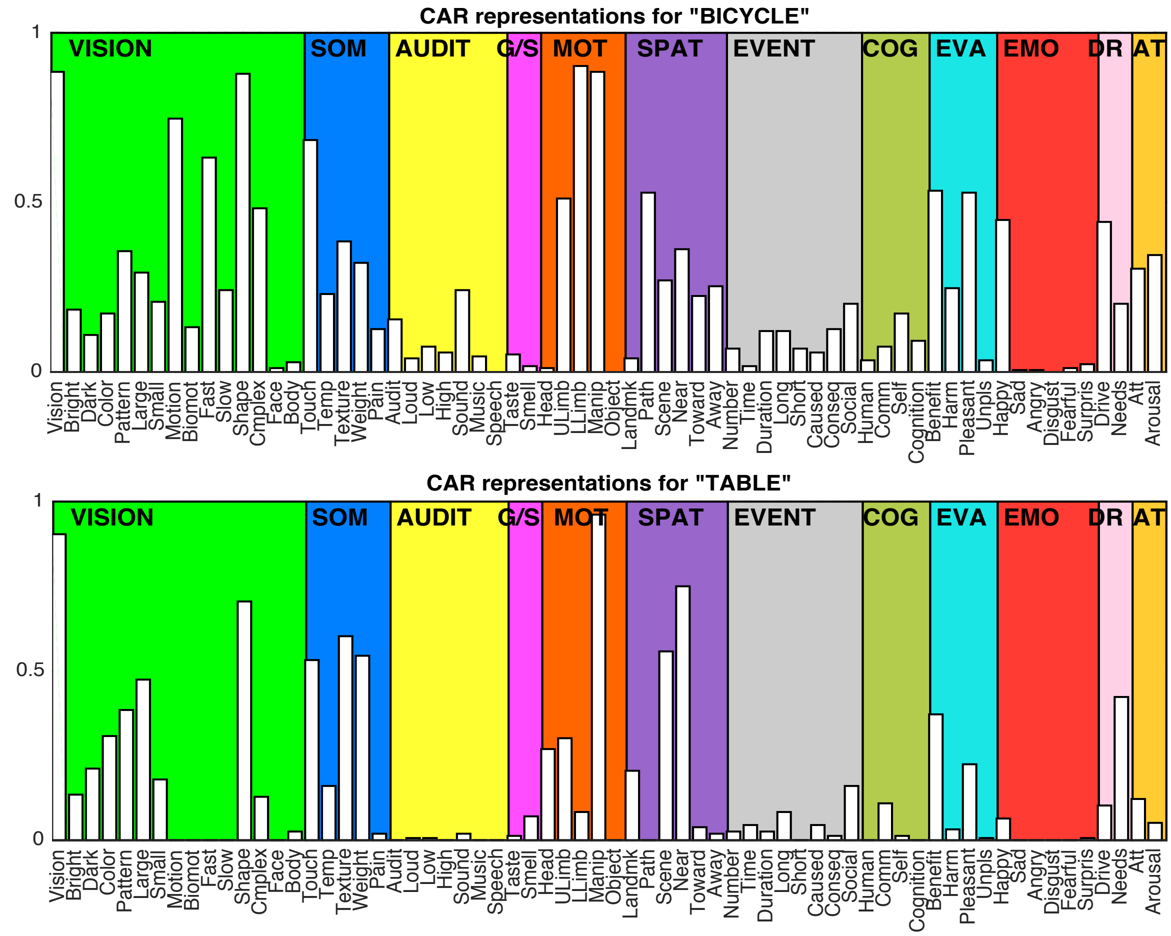



Figure 2.4: Bar plot of the 66 semantic features for the words *bicycle* and *table* (Binder et al., 2009, 2011, 2016). Given that both concepts are objects, they have low weightings on animate attributes such as Face, Body, Speech, Human, and emotions including Sad, and Fear and high weighting on attributes like Vision, Shape, Touch, and Manipulation. However, they also differ in expected ways, including stronger weightings in Motion, Fast, Lower Limb and Path for *bicycle* and stronger weightings in Smell, Scene, Near, and Needs for *table*. Weighted features for the words *bicycle* and *table*.

provide a powerful representation of abstract concepts.

During conceptual combination, CAR theory overlaps neural representations of two or more concepts, mutually enhancing features. This enhancement alters the similarity between them, resulting in some functional groupings or categorizations. This process is known as ad hoc categories (Barsalou, 1983), formed when concepts share the same context-related attribute enhancement. This type of combination corresponds to the attribute combination reviewed in Section 2.3.3 (e.g., *red apple*).

Other types of concept combinations illustrate how individual semantic factors allow words to combine. For example, *plastic bottle* is a bottle made out of plastic, but *baby bottle* is not a bottle made out of babies. There are some general principles that govern such combinations as part of people's world knowledge. In CAR theory, interactions occur when two concepts activate a similar set of brain systems, to the degree their features overlap. In the case of *baby bottle*, there is animacy involved, i.e., Biological motion, Affective, Social cognition, but *bottle* does not activate such systems. They do not have common semantic structures, therefore the meaning of the combination is strongly determined by attribute congruence. As a result, CARs cannot capture the thematic associations between concepts unless additional sources provide it Binder et al., (2009).

Chapter 3 will discuss the processes and materials used to instantiate the CAR theory through interviews of human subjects. For a more detailed account of feature selection and definition see Binder et al., (2009, 2011, 2016a, and 2016b). Next section will review the algorithmic details of the neural network architectures that inspired the model proposed in this dissertation, to quantify how word representations change to account for context.

## 2.6 Artificial Neural Networks

Artificial neural networks are computational models inspired by the brain. They consist of an input and output layers and one or more hidden layers of connected nodes that can receive and transfer activation to each other through weighted connections, like a population of neurons interconnected by synapses to carry out a specific function. Several kinds of neural networks exist and many variations have been proposed. In the following sections the design of a simple feed-forward neural network and a variation called FGREP (Forming Global Representations with Extended Backpropagation) will be reviewed.

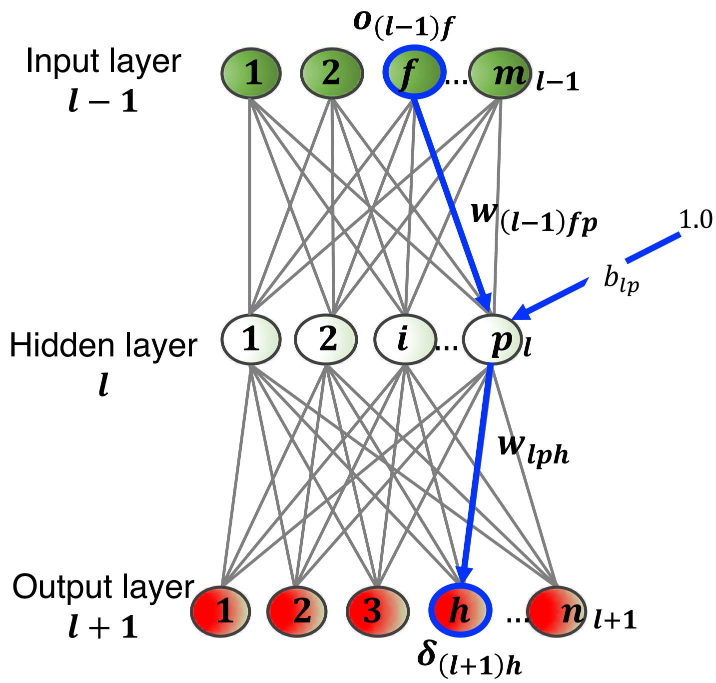



Figure 2.5: Three-layer feed-forward back-propagation neural network. Each node is connected to every other node in the next layer in a feedforward fashion. The input layer $l-1$ contain $m$ units or nodes, the hidden layer $l$ contains $p$ nodes and the output layer $l+1$ contains $n$ nodes. The input layer shows the term $o_{(l-1)f}$ that represents the output of unit $f$ in layer $l-1$. The connections shown between layers include $w_{(l-1)fp}$ which refers to the weight between unit $f$ in layer $(l-1)$ and unit $p$ in layer $l$. Similarly, $w_{lph}$ refers to the weight between unit $p$ in layer $l$ and unit $h$ in layer $(l+1)$. The term $b_{lp}$ is the bias of unit $lp$, and has the effect to adjust the output. The output layer shows the term $\delta_{(l+1)h}$ that stand for the error signal for node $h$ layer $(l+1)$. A basic three-layer neural network that maps input to output patterns.

### 2.6.1 Feedforward Back-propagation Neural Network (BPNN)

Figure 2.5 shows an example of a small neural network consisting of three layers. Each node is connected to every other node in the next layer in a feedforward fashion. Neural networks learn by examples (e.g., repetitive presentations of input and desired output patterns). The goal is to learn a mapping between the input and the desired output (target) patterns. Before learning, the set of input and target examples is prepared. Learning proceeds through modification of the weights that transfer the activation of the input nodes to the output nodes.

The algorithm proceeds as follows: The nodes or units receive one or more inputs. Each input is weighted and the sum is passed through a nonlinear function, known as activation function. Usually a Sigmoid function is used (it can be a Step function or other nonlinear function).

In the forward process, the input layer is loaded with the input patterns. In the subsequent layers, each unit computes its output value as

$$o_{lp} = g\left(y_{lp}\right) = g\left(\textstyle\sum_f w_{(l-1)fp} o_{(l-1)f} + b_{lp}\ \right), \tag{2.1}$$

where $o_{lp}$ is the output of unit $p$ in layer $l$, $y_{lp}$ refers to the weighted sum of its inputs, and $w_{(l-1)fp}$ is the weight between unit $f$ in layer $(l-1)$ and unit $p$ in layer $l$. The term $b_{lp}$ is the *bias* of unit $lp$, and has the effect of adjusting the response threshold of the unit; it allows to shift the activation function to the right or to the left. The logistic (sigmoid) activation function that limits the values between 0 and 1 is used to map the inputs of the unit to its outputs:

$$g\left(y_{lp}\right) = \frac{1}{1+e^{-y_{lp}}} \tag{2.2}$$

Once the entire pattern is produced for the output layer, it is compared to the target pattern. An error signal is formed for each output unit in the output layer $x$ as

$$\delta_{xm} = o_{xm}[1 - o_{xm}][t_m - o_{xm}], \tag{2.3}$$

where $t_m$ is the target activation value for the output unit $m$.

In the backward propagation process, the error signal for unit $p$ in a previous layer $l$ is formed by

$$\delta_{lp} = o_{lp}\left[1 - o_{lp}\right] \sum_h \delta_{(l+1)h} w_{lph}. \tag{2.4}$$

The learning process involves changing each connection weight in proportion to the error signal and the activation going through the connections. It thus implements gradient descent in the error. Often, a *momentum* term is included in order to reduce oscillations:

$$w_{lph}(t) = \eta \delta_{(l+1)h} o_{lp} + \alpha w_{lph}(t-1), \tag{2.5}$$

where $\eta$ is the learning rate.

This process is known as Supervised learning because each example is guided by the input and the target patterns. An epoch measures the number of times the full set of input and target patterns are presented to the network. For each epoch, error signals are usually collected to account for the total error over all patterns. Such global error as well as the epoch count, can be used as a stopping criteria, i.e., if the sum of all differences between the target pattern and the output, averaged across all patterns, is below some threshold value, or the number of epochs is above some threshold value.

BPNN is a very powerful statistical learning mechanism. It can be used for predictive tasks, adaptive control and data analysis applications (e.g., forecasting, airplane sensors, financial analysis, news classification). For a successful learning, these networks highly depend on the I/O encoding representations. Next section reviews the FGREP approach where global representations emerge automatically while at the same time the network learns the assigned task. This mechanism plays a central role in understanding the effect of context on word meanings in this dissertation.

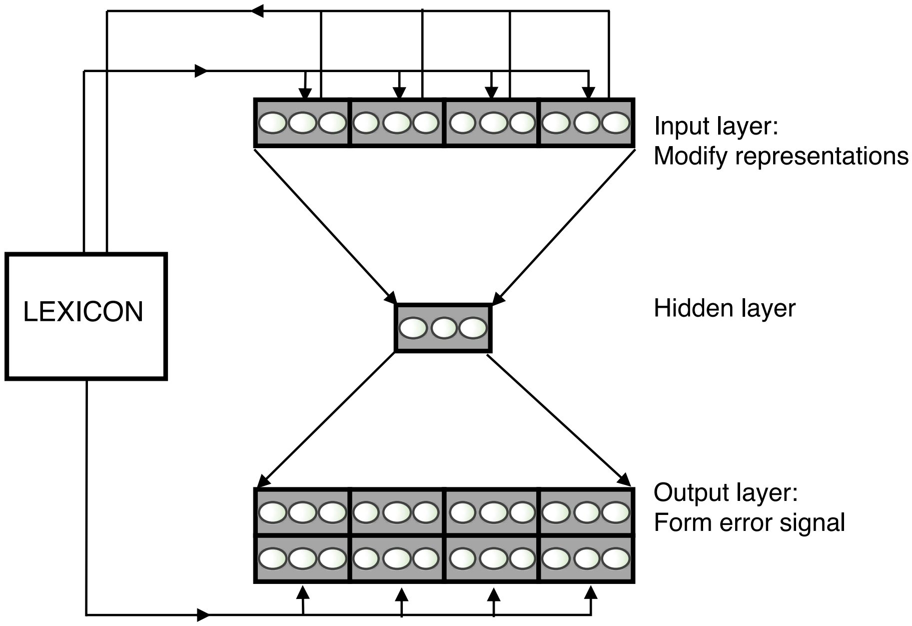



Figure 2.6: A simple FGREP architecture consists of a three-layer BPNN with an external lexicon that stores the I/O representations (Miikkulainen et al., 1988). The input and output patterns are fed into the network from the lexicon, and it learns the task by adapting the connection weights according to the standard backpropagation algorithm. At the end of each backpropagation cycle, the current input pattern is modified by extending backpropagation to the input layer. A neural network mechanism that forms global I/O representations with extended backpropagation.

### 2.6.2 The FGREP Mechanism

The similarities between artificial and biological neural networks have motivated researchers to use the artificial neural networks to explain several cognitive tasks. For example, Miikkulainen et al. (1988) designed a neural network with an additional mechanism called FGREP to develop meaningful distributed representations of words (Figure 2.6).

An FGREP network is similar to the three-layer BPNN. It follows the same dynamics and learns the task by adapting the connection weights as outlined above.

However, in order to develop meaningful word representations, the error signal was extended to the input layer.

FGREP was designed to meet two goals: (1) Learn the processing task by adapting the connection weights using standard backpropagation, and (2) Develop meaningful distributed representations in the process. In fact, both learning processes (task and representations) are done at the same time.

The representations are stored in an external lexicon. Initially, the lexical representations have random values. In processing a sentence, they are collected from the lexicon and used as input and target vectors in the BPNN. Standard backpropagation is used to update the weights in the network, but it is extended into the input vector: the component values in this vector are updated and stored back into the lexicon. Since the same words can appear both in the input and output, the learning is chasing a moving target.

The principle to change the representations is to consider the input units as ordinary nodes (i.e., hidden and output nodes). Since the input units have the characteristic that the activation function is the identity function, i.e., input signal is passed to output without any change, backpropagation can be extended one step further to update the input patterns, by treating them as an extra layer. Therefore, the error signal of an input unit is computed as

$$\delta_{1f} = \sum_h \delta_{2p} w_{1fp}, \tag{2.6}$$

where $\delta_{1f}$ represents the error signal for unit $f$ in the input layer 1, and $w_{1fp}$ is the weight between unit $f$ from the input layer and unit $p$ in the first hidden layer. Then, the representations are changed as

$$r_{cf} = \eta \delta_{1f}, \tag{2.7}$$

where $r_{cf}$ is the $f$th element of item $c$'s representation, $\delta_{1f}$ is the error signal of the associated input layer unit, and $\eta$ is the learning rate. Whereas weighted values are not limited, the I/O representations have maximum and minimum activation values of the units. The new value for the representation of item $c$ element $fth$ is obtained as

$$r_{cf}(t+1) = \max\left(o_L, \min\left(o_U, r_{cf}(t) + r_{cf}\right)\right), \tag{2.8}$$

where $o_L$ is the lower limit and $o_U$ is the upper limit for the unit activation, i.e., usually 0 and 1 with the standard sigmoid activation function.

FGREP is based on the philosophy that concepts are defined by the way they are used (Harris, 1970). Thus, the meanings of the concepts are encoded in their final representations and defined by all the contexts where the concept appeared. The representation as well as the meaning evolves continuously as more experience is gained. As a consequence, all aspects of a concept are distributed over the whole set of units making the system robust.

The FGREP mechanism has been used in several natural language processing tasks, such as assigning roles to sentence constituents (Miikkulainen and Dyer, 1991). It was also used as part of a script-based story understanding model (DISCERN, short for DIstributed Script processing and Episodic memory Network, Miikkulainen 1993), for disambiguating prepositional phrase attachment (Takahashi et al., 2001), and in developing language representations automatically, so that there is no need for humans to predefine feature vectors into the data set for an NLP system; the features automatically emerge during the FGREP process (Kleiweg & Nerbonne, 1998). In this dissertation, FGREP is used in a different role: to identify how the CAR weights should be changed to take context into account.

## 2.7 Discussion and Future Work

Concept and word meaning representations are the main focus of this dissertation. Several issues were reviewed in this Chapter: How are concepts represented in the brain? How is word meaning represented? How are concepts grounded in sensorimotor experiences? How do word meanings change in context? Where is the semantic system? What tools and approaches exist to model the brain's semantic organization?

Currently, the answers to all these questions are open to debate, i.e.,

- Whether concepts are strongly embodied (grounded in perception and motor systems), completely disembodied, or somewhere in the middle, i.e., Embodied Abstraction;
- Whether there is an integration of information in the brain from different modalities via hubs and convergence zones;
- Whether the organization of semantic knowledge is based on categories, prototypes, or feature representations; and
- Whether semantic features are better defined in terms of explainable features such as the neural-based features defined by Binder et al. (2009) or text-based models with distributed (unexplainable) representations.

Ongoing research is far from finding a unique representation model that can capture all kinds of semantic knowledge. This dissertation focuses on the last point, the neural-based semantic features. The research follows the theory proposed by Binder et al. that includes Embodied Abstraction, hierarchical multimodal integration, and uses explainable semantic features for prediction and interpretation of imaging data.

In the following chapters, the experiments will test the hypothesis that different properties of word meaning are activated in different contexts, and it is possible to capture those changes in the corresponding fMRI images using a neural network model.

To avoid confusion, throughout the rest of this dissertation the terms attributes, properties, and features will be used interchangeably to refer to properties of concepts such as color or size. Similarly, weights, activations, and values all refer to the CAR weighted attributes.

### CONCLUSION

As background for the overall focus of this dissertation, this chapter presented an overview of how semantic knowledge is represented in the brain, and which tools and approaches help in understanding such mechanism better. Semantic knowledge was defined in terms of its content (i.e., conceptual knowledge, word meaning, conceptual combinations, context effect), providing the psychological and neurological perspectives. Also, the neural correlates of the semantic system were outlined, based on experimental findings in the literature.

Four computational models of semantic representation were reviewed, addressing different ways of constructing semantic representations (e.g., text-based, & vision-based modalities). Additionally, a true multimodal representation that accounts for the full array of human senses was reviewed: The CAR theory of Binder et al. (2009, 2011), suggests that concept representations can be decomposed into a set features that are directly mapped to individual brain systems.

fMRI technology was reviewed as a non-invasive method to study semantic knowledge in vivo. Furthermore, simple backpropagation and FGREP neural networks were reviewed as part of the framework described in Chapter 3 for interpreting imaging data. In fact, the FGREP mechanism plays a central role quantifying the effect of context on word meanings.

In sum, this chapter not only presented a review of the literature on semantic representations, but also introduced the theoretical and experimental backgrounds adopted for the work presented in the subsequent chapters. Particularly, Binder's brain-based semantic model is the theoretical foundation, and the FGREP neural network the foundation for the computational experiments. Combined, they form the basis for characterizing how word meaning changes in the context of a sentence.

# Chapter 3

# A Computational Model to Account for Context Effects in the Brain

Building on the theory of grounded word representation, this research aims to understand how word meanings change depending on context. This chapter describes the computational model that characterizes such context-dependent meaning representations and the data needed to construct it. The model, CEREBRA, or Context-dependent mEaning REpresentation in the BRAin, is founded in the CAR theory (Section 2.5.3) and implemented using neural networks with FGREP (Section 2.6.2). The idea is to train a neural network to predict approximately what the sentence fMRI should be, based on the CARs, and then use FGREP to adjust the CARs so that the prediction becomes correct. The modified CARs then indicate how the meaning changed in context. In this Chapter, the CEREBRA model is first introduced, including a neural network that learns to predict fMRI sentences and the FGREP method that identifies the semantic changes on CAR words to account for context. Then, the sentence collection that forms the basis for the fMRI experiments, the fMRI data itself, and the CAR ratings obtained from human subjects, follows. At the end, a comparison between CEREBRA and the original FGREP mechanism is described.

## 3.1 THE CEREBRA MODEL

CEREBRA uses sentence fMRI patterns and the CAR semantic feature model of concept representations to characterize how word meanings are modulated within the context of a sentence. With CARs of words as input, the neural network is trained to generate first approximations of fMRI patterns of subjects reading sentences. Then, the FGREP mechanism is used to determine how the CARs would have to change to predict the fMRI patterns more accurately. These changes represent the effect of context; it is thus possible to track the brain dynamic meanings of words by tracking how the CARs feature-weightings change across contexts.

## Terminology

**CARWord:** The neural network input. CARWords are formed based on ratings by human subjects (Section 3.3). They are the original brain-based semantic representations of words, i.e., word without context. Each CARWord is a vector of 66 attributes.

**CARWordRevised:** The input of the neural network after FGREP. CARWordsRevised are formed by FGREP modifying the original CARWords. They are the context-dependent meaning representations of words for each sentence where they occurred**.** Each CARWordRevised is a vector of 66 attributes.

**$\varepsilon$ :** The error signal. The SynthSent is subtracted voxelwise from the fMRISent to produce an error signal. Each error is a vector of 396 changes.

**fMRISent:** The neural network target. They are the original brain data collected from human subjects using neuroimaging. Each fMRISent is a vector of 396 voxels.

**SyntSent:** The predicted fMRI sentence after training. The SynthWords in the sentence are averaged to form this prediction. Each SynthSent is a vector of 396 values.

**SyntSentRevised:** The modified SyntSent after applying the error signal changes. Each of these SynthSentRevised is a vector of 396 values.

**SyntWord:** The neural network target. They are derived by averaging the fMRISent. They are synthetic because individual fMRI data for words do not exist, thus they are obtained by averaging each fMRISent where the word occurred. Each SynthWord is a vector of 396 voxels.

**SyntWordRevised:** The target for the neural network after FGREP. They are derived from the SynthSentRevised using the error signal changes.

**W1..W3:** labels for each CARWord in a sentence.

**W'1..W'3:** labels for each SynthWord in a sentence.

### 3.1.1 System Design

The overall design of CEREBRA is shown in Figure 3.1. The neural network model serves two main tasks: Prediction and Interpretation. During the Prediction task, the model forms a predicted fMRI for each sentence, without the context effects. Each sentence is thus compared against the observed fMRI sentence to calculate an error signal. This error signal is used repeatedly by the Interpretation task. During the Interpretation task, the model is used to determine how the CARs should adjust to eliminate the remaining error. The error is used to change the inputs (CARs) using Extended-backpropagation (which is the FGREP method described in Section 2.6.2). The process iterates until the error goes to zero.

The following sections present a detailed description of the architecture at each stage of the system implementation. The specific terms to the CEREBRA model are denoted by abbreviations throughout the chapter (e.g., CARWord, fMRISent, SynthWord). For reference, they are described in the Terminology box.

### 3.1.2 Mapping CARs to Synthetic Words

The CEREBRA model is first trained to map the CARWord representations in each sentence to SynthWords (The "forward" side of Figure 3.1). It uses a standard three-layer backpropagation neural network (BPNN). Gradient descent is performed for each word, changing the connection weights of the network to learn this task (Rumelhart et al., 1986).

Algorithm 3.1 describes the model implementation and training in detail. A three-layer feed-forward BPNN with 66 input units, 66 hidden units and 396 output units was implemented to map CARs of words to fMRI of words. The training parameters included a learning rate of $\eta$=0.3, decreasing at a rate of 0.001 to 0.000001, to control how quickly

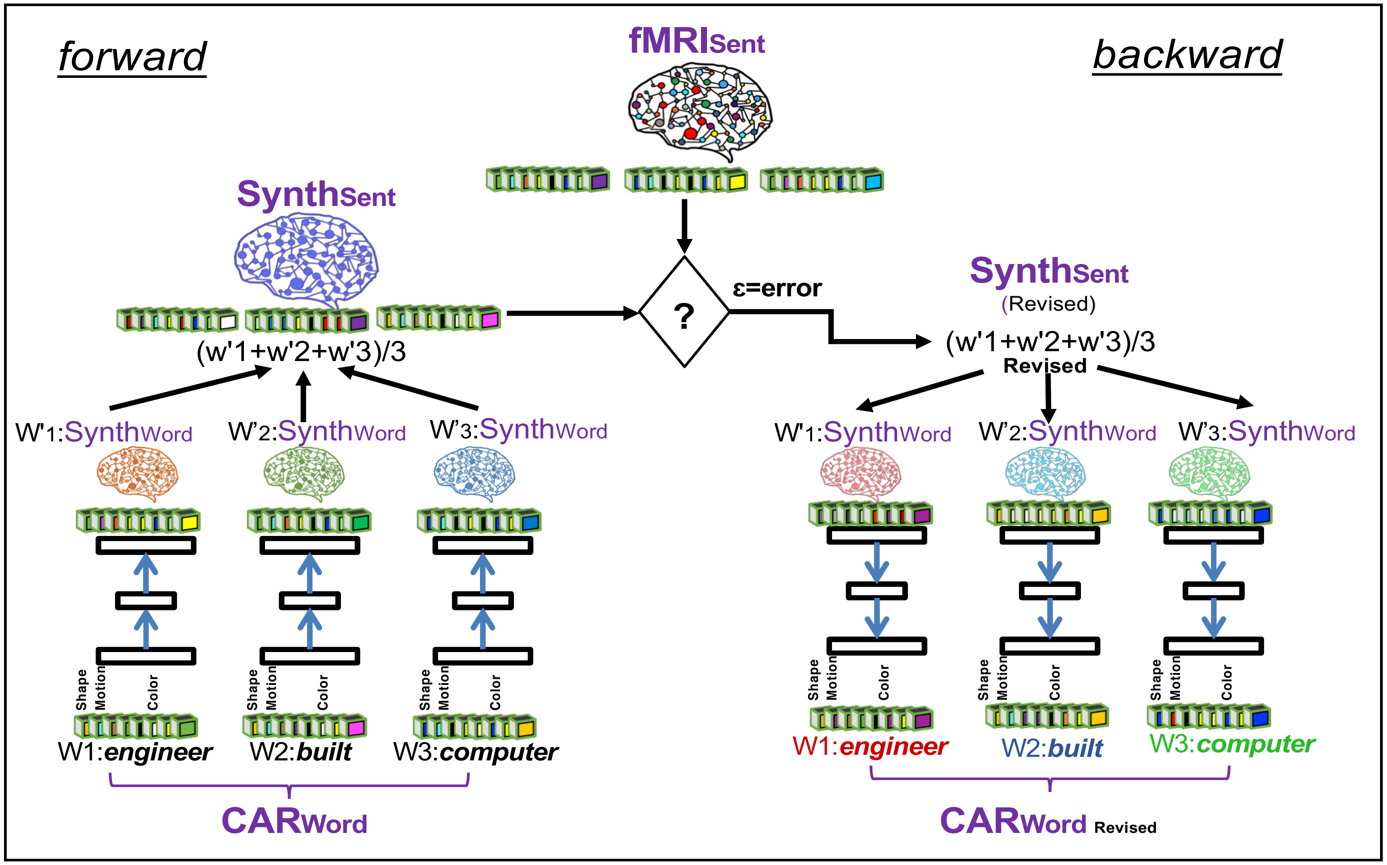



Figure 3.1: The CEREBRA model to account for context effects. (1) Propagate CARWords to SynthWords. (2) Construct SynthSent by averaging the SynthWords into a prediction of the sentence. (3) Compare SynthSent with the observed fMRI. (4) Backpropagate the error with FGREP for each sentence, freezing network weights and changing only CARWords. (5) Repeat until error reaches zero or CAR components reach their upper or lower limits. The modified CARs represent the word meanings in context. Thus, CEREBRA captures context effects by mapping brain-based semantic representations to fMRI sentence images.

the weights will change and avoid converging into a suboptimal solution; and a momentum rate of $\alpha$=0.3, to accelerate the training process by helping guide the weights toward the right direction (reducing oscillations). The neural network weighted connections and the bias were randomly initialized between -0.5 and 0.5. The BPNN was trained for each of the eleven fMRI subjects for a total of 20 repetitions each, using different random seeds.

The first part of the algorithm (Step 1 to 6) consists of training the BPNN to map CARWord representations (i.e., input) to SynthWord representations (i.e., target). After training is completed for each subject, it yields 20 different networks, plus 20 sets of 786 predicted SynthWord representations, that is, one word representation for each sentence where the word appears.

### 3.1.3 Predicting Sentences and Propagation Error Back to Words

The next segment of Algorithm 3.1 (Steps 7 to 14) describes the Prediction and Interpretation tasks mentioned at the beginning of this section. For the Prediction task, the sentences are assembled using the predicted SynthWords by averaging all the words that occur in the sentence (Step 9), yielding the prediction sentence called SynthSent. For the Interpretation task, in addition to the construction of the predicted sentence, further steps are required (Steps 10 to 14). First, the prediction error is calculated by subtracting the newly constructed predicted SynthSent from the original fMRISent. Then, the error is backpropagated to the inputs CARWords for each sentence (The "backward" side of Figure 3.1). The weights of the network no longer change. Instead, the error is used to adjust the CARWords in order for the prediction to become accurate.

Algorithm 3.1 Neural network to map CAR words to sentence fMRI & back to CARs

Using a three-layer feed forward back propagation neural network (BPNN) with 66 input units, 66 hidden units and 396 output units do as follow:

1: For Subject=1 to 11
2: For repetitions=1 to 20
3: Generate different random seeds to initialize the weights of the BPNN
4: Repeat for 1000 epochs (or until the average of the sum of all errors between the output and the target patterns is less than epsilon: $tss < \epsilon$, where $\varepsilon = 0.001$)
5: For ctxt_word=1 to 786 (for each word that occurs in every sentence)
6: Train the BPNN mapping CARWord (66 attributes) to SynthWord (396 voxels)

7: Repeat until the prediction error is very small (near zero) or no additional change is possible (CARWord already met their maximum or minimum values)
8: For ctxt_word =1 to 786 (for each word that occurs in every sentence)
9: Assemble the SynthWords into the 237 predicted SynthSent by averaging the appropriate words:
$SynthSent(sent, 1{:}396) = [W(a, 1{:}396) + W(b, 1{:}396) + \ldots + W(y, 1{:}396)]/n$
◊ $sent$ is number of sentences (1 to 237)
◊ $a, b, \ldots, y$ represent the correct word index for each SynthSent content words
◊ $n$ is the number of words for each SynthSent
10: Obtain the prediction error by subtracting the predicted SynthSent from the observed fMRISent for each voxel:
$PredictErr(sent, 1{:}396) = fMRI(sent, 1{:}396) - SynthSent(sent, 1{:}396)$
◊ $sent$ is number of sentences (1 to 237)
11: Use the error to change the values of the original SynthWord by replacing the target values with the adjusted SynthWord(ctxt_word,1:396).
12: Propagate the CARWord (ctxt_word,1:66) using the same trained network (from point 6)
13: With the adjusted SynthWord values calculate the new target minus output error using the BPNN
14: Backpropagate the error all the way to CARWord(ctxt_word,1:66) changing the inputs without changing the weights (with extended backpropagation, the FGREP method)

This process is performed until the prediction error is very small (near zero) or cannot be modified (CARWord already met their limits, between 0 and 1), which is possible since FGREP is run separately for each sentence.

These steps (7 to 14) are repeated 20 times for each subject. At the end, the average of the 20 representations is used to represent each of the 786 context-based words (CARWord Revised), for every single fMRI participant.

Eventually, the Revised CARWord represents the word meaning for the current sentence such that, when combined with other Revised CARWords in the sentence, the estimate of sentence fMRI becomes correct.

### 3.2 Sentence Collection

The sentence set was prepared for the fMRI study as part of the Knowledge Representation in Neural Systems (KRNS) project (Glasgow et al. 2016, www.iarpa.gov/index.php/researchprograms/krns), sponsored by the Intelligence Advanced Research Projects Activity (IARPA) under the White House BRAIN Initiative Program (BRAIN Initiative, 2013). The words used in the sentences stand for imaginable and concrete words such as:

1. Objects: Things that exist physically, can be animate or inanimate, natural or man-made. They are often nouns and can be count nouns or mass nouns. Examples: *ball*, *bicycle*, *dog*, and *water*.
2. Actions: Things that are done or experienced by living things. They are often verbs that describe moving, perceiving, feeling, and creating. Examples: *walked*, *ate*, *built*, and *drank*.

3. Settings: Locations where or when things happen. They are often nouns that describe indoor or outdoor locations, seasons, and time of day. Examples: *church*, *forest*, *spring,* and *morning*.
4. Roles: What people do or who they are. They are often nouns that describe vocations, professions, and kinship. Examples: *banker*, *doctor*, *minister*, and *family*.
5. State and emotions: Descriptive and characterizing words. They are often adjectives that portrays or typifies a noun. Examples: *hot*, *little, old*, *red*, and *sad*.
6. Events: Things that take place in space and time, such as human-organized encounters or natural incidents. They are often nouns that describe activities or situations. Examples: party, *flood,* and *hurricane*.

There were a total of 242 such words (141 nouns, 39 adjectives and 62 verbs) in the sentences. A total of 240 sentences were composed from two to five of those words. Sentences are in active voice and consist of a noun phrase followed by a verb phrase in past tense, with no relative clauses. Two hundred of these sentences contain an action verb and the remaining 40 contain the verb *was*. Examples of the sentences include *The family survived the powerful hurricane*, *The scientist spoke to the student*, *The diplomat negotiated at the embassy, The reporter interviewed the politician during the debate*, *The small church was near the school*. The complete collection is included in Appendix B.

### 3.3 CAR RATINGS

In a separate study Binder et al. (2009, 2016a, 2016b) collected CAR ratings for the original set of 242 words through Amazon Mechanical Turk. In a scale of 0-6, the participants were asked to assign the degree to which a given word is associated to a specific type of neural component of experience (e.g., “To what degree do you think of a *chair* as having a fixed location, as on a map?”). Figure 3.2 shows an example used during attribute rating collection. Participants responded by selecting a number where 0 indicates "not at all" and 6 indicates "very much". A "Not Applicable" option was also available to cover cases in which the participant felt the question has no logical relation to the word; these responses were coded as 0. Approximately 30 ratings were collected for each word in this manner. After averaging all ratings and removing outliers, the final attributes were transformed to unit length yielding a 66-dimensional feature vector such as the one shown in Figure 3.3 for the word *chair*. The concept *chair* is an object, therefore it was rated with low activations on animate attributes such as Face, Speech, Head, and emotions including Happy, Sad, and Angry, and high activations on attributes like Vision, Small (compared to a *table*), Shape, Touch, Lower-limb, Upper-limb, and Manipulation.

