Fachbereich Philosophie und Geisteswissenschaften

der Freien Universität Berlin

Brain Mechanisms of Language and Semantic Processing in Sighted and Congenitally Blind Populations: A Neurobiologically Constrained Model

Dissertation

zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

vorgelegt von

Rosario Tomasello

Berlin 2019

Erstgutachter:	Prof. Dr. Dr. Friedemann Pulvermüller
Zweitgutachter:	Prof. Dr. Angelo Cangelosi
Tag der Disputation:	8. Mai 2019

Table of Contents

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Friedemann Pulvermüller for the constant support and motivation for my doctoral project and related research. I could not imagine having a better mentor for my PhD study.

Beside this, I would also like to thank Dr Max Garagnani and Dr Thomas Wennekers for their precious support and guidance that made possible the present research work. My thanks go out also to my colleagues of the brain language laboratory (BLL) at the Freie Universität Berlin for the stimulating discussions during these years: Dr Luigi Grisoni, Dr Gugliemo Lucchese, Tally Miller, Konstantina Margiotoudi, Dr Felix Dreyer, Dr Malte Henningsen, Lea Doppelbauer, Isabella Boux, Verena Buscher and Verena Arndt. An exceptional thanks go to my cohort of the Berlin School of Mind and Brain at the Humboldt University that made this journey much more pleasant.

I would also like to thank my friends Dario, Enrique, Luca, Eva, Martina and my brother Antonio for encouraging me throughout this experience. Big thanks also go out to the dance community, where I found the perfect balance of everyday life. Special thanks go to Mascha for all her love and day by day support in every stage of this journey and for encouraging me in all my pursuits.

Last but not least, I would like to thank my parents that supported me in every stage of my personal and academic life to whom I dedicate this thesis.

Abstract

Understanding the meaning of words and its relationship with the outside world involves higher cognitive processes unique of the human brain. Despite many decades of research on the neural substrates of semantic processing, a consensus about the functions and components of the semantic system has not been reached among cognitive neuroscientists. This issue is mainly influenced by two sets of neurocognitive empirical findings that have shown (i) the existence of several regions acting as 'semantic hubs', where the meaning of all types of words is processed and (ii) the presence of other cortical regions specialised for the processing of specific semantic word categories, such as animals, tools, or actions. Further evidence on semantic meaning processing comes from neuroimaging and transcranial magnetic stimulation studies in visually deprived population that acquires semantic knowledge through non-sensory modalities. These studies have documented massive neural changes in the visual system that is in turn recruited for linguistic and semantic processing. On this basis, this dissertation investigates the neurobiological mechanism that enables humans to acquire, store and processes linguistics meaning by means of a neurobiologically constrained neural network and offers an answer to the following hotly debated questions: *Why* both semantic hubs and modality-specific regions are involved in semantic meaning processing in the brain? *Which* biological principles are critical for the emergence of semantics at the microstructural neural level and *how* is the semantic system implemented under deprived conditions, in particular in congenitally blind people?

First, a neural network model closely replicating the anatomical and physiological features of the human cortex was designed. At the micro level, the network was composed of 15,000 artificial neurons; at the large-scale level, there were 12 areas representing the frontal, temporal, and occipital lobes relevant for linguistic and semantic processing. The connectivity structure linking the different cortical areas was purely based on neuroanatomical evidence. Two models were used, each simulating the same set of cortical regions but at different level of details: one adopted a simple connectivity structure with a mean-field approach (i.e. graded-response neurons), and the other used a fully connected model with adaptation-based spiking cells. Second, the networks were used to simulate the process of learning semantic relationships between word-forms, specific object perceptions,

and motor movements of the own body in deprived and undeprived visual condition. As a result of Hebbian correlated learning, distributed word-related cell assembly circuits spontaneously emerged across the different cortical semantic areas exhibiting different topographical distribution. Third, the network was reactivated with the learned auditory patterns (simulating word recognition processes) to investigate the temporal dynamics of cortical semantic activation and compare them with real brain responses.

In summary, the findings of the present work demonstrate that meaningful linguistic units are represented in the brain in the form of cell assemblies that are distributed over both semantic hubs and category-specific regions spontaneously emerged through the mutual interaction of a single set of biological mechanisms acting within specific neuroanatomical structures. These biological principles acting together also offer an explanation of the mechanisms underlying massive neural changes in the visual cortex for language processing caused by blindness. The present work is a first step in better understanding the building blocks of language and semantic processing in sighted and blind populations by translating biological principles that govern human cognition into precise mathematical neural networks of the human brain.

Zusammenfassung

Um die Bedeutung von Wörtern und ihre Beziehung zur Außenwelt zu verstehen, müssen die kognitiven Prozesse betrachtet werden, die einzigartig für das menschliche Gehirn sind. Trotz jahrzehntelanger Forschungen an den neuronalen Substraten der semantischen Verarbeitung im menschlichen Gehirn wurde bisher kein Konsens über die Funktionen und Komponenten des semantischen Systems in den kognitiven Neurowissenschaftlern erreicht. Dieses Problem gründet darin, dass neurokognitive empirische Studien zumeist zu zwei Endergebnissen kamen: (i) der Existenz von mehrere Regionen, die als 'semantische Hubs' fungieren, in denen die Bedeutung aller Wortarten verarbeitet wird, und (ii) dem Vorhandensein weiterer kortikaler Regionen, die auf die Verarbeitung spezifischer semantischer Kategorien wie Tiere, Werkzeuge oder Aktionswörtern spezialisiert sind. Ein weiterer Beweis für die Verarbeitung semantischer Bedeutungen lässt sich aus Bildgebungsstudien und Studien mit transkranialer Magnetstimulation an visuell benachteiligten Probanden entnehmen, die die linguistische Bedeutung nicht durch sensorische Modalitäten erwerben. Diese Studien konnten massive neuronale Veränderungen im visuellen System dokumentieren, die wiederum für die sprachliche und semantische Verarbeitung verwendet werden. Die vorliegende Dissertation untersucht mittels eines biologischen neuronalen Netzwerkes jene kognitiven Prozesse, die es Menschen ermöglichen, linguistische Bedeutungen in der täglichen Kommunikation zu erfassen, zu speichern und zu verarbeiten. Sie schlägt Antworten auf die folgenden neurowissenschaftlich heiß diskutierten Fragen vor: *Warum* sind sowohl semantische Hubs als auch modalitätsspezifische Regionen relevant für die sprachliche und semantische Informationsverarbeitung im Gehirn? *Welche* biologischen Prinzipien sind von entscheidender Bedeutung für die Entstehung von Semantik auf mikrostruktureller neuronaler Ebene? Und *Wie* ist das semantische System unter benachteiligten Bedingungen repräsentiert?

Zunächst wurde ein neuronales Netzwerkmodell implementiert, das die anatomischen und physiologischen Merkmale des menschlichen Kortex präzise widerspiegelt. Auf der Mikroebene besteht das Netzwerkmodel aus 15.000 künstlichen Neuronen, auf der Großebene aus 12 Arealen der Frontal-, Temporal- und Okzipitallappen, die für die sprachliche und semantische Verarbeitung relevant sind. Die Verbindungsstruktur zwischen

den verschiedenen kortikalen Arealen wurde rein auf Grundlage von neuroanatomischen Befunden implementiert. Zwei Modelle wurden verwendet, die jeweils die gleichen kortikalen Regionen simulierten, allerdings in verschiedenen Varianten: Das erste Modell ging von einer einfachen Konnektivitätsstruktur mit einem Mean-field Ansatz (graded-response neurons) aus, während das zweite einen vollständig verbundenen Aufbau mit adaptionsbasierten Spiking-Zellen (Aktionspotential) verwendete. Anschließend dienten die neuronalen Netzwerke dazu, den Lernprozess der semantischen Verlinkung zwischen Wortformen, bestimmten Objektwahrnehmungen und motorischen Bewegungen des eigenen Körpers zu simulieren, sowohl in gesundem als auch in benachteiligtem Sehzustand. Als Ergebnis des Hebbschen Korrelationslernens traten spontan verteilte Neuronenverbindungen (cell assemblies) in den verschiedenen kortikalen semantischen Bereichen auf, die unterschiedliche topografische Verteilungen zeigten. Zuletzt wurde das Netzwerkmodell mit den erlernten auditorischen Mustern reaktiviert (Worterkennungsprozesse), um die zeitliche Dynamik kortikaler semantischer Aktivierung zu untersuchen und sie mit realen Gehirnantworten zu vergleichen.