The final collection of CAR words consists of 242 word vectors with a 66-dimensional attribute ratings that constitute the generic representation of the words, and is the first essential input to the CEREBRA model: These are the CARWords introduced in Section 3.1.

Note that this semantic feature approach builds its vector representations by directly mapping the conceptual content of a word (expressed in the questions) to the corresponding neural processes and systems for which the CAR dimensions stand (Binder et al., 2009, 2016a, 2016b). This approach thus contrasts with systems where the features are extracted from text corpora and word co-occurrence with no direct association to perceptual grounding (Baroni et al., 2010; Burgess, 1998; Harris, 1970; Landauer & Dumais, 1997).

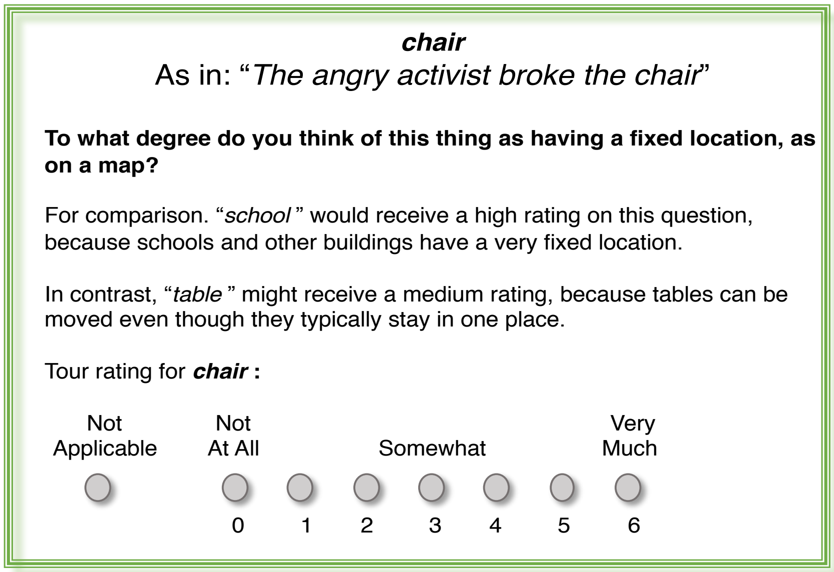



Figure 3.2: Example query for word *chair* addressing the attribute Landmark. Landmark in one of the dimensions of the Spatial brain zone that refers to large entities that have fixed location (e.g., mountains and buildings). They are critical for navigation. This is part of the questionnaire used to assemble each of the 66 dimensions for the 242 CAR word representations (Binder et al., 2009, 2011, 2016).

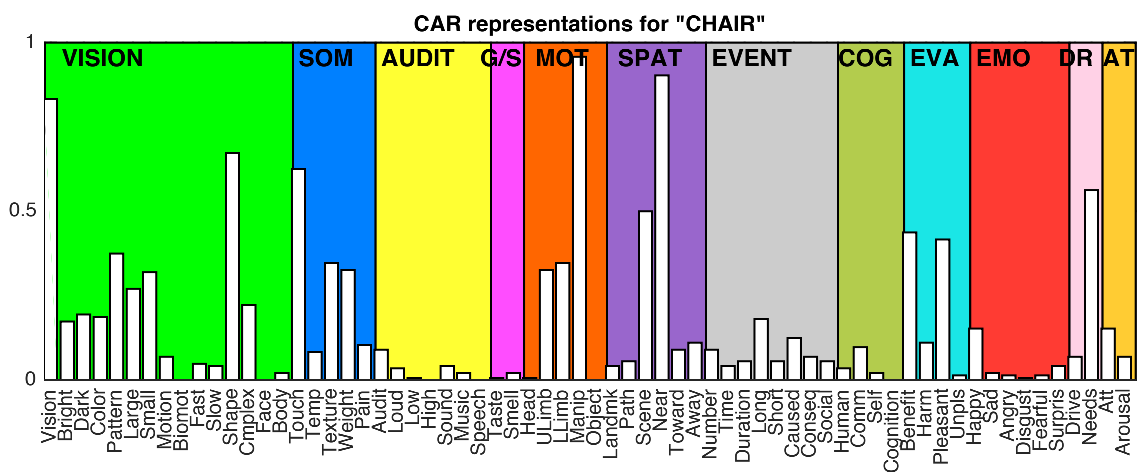



Figure 3.3: Bar plot of the 66 semantic features for the concept *chair* (Binder et al., 2009, 2011, 2016). The values represent average human ratings for each feature. Given that *chair* is an object, it gets low activations on animate attributes such as Face, Speech, Head, and emotions including Happy, Sad, and Angry, and high activations on attributes like Vision, Small (compared to a *table*), Shape, Touch, Lower-limb, Upper-limb, and Manipulation. Original CARs for the word *chair*.

## 3.4 NEURAL DATA COLLECTION

If indeed word meaning changes depending on context, it should be possible to see such changes by directly observing brain activity during word and sentence comprehension. As reviewed in Section 2.5, Binder et al. (2009, 2011, 2016b) identified a large-scale network with individual brain systems involved in the representation of specific attributes of conceptual knowledge (e.g., knowledge of actions, concrete and abstract concepts). Accordingly, Binder and his team collected brain imaging data from several subjects reading the sentences described in Section 3.2, by recording visual, sensory, motor, affective, and other brain systems contained in such network. The following sections describe the materials and methods used.

### 3.4.1 Neural fMRI Representation of Sentences

The study population consists of eleven healthy, right-handed, monolingual English-speaking adults, aged 20-60, with no history of neurological or psychiatric disorders. Each participant took part in this experiment producing 12 repetitions each.

To obtain the neural correlates of the 240 sentences, subjects viewed each sentence on a computer screen while in the fMRI scanner through a mirror attached to the head coil. The sentences were presented word-by-word using a rapid serial visual presentation paradigm. More specifically, images of nouns, verbs, adjectives, and prepositions were presented at the same spatial location for 400ms each, followed by a 200ms inter-stimulus interval. The mean sentence duration was 2.8 seconds. Participants were instructed to read the sentences and think about their overall meaning.

The fMRI patterns were acquired with a whole-body Three-Tesla GE 750 scanner at the Center for Imaging Research of the Medical College of Wisconsin (Anderson et al., 2016). The fMRI data were preprocessed using standard methods, including slice timing and head motion correction (AFNI software, Cox 1996). The most stable, active and discriminative voxels were then selected, and Principal Component Analysis and zero mean normalization were performed on them.

These transformed brain activation patterns were converted into a single-sentence fMRI representation per participant by taking the voxel-wise mean of all repetitions (Anderson et al., 2016; Binder et al., 2016, 2016b). The most significant 396 voxels per sentence were then chosen. The size selection mimics six case-role slots of content words consisting of 66 attributes each. The voxels were further scaled to [0.2..0.8]. This collection of 11 subject images for the 240 sentences constitutes the second essential input to the CEREBRA model: These images are the fMRISent representations introduced in Section 3.1.

### 3.4.2 Synthetic fMRI Representations of Words

The Mapping CARs task in CEREBRA (described in Section 3.1.2) requires fMRI images for words in isolation. Unfortunately, the neural data set does not include such images. A technique developed by Anderson et al. (2016) was adopted to approximate them. The voxel values for a word were obtained by averaging all fMRI images for the sentences where the word occurs. These vectors, called SynthWords, encode a combination of examples of that word along with other words that appear in the same fMRI sentences. Thus, the SynthWord representation for *mouse* (Figure 3.4) contains aspects of running, forest, man, seeing, and dead, from the sentences 56:*The mouse ran into the forest* and 60:*The man saw the dead mouse*.

Although the collection includes a small set of sentences, the CEREBRA process of mapping semantic CAR words to the synthetic words and further to fMRI sentences

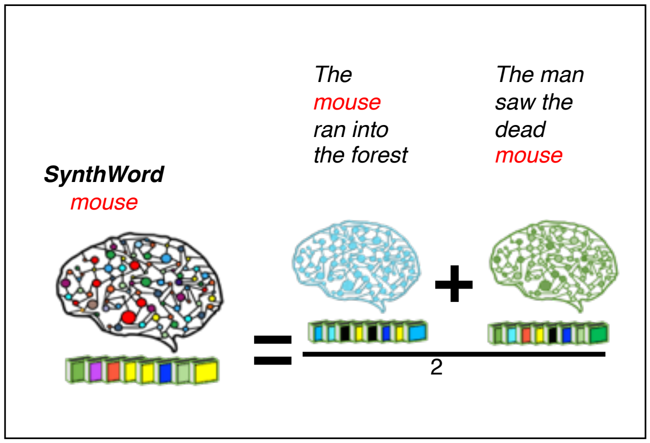



Figure 3.4: Example of SynthWord representation for the word *mouse* using the average of the two fMRI sentences where the word occurs. SynthWords encode a combination of examples of that word along with other words that appear in the same sentences, that is, the word *mouse* contains aspects of *ran*, *forest*, *man*, *saw*, and *dead*, by averaging the two fMRI sentence representations. SynthWord is derived by averaging the fMRI sentences where the word occurs.

helps refine the synthetic representations by removing noisy information. This process of combining contextual information is similar to many semantic models in computational linguistics (Baroni et al., 2010; Burgess, 1998; Landauer et al., 1997; Mitchell & Lapata, 2010). Additionally, in other studies, this approach has been used successfully to predict brain activation (Anderson et al., 2016; Binder et al., 2016a, 2016b; Just et al., 2017).

Due to the limited number of sentences, some SynthWords became identical and were excluded from the dataset. Therefore, the final collection includes 237 sentences and 236 words (138 nouns, 38 adjectives and 60 verbs). The list of words and sentences is included in Appendix B. This SynthWord collection represents the third essential input to the CEREBRA model: These are the SyntWord representations introduced in Section 3.1.

### 3.5 FGREP Tailored to Find Featured-based Semantic Representations

The original FGREP mechanism reviewed in Chapter 2 (Miikkulainen, 1988) was designed to meet two goals: (1) learn the processing task by adapting the connection weights using standard backpropagation and (2) develop meaningful distributed representations in the process. Most importantly, both learning processes (task and representations) are done simultaneously (Miikkulainen, 1988).

In CEREBRA, FGREP is applied in a different manner, and it carries different goals. CEREBRA uses (1) a neural network trained in the task of mapping words from CARs to fMRI word patterns (Section 3.1.2), and (2) based on an error signal at sentence level, FGREP modifies the baseline meaning of the words (Section 3.1.3). What is important here is that the task and the representations are learned separately.

On that account, the main difference is that CEREBRA develops representations through the error signal between sentences (i.e., fMRISent and SynthSent) calculated outside the neural network. In contrast, the original FGREP develops representations through the error signal from the neural network assigned task (i.e., case roles).

Ultimately, the original FGREP and CEREBRA were used for different purposes, and the way CEREBRA employs the FGREP mechanism is to characterize context-dependent meaning representations in the brain (i.e., from words to sentences and back). Additional dissimilarities are listed next:

1. FGREP starts with random input representations. The neural network develops its own representations. CEREBRA begins with CAR ratings that represent the generic meaning of words, they were collected from human subjects (described in Section 3.3).
2. In FGREP there are no identifiable microfeatures or categorizations in the representations. The meaning of a word is distributed over the entire set of

units. In CEREBRA, word representations are based on the CAR semantic model where each of the 66 features represents the association between a specific neural attribute and the word meaning (list of features described in Section 2.5.1).

3. In FGREP, the representation is determined by all the contexts where the word appeared, as a result it is a representation of all those contexts. In CEREBRA, each word representation is modified by a particular context and results in different representations for different contexts. It is used to find how each generic word in CARs should be adjusted for each context.

### 3.6 DISCUSSION AND FUTURE WORK

As reviewed in Section 2.5.3, the CAR theory proposes that conceptual knowledge can be decomposed into a set of attributes, and such attributes are mapped to brain systems that play a role during learning and recall. This theory hypothesizes that context (i.e., other concepts in a sentence), modify the baseline meaning of a concept. This hypothesis is central to this research. CEREBRA will test it by characterizing how CARs can be modified to account for the changes in the neural activation pattern of the concept.

CEREBRA decomposes sentence fMRI into words and words into embodied brain-based semantic features (CARs). Characterizing how words could change under the context of a sentence, this research will demonstrate that (1) context-dependent meaning representations are embedded in the sentence fMRI and (2) CARs semantic theory can be used as a foundation for modeling the neural representation of word meaning.

The next four chapters present three computational experiments as well a behavioral analysis to demonstrate CEREBRA's capability to characterize the effect of sentence context on word meanings. Chapter 4 will analyze interesting context effects for different shades of meaning (e.g., *dangerous flood* vs. *dangerous criminal*). Chapter 5 will focus on the different types of conceptual combinations and their effect on word meanings by analyzing statistically significant changes for individual sentence cases across multiple fMRI subjects (e.g., *boat crossed* vs. *car crossed*; *bird flew* vs. *plane flew*). Chapter 6 will demonstrate that the outcome is robust and general across the entire corpus of sentences and case roles. Chapter 7 will corroborate that these effects (changes) are actually meaningful to humans.

### CONCLUSION

This chapter introduced a computational model called CEREBRA that uses fMRI patterns of sentences, CAR words, and a neural network with the FGREP mechanism to characterize context-dependent meaning (i.e., implicit in the sentence fMRI). The essential inputs to CEREBRA were reviewed: the Glasgow word and sentence collection, the human-judgement based CAR ratings, and the neural data collection of the sentence fMRI. The chapter also highlighted how the FGREP mechanism was used to discover sentence-level meaning from word-level features.

# Chapter 4

# Characterizing the Context Effect on Word Meanings

Building on CAR theory, this chapter[1] characterizes the effect of grounding, i.e., how word meaning changes within the context of a sentence. This question will be addressed anecdotally by analyzing a few example cases. The goal is to predict sentence fMRI using two computational models to map CARs of words into fMRI data of subjects reading everyday sentences.

## 4.1 Motivation

Multimodal vector representations have been found to outperform text-based vector representations in the task of representing word meanings (Bruni et al., 2014; Silberer & Lapata, 2014, Vinyals et al., 2015). However, most of those multimodal approaches consider the different modalities to the same degree, even though each modality contributes in different ways to word meanings (Bruni et al., 2014; Silberer & Lapata, 2014; Silberer et al., 2017). For example, the concept of *mouse* is learned through perceptual experiences like visual, somatosensory, and auditory situations, compared to the abstract concept *dead*, which is learned from cognitive experiences such as events, states, or thoughts (Binder, 2016a, 2016b).

In contrast, for the brain-based semantic representations (CARs, Chapter 2, Binder et al., 2009), the meaning of a concept corresponds directly with the neural

[1] The content of this chapter was previously presented at the 39th Annual Meeting of the Cognitive Science Society (Aguirre-Celis & Miikulainen, 2017). Aguirre-Celis worked on experimental design, implementation and analysis; while Miikkulainen provided guidance and feedback through discussions.

processes. Concepts are represented as a set of experiential attributes that reflect the sensory, motor, affective, and other brain networks involved in concept learning.

Building on CAR theory, this chapter evaluates experimentally how word meaning changes across different sentences. It aims to explain these changes by testing whether Multiple Linear Regression approach and the CEREBRA nonlinear neural network can discriminate between sentences based on feature weightings. The experiments analyze a few example cases where word attributes are weighted differently in various contexts for verbs, nouns and adjectives. Each model identifies significant changes on CARs for the same word in different sentences. The experiments support the hypothesis that context adapts the meaning of words in the brain.

Next, the semantic model and the collection of contrasting sentences used to investigate the effect of context are presented. The general system framework and data flow as well as the mechanisms for determining the semantic changes, i.e., multiple linear regression and the CEREBRA model, follow. Three experiments are presented and the results analyzed, demonstrating how context affects word meanings.

## 4.2 THE SEMANTIC MODEL AND DATA SETS

The CAR theory was reviewed in Chapter 2. Essentially, it represents the basic components of meaning defined in terms of known neural processes and brain systems. Each word is represented as a collection of a 66-dimensional feature vector that captures the strength of association between each neural attribute and the word meaning. More detailed account of the attribute selection and definition is given by Binder et al. (2009, 2011, 2016a, and 2016b).

To run the experiments, several data sets are required, as described in detail in Chapter 3. As a reference these are: a sentence collection prepared by Glasgow et al.

(2016), the semantic vectors (CAR ratings) for the 236 words obtained via Mechanical Turk, and the fMRI images for the 237 sentences, both collected by the Medical College of Wisconsin (Anderson et al., 2016; Binder et al., 2016). Additionally, fMRI representations for individual words (called SynthWord) were synthesized by averaging the sentence fMRI producing 236 synthetic words.

The Glasgow sentence collection is not fully balanced and systematic, but instead aims to be a natural sample. To investigate the effect of context, finding mutual similarities between words or sentences sounds like a good approach. However, similarity alone is not enough, because anything is similar to anything else to some degree. Contrasting words or sentences is a better mechanism to address such effect. Therefore, several pairs of contrasting sentences were identified in this collection.

A group of 77 such sentences, with different shades of meaning for verbs, nouns and adjectives, as well as different contexts for nouns and adjectives was assembled (Table 4.1). This collection is used to prompt words of interest during the experimental process. These pairs include differences and similarities such as *live mouse* vs. *dead mouse*, *good soldier* vs. *soldier fighting*, *built hospital* vs. *damaged hospital*, and *playing soccer* vs. *watching soccer*. Such list will allow the computational models to evaluate distinctive attribute representations and consequently adjust the baseline meaning of a word to convey the effects of context and conceptual combination. Table 4.1 lists the contrasting sentences. It includes the semantic classification, the sentence number and the sentence itself. For example, the verb *flew* in sentences 200, 204 and 207 appears in two different contexts: animate (as in *bird* and *duck*) vs. inanimate (as in *plane*). Such contrasting sentences illustrate the idea of conceptual combination and provides the basis for computational models that characterize the effect of context.

Table 4.1: Collection of 77 contrasting sentences. Sentence examples with differences and similarities in meaning. For instance, the verb *kicked* in the first two sentences, is used in two different contexts, playing with a ball (as in a soccer game) vs. breaking the door (as an aggressive behavior). Such sentence pairs illustrate the idea of conceptual combination providing the basis for computational models that characterize the effect of context.

| SEMANTIC CONTRAST | No. | SENTENCES (verbs) |
|---|---|---|
| SOCCER | 236 | *The artist kicked the football* |
| | 62 | *The boy kicked the stone along the street* |
| BREAKING | 111 | *The soldier kicked the door* |
| ANIMAL | 200 | *The yellow bird flew over the field* |
| | 204 | *The duck flew* |
| PLANE | 207 | *The red plane flew through the cloud* |
| BLOCKING LIGHT | 99 | *The cloud blocked the sun* |
| BLOCKING PHYSICAL OBJECT | 209 | *The bicycle blocked the green door* |
| HUMAN COMMUNICATION | 89 | *The mayor listened to the voter* |
| | 90 | *The jury listened to the famous businessman* |
| | 24 | *The commander listened to the soldier* |
| NOISE FROM A MACHINE | 92 | *The lonely patient listened to the loud television* |
| DANGEROUS SITUATION | 81 | *The reporter interviewed the dangerous terrorist* |
| QUIET SITUATION | 82 | *The policeman interviewed the young victim* |
| INFORMATION FOCUS | 77 | *The author interviewed the scientist after the flood* |

| SEMANTIC CONTRAST | No. | SENTENCES (nouns) |
|---|---|---|
| GOOD | 93 | *The soldier delivered the medicine during the flood* |
| AGGRESSIVE | 111 | *The soldier kicked the door* |
| INFORMATION | 92 | *The lonely patient listened to the loud television* |
| OBJECT | 101 | *The dog broke the television* |
| | 118 | *The dangerous criminal stole the television* |
| PLAYING | 230 | *The young girl played soccer* |
| WATCHING | 234 | *The businessman watched soccer* |
| BAD | 29 | *The doctor stole the book* |
| GOOD | 115 | *The doctor helped the injured policeman* |
| | 164 | *The old doctor walked through the hospital* |
| OPAQUE | 99 | *The cloud blocked the sun* |
| TRANSPARENT | 207 | *The red plane flew through the cloud* |
| LIGHT | 199 | *The cloud was white* |
| DARK | 134 | *The old judge saw the dark cloud* |
| BLUE | 50 | *The feather was blue* |
| WHITE | 62 | *The white feather was under the tree* |
| EXPLOSION | 103 | *The accident destroyed the empty lab* |
| TRAFFIC | 112 | *The banker was injured in the accident* |
| SOLID | 31 | *The window was dusty* |
| BROKEN | 100 | *The baseball broke the window* |
| AGGRESSIVE | 102 | *The angry activist broke the chair* |
| PASSIVE | 117 | *The soldier arrested the injured activist* |
| PLANT | 51 | *The tree was green* |
| SHELTER | 202 | *The green duck slept under the tree* |

| SEMANTIC CONTRAST | No. | SENTENCES (adjectives) |
|---|---|---|
| HOT AIR | 208 | *The summer was hot* |
| HOT LIQUID | 224 | *The coffee was hot* |
| BAD PEOPLE | 118 | *The dangerous criminal stole the television* |
| | 151 | *The mob was dangerous* |
| NATURE | 98 | *The flood was dangerous* |
| STYLE OF PLAY | 217 | *The aggressive team took the baseball* |
| ANGER | 218 | *The duck was aggressive* |
| | 185 | *The diplomat bought the aggressive dog* |
| SMALL OBJECT | 42 | *The teacher broke the small camera* |
| YOUNG | 55 | *The small boy feared the storm* |
| LARGE OBJECT | 57 | *The boat crossed the small lake* |
| | 58 | *The army built the small hospital* |
| YELLOW FUR | 43 | *The yellow dog approached the friendly teacher* |
| YELLOW PAPER | 68 | *The magazine was yellow* |
| YELLOW METAL | 104 | *The accident damaged the yellow car* |
| SEE-THROUGH, LARGE | 31 | *The window was dusty* |
| SMALL | 63 | *The dusty feather landed on the highway* |
| LEAVES | 51 | *The tree was green* |
| FEATHERS | 202 | *The green duck slept under the tree* |

| DIFFERENT CONTEXTS | No. | SENTENCES (nouns) |
|---|---|---|
| DEAD | 60 | *The man saw the dead mouse* |
| ALIVE | 56 | *The mouse ran into the forest* |
| POSITIVE | 5 | *The parent watched the sick child* |
| NEGATIVE | 9 | *The parent shouted at the child* |
| POSITIVE, EMPATHY | 5 | *The parent watched the sick child* |
| NEGATIVE, DISCIPLINE | 21 | *The angry child threw the book* |
| NEGATIVE | 7 | *The priest approached the lonely family* |
| POSITIVE | 2 | *The family was happy* |
| | 3 | *The family played at the beach* |
| NEGATIVE | 218 | *The duck was aggressive* |
| ACTIVE | 204 | *The duck flew* |
| PEACEFUL | 202 | *The green duck slept under the tree* |
| NEGATIVE | 185 | *The diplomat bought the aggressive dog* |
| POSITIVE | 181 | *The dog ran in the park* |
| | 43 | *The yellow dog approached the friendly teacher* |
| ACTIVE | 157 | *The victim feared the criminal* |
| PASSIVE | 82 | *The policeman interviewed the young victim* |
| ACTIVE POSITIVE | 3 | *The family played at the beach* |
| | | |
| PASSIVE, NEGATIVE | 27 | *The beach was empty* |

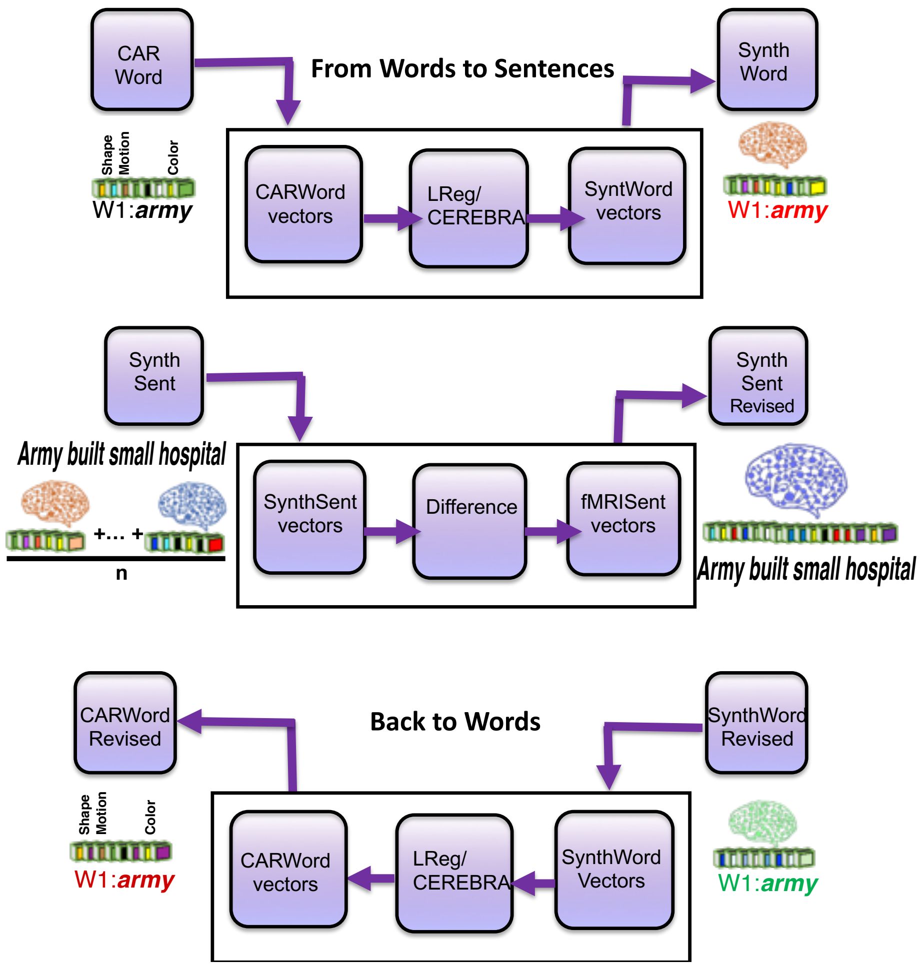



Figure 4.1: General System framework and data flow to predict sentence fMRI from CAR words. Mapping CARWord to SynthWord (top). Then SynthWord is combined by averaging to form SyntSent and to be compared to the actual fMRISent (middle). Invert the process to modify the CARWords via SynthWord revised (bottom). The Revised CARWord includes different word meaning across sentences.

### 4.3 Computational Models

A new mechanism was proposed in Chapter 3 that maps semantic vectors into sentence fMRI. It is grounded on CAR theory and implemented using the CEREBRA model (Chapter 3). In this chapter, CEREBRA is compared with Multiple Linear Regression (LReg) in this task.

The goal is to predict sentence fMRI from CAR words. Figure 4.1 presents the general system framework and data serving as a general description for the two approaches. The inputs are the concepts or neutral CARWords and the outputs are the context-based CARWords. It starts by mapping CARWord to SynthWords; gotten from the fMRI sentences (top of Figure 4.1). The SynthWord is then combined by averaging all words that occur in the sentence to form SyntSent for the predicted sentence. Next, the SynthSent is compared to the actual fMRISent (middle of Figure 4.1). The differences are included by modifying each of the SynthWord that map to fMRISent and by modifying each of the CARWord that map to the modified SynthWord (bottom of Figure 4.1). The resulting CARWord indicates how word meaning changes across sentences.

### 4.3.1 Multiple Linear Regression

At the word level, Multiple regression (LReg) is used to learn the mapping between CARWord and SynthWord voxels. The training set has attribute vectors of words as independent variables and the corresponding SynthWord vectors as the dependent variable, predicting one voxel at the time. Similarly, at the sentence level, the training contains assembled sentences (SynthSent) as independent and the corresponding observed fMRISent as the dependent variable. Once the prediction error is calculated, LReg is inverted (which is possible because it is linear), to determine what the CARWord values should have been to make the error zero. Algorithm 4.1 describes the LReg approach in detail. Note that the initial number of CARWords is 236 as explained in Chapter 3; however, for the analysis the list of words expands into a larger group of 786 context-based words or CARWord-revised. This expansion is indicated in Steps 8 and 13 of the algorithm.

The prediction task is run for each of the 11 subjects of the fMRI study. The Matlab function **fitml** is used to run multiple linear regression to map the CARWord to the SynthWord and the inverted linear process to map the SynthWord-revised to produce the CARWord-revised. It uses least squares to predict more than one dependent variable ($Y$) for one or more independent variables ($X$).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{i2} + \varepsilon \tag{4.1}$$

where $i$ is the number of observations (depending on the level of process, after 236 words or 237 sentences), $Y_i$ represents the dependent variable, $X_i$ represents the independent variable, $\beta_0$ represents $y$-intercept (constant term), $\beta_p$ is the slope coefficient for each independent variable, and $\varepsilon$ represents the error or residual.

Additional processes such as assembling the sentences (averaging all words in a sentence) and calculating the predicted and proportional errors are implemented in Matlab scripts.

### 4.3.2 Nonlinear Neural Network CEREBRA

It is possible that the linear prediction based on LReg is not powerful enough to account for the context effects. Therefore, the nonlinear approach based on the model proposed in previous chapter is tested as well. That is, a backpropagation neural network is trained to map CARWord to SynthWord for all sentences. Then, the words are averaged (as before) into a prediction of the sentence SynthSent. The prediction error is used (through backpropagation) to train the network.

After training, this network is used to determine how the CARWords should change to eliminate the error. That is, for each sentence, the CARWords are propagated and the error is formed as before, but during backpropagation, the network is no longer changed. Instead, the error is used to change the CARWords themselves (which is the

FGREP method; Miikkulainen et al., 1991). This modification can be carried out until the error goes to zero, or no additional change is possible (because the CAR values are already at their max or min limits). Ultimately, the revised CARWord represents the word meanings for the current sentence.

---

Algorithm 4.1 Linear Regression mapping CAR words to sentence fMRI & back to CARs

---

1: For Subject=1 to 11

2: For voxels=1 to 396

3: Run Matlab **fitlm** function using CARWord (66 attributes) as independent var X and SynthWord (one voxel) as dependent var Y

4: Concatenate all the single Y results for SynthWords

5: For sent=1 to 237

6: Assemble the SynthWords into the 237 SynthSent by averaging the appropriate words:

$SynthSent(sent, 1{:}396) = [W(a, 1{:}396) + W(b, 1{:}396) + \ldots + W(y, 1{:}396)]/n$

- ◊ $a, b, \ldots, y$ represent the correct word index for each SynthSent content words
- ◊ $n$ is number of words for each SynthSent

7: Obtain the prediction error by subtracting SynthSent from the observed fMRISent for each voxel:

$PredictErr(sent, 1{:}396) = fMRI\,(sent, 1{:}396) - SynthSent(sent, 1{:}396)$

8: For ctxt_word=1 to 786

9: Calculate the contribution of each word in the sentence:

$SynthWordcontrib(ctxt\_word, 1{:}396) =$

$SynthWord(a{:}n, 1{:}396)/[SynthWord(a, 1{:}396) + \cdots + SynthWord(n, 1{:}396)]$

- ◊ $n$ is number of words for each sentence

10: Change the values of the original SynthWord, distributing the proportional error:

$SynthWordrevised(ctxt\_word, 1{:}396) = SynthWord(ctxt\_word, 1{:}396) -$

$[SynthWordcontrib(ctx\_word, 1{:}396) * PredictErr(sent, 1{:}396)]$

- ◊ $sent$ changes according to which word belongs to the appropriate sentence

11: For attributes=1 to 66

12: Run Inverted LReg (**fitlm**) for SynthWord revised (396 Voxels) as independent var X and CARWord (one attribute) as dependent var Y

13: Concatenate all the single Y results for revised CARWords

---

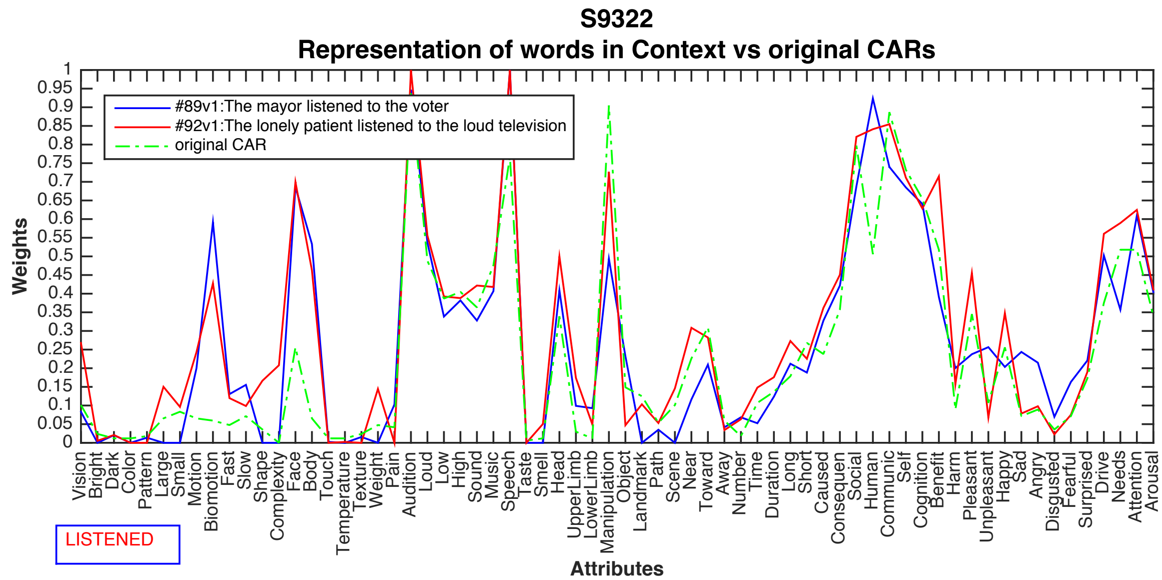
## 4.4 Experiments and Results

The two approaches LReg and CEREBRA were evaluated in a preliminary experiment to characterize the different meanings of the verb *listened*. LReg was found to be inadequate in this task and therefore in two subsequent experiments, focusing on the adjective *dangerous* and in the noun *mouse* only the CEREBRA approach was used. This analysis was performed on the individual subjects for which the fMRI data in general was most consistent.