Die vorliegende Arbeit kam zu folgenden Ergebnissen: Die neuronale Repräsentation linguistischer Bedeutung wird im Gehirn in Form von cell assemblies dargestellt, welche über semantische Hubs und modalitätsspezifische Regionen verteilt sind. Diese entstehen spontan durch die Interaktion einer Reihe von biologischen Mechanismen, die innerhalb spezifischer neuroanatomischer Strukturen wirken. Das Zusammenwirken dieser biologischen Prinzipien bietet zusätzlich eine Erklärung für jene Faktoren, die für die massiven neuronalen Veränderungen in der sprachlichen und semantischen Netzwerke bei Blindheit verantwortlich sind. Die in dieser Dissertation dokumentierten Studien sind ein erster Schritt in Richtung eines besseren Verständnisses der sprachlichen und semantischen Informationsverarbeitung bei sehenden und blinden Menschen, basierend auf einer Übersetzung der biologischen Prinzipien der menschlichen Kognition in präzise mathematische neuronale Netzwerke des menschlichen Gehirns.

List of publications

This dissertation is based on the following publications:

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., (2017). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia* 98, 111–129. doi:http://doi.org/10.1016/j.neuropsychologia.2016.07.004.

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., (2018). A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity. *Front. Comput. Neurosci.* 12, 88. doi:https://doi.org/10.3389/fncom.2018.00088

Tomasello, R., Wennekers, T., Garagnani, M., Pulvermüller, F., (2019). Recruitment of visual cortex for language processing in blind individuals is explained by Hebbian learning. *Scientific Reports* 9(1):3579. doi:https://doi.org/10.1038/s41598-019-39864-1

List of abbreviations

A1 = Primary auditory cortex

AB = Auditory belt area

AT = Anterior temporal area

ATL = Anterior temporal lobe

BA = Brodmann's area

CA = Cell assembly

DTI = Diffusion tensor imaging

DWI = Diffusion weighted imaging

EEG = Electroencephalography

E-cell = Excitatory cell

fMRI = Functional magnetic resonance imaging

I-cell = Inhibitory cell

ISI = Interstimulus interval

LTD = Long-term depression

LTP = Long-term potentiation

$M1_i$ = Inferior primary motor cortex

$M1_L$ = Lateral primary motor area

MEG = Magnetoencephalography

PB = Parabelt area

PF_i = Inferior prefrontal motor cortex

PM_i = Inferior premotor area

PM_L = Lateral premotor region

PF_L = Lateral prefrontal area

SD = Standard deviation

SE = Standard error

TMS = Transcranial magnetic stimulation

TO = Temporo-occipital area

V1 = Primary visual area

1. General Introduction

From the first year of life, information on the environment, everyday experiences, encounters with people, and factual data is acquired, stored, and processed in the human brain. Language is the main canal to disseminate this knowledge, and it does so in the form of strings of sounds or symbols referring to entities in the world. The mechanism in how this knowledge is mapped into words (linguistic symbols) has concerned philosophers since Plato, Aristotle, and Frege, all of whom contributed important reflections on language structure, word meaning, and linguistic categories (for an overview, see Runes, 1984). Although semantics has a long history in philosophy and linguistics, offering detailed theoretical models on the understanding of linguistic meaning, still little is known on the neural basis of semantic knowledge processing in the human brain. Patients with an acquired brain injury to the relevant language and semantic regions develop severe language impairments, and in turn, face difficulties in everyday skilled actions, ranging from planning complex activities to basic actions such as drinking a cup of coffee (Bak and Chandran, 2012; Damasio et al., 1996; Gainotti, 2010; Kemmerer et al., 2012; Pulvermüller and Fadiga, 2010). Hence, it is essential to determine the putative neural mechanisms of semantic processing in the human brain by seeking answers to the following long-standing debated questions in cognitive science and neuroscience:

1) *Where* is the cortical locus relevant for semantic processing in the human brain?
2) *How* is semantic meaning implemented at the neural circuit level and *which* are the relevant biological mechanisms behind it?
3) *When* in time the cortical semantic areas first emerge during meaning processing?
4) *Why* and *how* semantic representation is influenced by deprived conditions, such as the absence of visual experience?

This chapter provides an overview of the different philosophical and linguistic semantic theories and on what we know from decades of research on the representation and processing of semantic knowledge in the brain. It also introduces the specific questions that motivated this work. The last sections of this chapter discuss basic modelling assumptions and recent advances in neurocomputational models. It is revealed how realistic neural networks of the cortex, which are able to reproduce temporal and spatial aspects of brain

activity, are required to bridge the gap between semantic theories, brain data, and cognitive semantic behaviour.

Semantic theories

What is the meaning of the words '*run*', '*dog'* or '*house*'? Most likely a person will point out to what the words *refer* in the real world. The mapping between a word and its referent—namely how words, sign and symbols convey meaning—is a matter of longstanding debate among philosophers and linguists. A basic description of the relationship between a word and its referent was proposed by Saussure (1959) who divided the sign (defined as a basic linguistic unit) into two focal components: the *signifier*, the sound pattern (the word-form) described as an abstract entity, and the *signified,* the element indicated by the signifier. For Saussure, what makes meaning is the mutual and close relationship between these two components that cannot be separated. In contrast to the dyadic relationship, Ogden and Richards (1923) see language in terms of a triangle, arguing that the relationship between the sign 'dog' and its referent in the real world, is indirectly mediated by a mental representation (i.e., thought). Without the thought, the mapping between a word and its referent is not possible. Although this position has found many critics over the years (see for instance Alston, 1964), it has also found several followers proposing slightly different triadic models (Jackendoff, 1983; Lakoff, 1988). Note that the principle of arbitrariness of a linguistic sign is not a novel one: Aristotle in the *De Interpretatione* has already argued against a direct connection between a word and its signified (see, Ackrill, 1963), while in contrast, Plato in the *Cratylus* believed that words are correctly named because their meaning resides in the direct relationship with its referent in the real world (see, Sedley, 2003). However, all these positions share a belief that both aspects of meaning, concept and referent, are essential for the correct use of the sign in real-world interactions.

The science of words, signs and symbols has not always found a central position in the study of linguistics, which has focused mainly on syntax and grammar. For instance, the most influential linguist, Noam Chomsky, sees the study of meaning as peripheral to linguistic investigations, arguing that language structure (syntax) is driven by mechanisms independent from meaning (Chomsky, 1965). Based on Chomsky's theory of generative grammar, Katz and Fodor (1963) gave some space to semantics, including it as a component

of the generative grammar theory. The authors believed, however, that the meaning of a word is based on the semantic components or features of which a sign is composed. 'As a rule, the meaning of a word is a compositional function of the meanings of its parts …' (p. 191, Katz and Fodor, 1963). For example, the word 'dog' includes features of + animal, + four legs, + move, but not the feature of a human being (– human). Traditional cognitive semantic theories support the aforementioned compositionality theory, arguing that the meanings of signs and symbols are processed in a unique module in the human brain different from other cognitive processes (Fodor, 1983, 1975). In particular, once syntactic and grammatical structures are set, the full understanding of signs, words and sentences relies on domain general processes far from sensorimotor modules. The basis of this theory is on the restriction of information flow between the modules, in which each module relies on information stored on its own having no access to information outside of it (Ellis and Young, 1988; Fodor, 1983). Similarly, other semantic theories assume that the meaning of a word is stored in semantic networks based on the relationships between linguistic signs in a uniform format (Collins and Loftus, 1975; Quillian, 1969). Such approaches argue for an 'amodal symbolic system' in the brain, leaving out the importance of perceptual (or motor) experience in the real world (i.e., the referential context) for semantic knowledge.