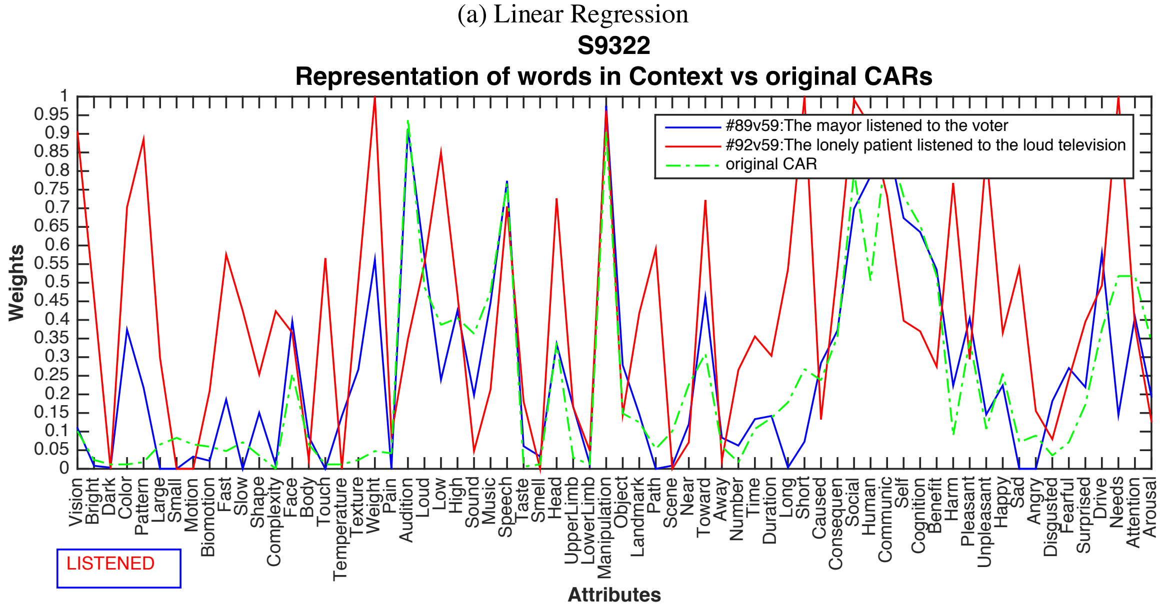
### 4.4.1 Different contexts for the verb "*listened* "

Both models were used in this experiment to compare the contrasting meanings of HUMAN COMMUNICATION vs. NOISE FROM A MACHINE for the word *listened* as expressed in 89: *The mayor listened to the voter*, 92: *The lonely patient listened to the loud television*. The top of Figure 4.2 shows the results for LReg between the original and transformed CARs. Although the CARs adjusted in all sentences, the changes were small and unprincipled, unable to characterize the difference between human communication versus noise from a machine. In contrast, the outcome for CEREBRA resulted in context-dependent changes as shown, for sentences 89 and 92 in the bottom of Figure 4.2.

CARs in Sentence 89 presented salient activations in human-related attributes like Face, and Body, Audition, and Speech, as well as Human, Communication, and Cognition, presumably denoting human verbal interaction. For Sentence 92, high activations on Vision, Bright, Color, Pattern, Large, Shape, Complexity, Touch, Temperature, Weight, Scene, Near, Harm, Unpleasant, Happy, and Angry describe a loud and large object such as a television. Other subjects and words were also evaluated (*injured, visited, watched, ate, kicked, accident, family*) with similar outcomes. These




(a) Linear Regression




(b) Neural Network CEREBRA

Figure 4.2: Results for the word *listened* in two contrasting sentences. LReg (top) did not capture context. All changes were insignificant to characterizing the context-dependent representations. The green line shows the original CARs for comparison. CEREBRA (bottom) did grasp context. The CARs for Sentence 89 have increased activations in human-related attributes like Face and Body, Auditory attributes, as well as Human, Communication and Cognition. In contrast, Sentence 92 activations on Vision, Color, Large, Shape, Complexity, Touch Temperature, High sound, and Unpleasant, depict a loud object such as a television.

results suggest that the linear mapping that LReg performs is not powerful enough to capture context, but the nonlinear mapping of CEREBRA is. A possible explanation is that the relations between the concept attributes and the voxels are too complex to be linearly separable. Therefore, the following experiments only used the CEREBRA method for characterizing the effect of context.

### 4.4.2 Different contexts for the adjective "*dangerous*"

This experiment compared the contrasting meanings of NATURE vs. BAD PEOPLE for the word "dangerous", as expressed in 98: *The flood was dangerous*, 118: *The dangerous criminal stole the television*. Figure 4.3 shows the differences resulting from the CEREBRA method. As with the verb *listened*, context-dependent changes did emerge.

CARs in Sentence 98 present changes on activation for Large, Motion, SOMS attributes Texture and Weight, and event attributes Time, Short, and Caused, reflecting moving water. The attributes Toward, Harm, Unpleasant, and the emotion of Angry, represent the experiential and personal nature of danger. Conversely, Sentence 118 shows high activation for Vision, Complexity, Face, and Speech, because they represent human types and roles such as a criminal. Motor attribute Lower Limb as well as evaluation attributes Benefit, Angry, Disgusted, and Fearful can be associated with a dangerous act by a criminal. The CEREBRA method, therefore, was largely able to differentiate between the contrasting relevant dimensions of *dangerous* act of nature and humans.

### 4.4.3 Different contexts for the noun "*mouse*"

This experiment compared the contrasting meanings of DEAD vs. ALIVE for the word *mouse* as expressed in sentences 56: *The mouse ran into the forest*, 60: *The man*

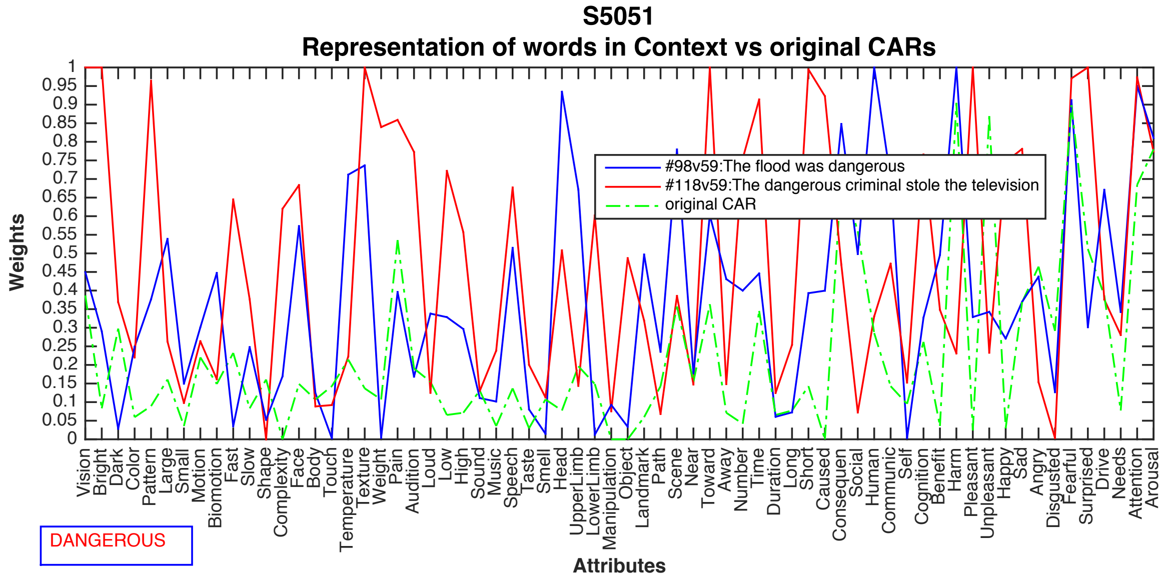



Figure 4.3: CEREBRA results for the adjective *dangerous* across two contrasting sentences. CARs in Sentence 98 changed activation for Large, Motion, Texture and Weight, Time, Short, and Caused, reflecting moving water. The attributes Toward, Harm, Unpleasant, and Angry, represent the experiential nature of danger. Sentence 118 shows high activation for Vision, Complexity, Face, and Speech, because they represent human types and roles. Lower Limb, Benefit, Angry, Disgusted and Fearful can be associated with a dangerous act by a criminal.

*saw the dead mouse*. Figure 4.4 shows the differences resulting from the CEREBRA method, which are again systematic and meaningful.

CARs in Sentence 56 have increased activation for Vision, Motion, Complexity, High, and Sound, possibly suggesting animate properties of the live mouse. Upper Limb, spatial attributes Path and Away, and event attributes Time, Duration, Short, and Consequence, symbolize activity such as running. Emotions of Fearful and Surprised may well be associated with seeing a live mouse. In contrast, Sentence 60 shows increased activation for Temperature, Weight, and Smell, as well as emotions Sad, Angry, Disgusted and Fearful, which may be associated to the dead mouse. These changes indicate different aspects of *mouse* in two contrasting contexts.

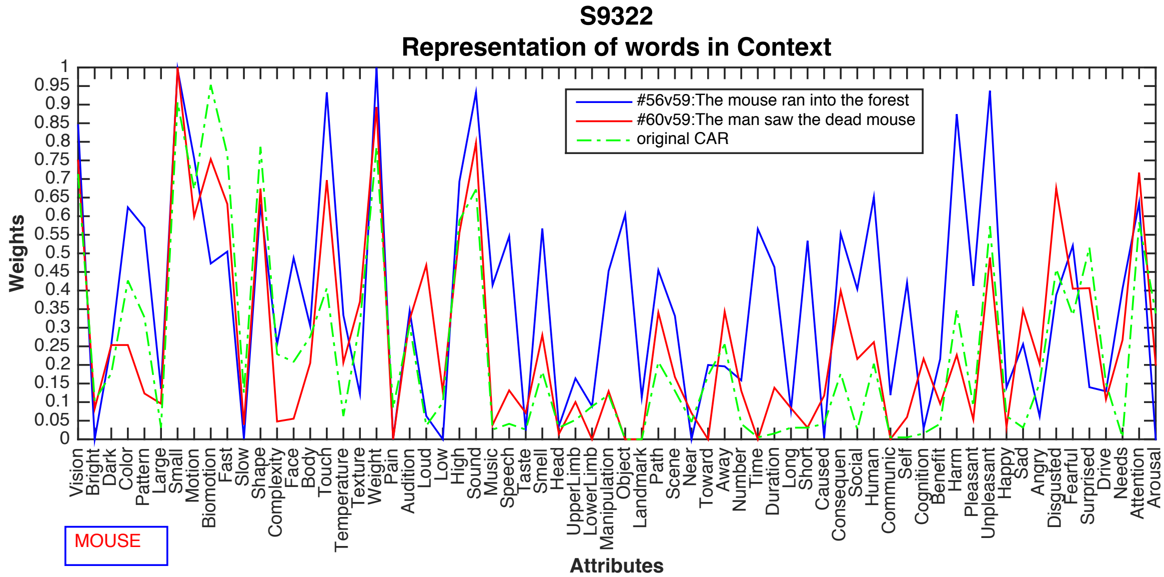


Figure 4.4: CEREBRA results for the noun *mouse* across two contrasting sentences. CARs in Sentence 56 increased activation for Vision, Motion, Complexity, High, and Sound, presumably to indicate the animate properties of the live mouse. Upper Limb, Path, Away, Time, Duration, Short, and Consequence, suggest activity such as running. In contrast, Sentence 60 shows increased activation for Temperature, Weight, and Smell, as well as Sad, Angry, Disgusted and Fearful, which can be associated to the dead mouse. These changes indicate different aspects of mouse in two contrasting contexts.

### 4.4.4 Analysis of the distribution of changes produced by CEREBRA

It is informative to analyze the overall distribution of changes produced by CEREBRA. Figure 4.5 shows that on average the new CAR values increased. The results suggests that the new representations are gaining content. This is a remarkable result, because a simple statistical learning system could be expected to instead regress to the mean. In contrast, the modified representations in CEREBRA become (1) more descriptive and (2) more distinctive, which provides a good foundation for understanding the structure of the semantic space and for building applications (as will be discussed in Chapter 8).

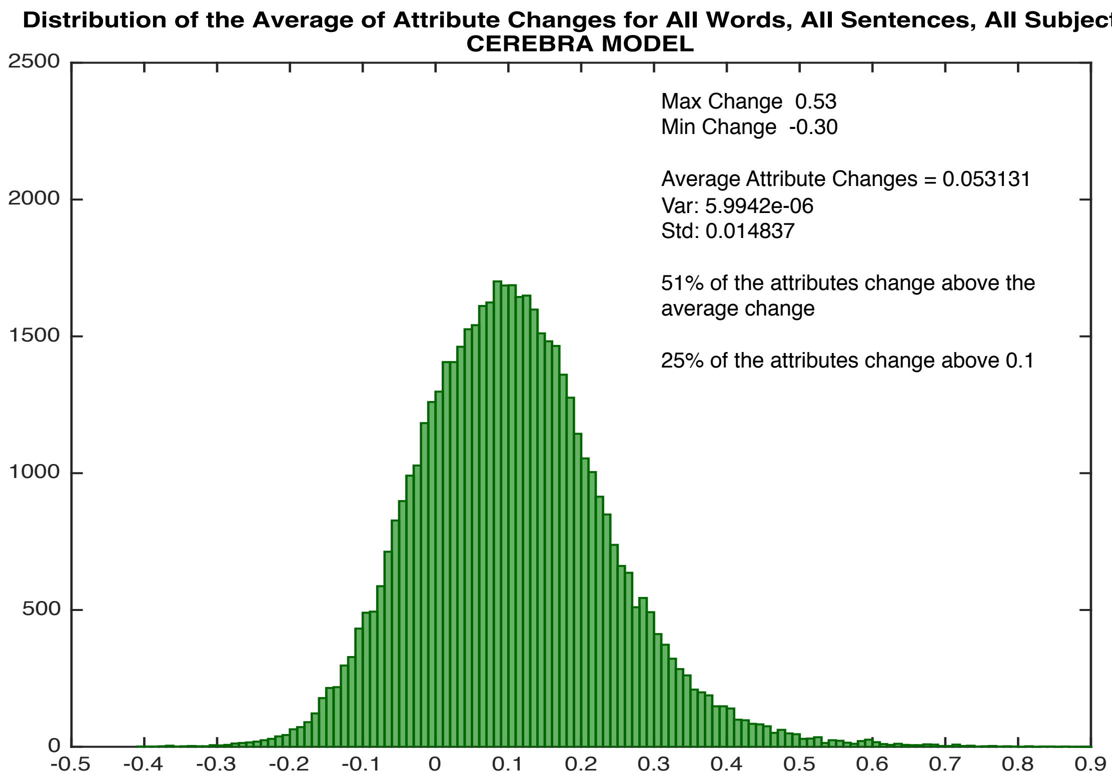



Figure 4.5: Distribution of the average changes produced by CEREBRA. The distribution is performed for all words, all sentences, and all subjects. The results show that on average, the new CAR values increased. This results suggest that the new CARs are gaining content. The new CARs are more descriptive and distinctive.

## 4.5 Discussion and Future Work

Experiments with available fMRI data show that the approach is feasible, demonstrating meaningful differences for e.g., *human communication* vs. *noise from a machine*; *dangerous storm* vs. *dangerous person*; *live mouse* vs. *dead mouse*. These changes are principled and can be captured by the neural network model. It may then be possible to create them dynamically and forming a basis for a more robust and grounded natural language processing system.

The results suggest that different aspects of word meaning are activated in different contexts, and it is possible to see those changes in the corresponding fMRI images using the CEREBRA model. These changes are likely to be nonlinear: The linear mapping approach (regression) tends to muddle them, but a nonlinear mapping (CEREBRA neural network) can tease them apart.

As the first step, only single subjects were analyzed in this chapter. In the next chapters, the analysis is extended to more subjects identifying which changes are consistent across subjects, and which ones are more unique. For instance, the subject in experiment 3 was Sad that the mouse was dead; another subject could show a different emotion.

## CONCLUSION

Concepts are always changing; their meaning depends on context and recent experience. In this chapter, word meaning was represented as a collection of attributes (CARs), grounded in observed brain systems. Multiple Linear Regression analysis and a nonlinear neural network CEREBRA were used to understand how the CARs could change to construct the actual sentence representations seen in fMRI images. The linear mapping approach yielded disorganized results, but the nonlinear mapping characterized the results in a meaningful manner. The results suggest that there are indeed systematic changes in CARs, and they make sense in each sentence context: Different features of word meaning are activated in different contexts.

# Chapter 5

# Exploring the Conceptual Combination Effect Under the Context of Sentences

Embodied theory of cognition represents concepts through weighted attributes and the weights may change in context. The challenge is to characterize these words' dynamic meanings by measuring how the attribute weighting changes across contexts. The CEREBRA model is used to capture those changes in the context of the sentence by combining the meaning of the individual words. This chapter[2] characterizes the context effect on word meanings by analyzing the changes produced by the CEREBRA model through different experiments: (1) Characterizing the effect of similar context on two different words, (2) Characterizing differences in two contexts, (3) Characterizing the effect of two different contexts on the same word, and (4) Characterizing the centrality of meaning. Each experiment analyzes the changes observed in the concept attributes across context and illustrates how individual conceptual combinations develop.

## 5.1 Motivation

Embodiment theories of knowledge representation reviewed in Chapter 2 suggest that word meaning consist of a collection of weighted attributes defined in terms of different neural systems, and the weights change in context. This chapter aims at quantifying such adaptive meanings using a computational modeling.

---

[2] The content of this chapter was previously presented at Brain Informatics 2018 (Aguirre-Celis & Miikulainen, 2018). Aguirre-Celis worked on experimental design, implementation and analysis; while Miikkulainen provided guidance and feedback through discussions.

An intriguing challenge to such theories is that concepts are dynamic, i.e., word meanings are not fixed entries or lists of attributes, but dynamically processed each time a word is encountered (Barclay et al., 1974; Barsalou, 1987; Gennari et al., 2007; Murphy, 1988; Pecher et al., 2004; Yee et al., 2016). For example, an art painter searching in a supermarket for a fresh *apple* with the perfect shape and glossy green color would invoke different aspects of the word *apple* depending on whether she will be painting or making and apple pie. When thinking about art, the emphasis will be on the *apple's* physical appearance, including perfect round shape and spotless green color. When thinking about baking, the emphasis will be on the apple's flavor and maturity, as in tart tasting and crispy texture (Pecher et al., 2004). It is possible to track such dynamic meanings of words by measuring how the attribute weighting changes across contexts.

In Chapter 4 these changes were reported anecdotally in two separate experiments. Multiple Linear Regression and the nonlinear CRERBRA model were used to map the CARs to the FMRI data in order to understand how the CARs could change to approximate the actual sentence representations seen in fMRI images. The results suggested that different features of word meaning were activated in different contexts. The linear mapping approach yielded disorganized results, but the nonlinear mapping characterized the results in a meaningful manner.

The experiments presented in this chapter analyze the CAR changes more systematically in two ways. First, they characterize the changes that occur when a word is used in the context of a sentence, and second, they explain how different conceptual combinations emerge from such context.

First, in CAR theory the activation of attribute representations is modulated continuously through attention and the interaction with context. Since the importance given to individual attributes of a word varies with context, four analyses to visualize those changes are included in this chapter:

(1) Characterizing the effect of similar context on two different words. The goal is to analyze the similarities and differences between the concepts *boat* and *car* across subjects, indicating how context emerges producing distinct members of the same category of vehicles.

(2) Characterizing differences in two contexts. The idea is to quantify the emotional context of *laughed* and *celebrated* by analyzing how context appears from thematic associations, and demonstrating how such cognitive content can be a powerful source of context beyond the more obvious physical context.

(3) Characterizing the effect of two different contexts on the same word. The purpose is to examine the conceptual noun-verb combination using the representations of *bird flew* vs. *plane flew*, to evaluate how context gives rise to different degrees of animacy.

(4) Characterizing the centrality of meaning. The point is to characterize centrality of meaning between the concepts *small camera* vs. *small hospital*, evaluating how the same attribute is true for both concepts, but more central to one than to the other.

Second, Chapter 2 reviewed the types of processes involved during conceptual combination. The first three experiments in this chapter illustrate Wisniewski's (1997): attribute combination (e.g., *red apple* is an apple of a red color), relational combination (e.g., *apple basket* is a basket that holds apples), and hybrid combination (e.g., *apple pie* is a pie made of apples). The last experiment illustrates the centrality effect (e.g., the attribute Small is more central for a *bird* than a *whale*) proposed by Medin & Shoben (1988).

The following sections describe the four different experiments, analyze statistical significance with context aggregation across subjects, and demonstrate how conceptual combinations emerge.

## 5.2 CHARACTERIZING EFFECTS OF SIMILAR CONTEXT

In the first experiment the salient attributes for the words *boat* and *car* are compared under the semantic category of transportation vehicles as expressed in 57: *The boat crossed the small lake* and 142: *The green car crossed the bridge*. In principle, *boat* and *car* should occur in the same sentence context, but due to data availability, the experiment is designed with sentences that are similar and typical of those nouns. Context draws attention to a subset of attributes, which are then enhanced, forming the basis for object categories. CEREBRA model quantifies such enhanced representations for *boat* and *car*, revealing common underlying properties in the transportation vehicle category (Binder, 2016). Other words were also considered (*dog* vs. *mouse; horse* vs. *fish; tea* vs. *water*), with comparable results.

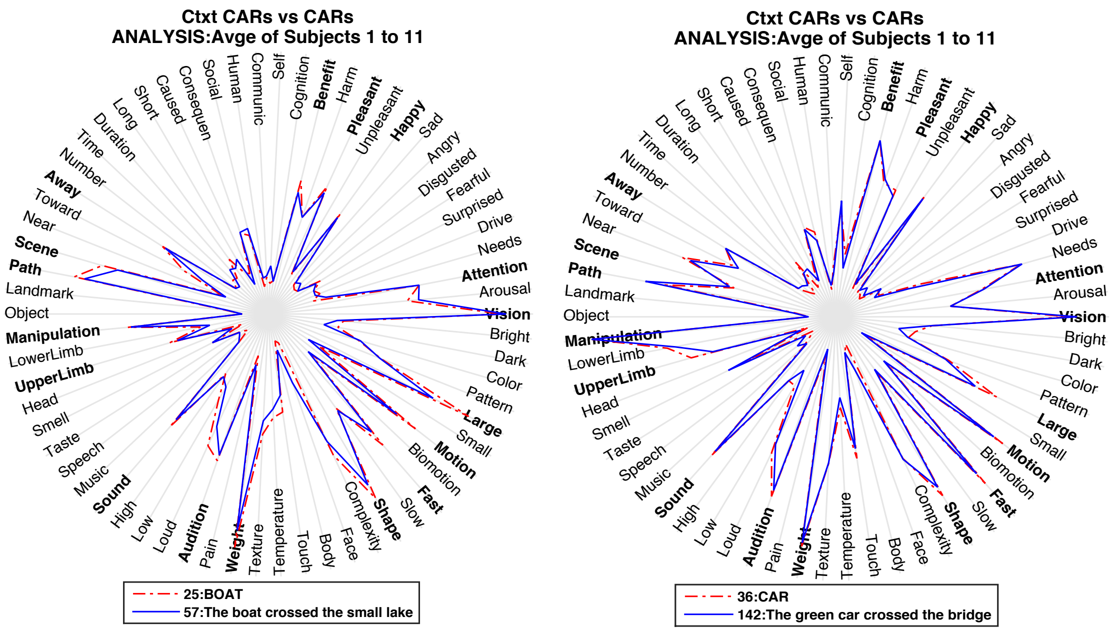
Figure 5.1 shows the results averaged across subjects. For *boat* in sentence 57, there are changes on Vision, Large, Motion, Shape, Complexity, Weight, Sound, Manipulation, Path and Scene and event attribute Away, reflecting a large moving object. Evaluation and Emotion attributes of Benefit, Pleasant and Happy represent the experiential and personal nature of using a boat. Similarly, *car* in sentence 142 shows analogous activation for the same brain areas. Since both belong to the same semantic category, they share similar context-related attribute enhancement. However, the distinctive weighting on these attributes sets them apart. The CEREBRA model was thus able to identify the effect of similar context on these two concepts across subjects.

By demonstrating the concept of vehicle categories, this experiment exhibits the attribute combination process (Wisniewski, 1997). One or more attributes from the combination are transferred to the other words in the sentence. For example, Motion, Manipulation, Path and Scene, Benefit, and Pleasant are some of the attributes shared by all these words before context: *boat*, *car*, *crossed*, *lake* and *bridge*. The conceptual combination produces overlapping dimensions that enhance context-relevant attributes of the concepts. This process can be seen as forming "ad hoc categories" (Barsalou, 1983). An ad hoc category is a novel category constructed spontaneously to achieve a goal relevant in the current situation (Barsalou, 1983). When concepts share the same context-related attribute enhancement, ad hoc categories are formed. This process may be considered as an instance of attributes combination (Wisniewski, 1997). The context enhancement of context-relevant attributes for one or more concepts alters the relative similarity between them, thus forming an ad hoc category.

### 5.3 Characterizing Differences in Two Contexts

The second experiment examines the common emotional context in Sentences 4: *The wealthy family celebrated at the party*, and 14: *The couple laughed at dinner*, by how such cognitive content can be an instrumental source of context and demonstrating how context develops from external relations.

Many concepts such as *celebrated* and *laughed* refer to affective states and emotions, and other cognitive experiences. One advantage on using CARs is that such experiences count as much as sensorimotor experiences in grounding conceptual



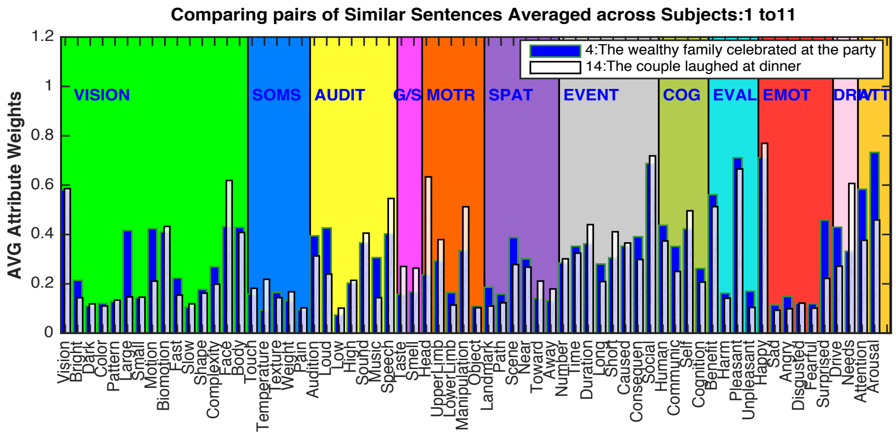

(a) Boat (b) Car

Figure 5.1: The effect of similar context for the words *boat* and *car* averaged across subjects. Results are shown for the new CARs as an average of all subjects. The dotted lines indicate the original CARs and solid lines specify the context-based representations. Both plots display similar changes, but the different weightings set them apart. The attribute combination process is validated (Wisniewski, 1997). One or more concepts share the same context-related attribute enhancement thus forming the vehicle category.

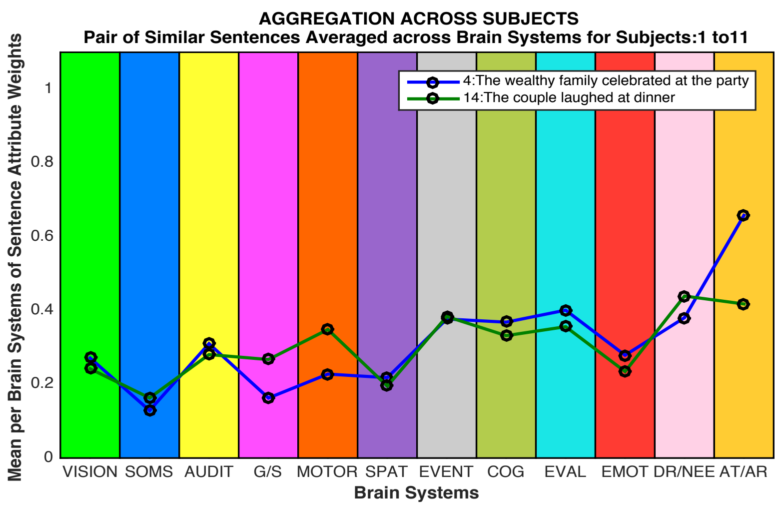
representations. When people "feel happy", they experience this phenomenon the same way as the sensory or motor events, except that the perception is internal. Similarly, to evaluate context in these sentences, CARs alone cannot capture the thematic associations between concepts (i.e., party, celebration, birthday cake, candles, laugh) unless additional sources provide it. Hence, the second experiment is designed to quantify that sort of context developed from external relations, i.e., spatial and temporal co-occurrence of events, captured by CEREBRA.

Figure 5.2 shows that these sentences resulted in very similar contexts, emphasizing Scenes, Events, and positive Emotions. Figure 5.2(a) shows the context CARs averaged for each sentence for all subjects. Both sentences are mostly similar on Spatial, Event, and Emotion attributes. Figure 5.2(b) aggregates these dimensions across the 12 corresponding brain areas according to the CAR theory. All subject brain signatures mainly differ in Gustatory, Motor, and Attention, possibly highlighting that laughing at dinner involves food and requires more head and upper body movements. In contrast, celebrating demands more Attention and Arousal. The results thus suggest that CEREBRA captured the thematic relations where the two contexts intersect semantically. They also validated that emotional content is a prominent and potentially powerful factor in sentence context, and there are subtle differences in it that can cause subtle differences in word meanings.

Finding how sentence meaning is represented in the brain remains a major challenge (Anderson et al., 2018; Just et al., 2017). The results in this experiment verify the relational combination process (Wisniewski, 1997). CEREBRA captures the thematic knowledge of the sentences by mapping the heteromodal semantic representations (CAR) to fMRI data. Context emerges from external relations, i.e., from the spatial and temporal co-occurrence of events. By using CEREBRA, it is possible to uncover the weightings of the brain systems for the entire sentence (as was done in Figure 5.2b), however, the thematic associations exposed by the model require further examination. CAR theory is capable of representing similar concepts across a large number of categories, but this classification of concepts become extremely similar (e.g., human roles --- lawyer, reporter, and judge), as a result, CARs is not able to capture the thematic associations. For example, the difference between judge and lawyer is related to the type of job they do, where they work etc. but the existing set of attributes in the CAR do not capture this type of relation. In the Discussion and Future Work Chapter this point will be addressed and a potential solution proposed.




(a) Celebrated vs Laughed




(b) Averaged

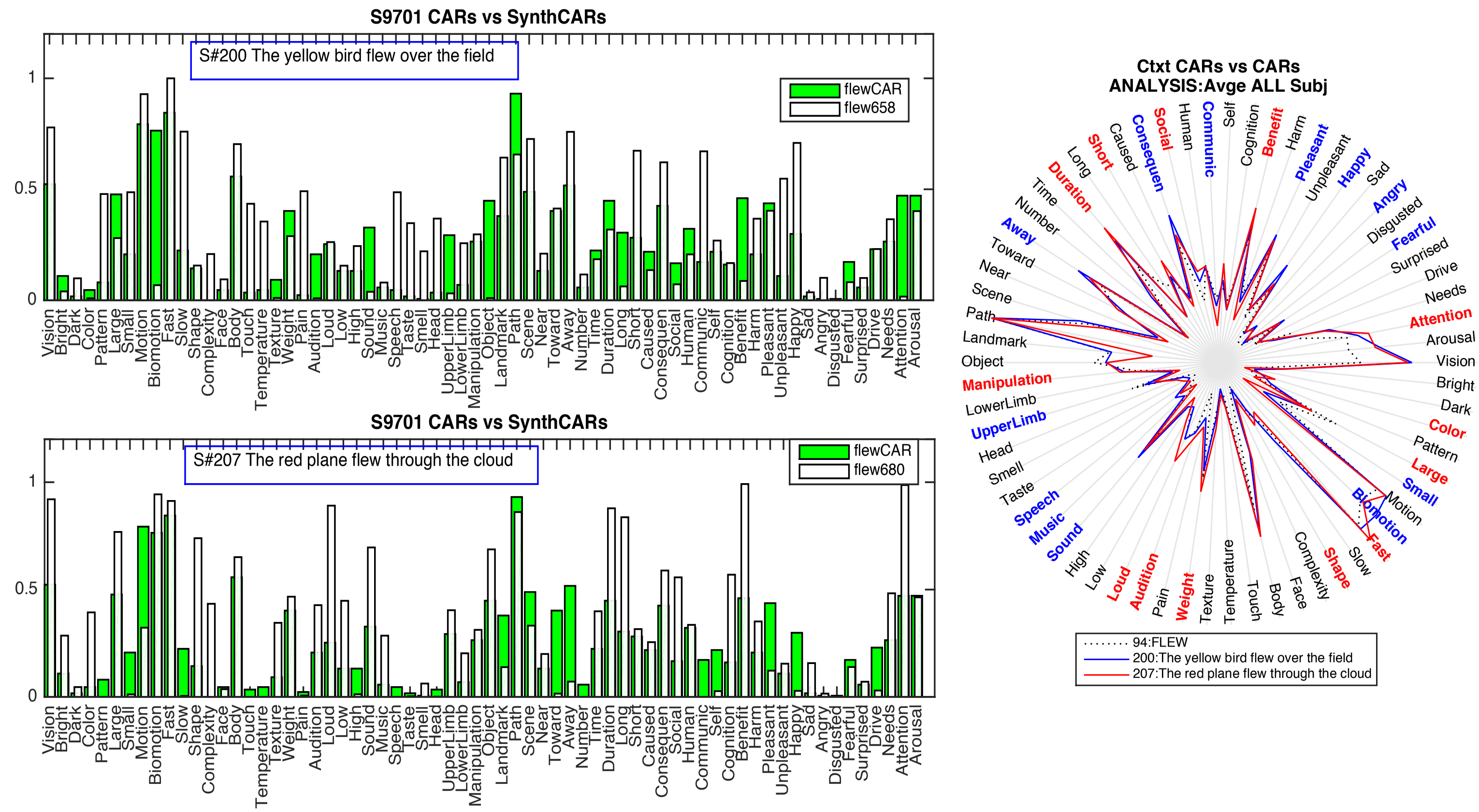
Figure 5.2: Results featuring differences between two contexts averaged across subjects. (a) A comparison of the averaged attributes for each sentence representing *celebrated* and *laughed*. (b) Aggregation analysis across subjects for each brain zones. These context-based representations differ mostly in the Gustatory, Motor, and Attention zones, possibly emphasizing that laughing at dinner involves food and requires more movement than celebrating at the party, but the latter demands more Attention and Arousal. The relational combination process is demonstrated (Wisniewski, 1997) by the way context emerges from external relations, i.e., spatial and temporal co-occurrence of events. The results in this experiment indicate that CEREBRA captures the thematic knowledge of the sentences by mapping the heteromodal semantic representations (CAR) to the fMRI data.

## 5.4 CHARACTERIZING EFFECTS OF DIFFERENT CONTEXTS

In the third experiment, the attributes of the noun-verb combination are analyzed for the word *flew*, as expressed in 200: *The yellow bird flew over the field*, and 207: *The red plane flew through the cloud*. According to the foundations of CAR theory (Section 2.5), noun-verb interaction arises within multiple brain systems, activating similar brain zones for both concepts, plus the interaction between concepts that activate similar attributes. These interactions determine the meaning of the conceptual combination (Binder, 2016). Since *bird* is a living thing, animate sensory, motor, affective, and cognitive experiences are activated, including attributes like Face and Speech. In contrast, *plane* has salient activations along animate dimensions such as Emotion, Cognition, and, Attention. These simple differences demonstrate the hybrid combination process (Wisniewski, 1997). Knowledge about animacy is a combination of several types of experiences including perception of causality, perception of motion, perception of individual's own internal affective and drive states, and the development of theory of mind (Binder, 2016). Such knowledge arises from interacting heteromodal brain networks in addition to the activation of similar attributes between concepts.