However, a caveat of these semantic theories, according to Harnad (1990) in his famous work '*The Symbol Grounding Problem'*, is best explained by the Chinese room thought experiment (Searle, 1980). If a native English speaker without knowing Chinese is locked in a room and Chinese symbols are given referring to other symbols following a set of rules, the person in question will not understand the meanings of words that refer to other symbols. This is because a circulatory system, which continuously refers to other symbols, fails to explain how these words refer to things in the outside world. In other words, meaning cannot be conveyed on its own in a purely symbolic module. Clearly there is the need to relate words (the word-form, 'dog') to what they refer in the world (the animal, dog), namely grounding words in sensory and motor experiences (Harnad, 1990). Grounded (or embodied) theories suggest that the meaning of words is not processed in a unique semantic module but grounded in action and perception systems of the human brain (e.g., Barsalou, 1999; Pulvermüller, 1999a). Semantics is believed to be represented in a distributed modality-specific format, and that sensory and motor information is what

constitutes a concept. Once a basis of lexicon has been acquired through grounding mechanisms, it can posit the foundation for learning an ample vocabulary from linguistic descriptions or indirect context, where the combinatorial power of language plays a key role (Cangelosi et al., 2002; Harnad, 2012, 1990; Stramandinoli et al., 2012a).

Brain representation of semantic processing

The debate between an amodal symbolic system and a grounded approach for meaning processing in the human brain is still central. Over the past 20 years, neuroscience investigations have measured various ways, in which semantic knowledge is constructed and processed in the brain, providing support for both amodal and modal grounded approaches (for a review see Binder and Desai, 2011; Pulvermüller, 2013).

Recent amodal semantic approach postulate a symbolic representations of meaning processing in the anterior temporal lobe (ATL), functioning as an integration centre (semantic hub), where the meanings of all types of signs and symbols are equally processed (Leshinskaya and Caramazza, 2016; Mahon and Caramazza, 2008; Patterson et al., 2007). One of the central pieces of evidence for the presence of such amodal representation comes from patients suffering from semantic dementia with focal brain damage in the ATL, which showed a widespread loss of all types of conceptual knowledge (Hodges et al., 1992; Patterson et al., 2007). However, functional imaging studies (fMRI) have documented the presence of other semantic hub regions for general semantic processing located in the anterior inferior parietal cortex (Binder et al., 2009; Binder and Desai, 2011; Pulvermüller, 2013) and the posterior inferior frontal cortex (Bookheimer, 2002; Carota et al., 2017; Posner and Pavese, 1998; Tate et al., 2014). Although, there are robust findings for the existence of abstract symbolic systems in the brain, why multiple regions having such a functional role is present in the human brain still needs to be clarified.

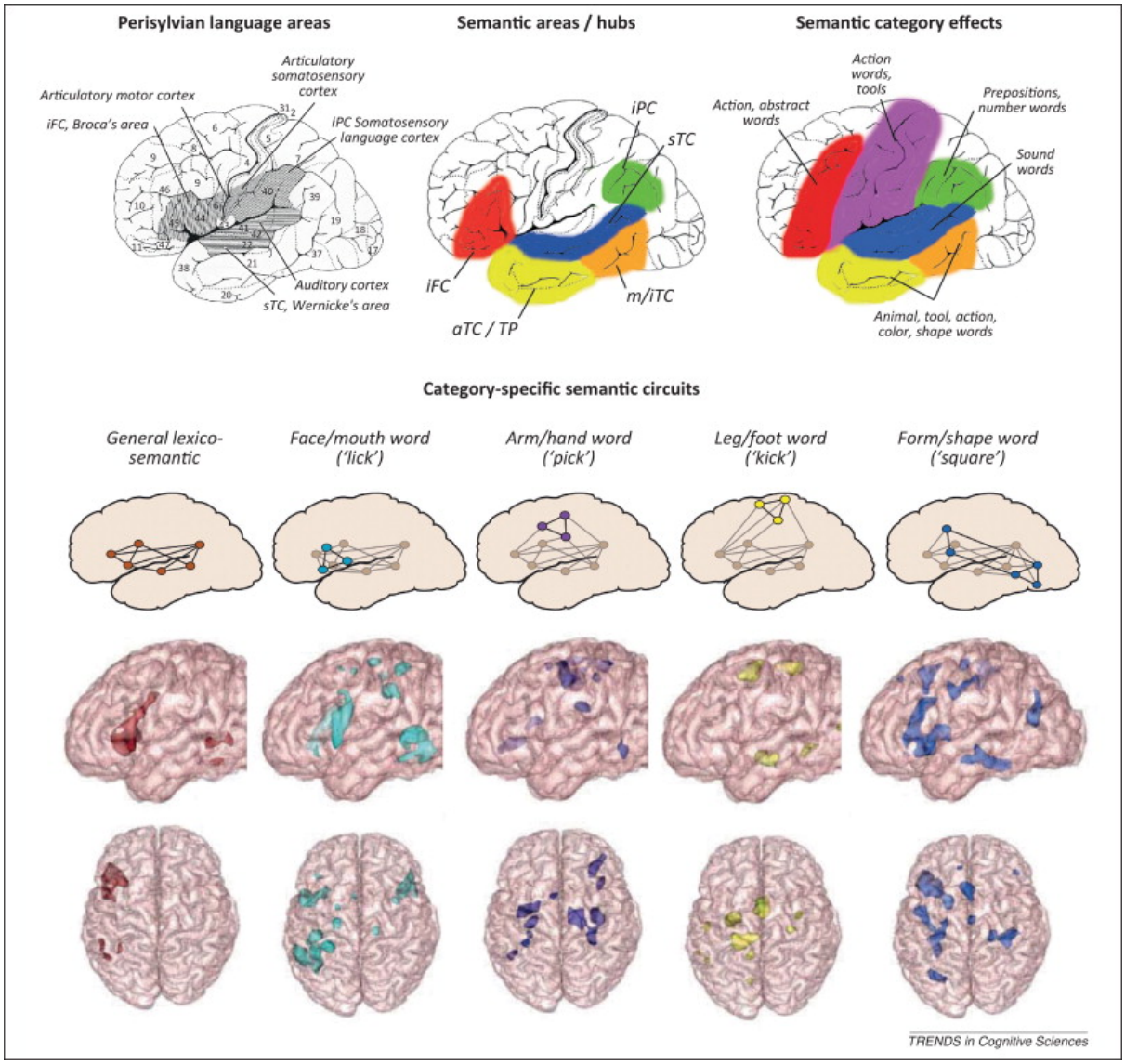

Figure 1.1 On the top, left perisylvian language area, semantic hubs and category-specific areas typically activated during word meaning processing reported in the literature. Bottom panel: fMRI study documenting category-specific semantic activation of 4 different semantic word types and their cell assembly representation. Abbreviations: iFC, inferior frontal cortex; iPC, inferior parietal cortex; sTC, superior temporal cortex; m/iTC, middle/inferior temporal cortex; aTC, anterior temporal cortex; TP, temporal pole (data adapted from Pulvermüller, 2013).

On the other hand, modal grounded approach has been supported by an increasing amount of neurocognitive studies, highlighting the significance of action and perception systems of the human brain for meaning processing. A number of empirical studies have focused on the importance of the motor cortex showing that perceiving an action-related verb such as 'run', 'grasp', or 'kiss' evokes activity in the motor and premotor cortices, intriguingly in a somatotopic manner (Grisoni et al., 2017; Hauk and Pulvermüller, 2004; Kemmerer, 2015; Shtyrov et al., 2004; Vukovic et al., 2017). Likewise, activity in the visual system for perceiving visually related words such as 'sun', 'dog', or 'house' is more pronounced (Chao et al., 1999; Kiefer, 2005; Moseley et al., 2013; Pulvermüller et al., 1999; Sim and Kiefer, 2005). Such grounded representation on modality-specific cortical areas has also been consistently documented for emotion, taste, smell, and colour words (e.g. Fernandino et al., 2015; Kemmerer, 2015; Martin, 2016; Pulvermüller, 1999; Simmons et al., 2005). An explanation of modality-specific cortical activation for specific semantic categories (e.g. visually-related, animals, tools, or action-related) has been proposed based on the Hebbian cell assembly (CA) mechanism (Hebb, 1949). In other words, the mapping between a sign and its reference in the outside world would lead to the formation of neuronal cell assemblies, distributed over perisylvian language networks and category-specific sensorimotor circuits by means of Hebbian synaptic modifications and correlation learning (Fig. 1.1, Pulvermüller, 2018, 2013, 2002, 1999).