Figure 5.3 shows the differences for *flew* in the two contexts. The top part displays all the 66 attributes for subject S9701, and the bottom part shows the average for all subjects with the statistically significant attributes highlighted. The analysis was done for all subjects using other conceptual combinations (*banker drank* vs. *dog drank*; *dangerous criminal* vs. *dangerous flood*; *injured horse* vs. *injured person*; *horse walked* vs. *person walked*) producing similar effects.

The results demonstrate context-dependent changes on Sentence 200 with salient activations on animate attributes like Face, Small, and Body, Audition, Music, Speech, Taste and Smell, as well as Communication. On the other hand, Sentence 207 yields




(a) Subject S9701

(b) Averaged across subjects

Figure 5.3 The effect of two different contexts for the word *flew*. (a) Showing subject S9701 changes of the original CAR vs. new CAR for all 66 attributes. (b) Averaged for all subjects. Showing the statistically significant attributes in blue for *bird flew* and in red for *plane flew,* and dotted lines represent the original concept *flew* before context. The new CARs for Sentence 200 have salient activations on animate features, presumably denoting *bird* properties like Speech, Small, and Communication. Sentence 207, has high activations on inanimate object features, describing a Loud, Large, and Heavy object such as a *plane*. This effect thus demonstrates hybrid combination (Wisniewski, 1997), which requires additional knowledge from complex and interacting brain networks along with the collaboration between concepts activating similar set of attributes in order to find the contexts of animacy and otherwise.

significant changes on Size, Color and Shape, Weight, Audition, Loud, Sound, Duration, Social, Benefit, and Attention. These results suggest that CEREBRA was able to determine the effect of two different contexts into the resulting CARs. As the context varies for each sentence representation, the overlap on neural representations creates a mutual enhancement: the attributes of the target concept (e.g., *bird flew*) that are relevant to the context (e.g., animacy), are those that overlap with the conceptual representation of the context, and this overlap results in a clear difference between animate and inanimate contexts. The hybrid combination process involves additional processing within the neural systems where attribute representations overlap (Wisniewski, 1997). In the CEREBRA model, the context of animacy has a conceptual representation that overlaps with the noun-verb combination of *bird flew*, thus demonstrating a hybrid combination process (Wisniewski, 1997).

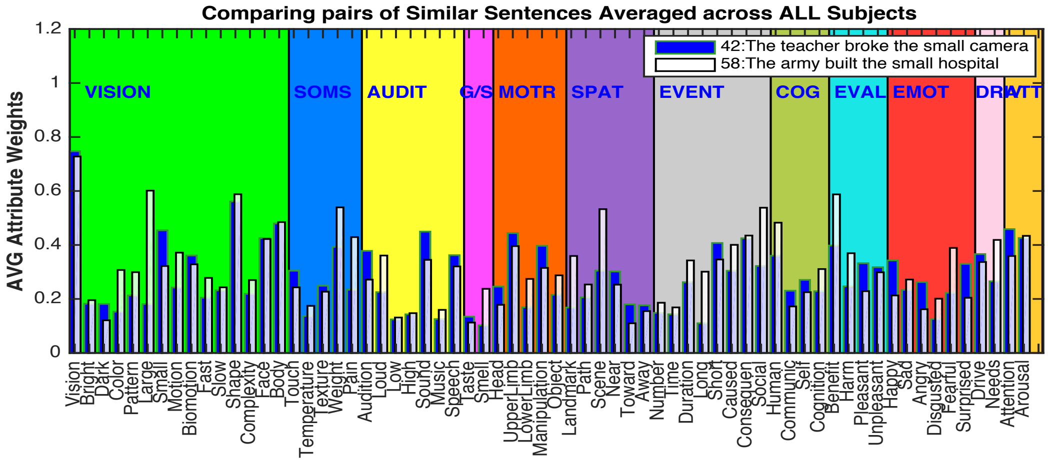
## 5.5 CHARACTERIZING THE CENTRALITY OF MEANING

In the fourth experiment, the attributes of the adjective-noun combinations are analyzed on the centrality effect for the word *small*, as expressed in 42: *The teacher broke the small camera*, and 58: *The army built the small hospital*. Centrality expresses the idea that some attributes are true to many different concepts, but they are more important to some concepts than others (Medin & Shoben, 1988). Figure 5.4 shows the differences for *small* in these two contexts. The top part displays all 66 attributes for the two sentences across subjects, and the bottom part displays the new *camera* and *hospital* representations averaged across all subjects.

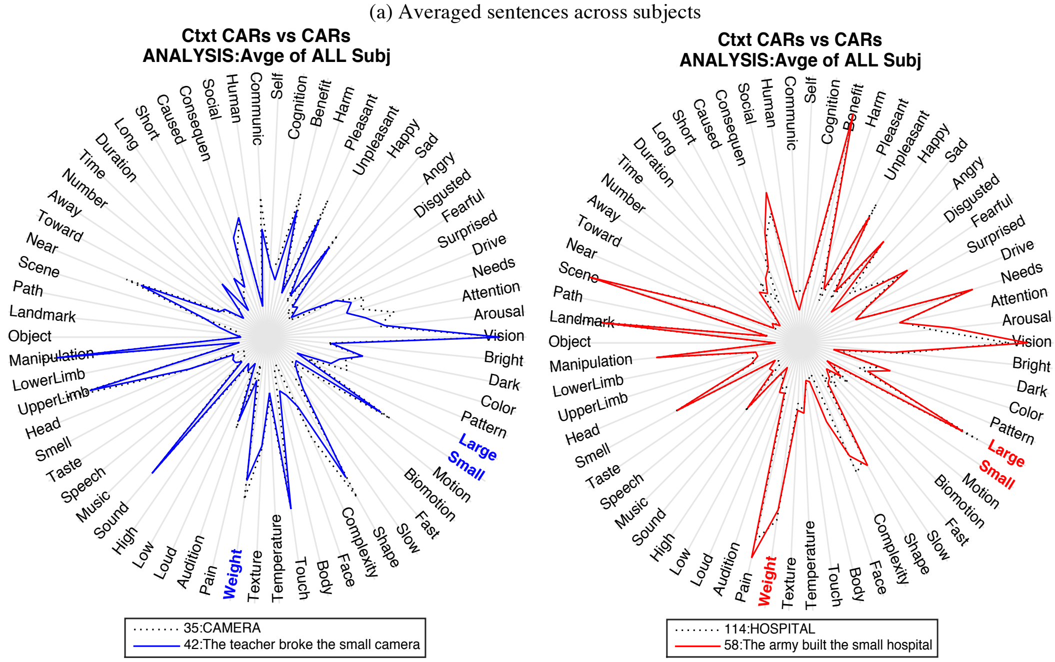
The size dimensions (e.g., Small and Large), demonstrated the centrality principle for these specific contexts. The top of the figure shows Sentence 42 (e.g., *small camera*) with salient activation for the central attribute Small and low activation for the non-central attribute Large. In contrast, Sentence 57 (e.g., *small hospital*) presents a low activation on the non-central Small attribute but a high activation on the central Large attribute. These findings suggest that these attributes are essential to small objects and big structures, respectively. However, the size dimension alone cannot represent the centrality effect completely. Adding other attributes such as Weight would help, but other cases may require additional knowledge. For example, *small boy* not only includes the size dimension but also young age; the Size and Weight dimensions do not represent such a concept completely. This type of conceptual combination requires additional world knowledge to determine the centrality for a particular adjective, and the relationships between the dimensions for various contexts.

Additionally, given that both *camera* and *hospital* are inanimate objects, the bottom part of the figure shows that they share low weightings on human-related attributes Biomotion, Face, Body, and Speech. However, they also differ in expected ways, including salient activations on Darkness, Color, Small and Large size, and Weight. As part of the sentence context, the activations include human-like attributes such as Social, Human, Communication, Pleasant, Happy, Sad and Fearful. Overall, each sentence representation moves towards their respective sentence context (e.g., *camera* or *hospital*).

These observations are robust and general: analysis was done for all subjects using other conceptual combinations (*small bird* vs. *small boy*; *small boy* vs. *small lake*; *small boy* vs. *small church*), producing comparable results.




(a) Averaged sentences across subjects




(b) *Camera* and *Hospital* averaged across subjects

Figure 5.4: The effect of centrality on two contexts for the word *small*. (a) Shows the average for all subjects for the two sentences. (b) Displays the new *camera* and *hospital* representations averaged for all subjects. In the top figure, the new CARs for Sentence 42 have salient activations on small object, denoting the *camera* properties like Dark, Small, Manipulation, Head, Upper Limb, Communication, and emotions such as Sad (e.g., *broke* the *camera*). Sentence 58 has high feature activations for large buildings describing a Large, and Heavy structure such as a *hospital*. In the bottom figures the central attributes are highlighted to emphasize how same attributes are more important to some concepts than others. Demonstrating the centrality effect (Medin & Shoben, 1988), each combination requires additional knowledge, and the interaction between concepts that activate similar set of attributes.

## 5.6 Discussion and Future Work

The experiments in this chapter suggest that different aspects of word meaning are weighted differently in distinct contexts, and it is possible to identify those changes for individual concepts, a combination of concepts, and for sentences by analyzing the corresponding fMRI images through the CEREBRA model. The changes in the CARs were averaged across subjects and found to be statistically significant.

The experiments demonstrated that different conceptual combinations include features that are not derived from their component parts only but require additional knowledge. The fMRI images possibly include such external knowledge, as was demonstrated in the third experiment by discovering the animate/inanimate differences for the noun-verb combinations *bird flew* vs. *plane flew*. Therefore, by using the CEREBRA model it is possible to characterize such conceptual combinations by observing the changes in CARs. What exactly is represented and how, is still unknown.

The results are significant considering that the dataset was not originally designed to answer the question of dynamic meaning. Limited by the data available, the experiments presented here address specific cases, however, by expanding the collection (e.g., identical contexts and contrasting contexts) the number of potential observations would increase, making it possible to test more systematically.

Synthetic words built by combining sentences where the word occurs, is similar to many semantic models in Computational Linguistics (Landauer & Dumais, 1997; Mitchell & Lapata, 2010; Vinyals et al., 2015). Also, synthetic words formed by fMRI sentence representations has been successful in cases like predicting brain activation (Anderson et al., 2016; 2018; Grand et al., 2018; Just et al., 2017). Although this study does not have a large set of sentences, the CEREBRA process of mapping semantic CAR words to the synthetic words and further to sentences fMRI refined the synthetic representations by removing noisy information. Still, fMRI images for individual words instead of having to synthesize them, should amplify the observed effects.

The next step is to aggregate the analysis across sentence contexts. The goal is to determine how similar sentences cause similar changes in word representations. The process starts by forming clusters of the 237 sentence representations. For each cluster, all new CARs with similar roles are identified and the changes between the new and the original CARs averaged and correlated with possibly the other words in the sentence. This experiment is the topic of the next chapter.

### CONCLUSION

Concepts are dynamic; their meaning depends on context and recent experience. In this chapter, word meaning was represented as a collection of attributes (CARs), grounded in observed brain systems. This chapter characterized the context effect on word meanings by analyzing the changes produced by the CEREBRA model through four different experiments. The experiments analyzed the CAR changes systematically and illustrated the different conceptual combination mechanisms. The changes in the CARs were demonstrated in many different sentences and averaged across subjects, and found to be statistically significant. Essentially, the CEREBRA model was able to capture the context of the sentence by measuring how the attribute weighting changed across context.

# Chapter 6

# Quantifying the Effect of Conceptual Combination on the Concept Attributes

This chapter focuses on the conceptual combination process. It describes how such a dynamic construction of concepts in the brain can be quantified. This idea was presented anecdotally in Chapter 5, by analyzing a few example cases of how the concept attributes are weighted differently in various sentence contexts. The current chapter[3] expands on this prior work by evaluating the robustness and generality of these conclusions across an entire corpus of sentences and semantic roles.

## 6.1 Motivation

What is the "glue" between words like *banana* and *pepper* that allows individuals to understand the resulting conceptual combination *banana pepper*? People weight concept features differently based on context. Whenever they encounter a combination of concepts, they start searching for a coherent explanation that put together the concepts usually by the relational combination or attribute combination. Conceptual combination was introduced in Chapter 2 and characterized anecdotally in Chapter 4. Essentially a relational combination forms some relation between two concepts to create a new one, and the attribute combination uses an attribute of one concept to describe another. Both combinations play an important role in the construction of new or complex concepts (Gagné & Shoben, 1997; Murphy 1990; Pecher, Zeelenberg, & Barsalou, 2004). This

[3] The content of this chapter was previously presented at the 41th Annual Meeting of the Cognitive Science Society (Aguirre-Celis & Miikulainen, 2019). Aguirre-Celis worked on experimental design, implementation and analysis; while Miikkulainen provided guidance and feedback through discussions.

chapter focuses on attribute combination. Here, the modifier features (e.g., noun, adjective) adapt other concepts in the combination to some degree, and as a result, the words involved are more alike (Wisniewski, 1998). For example, listeners must realize that *banana pepper* could mean just a fruit having a certain color by selecting salient features that dominate in the combination. The noun *pepper* is defined by color, size, shape, taste etc. and one or more of those dimensions will be modified during the attribute combination (i.e., curved shape and yellowish color resembles a banana).

This chapter demonstrates how such a dynamic construction of concepts in the brain can be quantified. Chapters 4 and 5 showed (1) that words in different contexts have different representations, and (2) these differences are determined by context. These effects were demonstrated by analyzing individual sentence cases across multiple fMRI subjects (Chapter 5). This study verifies these same conclusions in the aggregate through a statistical analysis across an entire corpus of sentences: It measures how the CARs of a word changes in different sentences, and correlates these changes to the CARs of the other words in the sentence. In other words, it quantifies the conceptual combination effect statistically across sentences and subjects.

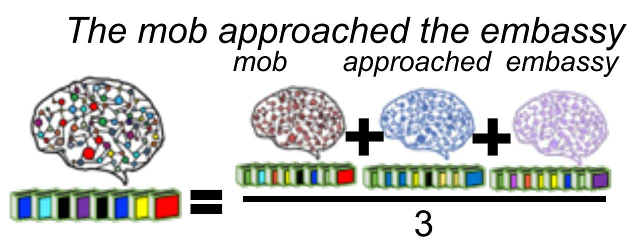
In the following sections an analysis across sentence contexts is described followed by an example on how conceptual combinations modify word meanings across clusters of sentences and roles. Then, the aggregation methodology is reviewed, and the results presented and discussed.

### 6.2 Analysis across Sentence Contexts

The aggregation study hypothesis is based on the idea that similar sentences have a similar effect, and this effect is consistent across all words in the sentence. In order to test this hypothesis, it is necessary to (1) form clusters of similar sentences for the entire collection, and (2) calculate the average changes on the words identified by the role they play for the same cluster of sentences. After that, it will be possible to demonstrate how similar sentences cause analogous changes in words that play identical roles in those sentences using correlations.

### 6.2.1 Sentence Clustering

First, synthetic sentences are assembled by using the new CARs produced by the CEREBRA model. The sentences are formed by averaging the representation of all words occurring in the sentences as described in Chapter 3, diagrammed concisely as:




Then, the 237 by 66 sentence representations are grouped into clusters (e.g., according to their context similarity) using the Matlab function ***linkage*** to form an agglomerative hierarchical cluster. It uses the Ward method to measure the distance between clusters and Euclidean metric to measure the distance between observations. Ward's linkage consists of computing the distance between clusters by the minimum variance algorithm. At each step, it finds the pair of clusters that leads to the minimum increase in total within-cluster variance after merging. The within-cluster sum of squares is a measure of the variability of the observations in each cluster. It is defined as the sum of the squared distances between all the observations inside the cluster and the centroid of the cluster. The sum of squares metric is equivalent to the following distance metric formula:

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}} \|\overline{x_r} - \overline{x_s}\|_2 \tag{6.1}$$

where $\| \quad \|_2$ is Euclidean distance, $\overline{x_r}$ and $\overline{x_s}$ are the centroids of clusters $r$ and $s$, $n_r$ and $n_s$ are the number of elements in clusters $r$ and $s$.

A series of cluster analyses are performed with the number of clusters ranging from 20-50. Thirty clusters are found to be the optimal number, displaying the most clear classifications and reflecting similar groupings to the semantic categories for the original CARs addressed by Binder et al. (2016).

Figure 6.1 shows a dendrogram example illustrating a few interesting contexts that emerged. Animals, which formed three clusters 28, 19, and 14, are shown at the top of the figure. It includes two sentences (marked in parenthesis) that do not have any animals in the constituents 162: *The magazine was in the car* and 172: *The car approached the river*. They do not fit naturally within the context but are closer to other sentences possibly due to some features that are shared with sentences like the change in location for *approached*, *ran*, and *flew* in cluster 14. Furthermore, Animals branched into a subgroup related to open locations as in clusters 19 and 14. Weather events, Accidents and Violence group together environmental and human phenomena in addition to some criminal violence as in *stole* and *damaged* objects or structures. Social Interaction, the cluster in the bottom of the figure, contained context associated with *interviewed*, *spoke*, *shouted*, *negotiated*, *met*, *listened*, and *protest*, featuring simple interaction between two or more people. For each of the 11 Subjects some distinctive clusters were formed.

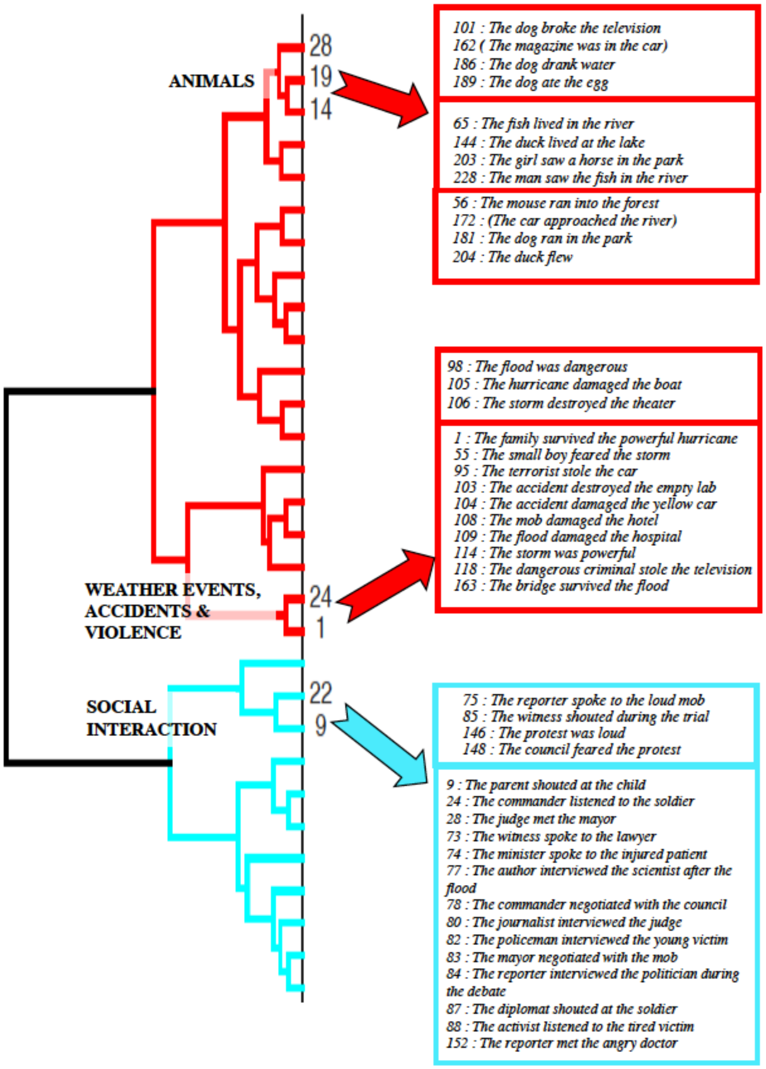



Figure 6.1: Dendrogram segments from the hierarchical cluster analysis using new sentence representations for Subject S9701. Animals, Weather events, Accidents, Violence, and Social Interaction, each clustered in separate branches. Only displays 7 clusters for better visualization. The sentences in parenthesis indicate that they do not fit intuitively in such context. Clusters of new sentence representations for the aggregation study.

### 6.2.2 Average Changes by Word Roles

Each cluster of sentences is expected to reveal similar changes in some of the dimensions. To recognize common patterns of changes modulated by context, the next step is to calculate the average of the changes for words with similar roles, e.g., *hospital*, *hotel*, and *embassy* (within the same cluster of sentences). That is, measure the difference between the new and the original CARs for each similar word roles and perform a statistical significance analysis using the Student's *t*-test to prove that those changes are meaningful.

Next, a detailed individual example of the conceptual combination effect on two different clusters is presented followed by the aggregation study and the results.

## 6.3 The Conceptual Combination Effect across Sentence Clusters

The aggregation analysis across sentence contexts targets a higher-level characterization by clustering sentence representations to find context. Hence, those clusters of similar sentences are expected to display equivalent changes on the words that play identical roles.

One way to identify the context of the sentences that cluster together and the effect on the roles is by looking at the centroid sentences for each cluster. The next three figures (Figure 6.2, 6.3 and 6.4), present an example that considers the concept interactions in the centroid Sentence 7: *The priest approached the lonely family* (Cluster 1, Subject S9324), and the centroid Sentence 150: *The mob approached the embassy* (Cluster 7, Subject S9726). Figure 6.2 displays the two clusters of sentences with similar contexts including the centroid sentences. Figure 6.3 shows the original CARs for the content words of the centroid sentences, and Figure 6.4 shows the word changes for the centroid sentences' concepts, playing the same role.

Figure 6.2 shows two different clusters representing a variety of sentences where the context similarity might be built upon the agent as in *priest* vs. *mob* or the patient/location as in *family* vs. *embassy*. For *priest* (top part), the cluster groups together context associated with nice people helping or interacting with other individuals (i.e., *minister*, *policeman*, *activist*, *woman*, and *doctor*). In contrast for *mob* (bottom part), the cluster groups together rather violent or aggressive people (i.e., *army*, *terrorist*, and *criminal*). Analogously for *family (top part),* animate dimensions cluster together various human beings (i.e., *patient*, *victim*, *tourist*, *policeman*), as opposed for *embassy (bottom part),* the inanimate dimensions cluster together objects or locations (i.e., *hospital*, *car*, *hotel*, and *television*).

Figure 6.3 shows the original CARs, displaying the weights of the generic representation (before context-based modifications) for each content word of the centroid sentences. This figure is important in order to verify how similar sentences cause analogous changes in word roles by comparing the generic meanings in the original CARs to the changes on the context-based representations for each of the word roles. The top part shows salient activations on dimensions related to agency for *priest* and *family* and relevant activations on dimensions associated with actions for motor, spatiotemporal, and events for the verb *approached*. In the same manner, the bottom part displays salient activations on dimensions related to agency and location for *mob* and *embassy*, respectively, and significant activations on dimensions associated with motion for motor, spatiotemporal, and events for the action word *approached*.

The top three plots in Figure 6.4 show only the statistically significant changes for the three roles on agent-like attributes like Small and Motion, Audition, Music, Speech, and Taste, as well as Cognition, Pleasant and Sad. Additionally, motor and spatiotemporal dimensions are overlapping in like manner among the word roles, sharing similar attributes related to the verbs in the cluster. For example, comparing the generic representation of the nouns *priest* and *family* (Figure 6.3), both have about the same activation for the attribute Path (0.2), however the verb *approached* has a very high activation (0.9). By analyzing the same attribute in the top three plots in Figure 6.4, the changes on each role display common modifications toward the context (i.e., *approached*). Similar effect was observed for other attributes. In the same manner the adjectives (e.g., *lonely*), which are not considered in this experiment but are part of the process of producing the new context representations, possibly influenced the changes in the new concept representations as well, but those changes will not be addressed here.

The bottom three plots in Figure 6.4 also show only the statistically significant changes for the three roles on agent-like attributes like Pain, Audition, Music, Speech, Taste and Smell, as well as Consequential, Short, Cognition, Self, and particularly negative emotions (e.g., Disgust) regarding the noun *mob*. On the other hand, for the inanimate nouns, attributes like Visual, Large, Landmark and Scene present congruent changes. Additionally, motor and spatiotemporal dimensions are overlapping in related manner among the word roles, sharing similar attributes related to the verbs in the cluster. For example, comparing the generic representation of the words *approached* and *embassy* (Figure 6.3), both have very low activation for the attribute Disgust (~0.01) compared to the noun *mob* (0.4). By analyzing the same attribute in the bottom three plots of Figure 6.4, the changes on each role display significant modifications toward the context (i.e., *mob*). Similarly, this effect was consistent for many other attributes.

These results illustrate the effect of conceptual combination on word meaning. As the context varies, the overlap on neural representations creates a mutual enhancement, producing a clear difference between good and bad people in addition to animate and inanimate contexts. The CEREBRA model encodes this effect into the CARs where it can be measured. Similar effect was observed for several other cluster pairs. In the next section this effect is quantified statistically across the entire corpus of sentences.

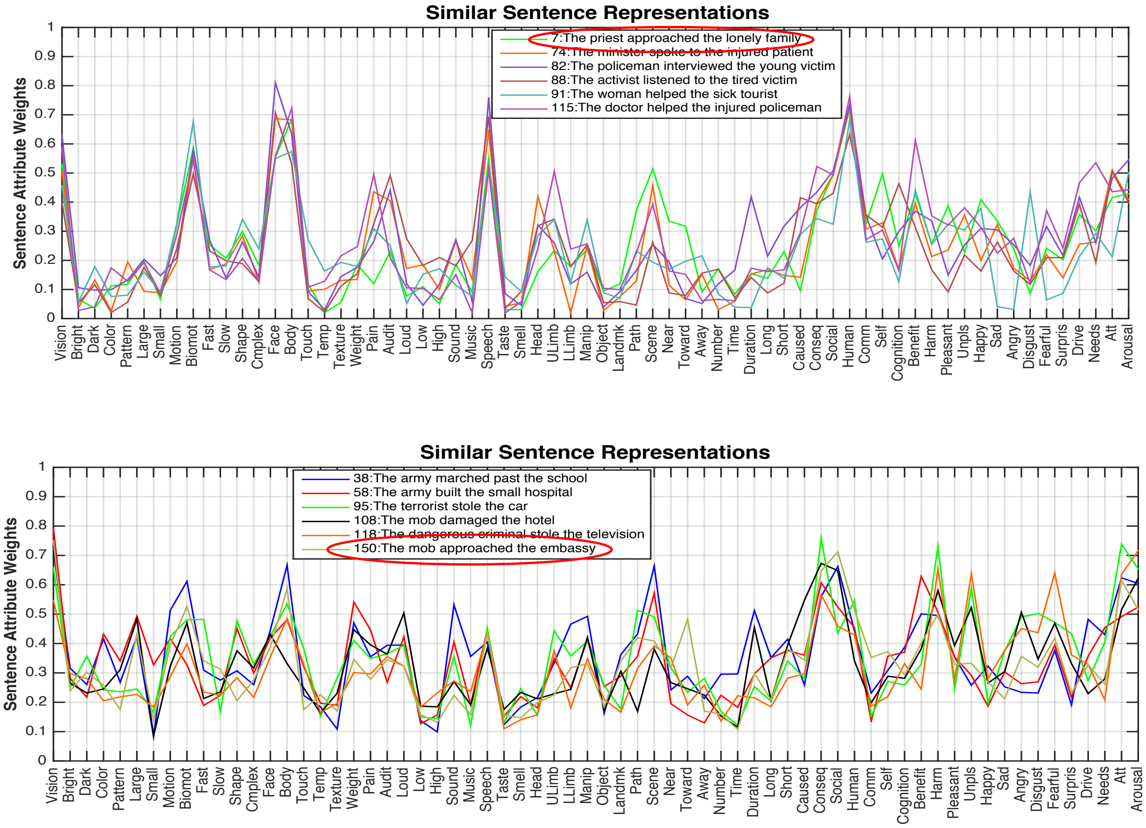



Figure 6.2: Sentences with similar contexts for clusters 1 and 7. The context similarity was organized around the agent as in *priest* vs. *mob* or the patient/location as in *family* vs. *embassy*. For *priest* (top part), the cluster groups together context associated with nice people helping or interacting with other individuals. In contrast, for *mob* (bottom part), the cluster groups together rather violent or aggressive people. Comparably for *family (top part),* animate dimensions cluster together different human beings, in the contrary for *embassy (bottom part),* the inanimate dimensions cluster together objects or locations. Plots display similar context sentences clustered together.

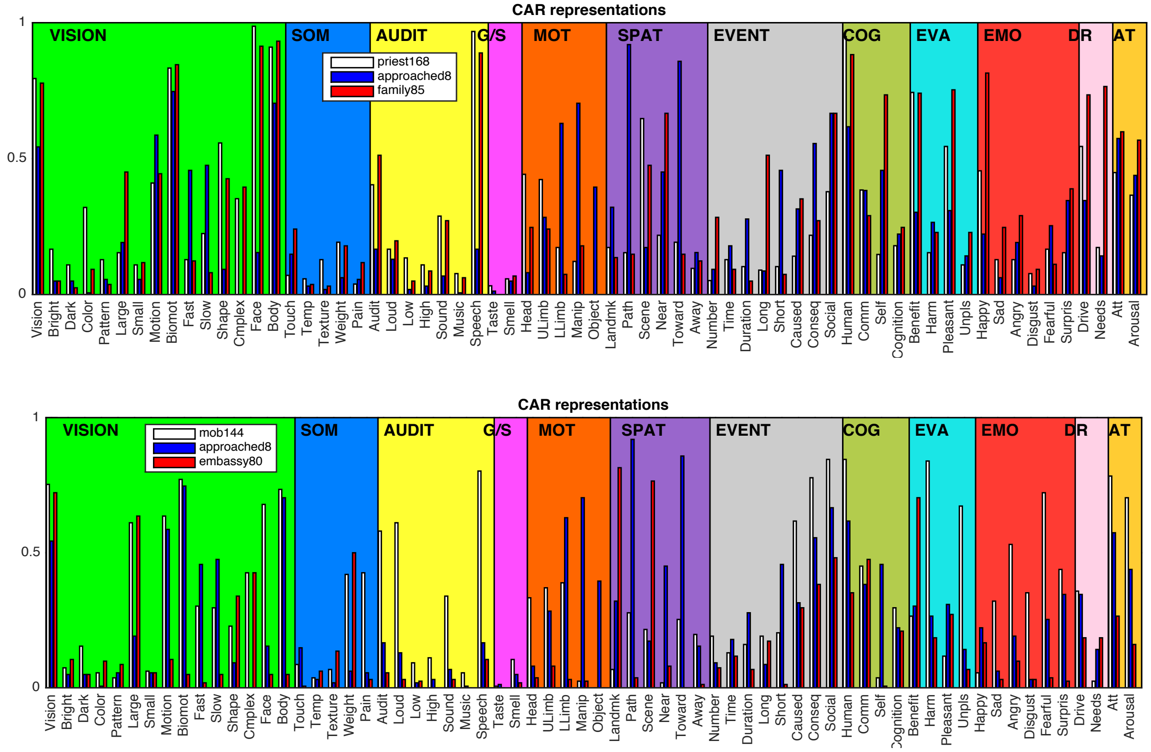



Figure 6.3: Original CARs displaying the weights of the generic representation (before context-based modifications) for each content word of the centroid sentences. The top part shows salient activations on dimensions related to agency for *priest* and *family* and activations on dimensions associated with the verb *approached*. The bottom part displays salient activations on dimensions related to agency and location for *mob* and *embassy* respectively, and activations on dimensions associated with motion for the word *approached*. Bar plots show the original CARs obtained by Binder (2016) for different words.

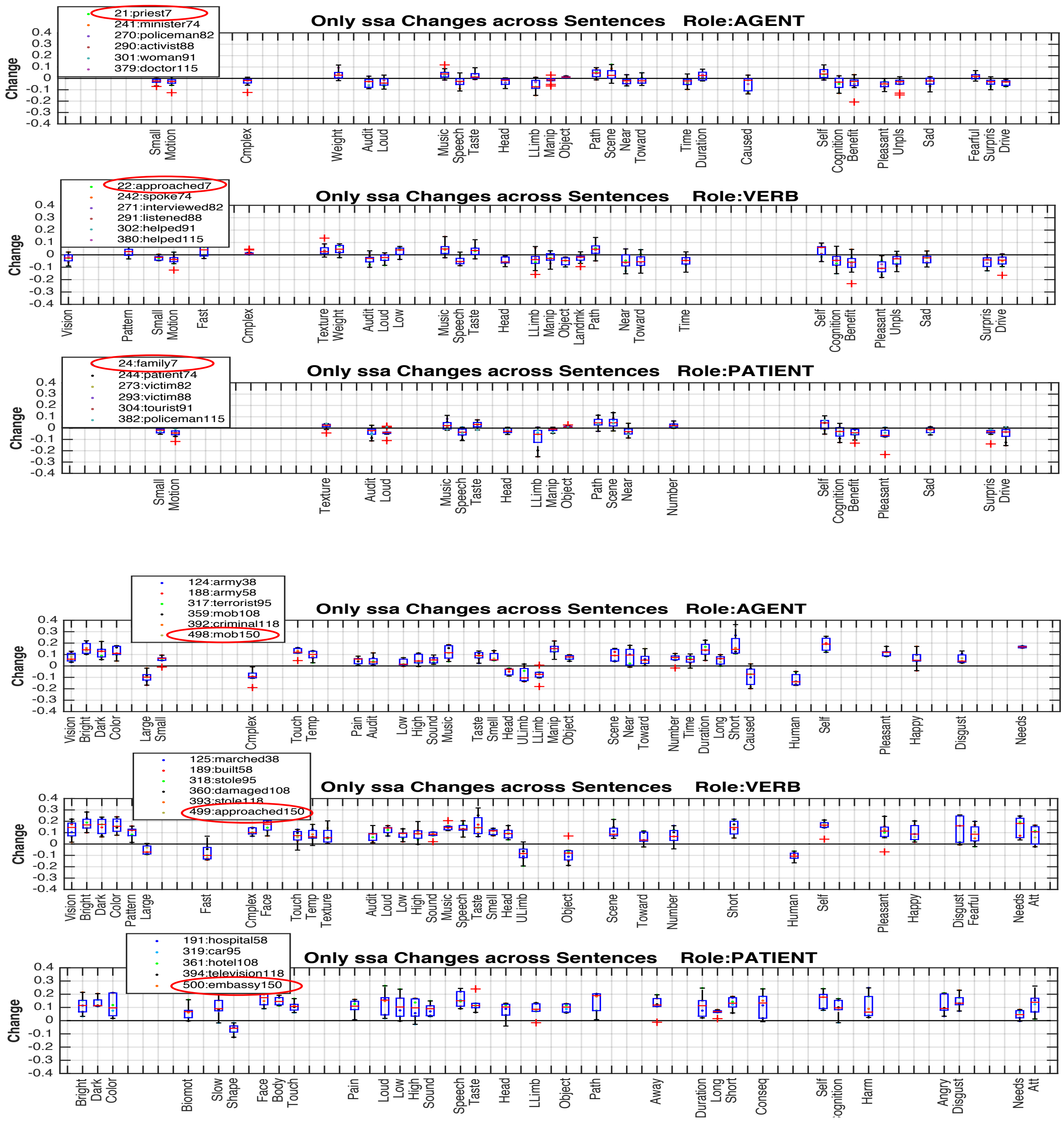



Figure 6.4: Word changes for the concepts playing the same role for the clusters. The top three plots display only the statistically significant changes for the three roles on agent-like attributes like Motion, Speech, and Taste, as well as Cognition, Benefit, Pleasant, and Surprise. The bottom three also display only the statistically significant changes for the three roles on agent-like attributes like Pain, Audition, Speech, Taste and Smell, as well as Consequential, and negative emotions (e.g., Disgust) regarding the noun *mob*. On the other hand, inanimate attributes (bottom plot) like Visual, Large, Complex, and Scene referring to objects. For the two clusters, motor and spatiotemporal dimensions are overlapping in like manner among the word roles. Box plots show only the statistically significant attribute changes for each role of the centroid sentences for clusters 1 and 7.