On this basis, the critical question posited by the present research is *why* there are both semantic hubs and modality-preferential cortical regions in the human brain? And *Which* biological principles are critical for the emergence of semantics at the microstructural neural level?

Newly integrative proposals have been often emphasized by different theoretical hydride models (Binder and Desai, 2011; Damasio et al., 1996; Pulvermüller, 2013; Ralph et al., 2017), which offered explanations for the presence of both category-general and category-specific functions for semantic knowledge in the human brain. Lambon Ralph and colleagues have proposed the so-called hub-and-spoke model of semantic cognition (Ralph et al., 2017) suggesting that modality-specific spokes are linked to a single hub region situated in the anterior temporal lobe (ATL) where semantic information converge. Although this model

explains some of the behavioural changes seen in semantic dementia patients, it leaves unexplained the presence of other reported semantic hubs and their functional roles during conceptualization. Interestingly, recent neurocognitive studies (Gainotti, 2012; Shebani et al., 2017; Silveri et al., 2018) have reported category-specific semantic impartments in semantic dementia patients with a pattern of atrophy in the anterior temporal lobe, which sit less well with a general category semantic processing there (Patterson et al., 2007; Ralph et al., 2017). Although, other theoretical proposals have integrated the full set of cortical areas (multimodal hubs and modality-specific regions) documented by empirical studies (Binder and Desai, 2011; Pulvermüller, 2013), a pure conceptual (or descriptive) model alone is not able to provide qualitative and quantitative mechanistic explanations of the biological principles underlying word meaning processing. In particular, how the mutual interaction of neural mechanisms give rise to the extensive reported category-general and category-specific semantic functions. Likewise, future empirical research on this issue might lead to offer additional traces of activity of the complex network underlying semantic processing without reaching the final goal of revealing the mechanism behind it. For instance, a recent fMRI study (Huth et al., 2016) employing a novel approach of voxel-wise modelling associated with neural encoding have produced an interesting atlas of semantic brain areas active during natural speech processing, nevertheless the underlying neural mechanisms and the functional role of the revealed cortical regions during meaning processing were left unexplained (for discussion see, Barsalou, 2017).

To bridge the gap between semantic theories and experimental brain data, biological-constrained computational models following precise neurobiological principles have been defined to be necessary to clarify the putative neural mechanisms underlying language and semantic processing, as well as making new testable and verifiable predictions (Barsalou, 2017; Breakspear, 2017; Pezzulo et al., 2013). The present thesis describes such a model and how it was applied to simulate word meaning acquisition, storage, and processing, in turn, to provide a unified neurobiological model of human semantic processing.