## 6.4 AGGREGATION ANALYSIS

So far, the conceptual combination effect has been demonstrated in a number of example cases, like the one presented in Section 6.3, and others in earlier chapters (Chapter 4 and 5). The aggregation methodology is reviewed next, and the goal is to demonstrate that the effect is robust and general across the entire corpus of sentences and case roles. Once more, the hypothesis that similar sentences have a similar effect, and this effect is consistent across all words in the sentence was verified in the following process:

1. For each subject, modified CARs for each word in each sentence were formed through CEREBRA as described in Chapter 3.
2. A representation for each sentence, SynthSent, was assembled by averaging the modified CARs (Chapter 3).
3. Agglomerative hierarchical clusters of sentences were formed using the set of SynthSents. The Ward method and Euclidean metric were used to measure the distance between clusters and observations respectively, as explained in Section 6.2.1. The process was stopped at 30 clusters, i.e., at the point where the granularity appeared most meaningful (e.g., sentences describing open locations vs. closed locations).
4. Each cluster of sentences is expected to reveal similar changes in some of the dimensions. To recognize such common patterns of changes, the next step is to calculate the average of the changes for words with similar roles, e.g., *hospital*, *hotel*, and *embassy* (within the same cluster of sentences). To that end, the differences between the modified and original CARs are measured separately for each CAR dimension in each word role, and their significance estimated using Student's *t*-test (Section 6.2.2).
5. The modified CARs of the other words in the sentence were averaged.

6. Pearson's Correlations were then calculated between the modified CARs and the average CARs of other words across all the dimensions.
7. Similarly, correlations were calculated for the original CARs.
8. These two correlations were then compared. If the modified CARs correlate with the CARs of other words in the sentence better than the original CARs, there is evidence of context effect based on conceptual combination.

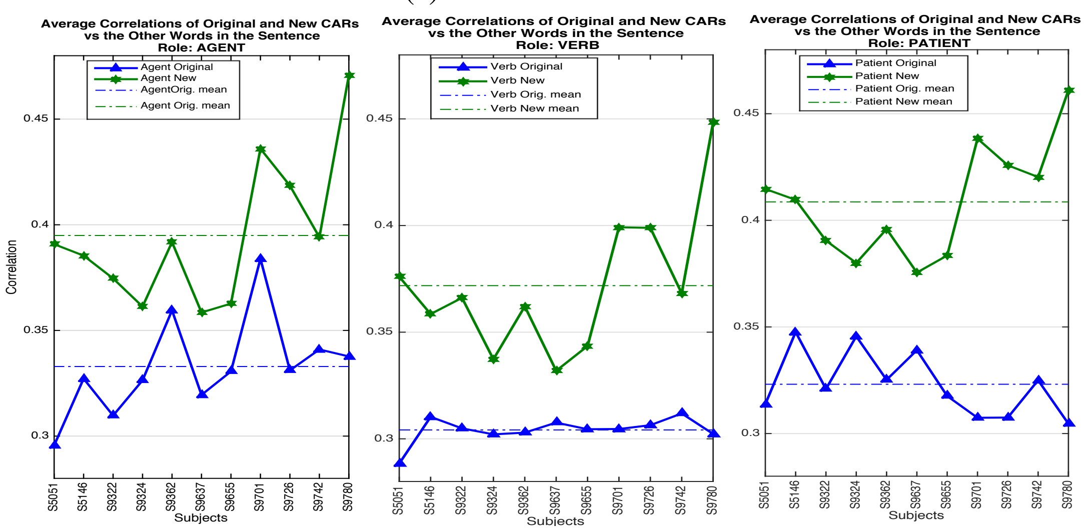
That is, this process aims to demonstrate that changes in a word CAR originate from the other words in the sentence. As in the example presented in the previous section *mob approached embassy* vs. *priest approached family, s*howed how some of the noun properties (good/bad, animate/inanimate) were transferred to the other words in the sentence, adapting the combination to the extent that the words share similar features. For example, if the other words in the sentence have high values in the CAR dimension for Landmark (e.g., *embassy* and *approached*), then that dimension in the modified CAR should be different than in the original CAR for the target word (e.g., *mob*). The correlation analysis measures this effect across the entire CARs. It measures whether the word meaning changes towards the context meaning.

### 6.5 RESULTS

The results are shown in detail in Figure 6.5. The top part of the figure presents the correlation results per subject and word roles, and the bottom part of the figure displays the results in graphic form. The correlations are significantly higher for new CARs than for the original CARs across all subjects and all roles. Additionally, the AGENT role represents a large part of the context in both analyses. In other words, the average correlations of the original and modified CARs are most similar in the Agent panel suggesting that this role encodes most of the context. It is important to note that the

| AVERAGE CORRELATIONS PER SUBJECT (3 ROLES) | | | | | | |
|---|---|---|---|---|---|---|
| | ORIGINAL | | | NEW | | |
| SUBJECTS | AGENT | VERB | PATIENT | AGENT | VERB | PATIENT |
| 5051 | 0.2956 | 0.2884 | 0.3138 | 0.3908 | 0.3760 | 0.4147 |
| 5146 | 0.3272 | 0.3103 | 0.3476 | 0.3854 | 0.3585 | 0.4096 |
| 9322 | 0.3097 | 0.3049 | 0.3209 | 0.3746 | 0.3661 | 0.3905 |
| 9324 | 0.3264 | 0.3021 | 0.3456 | 0.3613 | 0.3373 | 0.3800 |
| 9362 | 0.3595 | 0.3029 | 0.3252 | 0.3918 | 0.3621 | 0.3959 |
| 9637 | 0.3195 | 0.3076 | 0.3391 | 0.3585 | 0.3319 | 0.3755 |
| 9655 | 0.3306 | 0.3045 | 0.3176 | 0.3627 | 0.3435 | 0.3835 |
| 9701 | 0.3839 | 0.3046 | 0.3074 | 0.4360 | 0.3992 | 0.4383 |
| 9726 | 0.3311 | 0.3064 | 0.3075 | 0.4185 | 0.3989 | 0.4258 |
| 9742 | 0.3410 | 0.3119 | 0.3250 | 0.3941 | 0.3682 | 0.4203 |
| 9780 | 0.3377 | 0.3023 | 0.3046 | 0.4706 | 0.4483 | 0.4610 |
| MEAN | 0.33293 | 0.30417 | 0.32312 | 0.39494 | 0.37182 | 0.40865 |
| STDEV | 0.02364 | 0.00611 | 0.01525 | 0.03464 | 0.03355 | 0.02670 |

(a) Correlation results




(b) Correlation results in graphic form

Figure 6.5: Correlation results per subject cluster. The top part displays the correlation data per subject and word role, and the bottom part presents the same results in graphic form. Average correlations analyzed by word class for 11 subjects comparing the original and new CARs vs. the average of the other words in the sentence. A moderate to strong positive correlation was found between new CARs and the other words in the sentence suggesting that features on one word are transferred to other words in the sentence during conceptual combination. Interestingly, the original and new patterns are most similar in the AGENT panel, suggesting that this role encodes much of the context The results show that the conceptual combination effect occurs consistently across subjects and sentences.

clusters obtained for each subject's sentences in the aggregation analysis (Section 6.4), dictates the way the correlation analysis is conducted for the modified and the original CARs. Each subject produced a different arrangement of sentence clusters that is why the average correlations of the original CARs are different within each role (i.e., they depend on the subject's cluster organization), even though the original CARs include a single set of 236 words compared to the modified CARs that include eleven sets of 786 context-based words, or revised CARWords (Chapter 3).

Thus, the results indeed confirm that the conceptual combination effect occurs consistently across subjects and sentences, and it is possible to quantify it by analyzing the fMRI images using the CEREBRA model on the CARs. As a summary, the average correlation was 0.3201 (stdev 0.020) for original CARs and 0.3918 (stdev 0.034) for new CARs.

### 6.6 Discussion and Future Work

This study aimed to verify the hypothesis that during sentence comprehension, people adjust the word meanings according to the combination of the concepts that occur in the sentence. This effect had been demonstrated in individual cases before, and the goal was to demonstrate it more broadly across many subjects, an entire corpus of sentences, and different semantic case roles in those sentences. The correlation results indeed demonstrated that the effect is robust and can be quantified by analyzing fMRI images through the CEREBRA mechanism.

These findings are significant considering that the dataset was limited and was not designed to answer the question of dynamic effects in meaning. In the future, it may be possible to extend the data with identical contexts and contrasting contexts, such fully balanced stimuli could be used to test the hypothesis more systematically. Similarly, it would be desirable to extend the data with fMRI images of individual words to lead to stronger and clearer results.

One important advantage of CAR theory is that it is grounded on brain representations, and therefore a good choice when mapping semantic representations to fMRI. In the future, it would be interesting to compare whether similar effects can be observed with semantic representations based on co-occurrence in text corpora, or perhaps even a combination of the two.

## CONCLUSION

This chapter shows how word meanings adjust depending on context. All of these dynamic effects create a moving target that is (1) experience-based to allow representations to change over time, and (2) distributed such that various features can become active in different settings. Using CEREBRA as a mechanism it was possible to show that the difference between the expected and observed fMRI images can indeed be explained by a change in CARs. Across an entire corpus of sentences, the new CARs are more similar to the other words in the sentence than to the original CARs, demonstrating how features of the context are transferred to each word in the sentence.

# Chapter 7

# Mapping Brain to Behavior: Evaluating CEREBRA

Semantic feature models have become a popular tool for prediction and interpretation of fMRI data. In particular, prior work has shown that differences in the fMRI patterns in sentence reading can be explained by context-dependent changes in the semantic feature representations of the words. However, whether individuals are aware of such changes and agree with them has been an open question. This chapter[4] aims to answer this question through a human-subject study. In the survey, subjects were asked to judge how the word change from their generic meaning when the words were used in specific sentences. That is precisely what the CEREBRA model produces. Thus, the survey is intended to evaluate the results produced by the computational model.

## 7.1 MOTIVATION

Semantic feature theory suggests that a word meaning is instantiated by weighting its semantic attributes according to the context. (Barclay et al., 1974; Hampton, 1997; Kiefer & Pulvermüller 2012; Medin & Shoben, 1988; Mitchell & Lapata, 2010; Murphy, 1990; Wisniewsky, 1998). For example, when people think of the word *football*, they heavily weigh features like 'shape' and 'lower limbs' and features like 'smell' and 'size' lightly. In contrast, when they think of *forest*, the weighing on those features is likely to reverse. However, when the words appear in the context of a sentence such as *The team lost the football in the forest*, the context might bring up more unusual features like

[4] The content of this chapter was previously presented at the 6th Workshop on Cognitive Aspects of the Lexicon (Aguirre-Celis & Miikulainen, 2020). Aguirre-Celis worked on experimental design, implementation and analysis; while Miikkulainen provided guidance and feedback through discussions.

‘Landmark’, ‘Fearful’, and ‘Surprise’. Thus, when words share features, those aspects of the word representation that are relevant to the context are strengthened (Hampton, 1997; Kiefer & Pulvermüller 2012; Medin & Shoben, 1988; Mitchell & Lapata, 2010; Murphy, 1990; Wisniewsky, 1998).

If this theory is correct, it should be possible to see such changes in the fMRI patterns of subjects that are reading words in different contexts. The contextual effect of the words semantic encoding was demonstrated in earlier chapters. In Chapter 4, interesting context effects were observed for different shades of meaning (*dangerous flood* vs. *dangerous criminal*). In fact, the CEREBRA model captured the context of the sentence combining the meaning of the individual words. In Chapter 5, CEREBRA model was able to identify the effect of similar context on different concepts across subjects (*boat crossed* vs. *car crossed*), as well as the effect of different contexts on same concept (*bird flew* vs. *plane flew)*. In Chapter 6, the effect was quantified across a large corpus of sentences. It was demonstrated that the meaning of the sentence context is transferred to a degree, to each word in the sentence, i.e., the new CARs were more similar to the other words in the sentence than to the original CARs. In this manner, contextual modulation was characterized by the CEREBRA mechanism. What remains to be shown is that these effects (changes) are actually meaningful to the subjects, i.e., that they are aware of them and agree on the predictions of the model. To that end, a human subject study is presented in this chapter. Subjects were given words in context and asked to evaluate possible changes.

The hypothesis is that sentence context influences the interpretation of target words by modifying some of their semantic attributes. Consequently, if this attribute changes under the context of a sentence, the fMRI images should embed those changes. By using the CEREBRA model, the generic representation of the word attributes should adjust to reflect those changes as well. The human judgements can thus be compared to those predicted by the CEREBRA model.

In the next sections, the methods and results of the human subject study are described, followed by the methods and results of the computational study. The methods and results of comparing the human judgements and the computational model predictions concludes the chapter.

## 7.2 MEASURING HUMAN STUDY

The purpose of the survey is to evaluate the computational model predictions addressing the central question: How does the meaning of a word change in different sentences? As stated in Section 2.5.3, different attributes of the target word are weighted differently depending on context. Thus, the model is used to determine how the generic meaning of a word would have to change in order to account for the context. Specifically, the survey was designed to characterize these changes by asking the subject directly: In this context, how does this attribute change?

### 7.2.1 Materials and Design

The survey design was based on the fMRI subject data and sentence collection, the CEREBRA predictions, and the CAR's literal descriptions. A script was implemented to select the most representative subjects, sentences, words, and word attributes. To make the questions more understandable for the participants, the original descriptions of the 66 attributes (Table 2.1) were rephrased in simple English.

The data from the aggregation analysis described in Chapter 6 was used as the starting point and filtered further to make it systematic and uniform. Only the centroid non-copula sentences, only three-word classes, and only the top 10 statistically significant attribute changes for the target words (classes) were used. The final stimuli that met these criteria consisted of 64 different sentences from the Glasgow corpora containing the Agent, Verb, and Patient/Object/Location/Event (POLE). They contained 123 words: 38 Agents, 39 Verbs, and 46 POLE words. Table 7.1 lists the sentences. It includes the original sentence number from the Glasgow collection, the sentence itself, and the number of times each particular sentence was selected by the script (as a result of different subjects). Table 7.2 lists the three classes, the word in alphabetical order, and the word number from the original collection. Red indicates words used in two different roles in various sentences (e.g., *scientist* as Agent or Patient).

The complete survey is an array of 24 questionnaires that include 15 sentences each. For each sentence, the survey measures 10 attribute changes for each target word. Overall, each questionnaire thus contains 150 evaluations. For example, a questionnaire might measure changes on 10 specific attributes such as 'is visible', 'living thing that moves', 'is identified by sound', 'has a distinctive taste', for a specific word class such as *politician*, for 15 sentences such as *The politician celebrated at the hotel*. An example sentence questionnaire is shown in Figure 7.1.

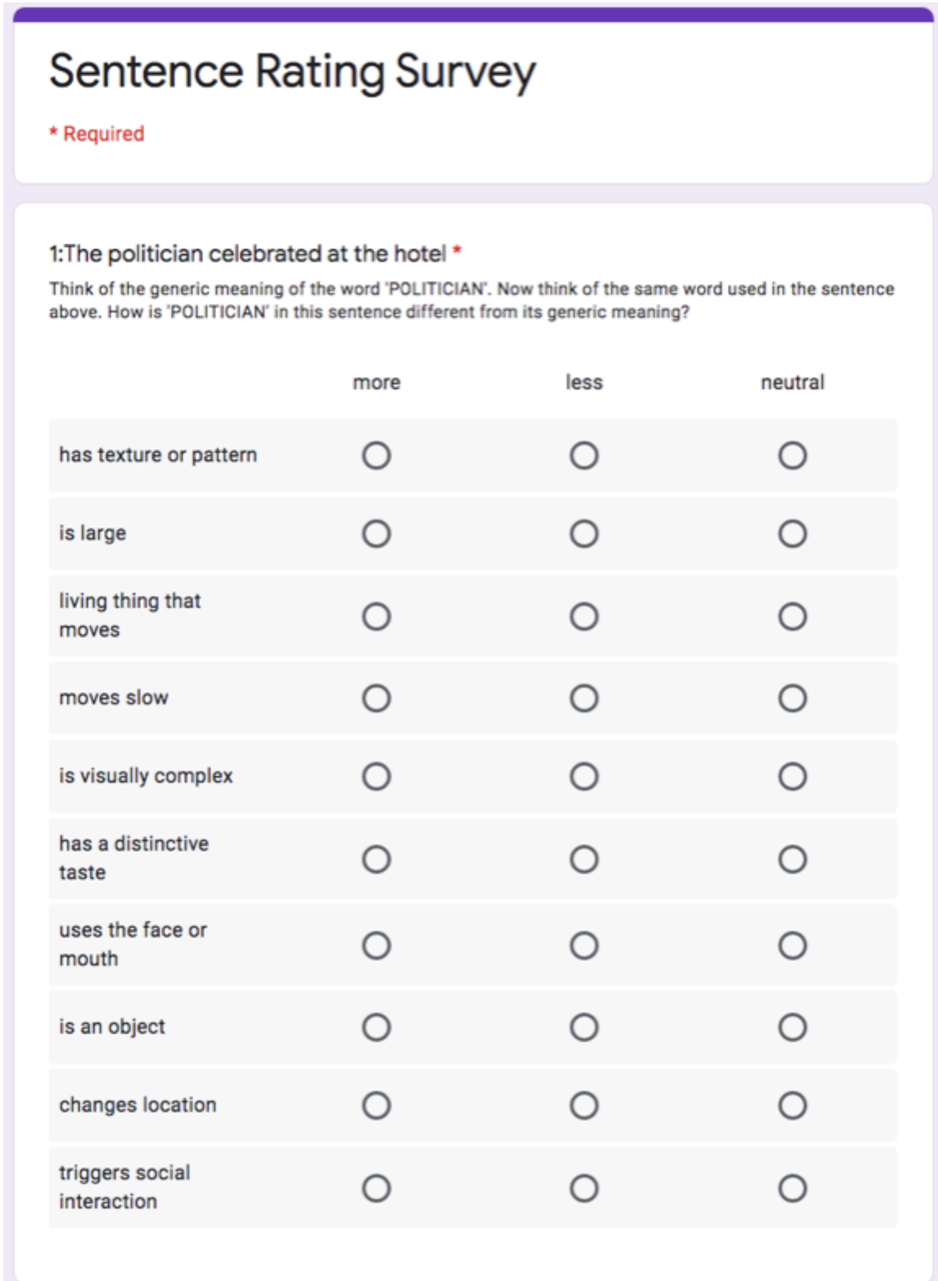
One way to select which attributes to test is to find those where the CEREBRA changes are the largest, for each subject, sentence, and role. Initially that choice appeared appropriate, however, the selection seemed incoherent for many sentences. For example, one attribute selected in this manner for the target word *played* in *The girl played in the forest* was 'has distinctive taste'. In this context, it did not make sense to measure the change in 'taste', even though in principle it is possible under feature sharing in concept combination and contextual modulation (*girl* has distinctive taste and *forest* has distinctive taste to a smaller degree, therefore, 'taste' might be transferred to the verb *played*). But such transfer seems to be taking it too far.

Table 7.1: The 64 different sentences used in the questionnaires. Shown are the original sentence number, the sentence itself, and the number of times each sentence was included in the survey (e.g., 120 total). In addition, the Subject, Verb, and Object were evaluated for each sentence, therefore the entire survey consisted of 360 sentences.

| Questionnaries Unique Sentences | | | | | |
|---|---|---|---|---|---|
| **No.** | **Sentence** | **Occ** | **No.** | **Sentence** | **Occ** |
| 10 | *The parent bought the magazine* | 3 | 113 | *The author kicked the desk* | 4 |
| 12 | *The couple planned the vacation* | 2 | 116 | *The injured horse slept at night* | 1 |
| 16 | *The wealthy couple left the theater* | 1 | 117 | *The soldier arrested the injured activist* | 2 |
| 17 | *The child broke the glass in the restaurant* | 3 | 119 | *The doctor bought the used boat* | 1 |
| 18 | *The happy child found the dime* | 2 | 123 | *The witness went to the trial* | 2 |
| 21 | *The angry child threw the book* | 2 | 131 | *The angry lawyer left the office* | 1 |
| 28 | *The judge met the mayor* | 4 | 144 | *The duck lived at the lake* | 1 |
| 39 | *The scientist spoke to the student* | 1 | 149 | *The banker watched the peaceful protest* | 1 |
| 40 | *The engineer gave a book to the student* | 4 | 150 | *The mob approached the embassy* | 2 |
| 43 | *The yellow dog approached the friendly teacher* | 1 | 154 | *The politician celebrated at the hotel* | 2 |
| 48 | *The mayor dropped the glass* | 3 | 157 | *The victim feared the criminal* | 1 |
| 54 | *The girl saw the small bird* | 1 | 168 | *The famous diplomat left the hospital* | 5 |
| 56 | *The mouse ran into the forest* | 1 | 170 | *The banker bought the expensive boat* | 2 |
| 57 | *The boat crossed the small lake* | 5 | 174 | *The farmer liked soccer* | 2 |
| 59 | *The judge lost the dime* | 1 | 177 | *The wealthy author walked into the office* | 2 |
| 61 | *The boy kicked the stone along the street* | 1 | 180 | *The tourist hiked through the forest* | 2 |
| 67 | *The woman bought medicine at the store* | 1 | 187 | *The tree grew in the park* | 1 |
| 73 | *The witness spoke to the lawyer* | 1 | 188 | *The commander ate chicken at dinner* | 4 |
| 75 | *The reporter spoke to the loud mob* | 2 | 195 | *The reporter ate at the new restaurant* | 3 |
| 77 | *The author interviewed the scientist after the flood* | 1 | 197 | *The boy held the football* | 1 |
| 78 | *The commander negotiated with the council* | 1 | 200 | *The yellow bird flew over the field* | 2 |
| 81 | *The reporter interviewed the dangerous terrorist* | 3 | 202 | *The green duck slept under the tree* | 1 |
| 82 | *The policeman interviewed the young victim* | 1 | 203 | *The girl saw a horse in the park* | 3 |
| 83 | *The mayor negotiated with the mob* | 2 | 205 | *The man lost the ticket to soccer* | 4 |
| 88 | *The activist listened to the tired victim* | 1 | 215 | *The tourist found a bird in the theater* | 1 |
| 89 | *The mayor listened to the voter* | 1 | 216 | *The old farmer ate at the expensive hotel* | 2 |
| 90 | *The jury listened to the famous businessman* | 2 | 226 | *The policeman read the newspaper* | 1 |
| 91 | *The woman helped the sick tourist* | 3 | 227 | *The criminal put the book on the desk* | 1 |
| 93 | *The soldier delivered the medicine during the flood* | 1 | 229 | *The happy girl played in the forest* | 3 |
| 108 | *The mob damaged the hotel* | 1 | 230 | *The young girl played soccer* | 1 |
| 109 | *The flood damaged the hospital* | 2 | 231 | *The old man threw the stone into the lake* | 1 |
| 111 | *The soldier kicked the door* | 1 | 235 | *The guard slept near the door* | 1 |

Table 7.2: The 123 words used in the questionnaires divided into 38 Agent, 39 Verb, and 46 Patient/Object/Location/Event (POLE) classes. Shown in alphabetical order. Words shown in red appeared in two different roles in separate sentences.

| Questionnaries Unique Words | | | | | |
|---|---|---|---|---|---|
| **No.** | **Agent** | **No.** | **Verb** | **No.** | **POLE** |
| 2 | *activist* | 8 | *approached* | 2 | *activist* |
| 13 | *author* | 10 | *arrested* | 21 | *bird* |
| 15 | *banker* | 12 | *ate* | 25 | *boat* |
| 21 | *bird* | 27 | *bought* | 26 | *book* |
| 25 | *boat* | 31 | *broke* | 33 | *businessman* |
| 28 | *boy* | 38 | *celebrated* | 41 | *chicken* |
| 42 | *child* | 56 | *crossed* | 52 | *council* |
| 48 | *commander* | 57 | *damaged* | 55 | *criminal* |
| 53 | *couple* | 62 | *delivered* | 63 | *desk* |
| 55 | *criminal* | 74 | *dropped* | 65 | *dime* |
| 67 | *diplomat* | 88 | *feared* | 70 | *door* |
| 68 | *doctor* | 94 | *flew* | 80 | *embassy* |
| 69 | *dog* | 99 | *found* | 92 | *field* |
| 75 | *duck* | 101 | *gave* | 95 | *flood* |
| 83 | *engineer* | 105 | *grew* | 97 | *football* |
| 87 | *farmer* | 109 | *held* | 98 | *forest* |
| 95 | *flood* | 110 | *helped* | 103 | *glass* |
| 102 | *girl* | 112 | *hiked* | 113 | *horse* |
| 106 | *guard* | 119 | *interviewed* | 114 | *hospital* |
| 113 | *horse* | 124 | *kicked* | 116 | *hotel* |
| 122 | *judge* | 130 | *left* | 126 | *lake* |
| 123 | *jury* | 131 | *liked* | 129 | *lawyer* |
| 129 | *lawyer* | 132 | *listened* | 137 | *magazine* |
| 138 | *man* | 133 | *lived* | 140 | *mayor* |
| 140 | *mayor* | 135 | *lost* | 141 | *medicine* |
| 144 | *mob* | 142 | *met* | 144 | *mob* |
| 147 | *mouse* | 148 | *negotiated* | 150 | *newspaper* |
| 155 | *parent* | 163 | *planned* | 151 | *night* |
| 165 | *policeman* | 164 | *played* | 152 | *office* |
| 166 | *politician* | 171 | *put* | 156 | *park* |
| 175 | *reporter* | 172 | *ran* | 170 | *protest* |
| 180 | *scientist* | 173 | *read* | 176 | *restaurant* |
| 188 | *soldier* | 178 | *saw* | 180 | *scientist* |
| 213 | *tourist* | 184 | *slept* | 186 | *soccer* |
| 214 | *tree* | 190 | *spoke* | 194 | *stone* |
| 219 | *victim* | 209 | *threw* | 198 | *student* |
| 231 | *witness* | 222 | *walked* | 204 | *teacher* |
| 232 | *woman* | 224 | *watched* | 207 | *terrorist* |
| | | 227 | *went* | 208 | *theater* |
| | | | | 210 | *ticket* |
| | | | | 213 | *tourist* |
| | | | | 214 | *tree* |
| | | | | 215 | *trial* |
| | | | | 218 | *vacation* |
| | | | | 219 | *victim* |
| | | | | 221 | *voter* |



# Sentence Rating Survey

* Required

1:The politician celebrated at the hotel *

Think of the generic meaning of the word 'POLITICIAN'. Now think of the same word used in the sentence above. How is 'POLITICIAN' in this sentence different from its generic meaning?

| | more | less | neutral |
|---|---|---|---|
| has texture or pattern | ○ | ○ | ○ |
| is large | ○ | ○ | ○ |
| living thing that moves | ○ | ○ | ○ |
| moves slow | ○ | ○ | ○ |
| is visually complex | ○ | ○ | ○ |
| has a distinctive taste | ○ | ○ | ○ |
| uses the face or mouth | ○ | ○ | ○ |
| is an object | ○ | ○ | ○ |
| changes location | ○ | ○ | ○ |
| triggers social interaction | ○ | ○ | ○ |

Figure 7.1: Example sentence in a questionnaire prepared to evaluate the computational model results. The sentence is *The politician celebrated at the hotel*, the target word is *politician* in the role of Agent. Ten different attribute changes are measured by selecting whether the attribute increased ("more"), decreased ("less") or remained "neutral". The human judgements were thus matched with those predicted by the CEREBRA model trained with the fMRI data.

An alternative and better selection was to: (1) use the sentences with at least 10 statistically significant attribute changes (ssa), (2) from the 25 attributes with the largest change, (or the number of ssa available), randomly select 10 within a sentence, (3) organize the attribute collection for each question using Binder's (Binder, 2016), original list arrangement.

The statistically significant attribute changes thus selected represent meaningful differences between the new and the original CARs. The point of the random selection within the top 25 was that: (1) there is a large number of potentially meaningful attributes, i.e., 25 at least, (2) for simplicity, the survey must not contain many questions, (3) the differences among the top 25 are not very large, and (4) it is necessary to get a varied selection of attributes. Choosing the top 10 instead would have resulted in too many visual features for most sentences, either because they frequently changed more, or because there is a significant number of visual attributes (i.e., 15 out of the 66). Two questionnaire examples and the link to find the complete set of 24 questionnaires are included in Appendix C.

The original descriptions of the CARs (i.e., the word attributes) were rephrased to make the questionnaires short and easy to read and to respond to, while retaining the meaning of the original descriptions elaborated by Binder et al. (Section 2.5.1). Table 7.3 lists these new descriptions.

### 7.2.2 Participants

Human judgements were crowdsourced using Google Forms in accordance with the University of Texas at Austin Institutional Review Board (2018-08-0114). The experiments were completed by 27 unpaid volunteers (nine females). The participants' ages ranged from 18 to 64 years, with the mean of 33. Nineteen of them were self-reported bilinguals (English as a second language) and eight English native speakers. Four subjects were affiliated with University of Texas at Austin; the rest of the population consisted of working people residing in different parts of north and central America (Texas, Seattle, California, Costa Rica, and Mexico). The subjects had no background in linguistics, psychology, or neurosciences.

Table 7.3: A rephrased description of the 66 attributes organized by domain, in order to maintain the questions' description short and easy to read and to respond to.

| BRAIN SYSTEMS | FEATURES | EXPLANATION | BRAIN SYSTEMS | FEATURES | EXPLANATION |
|---|---|---|---|---|---|
| VISION | Vision | is visible | SPAT | Landmark | has a fixed location |
| | Bright | is bright | | Path | changes location |
| | Dark | is dark | | Scene | is a particular setting |
| | Color | has a defining color | | Near | is near |
| | Pattern | has texture or pattern | | Toward | comes close |
| | Large | is large | | Away | goes away |
| | Small | is small | EVENT | Number | is countable |
| | Motion | moves | | Time | is an event in time |
| | Biomotion | living thing that moves | | Duration | has a certain duration |
| | Fast | moves fast | | Long | lasts a long time |
| | Slow | moves slow | | Short | lasts a short time |
| | Shape | has a defining shape | | Caused | caused by something |
| | Complexity | is visually complex | | Consequential | has consequences |
| | Face | has a face | | Social | triggers social interaction |
| | Body | has body parts | COG | Human | has intentions/plans/goals |
| SOMS | Touch | is identified by touch | | Communication | exchanges information |
| | Temperature | feels hot | | Self | relates to oneself |
| | Texture | feels smooth | | Cognition | increases mental activity |
| | Weight | is heavy | EVAL | Benefit | is beneficial |
| | Pain | is associated with pain | | Harm | is harmful |
| AUDIT | Audition | is audible | | Pleasant | is pleasant |
| | Loud | is loud | | Unpleasant | is unpleasant |
| | Low | makes a low pitch | EMOT | Happy | is happy |
| | High | makes a high pitch | | Sad | is sad |
| | Sound | is identified by sound | | Angry | causes anger |
| | Music | makes a musical sound | | Disgusted | is disgusting |
| | Speech | talks | | Fearful | causes fear |
| GS | Taste | has a distinctive taste | | Surprised | causes surprise |
| | Smell | has a distinctive smell | DR | Drive | causes to act in certain way |
| MOTOR | Head | uses the face or mouth | | Needs | is an essential part |
| | UpperLimb | uses the arm or hand | ATT | Attention | is a center of attention |
| | LowerLimb | uses the leg or foot | | Arousal | increases alertness |
| | Manipulation | is physically manipulated | | | |
| | Object | is an object | | | |

### 7.2.3 Procedures

The twenty-four questionnaires were designed using Google Forms. The respondents were asked to think how the meaning of a specific word changes within the context of a sentence compared to its generic meaning, by evaluating which word attributes change “more”, “less”, or stay the same.

Subjects were recruited by sending emails or text messages directly along with the survey link to access their assigned questionnaire. The data collection was done online, and the participants responded using their cell phone or personal computer. They did not need to login or have a Google account in order to participate. Each questionnaire consists of an Introduction, a Description of the Experiment, an Example, and the Survey. If the participant chose to be part of the experiment, she/he press the NEXT button from the Example section to the Survey section, then, they press the SEND button when they were done. Each questionnaire takes 15 minutes to complete.

Three of the participants responded to all of the 24 questionnaires. The entire survey consisted of a total of 3600 questions, so it took them four to seven days to complete this task at a pace of approximately four questionnaires an hour per day. The responses were considered correct whenever two out of the three participants agreed on their answers. Unfortunately, human raters do not often agree, and their judgement was influenced by their own perception or uncertainty. With three possible answers and three participants, the results were still inconclusive. This task was a lot of work, the fourth set of responses was obtained by distributing it among multiple raters: twenty-four additional participants were recruited to each respond to one of the 24 questionnaires.

### 7.2.4 Results

As stated before, the hypothesis was that sentence context influences the interpretation of target words by modifying some of their attributes. The survey directly asks for the direction of change in a particular word attribute (i.e., CAR dimension), in a particular word, in a particular sentence, compared to the words generic meaning. In other words, the responses directly measure the changes that CEREBRA derived from the fMRI. These changes can therefore be used to find the level of agreement between humans and CEREBRA.

Human responses were first characterized through data distribution analysis. Table 7.4 shows the number of answers “less” (-1), “neutral” (0), and “more” (1) for each respondent. Columns labeled P1, P2, and P3, show the responses of the three participants that were assigned the entire survey (24 questionnaires, 3600 answers). Column labeled P4 shows the combined answers of the 24 different participants responding to one questionnaire each. The top part of the table shows the distribution of the rater’s responses and the bottom part the level of agreement among them. As can be seen, the participants agreed only 47% of the time.

When inter-subject reliability is too low, it is not worth comparing system predictions vs. human judgements (Grand et al., 2018). However, since there were a lot of questions, it was possible to include only questions that were the most reliable, i.e., where all subjects agreed. There were 631 such questions or 18% of the total set of questions.

Table 7.4: Distribution analysis and inter-rater agreement. Top table shows human judgement distribution for the three possible questionnaire responses "less" (-1), "neutral" (0), and "more" (1). Bottom table shows percent agreement for the 4 raters. The task was difficult and the agreement low. Only those questions where all participants agreed were considered reliable and compared to the CEREBRA model.

| HUMAN RESPONSES DISTRIBUTION | | | | | | |
|---|---|---|---|---|---|---|
| **Resp/Part** | **P1** | **P2** | **P3** | **P4** | **AVG** | **%** |
| -1 | 2065 | 995 | 645 | 1185 | 1223 | 34.0% |
| 0 | 149 | 1120 | 1895 | 1270 | 1109 | 30.8% |
| 1 | 1386 | 1485 | 1060 | 1145 | 1269 | 35.3% |
| TOT | 3600 | 3600 | 3600 | 3600 | 3600 | 100% |

| PARTICIPANT AGREEMENT ANALYSIS | | | | | | |
|---|---|---|---|---|---|---|
| | **P1** | **P2** | **P3** | **P4** | **AVERAGE** | **%** |
| **P1** | 0 | 1726 | 1308 | 1650 | 1561 | 43% |
| **P2** | 1726 | 0 | 1944 | 1758 | 1809 | 50% |
| **P3** | 1308 | 1944 | 0 | 1741 | 1664 | 46% |
| **P4** | 1650 | 1758 | 1741 | 0 | 1716 | 48% |
| | | | | | | |
| | | | | **TOTAL** | **6751** | |
| | | | | **AVG xPAR** | **1688** | |
| | | | | | | |
| | | **AVERAGE** | Particip match each other | | | 47% |

## 7.3 Measuring Model Predictions

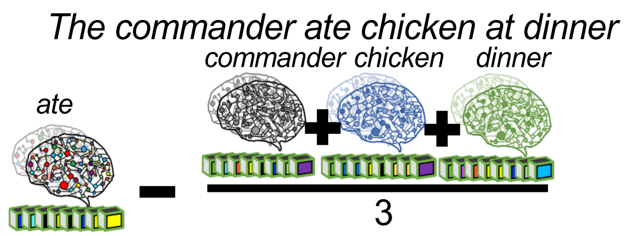
Three different approaches were designed to quantify the predictions of the CEREBRA model. Likewise, a model fitting procedure was implemented in order to measure the level of agreement between humans and CEREBRA.

### 7.3.1 Quantifying the CEREBRA Results

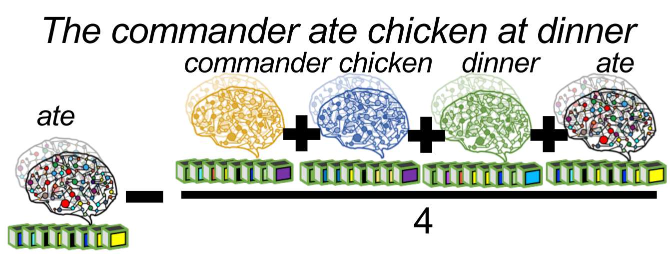
The survey directly asks for the direction of change of a specific word attribute in a particular sentence, compared to a generic meaning. Since the changes in the CEREBRA model range within (-1, 1), in principle that is precisely what CEREBRA produces. However, many times during the progress of this research, it was found that some word attributes always change (increase or decrease) and do so more in some contexts than others. This effect is well known in conceptual combination (Hampton, 1997; Wisniewsky, 1998), contextual modulation (Barclay, 1974), or what Medin and Shoben (1988) refer to as word attribute centrality: the same property is true for two different concepts but may be more central to one than to the other (e.g., it is more important for boomerangs to be curved than for bananas).

The direction of change in CEREBRA is therefore not a good predictor of human responses; instead, these changes need to be measured relative to changes in other words. Thus, the problem was addressed by three different approaches:

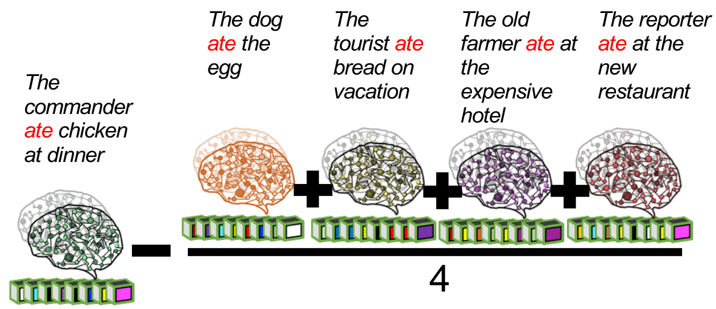
1) What is the effect of the rest of the sentence in the target word? This effect was measured by computing the average of the CEREBRA changes (i.e., new-original) of the other words in the sentence, and subtracting that average change from the change of the target word:




2) What is the effect of the entire sentence in the target word? This effect was measured by computing the average of the CEREBRA changes (i.e., new-original) of all the words in the sentence including the target word, and subtracting that average change from the change of the target word:



3) What is the effect of CARs used in context as opposed to CARs used in isolation? This effect was measured by computing the average of the CEREBRA changes (i.e., new-original) of the different representations of the same word in several contexts, and subtracting that average change from the change of the target word:




Although the first two approaches seem simple, a "bag-of-words" representation by averaging all the content words in the sentence, similar construction methods have proven to be effective in neural activity prediction (Anderson et al., 2016; Binder et al., 2016a, 2016b; Just et al., 2017) and computational linguistics (Baroni et al., 2010; Burgess, 1998; Landauer et al., 1997; Mitchell & Lapata, 2010). However, the third approach is motivated by neurological evidence suggesting that sentence comprehension involves a common core representation of multiple word meanings encoded into a network of regions distributed across the brain (Anderson et al., 2016; Gennari et al., 2007). In line with this view, a generic (or isolated) word representation can be formed by averaging the activity in multiple sentence contexts.

In each of these cases, the resulting vectors are expected to accurately represent the direction of change asked in the questionnaires. They are the ratings used in the evaluation procedure described in the following section.

### 7.3.2 Procedure

Starting from a different random seed, the CEREBRA model was trained 20 times for each of the eight best fMRI subjects (i.e., where the fMRI data in general was most consistent). Responses for each model where thus obtained for the 631 questions where all participants agreed. In order to demonstrate that the CEREBRA model has captured human performance, the agreements of the CEREBRA changes and human surveys need to be at least above chance. Therefore, a baseline model that generated random responses from the distribution of human responses was created. The chance model was queried 20 times for each of the 631 questions, for each of the eight subjects. In this manner, 20 means and variances for each of the eight subjects for both CEREBRA and chance were created.

To estimate the level of agreement of CEREBRA and chance models with humans, a single parameter in each model was fit to human data: the boundary value above which the change was taken to be an increase (i.e., "more") or decrease/no change (i.e., "less"/"neutral"). The "less" and "neutral" categories were combined because they were much smaller than the "more" category in human data. The optimal value for this parameter was found by simply sweeping through the range (-1..1) and finding the value that measured on the highest number of matching responses where the 631 questions are.

In the next section the experimental results between humans, the computational model, and the baseline (chance) are presented.

## 7.4 MATCHING PREDICTIONS WITH HUMAN JUDGEMENTS

The three approaches to measuring the predictions of the CEREBRA model, i.e., the context effect of the rest of the sentence, the context effect of the entire sentence, and the context effect of the word in different contexts, were implemented and fit to human data using single-boundary model fitting. The first two approaches produced very similar results. In fact, they achieved slightly better results than the third one (by about 2%). All three approaches are reported in this chapter.

The match results are presented in Table 7.5 and the statistical significance in Table 7.6. The CEREBRA change model for approaches one and two matches human responses in 77% of the questions, the CEREBRA change model for approach three matches human responses in 75% of the questions while the chance level is 68% - which is equivalent to always guessing “more”, i.e., the largest category of human responses. The differences shown in Table 7.6 are highly statistically significant for all of the eight subjects in the three approaches shown. These results indicate that the changes in word meanings due to sentence context that are observed in the fMRI and interpreted through semantic feature representations are real and meaningful to the subjects. More comprehensive analyses between humans and the models such as the analysis of the responses where three out of four participants agreed, are found in Appendix D.

Table 7.5: Matching CEREBRA predictions with human data (approaches one to three), compared to chance. The table shows the average agreement of the 20 repetitions across all subjects. CEREBRA approaches one and two agree with human responses 77%, CEREBRA approach three agrees 75%, while the chance level is 68%. Comparison agreement with human judgements.

| PARTICIPANTS AVERAGE AGREEMENT | | | | | |
|---|---|---|---|---|---|
| RATINGS | HUMAN | CEREBRA1 | CEREBRA2 | CEREBRA3 | CHANCE |
| -1/0 | 205 | 145 | 149 | 134 | 1 |
| 1 | 426 | 341 | 336 | 339 | 426 |
| TOTAL | 631 | 486 | 485 | 473 | 427 |
| AVERAGE | | 77% | 77% | 75% | 68% |

Table 7.6: Statistical analyses for CEREBRA approaches one to three, and chance. The table shows the means and variances of the CEREBRA change models and the chance model for each subject, and the p-values of the t-test, revealing that the differences are highly significant. Thus, the context-dependent changes are actionable knowledge that can be used to predict human judgements.

| SUBJECTS | CHANCE | | CEREBRA #1 | | CEREBRA #2 | | CEREBRA #3 | | p-value | p-value | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | MEAN | VAR | MEAN | VAR | MEAN | VAR | MEAN | VAR | CEREBRA #1 | CEREBRA #2 | CEREBRA #3 |
| S5051 | 427 | 0.91 | 486 | 46.74 | 486 | 56.42 | 466 | 152.98 | 5.42E-32 | 1.66E-30 | 1.17E-16 |
| S9322 | 427 | 1.10 | 481 | 32.62 | 480 | 21.54 | 466 | 105.61 | 1.67E-33 | 2.02E-36 | 2.30E-19 |
| S9362 | 426 | 0.57 | 486 | 42.58 | 485 | 37.85 | 480 | 39.29 | 6.50E-33 | 1.65E-33 | 6.22E-32 |
| S9655 | 427 | 1.69 | 486 | 21.95 | 486 | 27.73 | 481 | 32.62 | 1.46E-37 | 6.25E-36 | 2.55E-33 |
| S9701 | 427 | 1.71 | 490 | 57.00 | 488 | 57.09 | 470 | 89.12 | 3.80E-31 | 7.56E-31 | 8.82E-22 |
| S9726 | 427 | 2.87 | 486 | 44.06 | 484 | 34.04 | 469 | 80.66 | 6.59E-32 | 3.17E-33 | 6.29E-22 |
| S9742 | 427 | 2.77 | 489 | 24.77 | 489 | 21.21 | 483 | 54.05 | 3.09E-37 | 2.93E-38 | 1.62E-29 |
| S9780 | 427 | 1.67 | 480 | 75.78 | 480 | 54.22 | 471 | 92.68 | 1.82E-26 | 4.62E-29 | 5.56E-22 |

## 7.5 DISCUSSION AND FUTURE WORK

The study provides a missing piece on the theory of semantic feature representations: The context-dependent changes in them are actionable knowledge that can be used to predict human judgements. Given how noisy human responses data is, the 7% or 9% differences between the CEREBRA approaches and Chance is a strong result.

An interesting direction for future work would be to replicate the study on a more extensive data set with a fully balanced stimuli and with fMRI images of individual words. The differences should be even stronger and should be possible to uncover even more refined effects. Such data should also improve the survey, since it would be possible to identify questions where the effects can be expected to be more reliable. Inter-raters' reliability could also be improved by training the raters better so that they are comfortable with the concept of generic meaning and the concept of variable meanings. It may also be possible to design the questions such that they allow comparing alternatives which may be easier for the participants.

## CONCLUSION

The hypothesis was tested, the results suggest that the three different approaches were consistent to human judgements in average 76%. This consistency confirms three significant findings: (1) context-dependent meaning representations are embedded in the fMRI sentences, and (2) they can be characterized. Using brain-based semantic representations (CARs) together with the CEREBRA change model (3) such changes are real and meaningful to the subjects. It therefore takes a step towards understanding how the brain constructs sentence-level meanings from word-level attributes.

# Chapter 8

# Discussion and Future Work

This chapter reviews some of the key issues that emerged through the development of this dissertation. First, evaluating the soundness of the CAR theory as a measure of semantics is discussed. Second, limitations of the CEREBRA model are reviewed. Third, the semantic representations between CEREBRA and Word2vec are analyzed. Fourth, Understanding individual and cultural differences with CEREBRA is described, Fifth a potential application of an NLP system that includes CEREBRA's context-based representations is outlined. Sixth, a possible extension of CEREBRA is proposed by incorporating text-based representations to enhance the interpretation of the fMRI patterns.

## 8.1 Evaluating Soundness of CAR Theory

Research in semantic representation seeks answers to questions like: How are concepts represented in the brain? How is word meaning represented? How do word meanings change during concept combination or under the context of a sentence? What tools and approaches contribute to quantifying such meaning representation changes? The CEREBRA model (i.e., a tool) and the CAR theory (i.e., an approach) were combined to address these questions and to evaluate the soundness of CAR theory.

CAR theory has already been validated in many studies (Anderson 2016, 2018; Binder, 2016b; Fernandino, 2015), so this research took it as a starting point in building CEREBRA. However, whereas the original CAR concerns static representations, CEREBRA extends it to dynamic representations, and shows how they can modified based on context. Essentially, CEREBRA was able to capture dynamic meanings of words by tracking how the CARs feature-weightings changed across sentences. It demonstrated that (1) context-dependent meaning representations are embedded in the sentence fMRI, and (2) CAR theory is well-grounded and can be used as the foundation for modeling neural representations of word meaning. Thus, the computational and behavioral studies presented in this dissertation provided useful insight into the central issues in semantic representation listed above.

The first set of CEREBRA experiments found significant context effects for different shades of meaning for individual subjects. Different types of conceptual combinations were studied systematically, and the changes were found to be statistically significant across the entire corpus, and actionable knowledge across human judgements. Interestingly, some of the experiments showed that some combination of concepts have context-based changes that could not have been derived from the concept's features alone (i.e., *celebrated*), but only from external knowledge to form thematic relations. Then, how could CEREBRA find such thematic associations? On the one hand, it was expected that the fMRI images embody such knowledge, and CEREBRA verified it by systematically changing the CARs. On the other hand, many abstract words (i.e., *happy*, *survived*, *injured*, *dangerous*, *friendly*, *celebrated*, *laughed*, *feared*) contained affective connotations in the corpus; CEREBRA was able to encode different contexts for these words, confirming that abstract concepts are grounded on affective experiences (Vigliocco et al., 2014). However, CEREBRA is not well suited to capturing the different meanings of words with a higher level of abstraction (e.g., *fiction*, *interest*, *purpose*), which are described in terms of their relationships to other words. Such words lack physical referents and in many cases an emotion or an internal state to which their meaning can be grounded. Essentially, CEREBRA was able to address the differences on meaning for words that can be grounded in the twelve brain systems outlined in the original CARs, but for words with no relation to the physical world, in order to be able to disambiguate their different meanings, it would require additional linguistic knowledge; that is, a full account of ambiguity requires a hybrid model of linguistic and experiential knowledge (as will be discussed in Sections 8.4 and 8.5).

## 8.2 Limitations of the CEREBRA Model

The CEREBRA model generates good interpretations of word meanings considering that the dataset was limited and was not originally designed to address the dynamic effects in meaning. In future work, it would be interesting to replicate the studies on a more extensive data set. A fully balanced stimuli including sentences with identical contexts (e.g., *The yellow bird flew over the field* vs. *The yellow plane flew over the field*) and contrasting contexts (e.g., *The vicious dog chased the boy* vs. *The friendly dog chased the boy*), could help test the hypothesis more systematically. The context-based changes should be even stronger, and it should be possible to uncover more refined effects.

Similarly, it would be desirable to extend the fMRI data with images of individual words. The current approach of synthetizing words (SynthWords) is an approximation often used in computational linguistic (Baroni et al., 2010; Burgess, 1998; Landauer et al., 1997; Mitchell & Lapata, 2010) and neural activity prediction research (Anderson et al., 2016; Binder et al., 2016a, 2016b; Just et al., 2017). The CEREBRA process of mapping semantic CARs to SynthWords and further to sentence fMRI refines the synthetic representations by removing noise. However, such representations blend together the meanings of many words in many sentences. Thus, by acquiring actual word fMRI, the observed effects should become even more clear.

Also, CEREBRA does not consider word order in assembling sentences. Instead, it averages all words that occur in the sentence and uses the representation to find the difference between the predicted (SynthSent) and observed fMRI (fMRISent). Although the approach seems simple, similar "bag-of-words" methods have proven effective in neural activity prediction (Anderson et al., 2016; Binder et al.,2016a, 2016b; Grand et al., 2018; Just et al., 2017) and computational linguistics (Baroni et al., 2010; Mitchell & Lapata, 2010).

Instead of averaging, it would be interesting to use case-role representations to organize the content words. Case-role slots could include Agent, Verb, Patient, Instrument, Location, Agent-Modifier, and Patient-Modifier. Each slot would be assigned a role, and for every sentence, the word representation placed in the correct slot according to the word's role. For example, for *The minister interviewed the injured policeman*, the role assignments could be Agent (*minister*), Verb (*interviewed*), Patient-Modifier (*injured*), and Patient (*policeman*). The case-role set must fit all types of sentences. One drawback is that it would produce a more high-dimensional representation of the sentence. Consequently, the original fMRISent voxel representation would need to be large enough in order to apply the pairwise comparison. Usually, the fMRI voxel representations run well over the thousand voxels which should be enough. In case there are not enough voxels, a dimensionality reduction approach (e.g., PCA) could be adopted before developing the case-role representations.

CEREBRA was trained using all content words in the sentence collection, however for the aggregation analysis only those sentences that included the following three content words were considered: Agent, Verb, and Patient/Object/Location/Event. Undoubtedly, with a broader data set, this analysis could be expanded. Likewise, the fMRI data was collected for eleven subjects, but uncontrolled situations (i.e., subjects not focused), and noise (i.e., head movement) precluded using four of them. Only S5051, S9322, S9362, S9655, S9701, S9726, S9742, S9780 were eventually used. All experiments could be expanded with a larger collection of fMRI subjects.

### 8.3 How do Distributional Semantic models map the semantic space of the Brain?

How distributional semantic representations compare to CARs in terms of mapping the semantic space of the brain? CAR theory enables direct correspondence between conceptual content and neural representations. Conceptual knowledge is distributed across a small set of modality-specific neural systems that are engaged when instances of the concept are experienced. In contrast, distributional semantic models (DSM) construct conceptual knowledge from text co-occurrence. They are not grounded on perception and motor mechanisms, instead their representations reflect the semantic knowledge acquired through a lifetime of linguistic experience. Brain mapping analysis can be performed by correlating the fMRI and the DSM word representations, on the whole brain and for each anatomical region (frontal, parietal, temporal and occipital lobes). Although such partitioning is very coarse, since each lobe is large and serves many different functions, each lobe has specializations that can be used for interpretation. DSMs based on co-occurrence have been successfully correlated to neural semantic representations in this manner (Anderson et al., 2013; Bruni et al., 2014; Mitchell, 2008; Mitchell & Lapata, 2010).

A different DSM based on prediction, Word2vec, has gained popularity due to its simplicity and remarkable performance on computational semantic tasks (Mikolov et al., 2013). It is based on the recurrent neural network proposed by Elman (1990), while using different training methods to scale up to extremely large corpora. Particularly, skip-gram model is one of the two Word2vec semantic models that has been widely used on neural activity prediction research (Anderson et al., 2019; Bulat et al., 2017; Ruan et al., 2016; Silberer et al., 2017).

Many studies agreed that skip-gram is moderately to strongly correlated with neural activations in the fMRI patterns. For instance, Ruan et al. (2016) found that skip-gram word representations are significantly correlated with the fMRI data of all brain lobes except the occipital lobe. Previous neuroscience research has revealed that the frontal, temporal, and parietal lobes play important roles in semantic cognition, such as high-level and abstract knowledge processing, integration of lexical information, speech comprehension, and knowledge retrieval. These findings confirm that the skip-gram model can partly account for the semantic processing in the cortex and contain little visual information about words.

In contrast, the richness and complexity of the representations in the CAR theory is based on a direct mapping between the conceptual content of a word and the corresponding neural representations. Distributing conceptual knowledge across modality-specific neural systems offers a powerful model to further explore the semantic space of the brain. To this end, experiments in this dissertation demonstrated that CARs can capture fine distinctions in meaning, therefore creating many possibilities of improvements of the theory itself.

### 8.4 Understanding Individual and Cultural Differences with CEREBRA

CEREBRA can be used to further understand how semantic knowledge is represented across cultures and individuals. Individual experiences are embedded into the fMRI patterns; thus CEREBRA could be used to study how each individual perceives the world (i.e., cultural differences). Note that such an analysis is not possible with the DSM representations obtained from averaging millions of documents from millions of individuals. Further, CEREBRA's representations could be used to analyze whether groups of individuals with a similar cultural background modify CARs the same way and differently from other cultural backgrounds.

For example, during the analysis of individual words Aguirre-Celis & Miikkulainen (2017) found that for the CEREBRA's context-modified representation of the word *mouse*, used in the sentence *The man saw the dead mouse*, some subjects showed salient activation on the emotional attribute Sad (because the mouse was dead), compared to other subjects. Thus, CEREBRA could be a useful tool to identify individualities across subjects and groups, and to find where those differences in conceptual content occur.

### 8.5 Building an NLP Application Using CEREBRA Representations

Language grounding refers to understanding the meaning of words as it applies to the physical world. It assumes that the perceiver is aware of the world, the context, and the communication techniques (e.g., oral, written, visual). To build real-world applications (i.e., Natural Language Processing systems) it is crucial to use conceptual grounding, and multimodal CEREBRA representations could be used to make such applications more robust.

Many studies (Anderson et al., 2019; Andrews et al., 2009; Martin, 2007) have emphasized that DSMs capture encyclopedic, functional and discourse-related properties of word meanings (e.g., a carrot is a root vegetable, usually orange, Dutch invented the orange carrots, it contains high carotene, human body turns carotene into vitamin A), but tend to miss their concrete aspects (e.g., a carrot refers to an object whose attributes describes it as orange, conical/cylindrical, juicy, crispy, sweet). Current text-based NLP applications could be improved by combining experiential and linguistic data. In this section a neural network model that learns to represent context simultaneously from both large corpora and the multimodal CEREBRA representations will be outlined. This model could be used as part of a natural language understanding system for service robot applications (i.e., Agriculture, Medicine, Security).

The proposed model expands that of Melamud et al. (2016). It consists of a recurrent neural network for learning generic sentence context representations (context2vec). Specifically, it learns from large text corpora of sentence contexts and target words through bidirectional LSTM. CEREBRA representations, which provide a different kind of context (i.e., experiential context), can be added to it to provide supplementary knowledge that should improve the natural language understanding process.

This model resembles the original FGREP mechanism (Miikkulainen et al., 1988) in regard to the learning process: it develops the target word and the sentence context representations at the same time. Context2vec uses a neural network based on word2vec's CBOW architecture (Mikolov et al., 2013). Melamud et al. (2016) replaces the context modeling of averaging word representations in a fixed window with a bidirectional LSTM. The context2vec architecture with the CEREBRA extension (red box) is outlined in Figure 8.1.

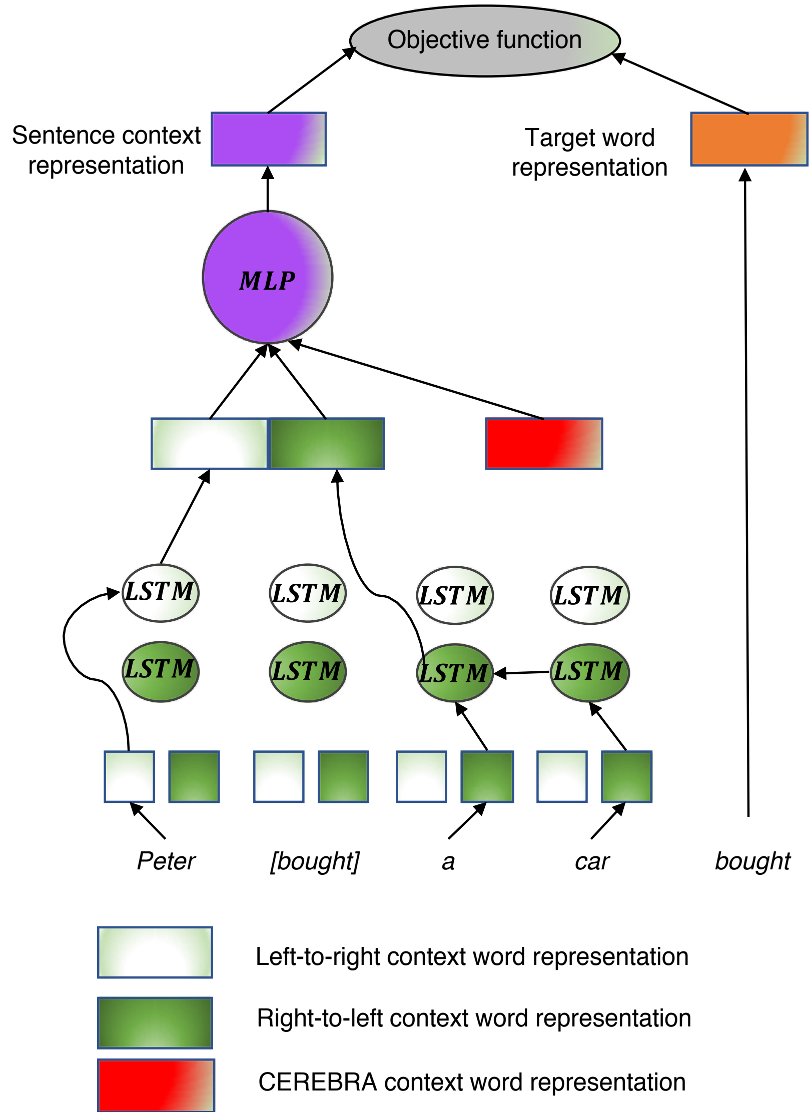



Figure 8.1: Context2vec neural network model enhanced with multimodal CEREBRA representations. Context representations for linguistic knowledge are developed using a bidirectional LSTM. One LSTM network reads words from left to right, and another from right to left. The outputs of the two LSTM networks are concatenated and fed into a multilayer perceptron (MLP) together with the CEREBRA vector. At the same time, the target word is represented with the same dimensionality as that of the sentence contexts. The output of this layer is the representation of the combined sentence context. Thus, the architecture learns generic context representations by integrating linguistic and experiential knowledge for NLP.

In a bidirectional LSTM, one LSTM network reads words from left to right, and another from right to left. The parameters of these two networks are independent, including the left-to-right and right-to-left context word representations. For example, to represent the context of a target word in the sentence *Peter [bought] a car*, the architecture concatenates the LSTM output vector representing its left-to-right context (*Peter*) with the one representing its right-to-left context (*a car*). This concatenated vector and the CEREBRA vector are fed into a multilayer perceptron (MLP) to learn dependencies between the different contexts. At the same time, the target word is represented with the same dimensionality as the sentence contexts. The output of this layer is thus the representation of the combined sentence context. To learn the parameters of the network, the word2vec's negative sampling objective function is used. A positive pair consists of a target word and its entire sentential context, and *k* negative pairs consist of random target words sampled from a (smoothed) unigram distribution over the vocabulary and paired with the same context (Melamud et al., 2016).

The difference between context2vec and the proposed extension is the incorporation of the CEREBRA representations. Although a simple modification, together the context2vec and CEREBRA representations make an important contribution: Not only their representations contain a full sentence context, but also, they bring experiential knowledge into the representations. This knowledge contrasts with other models such as word2vec that usually consider the representations of neighboring words with a strict window size. As a result, they carry only limited information regarding the relations between the target word and the entire sentence context, and use only the type of information related to DSMs.

In conclusion, the proposed model combines linguistic and experiential information at the contextual level. Thus, the CEREBRA representations would provide the experiential-based data (i.e., concrete words) and the text-based representations would provide the association-based data (i.e., abstract words), leading to a better performance. For instance, agricultural service robots with such representations would have the capability to understand natural language commands (i.e., watering plants), to have encyclopedic knowledge (i.e., to make decisions regarding weed or pest control), to ground language by adapting to the environment (i.e., object recognition and location to plant seeds, prune, or harvest), and by understanding novel concepts (i.e., "rain water"). For dynamic environments, such robots could accommodate an additional mechanism (e.g., a library of modified CARs) in an attempt to adapt to new contexts, by using the closest new CAR representation to deliver a possible solution. Next section presents a promising extension of the CEREBRA model inspired by this hybrid model application.

### 8.6 Integrating Text-based and Brain-based representations

Many semantic memory researchers would agree that most of the semantic knowledge comes from direct experience with objects and actions. However, a great deal of semantic knowledge is also acquired from spoken and written language. People have knowledge of exotic animals, facts, places, science fiction books, and historic events that they have never experienced. For example, they might know about polar bears, Apollo 11, the Great Wall of China, Frankenstein, and the Jewish Holocaust without ever seen or experienced them. Recent work on DSMs reported that these models successfully capture many aspects of human semantic memory. A new type of semantic models that integrate linguistic and multimodal knowledge suggest that fMRI activations are more correlated to the combined model than to each model in isolation (Anderson et al., 2019; Bulat et al., 2017; Ruan et al., 2016; Silberer et al., 2017).

A natural next step would be to adapt CEREBRA to use such combination of linguistic and multimodal approaches (i.e., text-based and brain-based representations) to explore whether similar effects or improved differences of meaning are observed. To modify CEREBRA in this manner, it is necessary to add two extra steps: (1) An additional neural network trained to map distributional semantic representations (text-based representations) of words (DSMWords) to SynthWord (fMRI synthetic words).
This network has the same role as the network that maps CARWords to SynthWords. As before, the SynthWords are combined into the predicted sentence (SyntSent), by averaging all words in the sentence. However, this step is performed for each kind of semantic input, and therefore there will be two predicted sentences (SynthSentCAR & SynthSentDSM). Each SynthSent is compared to the actual fMRISent (original fMRI data) to form two error signals. (2) These error signals, obtained independently from the two different networks, are averaged and used to backpropagate the proportional error accordingly. Subsequently, the trained neural networks are utilized to determine how the CARWords and the DSMWords should change in the context of the sentence. As before, the networks no longer change; the averaged error is used to change the CARWords and DSMWords through the FGREP method (Miikkulainen et al., 1988). These modifications can be carried out until the error goes to zero, or no additional change is possible (because the CAR attributes are already at their max or min limits). CAR limits control the stopping criteria. During the prediction process, both semantic inputs are changed to calculate the two errors, but in the end, only the revised CARWord would include the explainable feature representations of the word for each sentence. The DSWord representations could not be interpreted.

Essentially, the process consists of adding linguistic knowledge to help predict the fMRI sentences and consequently fully interpret the neural activation patterns associated with meaning. DSMs have been found to contribute linguistic abstract concepts to sentence interpretation; in particular the kind of abstract concepts that are not clearly associated with emotion or sensorimotor systems and are difficult to identify with the CAR theory, e.g., *purpose* (Anderson et al., 2017; 2019). The proposed modification to the CEREBRA model should thus improve interpretation of such concepts.

## CONCLUSION

This chapter reviewed the CEREBRA approach from the point of view of the central questions in semantic theory. The different experimental analyses of CEREBRA's context-based representations serve to assess the soundness of the CAR theory. The alternative skip-gram semantic model is limited in its mapping to brain semantic space, suggesting that CAR theory and CEREBRA together are a better approach to learn about concept representation in the brain. A computational model can be devised to use CEREBRA's context-based representations to enhance NLP applications. A further extension may include distributional semantic representations as part of the model inputs, resulting in a more comprehensive interpretation of the neural activation patterns associated with meaning.

# Chapter 9

# Conclusion

In this dissertation, the CEREBRA model was developed to characterize the effect of sentence context on word meanings. It uses CAR theory and FGREP mechanism to take a step forward in understanding how the brain constructs sentence-level meanings from word-level features. This chapter reviews the main contributions of this dissertation and concludes with a general perspective: How CEREBRA provides a framework for novel experiments and applications that could benefit future research on brain diseases (e.g., dementia), disorders (e.g., dyslexia), and head injuries, as well as to further explore how the brain functions.

## 9.1 CONTRIBUTIONS

Chapter 3 introduced a computational model to account for context effects in the brain. The chapter described the CEREBRA approach, grounded in CAR theory and implemented using neural networks with the FGREP mechanism. CAR theory assumes that context modifies the baseline meaning of a concept, and CEREBRA tested it, by characterizing how different parts of the concept attribute representation can be modified to account for changes in the neural activation patterns. Particularly, CEREBRA used a neural network to predict approximately what the sentence fMRI should be, based on the CARs, and then used FGREP to adjust the CARs so that the prediction became correct. The modified CARs then indicate how the meaning changed in context.

CEREBRA's capability to characterize the effect of sentence context on word meanings was demonstrated through several computational experiments, presented in Chapters 4 to 6, as well as a behavioral analysis in Chapter 7.

Chapter 4 analyzed context effects for different shades of meaning. Image (fMRI) data for individual subjects were analyzed in this chapter. Experiments showed that the approach is feasible, demonstrating meaningful differences for e.g., *human communication* vs. *noise from a machine*; *dangerous natural disaster* vs. *dangerous person*; *live animal* vs. *dead animal*. The results suggest that different aspects of word meaning are activated in different contexts, and it is possible to see those changes in the corresponding fMRI images using the CEREBRA model. The linear mapping approach (regression) was disorganized, but the nonlinear mapping (CEREBRA) identified the relevant changes.

Chapter 5 examined the different types of conceptual combinations and their effect on word meanings by analyzing statistically significant changes for individual sentences across multiple fMRI subjects. Four experiments in this chapter aimed to: (1) characterize the changes that occur when a word is used in the context of a sentence, and (2) explain how different types of conceptual combinations emerge from such context.

The first experiment focused on a of similar context for two different words such as *boat crossed* vs. *car crossed*. The effect on *boat* and *car* produced parallel changes with different weightings. Because word attributes and context attributes changed in similar ways, they were associated with the vehicle category. The second experiment characterized differences in two contexts such as *laughed* vs. *celebrated*. In this case, the effect of context developed from external relations. The results suggested that CEREBRA captures the thematic relations where the two contexts intersect. Also, this experiment demonstrated that an affective state of emotions is a relevant experience just like any sensorimotor experience based on CAR theory. The third experiment characterized the effect of two different contexts on the same word such as *bird flew* vs. *plane flew*. There was a large overlapping effect within the attributes of the context (e.g., animacy) and the attributes of the target concept (e.g., *bird flew*), resulting in a clear difference between animate and inanimate contexts. The fourth experiment characterized the centrality of meaning such as *small camera* vs. *small hospital*. The Size dimensions is central in this case distinguishing between a small object and a large structure. However, each conceptual combination required additional knowledge to determine what is central for a particular adjective.

Each experiment thus analyzed the changes observed in the concept attributes across contexts and illustrated how unique conceptual combinations develop. A further intriguing result emerged in the last three experiments: The process of interpreting conceptual combinations involves additional external knowledge. The fMRI images include such knowledge, and consequently, CEREBRA was able to capture the changes in CARs. However, with the current set of attributes, thematic associations are not captured accurately or completely. A solution was proposed in Chapter 8 to extend CAR with a set of attributes that target convergence zones that integrate distributed representations into a more general knowledge.

Chapter 6 aggregated the analysis across sentence contexts. CEREBRA showed that the difference between the expected and observed fMRI images can indeed be explained by changes in CARs. Across the entire corpus of sentences, the new CARs were more similar to the other words in the sentence than the original CARs were, demonstrating how features of the context were transferred to each word in the sentence. This result is robust and general across the entire corpus of sentences and case roles.

Correlation analysis demonstrated that the conceptual combination effect is consistent and can be quantified by analyzing fMRI images through the CEREBRA mechanism.

Chapter 7 evaluated CEREBRA's context-based representations via human judgements. First, CEREBRA was used to characterize the changes between generic and contextual representations of words in a number of sentences. The survey was then designed to characterize these changes in human subjects. The results confirmed that the changes produced by CEREBRA were actually meaningful to humans. The study provided the last piece on the theory of semantic feature representations: The CEREBRA's context-dependent changes represent structured semantic knowledge that can be used to predict human judgements.

In sum,

1. CEREBRA allows discerning how concepts are dynamically encoded in the brain.
2. CEREBRA demonstrated the soundness of the CAR theory by explaining the observed fMRI sentence representations based on context-dependent changes in CARs.
3. The experiments demonstrated that CARs can capture fine distinctions in meaning, therefore creating possibilities for improving the theory itself.
4. CEREBRA validates different types of conceptual combinations originally performed in behavioral studies as contextual effect in the brain.
5. CEREBRA demonstrated that the changes are meaningful to human subjects.
6. With CEREBRA, domain experts can gain insights and form new hypotheses about the functioning of the brain through fMRI data.
7. CEREBRA extends CARs static representations by showing that the CARs can be dynamic

8. CEREBRA's context-modified CARs could be used for building artificial natural language processing systems by dynamically adapting the vector representations to fit context (i.e., thus adding experiential knowledge).

### 9.2 General Conclusion

A great deal of research in neuroscience has focused on how the brain creates semantic memories, and what brain regions are responsible for the storage and retrieval of the semantic knowledge. Early studies focused on behavior of individuals with brain damage and with various types of semantic disorders, but more modern studies employed neuroimaging techniques to learn how the brain creates, stores and integrates semantic memories. Despite recent success in text-based semantic modeling and multimodal meaning representations, there is still a great deal of disagreement about how semantic knowledge is represented, and how these models correlate and reflect the semantic space of the brain. While all of these studies report correlation between semantic models and neural activations, they use different datasets and prediction methods which make the results difficult to compare. This dissertation followed the theory proposed by Binder et al. (2011) to develop a computational framework that addresses several issues on semantic memory. CEREBRA is an instrumental tool for interpreting fMRI patterns. It uses experiential-based distributed semantic features that are directly grounded on brain systems involved during semantic processing. The distributed features allow for a hierarchical multimodal integration through convergence zones or hubs. The experimental findings suggest that the representations are dynamic, changing with context. Building on this foundation, the CEREBRA approach makes novel experiments and applications possible. It provides a way to understand how word meanings change in the context of a sentence.

At present time, behavioral and neuropsychological evidence emphasizes how semantic representations are constructed from a lifetime of linguistic data and perceptual experience. There are obvious problems with the different views that prefer one data type over the other. Knowledge acquired experientially is not sufficient to fully account for the brain's semantic representations. In contrast, distributional approaches are disembodied from the physical world. How, then, can linguistic knowledge be grounded? CEREBRA's context-based representations capture information embedded in the fMRI patterns, which are based on the semantic space of the brain, and are multimodal, therefore they could be used to ground linguistic knowledge. Consequently, they can be used to interpret neural activation patterns associated with meaning.

Although the work on this dissertation does not reveal how sentence context is constructed, it provides a glimpse of meaning in action: (1) Contextual modulation is embedded in the fMRI sentences; (2) Every single attribute points to neural brain systems involved in contextual modulation.

Overall, this dissertation is expected to contribute to the development of a unified theory of concepts, the organization of the semantic space, and the processes involved on concept representation. In addition, this dissertation should serve as a useful tool for researchers studying brain diseases, disorders, or injuries related to language processing, as well as to provide enhanced context-based representations for systems such as Siri, OK Google, and Alexa, to advance grounded natural language understanding systems supporting service robot applications.

# Appendix A

**Terminology of Brain Areas and Function**

**Occipital lobe**: This is found in the back of the brain. The area is involved with the brain's ability to recognize objects. It is responsible for our vision.

**Temporal lobe**: The temporal lobes are found on either side of the brain and just above the ears. The temporal lobes are responsible for hearing, memory, meaning, and language. They also play a role in emotion and learning. The temporal lobes are concerned with interpreting and processing auditory stimuli.

**Parietal lobe**: The parietal lobes are found behind the frontal lobes, above the temporal lobes, and at the top back of the brain. They are connected with the processing of nerve impulses related to the senses, such as touch, pain, taste, pressure, and temperature. They also have language functions.

**Frontal lobe**: It is concerned with emotions, reasoning, planning, movement, and parts of speech. It is also involved in purposeful acts such as creativity, judgment, and problem solving, and planning

**Cerebral cortex**: The cerebral cortex controls your thinking, voluntary movements, language, reasoning, and perception. In higher mammals the cortex looks like it has lots of wrinkles, grooves and bumps.

**Cerebellum**: controls your movement, balance, posture, and coordination. New research has also linked it to thinking, novelty, and emotions. *The limbic system*, often referred to as the "emotional brain", is found buried within the cerebrum.

**Hypothalamus**: controls your body temperature, emotions, hunger, thirst, appetite, digestion and sleep. The hypothalamus is composed of several different areas and is located at the base of the brain. It is only the size of a pea (about 1/300 of the total brain weight), but is responsible for some very important behaviors.

**Thalamus**: controls your sensory integration and motor integration. Receives sensory information and relays it to the cerebral cortex. The cerebral cortex also sends information to the thalamus which then transmits this information to other parts of the brain and the brain stem.

**Pituitary gland**: it controls your hormones and it helps to turn food to energy. Without this gland you could eat but you wouldn't get any energy from the food.

**Pineal gland**: This part controls your growing and maturing. It is activated by light so if you were born and lived all your life in a place without a trace of light your pineal gland would never start to work.

**Amygdala**: The amygdala (there are two of them) control your emotions such as regulating when you're happy or mad. Your amygdala is very important. Without it you could win the lottery and feel nothing. You wouldn't be happy.

The brain

**References**

*http://serendip.brynmawr.edu/bb/kinser/definitions/def-medulla.html*
www.library.thinkquest.org/J002391/functions.html
http://www.ninds.nih.gov/disorders/brain_basics/know_your_brain.htm

**Terminology of Brain Areas and Function (continue)**

---

**Amygdala** is one of two almond-shaped clusters of nuclei located deep and medially within the temporal lobes of the brain in complex vertebrates. Shown to perform a primary role in the processing of memory, decision-making and emotional responses (including fear, anxiety, and aggression), the amygdalae are considered part of the limbic system.

**Angular Gyrus**: is a region of the brain lying mainly in the anterolateral region of parietal lobe, that lies near the superior edge of the temporal lobe, and immediately posterior to the supramarginal gyrus. Its significance is in transferring visual information to Wernicke's area, in order to make meaning out of visually perceived words. It is also involved in a number of processes related to language, number processing and spatial cognition, memory retrieval, attention, and theory of mind.

**Fusiform Gyrus:** also known as the lateral occipitotemporal gyrus,[1] is part of the temporal lobe and occipital lobe in Brodmann area 37.[2] The fusiform gyrus is located between the lingual gyrus and parahippocampal gyrus above, and the inferior temporal gyrus below. Though the functionality of the fusiform gyrus is not fully understood, it has been linked with various neural pathways related to recognition. Additionally, it has been linked to various neurological phenomena such as synesthesia, dyslexia, and prosopagnosia.

**Hippocampus** is a major component of the brain of humans and other vertebrates. Humans and other mammals have two hippocampi, one in each side of the brain. The hippocampus is part of the limbic system, and plays important roles in the consolidation of information from short-term memory to long-term memory, and in spatial memory that enables navigation.

**Middle temporal Gyrus** is a gyrus in the brain on the Temporal lobe. It is located between the superior temporal gyrus and inferior temporal gyrus.

**Posterior Cingulate Cortex** (PCC) is the caudal part of the cingulate cortex, located posterior to the anterior cingulate cortex. This is the upper part of the "limbic lobe". The cingulate cortex is made up of an area around the midline of the brain. It has been shown to communicate with various brain networks simultaneously and is involved in diverse functions.[1] Along with the precuneus, the PCC has been implicated as a neural substrate for human awareness in numerous studies of both the anesthesized and vegetative (coma) states. Imaging studies indicate a prominent role for the PCC in pain and episodic memory retrieval.

**Precuneus** is the portion of the superior parietal lobule on the medial surface of each brain hemisphere. It is located in front of the cuneus (the upper portion of the occipital lobe). The precuneus is bounded in front by the marginal branch of the cingulate sulcus, at the rear by the parietooccipital sulcus, and underneath by the subparietal sulcus. It is involved with episodic memory, visuospatial processing, reflections upon self, and aspects of consciousness.

**Terminology of Brain Areas and Function (continue)**

**Subgenual Cingulate** Brodmann area 25 (BA25); area in the cerebral cortex of the brain. This region is extremely rich in serotonin transporters and is considered as a governor for a vast network involving areas like he hippocampus, which plays an important role in memory formation; and some parts of the frontal cortex responsible for self-esteem. This region is particularly implicated in the normal processing of sadness.

**References**

https://en.wikipedia.org/wiki/

# Appendix B

Glasgow sentence and word Collections. Sentences 35, 157, and 183 and words 37, 95, 109, 138, 146, 239, marked in red, were not included in this dissertation for reasons explained in Section 3.4.2.

| No. | Glasgow Sentence Collection (Part I) |
|---|---|
| 1 | The family survived the powerful hurricane. |
| 2 | The family was happy. |
| 3 | The family played at the beach. |
| 4 | The wealthy family celebrated at the party. |
| 5 | The parent watched the sick child. |
| 6 | The politician visited the family. |
| 7 | The priest approached the lonely family. |
| 8 | The parent visited the school. |
| 9 | The parent shouted at the child. |
| 10 | The parent bought the magazine. |
| 11 | The happy couple visited the embassy. |
| 12 | The couple planned the vacation. |
| 13 | The parent took the cellphone. |
| 14 | The couple laughed at dinner. |
| 15 | The couple read on the beach. |
| 16 | The wealthy couple left the theater. |
| 17 | The child broke the glass in the restaurant. |
| 18 | The happy child found the dime. |
| 19 | The child gave the flower to the artist. |
| 20 | The child held the soft feather. |
| 21 | The angry child threw the book. |
| 22 | The girl dropped the shiny dime. |
| 23 | The actor gave the football to the team. |
| 24 | The commander listened to the soldier. |
| 25 | The soldier crossed the field. |
| 26 | The editor drank tea at dinner. |
| 27 | The beach was empty. |
| 28 | The judge met the mayor. |
| 29 | The doctor stole the book. |
| 30 | The artist drew the river. |
| 31 | The window was dusty. |
| 32 | The teacher worked at the new school. |
| 33 | The school was famous. |
| 34 | The school was empty during the summer. |
| 35 | The student walked along the long hall. |
| 36 | The young student read at the desk. |
| 37 | The small church was near the school. |
| 38 | The teacher used the computer. |
| 39 | The army marched past the school. |
| 40 | The scientist spoke to the student. |

| No. | Glasgow Sentence Collection (Part II) |
|---|---|
| 41 | The engineer gave a book to the student. |
| 42 | The student planned the protest. |
| 43 | The teacher broke the small camera. |
| 44 | The yellow dog approached the friendly teacher. |
| 45 | The teacher visited the beach in summer. |
| 46 | The red pencil was on the desk. |
| 47 | The team played soccer in spring. |
| 48 | The council read the agreement. |
| 49 | The mayor dropped the glass. |
| 50 | The street was dark. |
| 51 | The feather was blue. |
| 52 | The tree was green. |
| 53 | The diplomat was wealthy. |
| 54 | The dime was new. |
| 55 | The girl saw the small bird. |
| 56 | The small boy feared the storm. |
| 57 | The mouse ran into the forest. |
| 58 | The boat crossed the small lake. |
| 59 | The army built the small hospital. |
| 60 | The judge lost the dime. |
| 61 | The man saw the dead mouse. |
| 62 | The boy kicked the stone along the street. |
| 63 | The white feather was under the tree. |
| 64 | The dusty feather landed on the highway. |
| 65 | The cellphone was black. |
| 66 | The fish lived in the river. |
| 67 | The activist dropped the new cellphone. |
| 68 | The woman bought medicine at the store. |
| 69 | The magazine was yellow. |
| 70 | The minister found cash at the airport. |
| 71 | The businessman laughed in the theater. |
| 72 | The big horse drank from the lake. |
| 73 | The pilot was friendly. |
| 74 | The witness spoke to the lawyer. |
| 75 | The minister spoke to the injured patient. |
| 76 | The reporter spoke to the loud mob. |
| 77 | The young author spoke to the editor. |
| 78 | The author interviewed the scientist after the flood. |
| 79 | The commander negotiated with the council. |
| 80 | The diplomat negotiated at the embassy. |

| No. | Glasgow Sentence Collection (Part III) |
| --- | --- |
| 81 | The journalist interviewed the judge. |
| 82 | The reporter interviewed the dangerous terrorist. |
| 83 | The policeman interviewed the young victim. |
| 84 | The mayor negotiated with the mob. |
| 85 | The reporter interviewed the politician during the debate. |
| 86 | The witness shouted during the trial. |
| 87 | The artist shouted in the hotel. |
| 88 | The diplomat shouted at the soldier. |
| 89 | The activist listened to the tired victim. |
| 90 | The mayor listened to the voter. |
| 91 | The jury listened to the famous businessman. |
| 92 | The woman helped the sick tourist. |
| 93 | The lonely patient listened to the loud television. |
| 94 | The soldier delivered the medicine during the flood. |
| 95 | The engineer built the computer. |
| 96 | The terrorist stole the car. |
| 97 | The artist found the red ball. |
| 98 | The scientist watched the duck. |
| 99 | The flood was dangerous. |
| 100 | The cloud blocked the sun. |
| 101 | The baseball broke the window. |
| 102 | The dog broke the television. |
| 103 | The angry activist broke the chair. |
| 104 | The accident destroyed the empty lab. |
| 105 | The accident damaged the yellow car. |
| 106 | The hurricane damaged the boat. |
| 107 | The storm destroyed the theater. |
| 108 | The editor damaged the bicycle. |
| 109 | The mob damaged the hotel. |
| 110 | The flood damaged the hospital. |
| 111 | The horse kicked the fence. |
| 112 | The soldier kicked the door. |
| 113 | The banker was injured in the accident. |
| 114 | The author kicked the desk. |
| 115 | The storm was powerful. |
| 116 | The doctor helped the injured policeman. |
| 117 | The injured horse slept at night. |
| 118 | The soldier arrested the injured activist. |
| 119 | The dangerous criminal stole the television. |
| 120 | The doctor bought the used boat. |

| No. | Glasgow Sentence Collection (Part IV) |
|---|---|
| 121 | The guard opened the window. |
| 122 | The egg was blue. |
| 123 | The glass was cold. |
| 124 | The witness went to the trial. |
| 125 | The trial ended in spring. |
| 126 | The politician watched the trial. |
| 127 | The reporter wrote about the trial. |
| 128 | The activist marched at the trial. |
| 129 | The tired jury left the court. |
| 130 | The jury watched the witness. |
| 131 | The lawyer was friendly. |
| 132 | The angry lawyer left the office. |
| 133 | The tired lawyer visited the island. |
| 134 | The lawyer drank coffee. |
| 135 | The old judge saw the dark cloud. |
| 136 | The judge stayed at the hotel during the vacation. |
| 137 | The policeman arrested the angry driver. |
| 138 | The tired patient slept in the dark hospital. |
| 139 | The man read the newspaper in church. |
| 140 | The criminal wanted cash. |
| 141 | The clever scientist worked at the lab. |
| 142 | The editor gave cash to the driver. |
| 143 | The green car crossed the bridge. |
| 144 | The vacation was peaceful. |
| 145 | The duck lived at the lake. |
| 146 | The bird landed on the bridge. |
| 147 | The protest was loud. |
| 148 | The voter went to the protest. |
| 149 | The council feared the protest. |
| 150 | The banker watched the peaceful protest. |
| 151 | The mob approached the embassy. |
| 152 | The mob was dangerous. |
| 153 | The reporter met the angry doctor. |
| 154 | The voter read about the election. |
| 155 | The politician celebrated at the hotel. |
| 156 | The wealthy politician liked coffee. |
| 157 | The worker fixed the door at the church. |
| 158 | The corn grew in spring. |
| 159 | The victim feared the criminal. |
| 160 | The young engineer worked in the office. |

| No. | Glasgow Sentence Collection (Part V) |
|---|---|
| 161 | The tourist was friendly. |
| 162 | The baseball was in the office. |
| 163 | The used book was on the table. |
| 164 | The magazine was in the car. |
| 165 | The bridge survived the flood. |
| 166 | The old doctor walked through the hospital. |
| 167 | The patient survived. |
| 168 | The patient put the medicine in the cabinet. |
| 169 | The medicine was on the table. |
| 170 | The famous diplomat left the hospital. |
| 171 | The commander opened the heavy door. |
| 172 | The banker bought the expensive boat. |
| 173 | The storm ended during the morning. |
| 174 | The car approached the river. |
| 175 | The door was blue. |
| 176 | The farmer liked soccer. |
| 177 | The engineer walked in the peaceful park. |
| 178 | The horse walked through the green field. |
| 179 | The wealthy author walked into the office. |
| 180 | The young policeman walked to the theater. |
| 181 | The artist hiked along the mountain. |
| 182 | The tourist hiked through the forest. |
| 183 | The editor carried the magazine to the meeting. |
| 184 | The dog ran in the park. |
| 185 | The woman took the flower from the field. |
| 186 | The street was empty at night. |
| 187 | The wealthy farmer fed the horse. |
| 188 | The diplomat bought the aggressive dog. |
| 189 | The dog drank water. |
| 190 | The tree grew in the park. |
| 191 | The commander ate chicken at dinner. |
| 192 | The dog ate the egg. |
| 193 | The computer was new. |
| 194 | The company delivered the computer. |
| 195 | The computer was on the desk. |
| 196 | The businessman lost the computer at the airport. |
| 197 | The expensive camera was in the lab. |
| 198 | The reporter ate at the new restaurant. |
| 199 | The minister lost the spiritual magazine. |
| 200 | The boy held the football. |

| No. | Glasgow Sentence Collection (Part VI) |
|---|---|
| 201 | The bird was red. |
| 202 | The cloud was white. |
| 203 | The yellow bird flew over the field. |
| 204 | The flower was yellow. |
| 205 | The green duck slept under the tree. |
| 206 | The girl saw a horse in the park. |
| 207 | The duck flew. |
| 208 | The man lost the ticket to soccer. |
| 209 | The team celebrated. |
| 210 | The red plane flew through the cloud. |
| 211 | The summer was hot. |
| 212 | The bicycle blocked the green door. |
| 213 | The park was empty in winter. |
| 214 | The driver wanted cold tea. |
| 215 | The minister visited the prison. |
| 216 | The tourist ate bread on vacation. |
| 217 | The tourist went to the restaurant. |
| 218 | The tourist found a bird in the theater. |
| 219 | The old farmer ate at the expensive hotel. |
| 220 | The aggressive team took the baseball. |
| 221 | The duck was aggressive. |
| 222 | The chicken was expensive at the restaurant. |
| 223 | The artist liked chicken. |
| 224 | The restaurant was loud at night. |
| 225 | The woman left the restaurant after the storm. |
| 226 | The banker drank cold water. |
| 227 | The coffee was hot. |
| 228 | The boy threw the baseball over the fence. |
| 229 | The policeman read the newspaper. |
| 230 | The criminal put the book on the desk. |
| 231 | The man saw the fish in the river. |
| 232 | The happy girl played in the forest. |
| 233 | The young girl played soccer. |
| 234 | The old man threw the stone into the lake. |
| 235 | The team lost the football in the forest. |
| 236 | The businessman slept on the expensive bed. |
| 237 | The businessman watched soccer. |
| 238 | The guard slept near the door. |
| 239 | The artist kicked the football. |
| 240 | The ticket was on the red desk. |

| No. | Word | No. | Word | No. | Word | No. | Word | No. | Word | No. | Word |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | accident | 42 | chicken | 83 | ended | 124 | journalist | 165 | pencil | 206 | survived |
| 2 | activist | 43 | child | 84 | engineer | 125 | judge | 166 | pilot | 207 | table |
| 3 | actor | 44 | church | 85 | expensive | 126 | jury | 167 | plane | 208 | tea |
| 4 | aggressive | 45 | clever | 86 | family | 127 | kicked | 168 | planned | 209 | teacher |
| 5 | agreement | 46 | cloud | 87 | famous | 128 | lab | 169 | played | 210 | team |
| 6 | airport | 47 | coffee | 88 | farmer | 129 | lake | 170 | policeman | 211 | television |
| 7 | angry | 48 | cold | 89 | feared | 130 | landed | 171 | politician | 212 | terrorist |
| 8 | approached | 49 | commander | 90 | feather | 131 | laughed | 172 | powerful | 213 | theater |
| 9 | army | 50 | company | 91 | fed | 132 | lawyer | 173 | priest | 214 | threw |
| 10 | arrested | 51 | computer | 92 | fence | 133 | left | 174 | prison | 215 | ticket |
| 11 | artist | 52 | corn | 93 | field | 134 | liked | 175 | protest | 216 | tired |
| 12 | ate | 53 | council | 94 | fish | 135 | listened | 176 | put | 217 | took |
| 13 | author | 54 | couple | 95 | fixed | 136 | lived | 177 | ran | 218 | tourist |
| 14 | ball | 55 | court | 96 | flew | 137 | lonely | 178 | read | 219 | tree |
| 15 | banker | 56 | criminal | 97 | flood | 138 | long | 179 | red | 220 | trial |
| 16 | baseball | 57 | crossed | 98 | flower | 139 | lost | 180 | reporter | 221 | used |
| 17 | beach | 58 | damaged | 99 | football | 140 | loud | 181 | restaurant | 222 | used |
| 18 | bed | 59 | dangerous | 100 | forest | 141 | magazine | 182 | river | 223 | vacation |
| 19 | bicycle | 60 | dark | 101 | found | 142 | man | 183 | saw | 224 | victim |
| 20 | big | 61 | dead | 102 | friendly | 143 | marched | 184 | school | 225 | visited |
| 21 | bird | 62 | debate | 103 | gave | 144 | mayor | 185 | scientist | 226 | voter |
| 22 | black | 63 | delivered | 104 | girl | 145 | medicine | 186 | shiny | 227 | walked |
| 23 | blocked | 64 | desk | 105 | glass | 146 | meeting | 187 | shouted | 228 | wanted |
| 24 | blue | 65 | destroyed | 106 | green | 147 | met | 188 | sick | 229 | watched |
| 25 | boat | 66 | dime | 107 | grew | 148 | minister | 189 | slept | 230 | water |
| 26 | book | 67 | dinner | 108 | guard | 149 | mob | 190 | small | 231 | wealthy |
| 27 | bought | 68 | diplomat | 109 | hall | 150 | morning | 191 | soccer | 232 | went |
| 28 | boy | 69 | doctor | 110 | happy | 151 | mountain | 192 | soft | 233 | white |
| 29 | bread | 70 | dog | 111 | heavy | 152 | mouse | 193 | soldier | 234 | window |
| 30 | bridge | 71 | door | 112 | held | 153 | negotiated | 194 | spiritual | 235 | winter |
| 31 | broke | 72 | drank | 113 | helped | 154 | new | 195 | spoke | 236 | witness |
| 32 | built | 73 | drew | 114 | highway | 155 | newspaper | 196 | spring | 237 | woman |
| 33 | businessman | 74 | driver | 115 | hiked | 156 | night | 197 | stayed | 238 | worked |
| 34 | cabinet | 75 | dropped | 116 | horse | 157 | office | 198 | stole | 239 | worker |
| 35 | camera | 76 | duck | 117 | hospital | 158 | old | 199 | stone | 240 | wrote |
| 36 | car | 77 | dusty | 118 | hot | 159 | opened | 200 | store | 241 | yellow |
| 37 | carried | 78 | editor | 119 | hotel | 160 | parent | 201 | storm | 242 | young |
| 38 | cash | 79 | egg | 120 | hurricane | 161 | park | 202 | street | | |
| 39 | celebrated | 80 | election | 121 | injured | 162 | party | 203 | student | | |
| 40 | cellphone | 81 | embassy | 122 | interviewed | 163 | patient | 204 | summer | | |
| 41 | chair | 82 | empty | 123 | island | 164 | peaceful | 205 | sun | | |

# Appendix C

For illustration, two questionnaires are included in the next pages. Each questionnaire is composed of the Introduction, an Example and the list of 15 questions. To find the entire set follow this link:

https://drive.google.com/drive/folders/1jDCqKMuH-SyTxcJ7oJRbr7mYV6WNNEWH?usp=sharing

# Sentence Rating Survey

How does the meaning of a word change in different sentences? For example, a subject would invoke different properties of the word PIANO depending on what he read. The concept of piano is composed of visual features like size and weight, auditory features like sound, and motor features like manipulation, and so on. Depending on context, the weighting of those properties will show different activation. When thinking about "moving the piano", the focus of attention will be on size, shape, weight, and larger limbs movement . If thinking about "playing the piano", the emphasis will be on the piano's function such as sound and fine hand movement properties.

Brain imaging tools provide a new approach to understanding this phenomenon by directly observing brain activity during word and sentence comprehension. We collected brain image data from 11 subjects reading everyday sentences. Brain activity was recorded on their visual, sensory, motor, and other brain systems during the experiments.

Our research aims at explaining this data in a computational model. In the model, different properties of words are weighted differently depending on context. The model is then used to determine how the generic meaning of a word would have to change in order to account for context. This survey is intended to evaluate the results produced by the computational model. You will make judgements about which properties changed, and your judgement will be compared to those suggested by the model.

* Required

EXPERIMENT DESCRIPTION

You will be given 15 sentences and asked to evaluate 10 properties of one of the words in the sentence. Specifically, you will decide if the meaning of the word contains MORE or LESS of that property or remains NEUTRAL compared to the word's generic meaning (i.e., the meaning of that word without specific context), such as, its general, common, nonspecific meaning.

As an example, this page shows the analysis of 6 properties relevant to the word 'MAN" in the sentence "The old man rowed the boat during the summer". The explanations to the responses are shown in parenthesis. We suggest to follow that kind of reasoning to help you select your answers during the survey.

THANK YOU!

## The old man rowed the boat during the summer

Think of the generic meaning of the word 'MAN'. Now think of the same word used in the sentence above. How is 'MAN' in this sentence different from its generic meaning?

| | more | less | neutral |
|---|---|---|---|
| is dark... ('less' since the event occurs during summer) | ○ | ◉ | ○ |
| moves slow ('more' since is an old man) | ◉ | ○ | ○ |
| talks... ('neutral' because the sentence doesn't address the activity of talking) | ○ | ○ | ◉ |
| uses the leg or foot... ('less' because the rowing activity involves the upper body) | ○ | ◉ | ○ |
| changes location ('more' because the boat moved) | ◉ | ○ | ○ |
| is an essential part... ('more' because the man is necessary for rowing the boat) | ◉ | ○ | ○ |

## Press NEXT to start the survey

1. 1:The banker bought the expensive boat *

Think of the generic meaning of the word 'BANKER'. Now think of the same word used in the sentence above. How is 'BANKER' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ⬭ | ⬭ | ⬭ |
| feels hot | ⬭ | ⬭ | ⬭ |
| feels smooth | ⬭ | ⬭ | ⬭ |
| makes a musical sound | ⬭ | ⬭ | ⬭ |
| has a distinctive taste | ⬭ | ⬭ | ⬭ |
| uses the leg or foot | ⬭ | ⬭ | ⬭ |
| is an object | ⬭ | ⬭ | ⬭ |
| lasts a short time | ⬭ | ⬭ | ⬭ |
| is happy | ⬭ | ⬭ | ⬭ |
| causes fear | ⬭ | ⬭ | ⬭ |

2. 2:The author interviewed the scientist after the flood *

Think of the generic meaning of the word 'INTERVIEWED'. Now think of the same word used in the sentence above. How is 'INTERVIEWED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ⬭ | ⬭ | ⬭ |
| has a face | ⬭ | ⬭ | ⬭ |
| is identified by touch | ⬭ | ⬭ | ⬭ |
| is heavy | ⬭ | ⬭ | ⬭ |
| makes a musical sound | ⬭ | ⬭ | ⬭ |
| has a distinctive taste | ⬭ | ⬭ | ⬭ |
| triggers social interaction | ⬭ | ⬭ | ⬭ |
| is sad | ⬭ | ⬭ | ⬭ |
| causes anger | ⬭ | ⬭ | ⬭ |
| increases alertness | ⬭ | ⬭ | ⬭ |

3. 3:The scientist spoke to the student *

Think of the generic meaning of the word 'STUDENT'. Now think of the same word used in the sentence above. How is 'STUDENT' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | | | |
| has a face | | | |
| is associated with pain | | | |
| makes a musical sound | | | |
| is an object | | | |
| goes away | | | |
| has consequences | | | |
| causes anger | | | |
| causes surprise | | | |
| is an essential part | | | |

4. 4:The judge met the mayor *

Think of the generic meaning of the word 'JUDGE'. Now think of the same word used in the sentence above. How is 'JUDGE' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has texture or pattern | | | |
| is identified by touch | | | |
| makes a high pitch | | | |
| makes a musical sound | | | |
| talks | | | |
| has a distinctive smell | | | |
| uses the leg or foot | | | |
| has a fixed location | | | |
| has a certain duration | | | |
| causes anger | | | |

5. 5:The engineer gave a book to the student *

Think of the generic meaning of the word 'GAVE'. Now think of the same word used in the sentence above. How is 'GAVE' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is dark | | | |
| has a face | | | |
| is heavy | | | |
| is audible | | | |
| makes a musical sound | | | |
| has a distinctive smell | | | |
| is physically manipulated | | | |
| goes away | | | |
| causes anger | | | |
| is disgusting | | | |

6. 6:The criminal put the book on the desk *

Think of the generic meaning of the word 'BOOK'. Now think of the same word used in the sentence above. How is 'BOOK' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | | | |
| has texture or pattern | | | |
| moves slow | | | |
| has body parts | | | |
| is associated with pain | | | |
| makes a high pitch | | | |
| has a distinctive taste | | | |
| is an object | | | |
| changes location | | | |
| causes anger | | | |

7. 7:The girl saw a horse in the park *

Think of the generic meaning of the word 'GIRL'. Now think of the same word used in the sentence above. How is 'GIRL' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has a defining shape | ◯ | ◯ | ◯ |
| makes a musical sound | ◯ | ◯ | ◯ |
| is an object | ◯ | ◯ | ◯ |
| comes close | ◯ | ◯ | ◯ |
| goes away | ◯ | ◯ | ◯ |
| is countable | ◯ | ◯ | ◯ |
| has a certain duration | ◯ | ◯ | ◯ |
| is sad | ◯ | ◯ | ◯ |
| causes fear | ◯ | ◯ | ◯ |
| causes surprise | ◯ | ◯ | ◯ |

8. 8:The yellow dog approached the friendly teacher *

Think of the generic meaning of the word 'APPROACHED'. Now think of the same word used in the sentence above. How is 'APPROACHED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ◯ | ◯ | ◯ |
| moves | ◯ | ◯ | ◯ |
| feels smooth | ◯ | ◯ | ◯ |
| is heavy | ◯ | ◯ | ◯ |
| is associated with pain | ◯ | ◯ | ◯ |
| makes a high pitch | ◯ | ◯ | ◯ |
| has a certain duration | ◯ | ◯ | ◯ |
| lasts a short time | ◯ | ◯ | ◯ |
| causes fear | ◯ | ◯ | ◯ |
| is a center of attention | ◯ | ◯ | ◯ |

9. 9:The man lost the ticket to soccer *

Think of the generic meaning of the word 'TICKET'. Now think of the same word used in the sentence above. How is 'TICKET' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ◯ | ◯ | ◯ |
| has body parts | ◯ | ◯ | ◯ |
| makes a musical sound | ◯ | ◯ | ◯ |
| uses the face or mouth | ◯ | ◯ | ◯ |
| is an object | ◯ | ◯ | ◯ |
| goes away | ◯ | ◯ | ◯ |
| has intentions | ◯ | ◯ | ◯ |
| exchanges information | ◯ | ◯ | ◯ |
| is beneficial | ◯ | ◯ | ◯ |
| causes anger | ◯ | ◯ | ◯ |

10. 10:The boat crossed the small lake *

Think of the generic meaning of the word 'CROSSED'. Now think of the same word used in the sentence above. How is 'CROSSED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is visually complex | ◯ | ◯ | ◯ |
| is heavy | ◯ | ◯ | ◯ |
| makes a high pitch | ◯ | ◯ | ◯ |
| has a distinctive taste | ◯ | ◯ | ◯ |
| is physically manipulated | ◯ | ◯ | ◯ |
| has a certain duration | ◯ | ◯ | ◯ |
| lasts a short time | ◯ | ◯ | ◯ |
| is sad | ◯ | ◯ | ◯ |
| causes surprise | ◯ | ◯ | ◯ |
| is a center of attention | ◯ | ◯ | ◯ |

11. 11:The victim feared the criminal *

Think of the generic meaning of the word 'FEARED'. Now think of the same word used in the sentence above. How is 'FEARED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is visible | | | |
| is visually complex | | | |
| has a distinctive taste | | | |
| has a distinctive smell | | | |
| is countable | | | |
| lasts a short time | | | |
| triggers social interaction | | | |
| is unpleasant | | | |
| is happy | | | |
| is an essential part | | | |

12. 12:The boy held the football *

Think of the generic meaning of the word 'FOOTBALL'. Now think of the same word used in the sentence above. How is 'FOOTBALL' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| moves slow | | | |
| has a face | | | |
| has body parts | | | |
| talks | | | |
| has a distinctive taste | | | |
| has a certain duration | | | |
| lasts a short time | | | |
| has intentions | | | |
| relates to oneself | | | |
| is disgusting | | | |

13. 13:The famous diplomat left the hospital *

Think of the generic meaning of the word 'DIPLOMAT'. Now think of the same word used in the sentence above. How is 'DIPLOMAT' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | | | |
| has texture or pattern | | | |
| living thing that moves | | | |
| has a distinctive smell | | | |
| uses the leg or foot | | | |
| has a fixed location | | | |
| is a particular setting | | | |
| is near | | | |
| goes away | | | |
| has a certain duration | | | |

14. 14:The injured horse slept at night *

Think of the generic meaning of the word 'SLEPT'. Now think of the same word used in the sentence above. How is 'SLEPT' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | | | |
| has texture or pattern | | | |
| is large | | | |
| is associated with pain | | | |
| is loud | | | |
| makes a musical sound | | | |
| is physically manipulated | | | |
| has a fixed location | | | |
| caused by something | | | |
| is a center of attention | | | |

15. 15:The commander ate chicken at dinner *

Think of the generic meaning of the word 'CHICKEN'. Now think of the same word used in the sentence above. How is 'CHICKEN' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has texture or pattern | ◯ | ◯ | ◯ |
| has a face | ◯ | ◯ | ◯ |
| is heavy | ◯ | ◯ | ◯ |
| is an object | ◯ | ◯ | ◯ |
| has a certain duration | ◯ | ◯ | ◯ |
| lasts a short time | ◯ | ◯ | ◯ |
| is unpleasant | ◯ | ◯ | ◯ |
| is happy | ◯ | ◯ | ◯ |
| is sad | ◯ | ◯ | ◯ |
| is a center of attention | ◯ | ◯ | ◯ |

# Sentence Rating Survey

How does the meaning of a word change in different sentences? For example, a subject would invoke different properties of the word PIANO depending on what he read. The concept of piano is composed of visual features like size and weight, auditory features like sound, and motor features like manipulation, and so on. Depending on context, the weighting of those properties will show different activation. When thinking about "moving the piano", the focus of attention will be on size, shape, weight, and larger limbs movement . If thinking about "playing the piano", the emphasis will be on the piano's function such as sound and fine hand movement properties.

Brain imaging tools provide a new approach to understanding this phenomenon by directly observing brain activity during word and sentence comprehension. We collected brain image data from 11 subjects reading everyday sentences. Brain activity was recorded on their visual, sensory, motor, and other brain systems during the experiments.

Our research aims at explaining this data in a computational model. In the model, different properties of words are weighted differently depending on context. The model is then used to determine how the generic meaning of a word would have to change in order to account for context. This survey is intended to evaluate the results produced by the computational model. You will make judgements about which properties changed, and your judgement will be compared to those suggested by the model.

* Required

EXPERIMENT DESCRIPTION

You will be given 15 sentences and asked to evaluate 10 properties of one of the words in the sentence. Specifically, you will decide if the meaning of the word contains MORE or LESS of that property or remains NEUTRAL compared to the word's generic meaning (i.e., the meaning of that word without specific context), such as, its general, common, nonspecific meaning.

As an example, this page shows the analysis of 6 properties relevant to the word 'MAN" in the sentence "The old man rowed the boat during the summer". The explanations to the responses are shown in parenthesis. We suggest to follow that kind of reasoning to help you select your answers during the survey.

THANK YOU!

## The old man rowed the boat during the summer

Think of the generic meaning of the word 'MAN'. Now think of the same word used in the sentence above. How is 'MAN' in this sentence different from its generic meaning?

| | more | less | neutral |
|---|---|---|---|
| is dark... ('less' since the event occurs during summer) | ○ | ● | ○ |
| moves slow ('more' since is an old man) | ● | ○ | ○ |
| talks... ('neutral' because the sentence doesn't address the activity of talking) | ○ | ○ | ● |
| uses the leg or foot... ('less' because the rowing activity involves the upper body) | ○ | ● | ○ |
| changes location ('more' because the boat moved) | ● | ○ | ○ |
| is an essential part... ('more' because the man is necessary for rowing the boat) | ● | ○ | ○ |

Press NEXT to start the survey

1. 1:The happy girl played in the forest *

Think of the generic meaning of the word 'GIRL'. Now think of the same word used in the sentence above. How is 'GIRL' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is dark | ○ | ○ | ○ |
| feels hot | ○ | ○ | ○ |
| makes a high pitch | ○ | ○ | ○ |
| has a distinctive taste | ○ | ○ | ○ |
| is an object | ○ | ○ | ○ |
| has a fixed location | ○ | ○ | ○ |
| is near | ○ | ○ | ○ |
| exchanges information | ○ | ○ | ○ |
| causes to act in certain way | ○ | ○ | ○ |
| is an essential part | ○ | ○ | ○ |

2. 2:The jury listened to the famous businessman *

Think of the generic meaning of the word 'LISTENED'. Now think of the same word used in the sentence above. How is 'LISTENED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has texture or pattern | ○ | ○ | ○ |
| moves fast | ○ | ○ | ○ |
| is visually complex | ○ | ○ | ○ |
| has a face | ○ | ○ | ○ |
| makes a high pitch | ○ | ○ | ○ |
| has a distinctive taste | ○ | ○ | ○ |
| has a distinctive smell | ○ | ○ | ○ |
| changes location | ○ | ○ | ○ |
| lasts a short time | ○ | ○ | ○ |
| is an essential part | ○ | ○ | ○ |

3. 3:The mob damaged the hotel *

Think of the generic meaning of the word 'HOTEL'. Now think of the same word used in the sentence above. How is 'HOTEL' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| moves fast | | | |
| is identified by touch | | | |
| is loud | | | |
| has a distinctive taste | | | |
| changes location | | | |
| is near | | | |
| is an event in time | | | |
| lasts a short time | | | |
| has intentions | | | |
| is sad | | | |

4. 4:The parent bought the magazine *

Think of the generic meaning of the word 'PARENT'. Now think of the same word used in the sentence above. How is 'PARENT' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | | | |
| has a defining color | | | |
| living thing that moves | | | |
| has a defining shape | | | |
| feels hot | | | |
| talks | | | |
| has a distinctive taste | | | |
| uses the leg or foot | | | |
| has a certain duration | | | |
| is an essential part | | | |

5. 5:The green duck slept under the tree *

Think of the generic meaning of the word 'SLEPT'. Now think of the same word used in the sentence above. How is 'SLEPT' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is visible | | | |
| is dark | | | |
| feels hot | | | |
| is loud | | | |
| has a distinctive taste | | | |
| has a distinctive smell | | | |
| has a certain duration | | | |
| lasts a short time | | | |
| has consequences | | | |
| is disgusting | | | |

6. 6:The yellow bird flew over the field *

Think of the generic meaning of the word 'FIELD'. Now think of the same word used in the sentence above. How is 'FIELD' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has a defining color | | | |
| is large | | | |
| moves fast | | | |
| has body parts | | | |
| is associated with pain | | | |
| has a distinctive taste | | | |
| is an event in time | | | |
| has consequences | | | |
| increases mental activity | | | |
| causes surprise | | | |

7. 7:The wealthy author walked into the office *

Think of the generic meaning of the word 'AUTHOR'. Now think of the same word used in the sentence above. How is 'AUTHOR' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has a defining shape | ⬭ | ⬭ | ⬭ |
| has body parts | ⬭ | ⬭ | ⬭ |
| is loud | ⬭ | ⬭ | ⬭ |
| has a distinctive taste | ⬭ | ⬭ | ⬭ |
| has a distinctive smell | ⬭ | ⬭ | ⬭ |
| has a certain duration | ⬭ | ⬭ | ⬭ |
| caused by something | ⬭ | ⬭ | ⬭ |
| exchanges information | ⬭ | ⬭ | ⬭ |
| causes to act in certain way | ⬭ | ⬭ | ⬭ |
| is an essential part | ⬭ | ⬭ | ⬭ |

8. 8:The child broke the glass in the restaurant *

Think of the generic meaning of the word 'BROKE'. Now think of the same word used in the sentence above. How is 'BROKE' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has a defining color | ⬭ | ⬭ | ⬭ |
| is visually complex | ⬭ | ⬭ | ⬭ |
| is loud | ⬭ | ⬭ | ⬭ |
| is physically manipulated | ⬭ | ⬭ | ⬭ |
| has a fixed location | ⬭ | ⬭ | ⬭ |
| is an event in time | ⬭ | ⬭ | ⬭ |
| has a certain duration | ⬭ | ⬭ | ⬭ |
| causes fear | ⬭ | ⬭ | ⬭ |
| causes surprise | ⬭ | ⬭ | ⬭ |
| is an essential part | ⬭ | ⬭ | ⬭ |

9. 9:The happy child found the dime *

Think of the generic meaning of the word 'DIME'. Now think of the same word used in the sentence above. How is 'DIME' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is small | ⬭ | ⬭ | ⬭ |
| moves fast | ⬭ | ⬭ | ⬭ |
| is visually complex | ⬭ | ⬭ | ⬭ |
| is identified by touch | ⬭ | ⬭ | ⬭ |
| is loud | ⬭ | ⬭ | ⬭ |
| has a certain duration | ⬭ | ⬭ | ⬭ |
| lasts a short time | ⬭ | ⬭ | ⬭ |
| increases mental activity | ⬭ | ⬭ | ⬭ |
| causes anger | ⬭ | ⬭ | ⬭ |
| is disgusting | ⬭ | ⬭ | ⬭ |

10. 10:The reporter interviewed the dangerous terrorist *

Think of the generic meaning of the word 'REPORTER'. Now think of the same word used in the sentence above. How is 'REPORTER' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| living thing that moves | ⬭ | ⬭ | ⬭ |
| moves fast | ⬭ | ⬭ | ⬭ |
| is identified by touch | ⬭ | ⬭ | ⬭ |
| feels hot | ⬭ | ⬭ | ⬭ |
| makes a high pitch | ⬭ | ⬭ | ⬭ |
| is an object | ⬭ | ⬭ | ⬭ |
| has a fixed location | ⬭ | ⬭ | ⬭ |
| lasts a short time | ⬭ | ⬭ | ⬭ |
| relates to oneself | ⬭ | ⬭ | ⬭ |
| is beneficial | ⬭ | ⬭ | ⬭ |

11. 11:The old man threw the stone into the lake *

Think of the generic meaning of the word 'THREW'. Now think of the same word used in the sentence above. How is 'THREW' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is visible | ⬭ | ⬭ | ⬭ |
| has a defining color | ⬭ | ⬭ | ⬭ |
| is visually complex | ⬭ | ⬭ | ⬭ |
| has a distinctive taste | ⬭ | ⬭ | ⬭ |
| uses the arm or hand | ⬭ | ⬭ | ⬭ |
| is near | ⬭ | ⬭ | ⬭ |
| goes away | ⬭ | ⬭ | ⬭ |
| lasts a long time | ⬭ | ⬭ | ⬭ |
| caused by something | ⬭ | ⬭ | ⬭ |
| is a center of attention | ⬭ | ⬭ | ⬭ |

12. 12:The author kicked the desk *

Think of the generic meaning of the word 'DESK'. Now think of the same word used in the sentence above. How is 'DESK' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is dark | ⬭ | ⬭ | ⬭ |
| moves fast | ⬭ | ⬭ | ⬭ |
| is heavy | ⬭ | ⬭ | ⬭ |
| is loud | ⬭ | ⬭ | ⬭ |
| is an object | ⬭ | ⬭ | ⬭ |
| changes location | ⬭ | ⬭ | ⬭ |
| has a certain duration | ⬭ | ⬭ | ⬭ |
| has intentions | ⬭ | ⬭ | ⬭ |
| is disgusting | ⬭ | ⬭ | ⬭ |
| causes fear | ⬭ | ⬭ | ⬭ |

13. 13:The flood damaged the hospital *

Think of the generic meaning of the word 'DAMAGED'. Now think of the same word used in the sentence above. How is 'DAMAGED' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ⬭ | ⬭ | ⬭ |
| is visually complex | ⬭ | ⬭ | ⬭ |
| has a face | ⬭ | ⬭ | ⬭ |
| is heavy | ⬭ | ⬭ | ⬭ |
| talks | ⬭ | ⬭ | ⬭ |
| comes close | ⬭ | ⬭ | ⬭ |
| has a certain duration | ⬭ | ⬭ | ⬭ |
| lasts a short time | ⬭ | ⬭ | ⬭ |
| caused by something | ⬭ | ⬭ | ⬭ |
| is happy | ⬭ | ⬭ | ⬭ |

14. 14:The witness spoke to the lawyer *

Think of the generic meaning of the word 'SPOKE'. Now think of the same word used in the sentence above. How is 'SPOKE' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| is bright | ⬭ | ⬭ | ⬭ |
| is dark | ⬭ | ⬭ | ⬭ |
| has a face | ⬭ | ⬭ | ⬭ |
| is identified by touch | ⬭ | ⬭ | ⬭ |
| feels hot | ⬭ | ⬭ | ⬭ |
| has a distinctive smell | ⬭ | ⬭ | ⬭ |
| changes location | ⬭ | ⬭ | ⬭ |
| lasts a short time | ⬭ | ⬭ | ⬭ |
| caused by something | ⬭ | ⬭ | ⬭ |
| is beneficial | ⬭ | ⬭ | ⬭ |

15. 15:The commander ate chicken at dinner *

Think of the generic meaning of the word 'CHICKEN'. Now think of the same word used in the sentence above. How is 'CHICKEN' in this sentence different from its generic meaning?

*Mark only one oval per row.*

| | more | less | neutral |
|---|---|---|---|
| has a face | ◯ | ◯ | ◯ |
| makes a high pitch | ◯ | ◯ | ◯ |
| makes a musical sound | ◯ | ◯ | ◯ |
| talks | ◯ | ◯ | ◯ |
| changes location | ◯ | ◯ | ◯ |
| has intentions | ◯ | ◯ | ◯ |
| is unpleasant | ◯ | ◯ | ◯ |
| is disgusting | ◯ | ◯ | ◯ |
| causes fear | ◯ | ◯ | ◯ |
| is an essential part | ◯ | ◯ | ◯ |

# Appendix D

The following results represent the analyses between humans, the three CEREBRA approaches, and chance. They are similar to the those presented in Section 7.4. However, the evaluation here was done using 1966 responses (i.e., where three out of four participants agreed) from the 3600 questions posed by the questionnaires.

Table D.1: Matching CEREBRA predictions with human data (approaches one to three), compared to chance. The table shows the average agreement of the 20 repetitions across all subjects. CEREBRA approaches one and two, agree with human responses 55%, CEREBRA approach three agrees 54%, while the chance level is 45%. Comparison agreement with human judgements.

| PARTICIPANTS AVERAGE AGREEMENT | | | | | |
|---|---|---|---|---|---|
| RATINGS | HUMAN | CEREBRA#1 | CEREBRA#2 | CEREBRA#3 | CHANCE |
| -1/0 | 1074 | 480 | 486 | 466 | 8 |
| 1 | 892 | 608 | 599 | 587 | 886 |
| TOTAL | 1966 | 1088 | 1085 | 1053 | 894 |
| AVERAGE | | 55% | 55% | 54% | 45% |

Table D.2: Statistical analyses for CEREBRA approaches one to three, and chance. The table shows the means and variances of the CEREBRA change models and the chance model for each subject, and the p-values of the t-test, revealing that the differences are highly significant.

| SUBJECTS | CHANCE | | CEREBRA #1 | | CEREBRA #2 | | CEREBRA #3 | | p-value | p-value | p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | MEAN | VAR | MEAN | VAR | MEAN | VAR | MEAN | VAR | CEREBRA #1 | CEREBRA #2 | CEREBRA #3 |
| S5051 | 894 | 6.01 | 1082.5 | 149.0 | 1083 | 131.32 | 1033 | 707.25 | 2.94E-41 | 2.99E-42 | 3.92E-24 |
| S9322 | 894 | 7.21 | 1076.8 | 199.0 | 1073 | 128.31 | 1035 | 233.91 | 2.15E-38 | 1.80E-41 | 6.10E-33 |
| S9362 | 894 | 11.52 | 1089.4 | 186.6 | 1086 | 166.91 | 1063 | 224.41 | 8.89E-40 | 2.48E-40 | 5.22E-36 |
| S9655 | 894 | 7.21 | 1086.7 | 39.0 | 1087 | 36.64 | 1077 | 94.79 | 1.51E-51 | 5.06E-52 | 3.89E-44 |
| S9701 | 895 | 12.03 | 1099.1 | 183.8 | 1097 | 157.71 | 1048 | 252.79 | 1.19E-40 | 1.12E-41 | 1.83E-33 |
| S9726 | 894 | 4.62 | 1088.0 | 179.5 | 1082 | 161.88 | 1048 | 205.82 | 2.64E-40 | 1.24E-40 | 1.73E-35 |
| S9742 | 895 | 7.21 | 1097.6 | 64.1 | 1096 | 41.73 | 1075 | 216.77 | 8.52E-49 | 8.54E-52 | 1.65E-37 |
| S9780 | 894 | 2.52 | 1079.6 | 229.6 | 1077 | 129.91 | 1039 | 366.06 | 1.09E-37 | 5.10E-42 | 6.10E-30 |

The next results include the analyses between humans, the three CEREBRA approaches, and chance. They are similar to Table 7.5 presented in Chapter 7 and Table D.1 in this appendix. The results are consistent to those presented before. However, the evaluation in this case was done using the three ratings (-1, 0, 1) to show more details for each category used in the questionnaires (decrease, no change, increase).

Table D.3: Matching CEREBRA predictions with human data (approaches one to three), compared to chance and considering all category ratings (decrease, no change, increase). Responses for each model where thus obtained for the 631 questions where all participants agreed. The table shows the average agreement of the 20 repetitions across all subjects. CEREBRA approaches one and two, agree with human responses 77%, CEREBRA approach three agrees 75%, while the chance level is 68%. None of the approaches matched any of the neutral responses. Comparison agreement with human judgements.

| ALL 4 PARTICIPANTS AVERAGE AGREEMENT (3 Ratings) | | | | | |
|---|---|---|---|---|---|
| RATINGS | HUMAN | CEREBRA#1 | CEREBRA#2 | CEREBRA#3 | CHANCE |
| -1 | 190 | 145 | 149 | 134 | 1 |
| 0 | 15 | 0 | 0 | 0 | 0 |
| 1 | 426 | 341 | 336 | 339 | 426 |
| TOTAL | 631 | 486 | 485 | 473 | 427 |
| AVERAGE | | 77% | 77% | 75% | 68% |

Table D.4: Matching CEREBRA predictions with human data (approaches one to three), compared to chance. Responses for each model where thus obtained for the 1966 questions where 3 out of 4 participants agreed. The table shows the average agreement of the 20 repetitions across all subjects. CEREBRA approaches one and two, agree with human responses 55%, CEREBRA approach three agrees 54%, while the chance level is 45%. The three CEREBRA approaches matched some of the neutral responses to some degree. Comparison agreement with human judgements.

| 3 OF 4 PARTICIPANTS AVERAGE AGREEMENT (3 ratings) | | | | | |
|---|---|---|---|---|---|
| RATINGS | HUMAN | CEREBRA#1 | CEREBRA#2 | CEREBRA#3 | CHANCE |
| -1 | 618 | 478 | 484 | 463 | 8 |
| 0 | 456 | 2 | 2 | 3 | 0 |
| 1 | 892 | 608 | 599 | 587 | 886 |
| TOTAL | 1966 | 1088 | 1085 | 1053 | 894 |
| AVERAGE | | 55% | 55% | 54% | 45% |

# Bibliography


Aguirre-Celis, N., Miikkulainen, R. (2017). From Words to Sentences & Back: Characterizing Context-dependent Meaning Representations in the Brain. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, London, UK, pp. 1513-1518.

Aguirre-Celis N., Miikkulainen, R. (2018) Combining fMRI Data and Neural Networks to Quantify Contextual Effects in the Brain. In: Wang S. et al. (Eds.). *Brain Informatics*. BI 2018. Lecture Notes in Computer Science. 11309, pp. 129-140. Springer, Cham.

Aguirre-Celis, N., Miikkulainen, R. (2019). Quantifying the Conceptual Combination Effect on Words Meanings. *Proceedings of the 41th Annual Conference of the Cognitive Science Society*, Montreal, CA. 1324-1331.

Aguirre-Celis, N., Miikkulainen, R. (2020). Characterizing the Effect of Sentence Context on Word Meanings: Mapping Brain to Behavior. *Computation and Language*. arXiv:2007.13840.

Anderson AJ, Bruni E, Bordignon, U, Poesio M, Baroni M. (2013). Of words, eyes and brains: Correlating image-based distributional semantic models with neural representations of concepts. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP 2013); Seattle, WA: Association for Computational Linguistics. pp. 1960–1970.

Anderson, A. J., Binder, J. R., Fernandino, L., Humpries C. J., Conant L. L., Aguilar M., Wang X., Doko, S., Raizada, R. D. (2016). Predicting Neural activity patterns associated with sentences using neurobiologically motivated model of semantic representation. *Cerebral Cortex*, pp. 1-17. DOI:10.1093/cercor/bhw240

Anderson, A.J., Kiela, D., Clark, S., and Poesio, M. (2017). Visually Grounded and Textual Semantic Models Differentially Decode Brain Activity Associated with Concrete and Abstract Nouns. *Transaction of the Association for Computational Linguistics* 5: 17-30.

Anderson, A. J., Lalor, E. C., Lin, F., Binder, J. R., Fernandino, L., Humpries C. J., Conant L., Raizada, R. D., Grimm, S., Wang, X. (2018). Multiple Regions of a Cortical Network Commonly Encode the Meaning of Words in Multiple Grammatical Positions of Read Sentences. *Cerebral Cortex*, pp. 1-16. DOI:10.1093/cercor/bhy110.

Anderson, A.J., Binder, J.R., Fernandino, L., Humphries, C.J., Conant, L.L., Raizada, R.D., Lin, F., & Lalor, E.C. (2019). An integrated neural decoder of linguistic and experiential meaning. *The Journal of neuroscience: the official journal of the Society for Neuroscience*.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116(3):463–498.

Barclay, J.R., Bransford, J.D., Franks, J.J., McCarrell, N.S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*:471–481.

Baroni, M., Murphi, B., Barbu, E., Poesio, M. (2010). Strudel: A Corpus-Based Semantic Model Based on Properties and Types. *Cognitive Science,* 34(2):222-254.

Barsalou, L. W. (1982). Context-independent and context-dependent: information in concepts. *Memory and Cognition*, 10:82–93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*(3):211–227. https://doi.org/10.3758/BF03196968

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, England: Cambridge University Press.

Barsalou, L., Yeh, W., Luka, B., Olseth, K., Mix, K., & Wu, L. (eds.). (1993). Concepts and Meaning. *Chicago Linguistic Society 29: Papers From the Parasession on Conceptual Representations*, 23-61. University of Chicago.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577-660.

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

Bergen, B. and Feldman, J. (2006). It's the body, stupid! Concept Learning According to Cognitive Science. *Technical report, ICSI*. TR-06-002

Binder, J. R., Desai, R. H., Graves, W. W., Conant, L. L. (2009). Where is the semantic system? A critical review of 120 neuroimaging studies. *Cerebral Cortex*, 19:2767-2769.

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

Binder, J. R. (2016a). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, 23. doi:10.3758/s13423-015-0909-1

Binder, J. R., Conant L. L., Humpries C. J., Fernandino L., Simons S., Aguilar M., Desai R. (2016b). Toward a brain-based Componential Semantic Representation. *Cognitive Neuropsychology*, 33(3-4):130-174.

Borghesani, V. (2017). The neuro-cognitive representation of word meaning resolved in space and time. *Neurons and Cognition* [q-bio.NC]. Université Pierre et Marie Curie - Paris VI; Università degli studi. Trente, Italie. NNT : 2017PA066091.

Borghi, A. M., Glenberg, A., and Kaschak, M. (2004). Putting words in perspective. *Memory and Cognition* 32:863–873.

Brain Initiative. (2013). https://braininitiative.nih.gov/

Brants, T. and Franz, A. (2006). (Linguistic Data Consortium), Philadelphia, PA. www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2006T13

Brooks L. (1978). Nonanalytic concept formation and memory for instances. *In Cognition and Categorization*. ed. Rosch & Lloyd, 169–211. Hillsdale, NJ.

Bruni, E., Tran, N., Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research (JAIR),* 49:1-47.

Bulat, L., Clark, S., and Shutova, E. (2017). Speaking, Seeing, Understanding: Correlating semantics models with conceptual representation in the brain. In *Proceedings of EMNLP*, Copenhagen, Denmark.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning with HAL. *Behavior Research Methods, Instruments, & Computers*, 30:188–198.

Choi, H., Cho, K., Bengio, Y. (2016). Context-Dependent Word Representation for Neural Machine Translation. *Computation and Language*. arXiv:1607.00578v1

Cohen, B., Murphy, G. (1984). Models of Concepts. *Cognitive Science 8:*25-78.

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.,* 29:162-173.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*(2):163–201. https://doi.org/10.1037/0096-3445.132.2.163

Cuccio, V., Ambrosecchia, M., Ferri, F., Carapezza, M., Lo Piparo, F., Fogassi, L., Gallese, V. (2014). How the Context Matters: Literal and Figurative Meaning in the Embodied Language Paradigm. *PLoS ONE,* 9(12):e115381.

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

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92:179-229.

Deng, J., Dong,W., Socher,R., Li, L., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 248–255. Miami, Florida.

Devlin, H. (2018). What is Functional Magnetic Resonance Imaging (fMRI)? https://psychcentral.com/lib/what-is-functional-magnetic-resonance-imaging-fmri/

Devlin, J., Chang, M-W, Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Computation and Language*. arXiv:1810.04805

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.

Feng, Y. and Lapata, M. (2010). Visual Information in Semantic Representation. *In Human Language Technologies*: *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 91–99. Los Angeles, California.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W., Seidenberg, M. S. (2015). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*. doi:10.1093/cercor/bhv020.

Fodor J. A. (1983) The modularity of mind: An essay on faculty psychology. *MIT Press*. Cambridge, MA.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1):71–87. https://doi.org/10.1037/0278-7393.23.1.71

Gennari, S., MacDonald, M., Postle, B., Seidenberg, S. (2007). Context-dependent interpretation of words: Evidence for interactive neural processes. *NeuroImage*, 35(3):1278-1286.

Glasgow, K., Roos, M., Haufler, A. J., Chevillet, M., A., Wolmetz, M. (2016). Evaluating semantic models with word-sentence relatedness. arXiv:1603.07253

Grand, G., Blank, I., Pereira, F., Fedorenko, E. (2018). Semantic Projection: Recovering Human Knowledge of Multiple, Distinct Object Features from Word Embeddings. arXiv:1802.01241v2

Hampton, J.A. (1991). The combination of prototype concepts. In P. Schwanenflugel (Ed.), *The psychology of word* meanings, 91-116. Hillsdale, NJ: Erlbaum.

Hampton, J. (1997). Conceptual combination. In K. Lamberts & D. R. Shanks (Eds.), *Studies in cognition*. *Knowledge, concepts and categories*, 133–159. MIT Press.

Harnad, S. (1990). The Symbol Grounding Problem, *Physica D*, 42:335-346.

Harris, Z. (1970). Distributional Structure. *In Papers in Structure and Transformational Linguistics,* 775-794.

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin and Review*, 23(6), 1744–1756.

Humphries, C, Binder, JR, Medler, DA, Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage,* 36 (3):924-932.

Humphreys, G.W., Price, C.J., & Riddoch, M.J. (1999). From objects to names: A cognitive neuroscience approach. *Psychological Research*, 62:118-130.

Janetzko, D. (2001). Conceptual Combination as Theory Formation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 23.

Just MA, Wang J, Cherkassky VL. (2017). Neural representations of the concepts in simple sentences: concept activation prediction and context effect. Neuroimage, 157:511–520.

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

Kintsch, W. & Mangalath, P. (2011). The Construction of Meaning. *Topics in Cognitive Science*, 3 (2):346-370

Kleiweg, P. & Nerbonne, J. (1998). An FGREP Investigation into Phonotactics. *Computational Linguistics in the Netherlands*: Selected Papers from the Ninth CLIN Meeting. p. 37.

Kumar, U. (2018) The neural realm of taxonomic and thematic relation: an fMRI study. *Language, Cognition and Neuroscience*, 33(5):648-658, DOI:10.1080/23273798.2017.1411962

Landau, B., Smith, L., Jones, S. (1998). Object Perception and Object Naming in Early Development. Trends in Cognitive Science, 27: 19-24.

Landauer, T. K., Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory. *Psychological Review*, 104:211-240.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Revista di Linguistica*. 20(1):1-31.

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558– 564.

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

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*:547-559.

McRae, K., & Jones, M. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*, 206–219. Oxford U. Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0014

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20:158-190.

Melamud, O., Goldberger, J., and Dagan, I. (2016). context2vec: Learning Generic Context Embedding with Bidirectional LSTM. *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL)*, pp 51–61, Berlin, Germany.

Meteyard, L., Rodriguez Cuadrado, S., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48:788–804.

Middleton, E. L.; Rawson, K. A.; Wisniewski, E. J. (April 2011). "How do we process novel conceptual combinations in context?". *Quarterly Journal of Experimental Psychology*. **64** (4): 807–822.

Miikkulainen, R. and Dyer, M. G. (1988). Forming global representations with extended backpropagation. *In Proceedings of the IEEE International Conference on Neural Networks*. Piscataway, NJ: IEEE.

Miikkulainen, R. (1990). A PDP architecture for processing sentences with relative clauses. In Karlgren, H., editor, *Proceedings of the 13th International Conference on Computational Linguistics*, 201-206, Helsinki, Finland: Yliopistopaino

Miikkulainen, R., Dyer, M., G. (1991). Natural Language Processing with Modular PDP Networks and Distributed Lexicon. Cognitive Science, 15: 343-399.

Miikkulainen, R. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*, Cambridge, MA: MIT Press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 3111–3119.

Mitchell, T.M., Shinkareva S.V., Carlson A., Chang K-M, Malave V.L., Mason R.A., and Just M.A. 2008. Predicting human brain activity associated with the meaning of nouns. *Science*, 320:1191–1195.

Mitchell, J., and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, 38(8):1388–1439. DOI: 10.1111/j.1551-6709.2010.01106.x

Morris, L.S., Kundu, P., Costi, S., Collins, A., Schneider, M., Verma, G., Balchandani, P., Murrough, J. (2019). Ultra-high field MRI reveals mood-related circuit disturbances in depression: a comparison between 3-Tesla and 7-Tesla. *Transl Psychiatry* 9, 94. https://doi.org/10.1038/s41398-019-0425-6

Murphy, B, Talukdar, P., and Mitchell. T. (2012). Selecting corpus-semantic models for neurolinguistic decoding. *In Proceedings of the First Joint Conference on Lexical and Computational Semantics*, 114–123, Montreal, Canada.

Murphy, G. (1988). Comprehending complex concepts. *Cognitive Science*, 12: 529-562.

Murphy, G. (1990). Noun Phrase Interpretation and Conceptual Combination. *Journal of Memory and Language*, 29: 259-288.

Patterson, K, Nestor PJ, Rogers, TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8:976–987.

Pecher, D., Zeelenberg, R., Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, 11: 164-167.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018). Deep contextualized word representations. *Computation and Language*. arXiv:1802.05365

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

Price A., Peelle J., Bonner M., Grossman M., Hamilton R. (2016). Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation (HD-tDCS). *Journal of Neuroscience*, 36(13):3829 –3838.

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

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics, 17*, 409-441.

Pylyshyn ZW (1984) Computation and cognition. *MIT Press*, Cambridge, MA.

Regier, T. (1996). The Human Semantic potential. *MIT Press*, Cambridge, MA.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Ruan, Y., Ling, Z., and Hu, Y. (2016). Exploring semantic representation in brain activity using word embeddings. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 669–679, Austin, Texas, November. Association for Computational Linguistics.

Rudy, J. W. (2009). Context representations, context functions, and the parahippocampal-hippocampal system. *Learning & memory*, 16(10), 573–585.

Rumelhart, D. E., McClelland, J. L., and PDP Research Group (1986) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Volume 1: Foundations. Cambridge, MA: MIT Press.

Russell, S. J., & Norvig, P. (2004). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., Flickinger, D. (2001). Multiword expressions: A pain in the neck for NLP. *In International conference on intelligent text processing and computational linguistics*, 1-15. Springer, Berlin, Heidelberg.

Silberer, C., Lapata, M. (2012). Grounded Models of Semantic Representation. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1423-1433. Jeju Island, Korea.

Silberer, C., V. Ferrari, and M. Lapata. (2013). Models of Semantic Representation with Visual Attributes. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria, 572–582.

Silberer, C., Lapata, M. (2014). Learning Grounded Meaning Representations with Autoencoders. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 721-732.

Silberer, C., Ferrari V., and Lapata, M. (2017). "Visually Grounded Meaning Representations," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2284-2297.

Takahashi, N., Motoki, M., Shimazu, Y., Tomiura, Y., and Hitaka, T. (2001). Pp-attachment Ambiguity Resolution Using a Neural Network with Modified FGREP Method. *In Proceedings of the Second Workshop on Natural Language Processing and Neural Networks*, Tokyo.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychologia*, 41(3), 280–292.

Vigliocco, G., Vinson, D. P., Lewis, W. & Garrett, M. F. (2004). The meanings of object and action words. *Cognitive Psychology*, 48, 422-488.

Vigliocco, G. & Vinson, D. P. (2007). Semantic representation. In G. Gaskell (ed.), *Handbook of Psycholinguistics*. Oxford: Oxford University Press.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 219-247. doi:10.1515/LANGCOG.2009.011

Vigliocco G, Kousta ST, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF (2014). The neural representation of abstract words: the role of emotion. *Cereb Cortex* 24:1767–1777.

Vinyals, O., Toshev, A., Bengio, S., Erhan, D. (2015). Show and Tell: A New Image Caption Generator. arXiv:1506.03134v2

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*(5), 719–736.

Wisniewski, E. J. (1997). When concepts combine. Psychonomic Bulletin & Review, 4, 167–183.

Wisniewski, E. (1998). Property Instantiation in Conceptual Combination. *Memory & Cognition*, 26, 1330-1347.

Yacoub, E., Shmuel, A., Pfeuffer, J., Moortele, P-F, Adriany, G., Andersen, P., Vaughan, J.T., Merkle, H., Ugurbil, K., Hu, X. (2001). "Imaging Brain Function in Humans at 7 Tesla", Magn Reson Med; 45: 588-594.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2013). Semantic Memory. In K. Ochsner & S. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience,* (1), pp. 353–374. Oxford, United Kingdom: Oxford University Press.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, 23, 1015–1027.

Yee, E., Jones, M. N., & McRae, K. (2018). Semantic Memory. In J. T. Wixted & S. L. Thompson-Schill (Eds), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, Vol. 3: Language and Thought, pp. 319-356. New York: Wiley.

# Vita

Nora Elsa Aguirre Sampayo (a.k.a. Nora E Celis, Nora Aguirre-Celis) was born in Monterrey, Mexico. She received her bachelor's and master's degrees in Computational Technology from ITESM, and her master's degree in Computer Science from Indiana University. She worked at ITESM for fifteen years as an assistant professor. She holds a visiting scholar position in the Department of Computer Science at the University of Texas in Austin under the supervision of Prof. Risto Miikkulainen. She participated in the Brain Initiative Project (USA) to develop innovative technologies to understand brain function (2013-2015). She has published papers in the Cognitive Science Conference, Computational Neuroscience Meeting, Brain Informatics Conference, COLING-CogAlex Workshop and Semantic Spaces Workshop.

This dissertation was typed by the author.