Semantic processing under visually deprived condition

Recent evidence of semantic processing, in particular on the association between perception, sensory, and meaning processing, comes from blind individuals who acquire their vocabulary exclusively through non-visual inputs. Loss of vision from an early age has been shown to cause neural plastic changes in the visual system, affecting semantic representation in the brain (Amedi et al., 2004, 2003; Burton et al., 2002; Raz et al., 2005; Struiksma et al., 2011; see also Burton et al., 2003). Lesion and functional imaging studies have documented the activation of visual cortices (including the V1 area) in blind individuals during semantic retrieval in a verb production task (upon hearing the noun 'cake', they would produce 'bake') (Amedi et al., 2003; Burton, 2002; Struiksma et al., 2011). This corresponds to similar visual cortical areas being responsive for visual/object-related word processing in sighted people (Chao et al., 1999; Kiefer, 2005; Moseley et al., 2013a; Sim and Kiefer, 2005). Likewise, virtual disruption of the deprived primary visual area (V1) using transcranial magnetic stimulation (TMS) leads to verb production impairments with higher semantic errors in blind, but not in sighted individuals (Amedi et al., 2004). An explanation for the involvement of visual areas in sighted individuals during language processing is best described by the grounding perspective of semantic processing described above, in which concepts with high visual features, such as 'sun', 'dog' or 'cup', are grounded in areas relevant for the processing of sensory information (Barsalou et al., 2003; Harnad, 1990; Pulvermüller, 2018a). However, this grounding mechanism cannot explain why the deprived visual cortex is active during the retrieval and generation of a verb that typically describes actions, as blind people are unable to visualise the objects/actions that are used to speak about. Likewise, it is difficult to see how amodal symbolic approach can provide an explanation for the aforementioned plastic changes in the blind brain, as if semantic knowledge is processed in a unique abstract brain region far from sensorimotor regions, it should be so for both blind and sighted individuals. Hence, the critical questions posit by the present research work are as follow: *what* are the critical factors for the takeover of the visual cortex for linguistic and semantic processing under visual deprivation? *How* its emerge at the neural circuit level in the human brain?

A possible explanation for the involvement of the visual cortex in blindness could be based on different anatomical projections between visual and the relevant language

regions, however a number of diffusor tensor imaging (DTI) studies have failed to find any increase anatomical connectivity in blind compared to sighted individuals (Noppeney et al., 2005; Shimony et al., 2005; Shu et al., 2009a, 2009b). While it is clear that the neural changes caused by blindness are mediated by the anatomical input from frontal, temporal and parietal cortical regions (Bedny, 2017), little is known how it affects specifically the neural changes in the deprived visual areas. Another possible explanation is that under sensory deprivation no competing inputs are present in early visual cortices during development, hence leading to the functional recruitment of such areas by the remaining modality, such as language, however, the neuromechanism principles guiding these processes remain debated (Amedi et al., 2017; Bedny, 2017; Heimler et al., 2015). Providing an answer to the aforementioned questions of the neural reorganization of the visual system functionally recruited for language and semantics in the absence of visual experience is a challenge of neurobiological models of the human brain. Importantly, it could provide further critical insights in how semantic knowledge is represented in the brain under deprived conditions and in particular, could help in deciding how artificial retinas should be implanted to restore sensory loss.

How neurocomputational models can help

Computational models are precise mathematical models of artificial neurons used to simulate and investigate the dynamic processes behind the behaviour of complex systems, such as language processing. Artificial neural models of the cortex can serve a variety of purposes, ranging from testing hypotheses, generating specific and quantitative predictions, to replicating (i.e., interpreting) experimental results. A simulation obviously cannot replace experimental data, but they are useful tools to demonstrate whether or not a proposed idea, mechanism, or detailed verbal description of a brain function is valid. Importantly, to ensure the outputs and explanations of the model are biologically plausible, the neural architecture needs to mimic the well-known processes of the behaviour it is based on. Hence, neuroanatomical and neurophysiological details should be included in the neural architecture, as structural and functional network properties have been defined essential features for modelling higher-order cognitive processes (Breakspear, 2017; O'Reilly, 1998; Pezzulo et al., 2013).

Traditional neural networks started to model single neurons to investigate the basic principle and dynamics of the human brain (e.g., McCulloch and Pitts, 1943; Palm, 1982; Willshaw et al., 1969). Although modelling at the single level is important to understand the communication between minimal parts of the cortex (i.e. neurons), this method is far from understanding higher cognitive functions that are associated with widely distributed neural activation involving numerous cortical and subcortical regions (Bressler and Menon, 2010; Sporns et al., 2004). It is widely accepted that larger groups of interconnected neurons distributed across the cortex carry information underlying complex cognitive processes, especially in the domains of language processing, visual and attention (Bressler and Kelso, 2001; Bressler, 2002; Damasio and Damasio, 1994; Jirsa, 2004; McIntosh, 2000). The network types most used to simulate language processing are the so-called connectionist models, which are made up of layers of connected neurons that are weighted depending on the learning rule. These models are inspired and derived by biological systems, such as the human brain, and used to simulate how activity propagates across interconnected units of neurons and in turn to investigate how cognition works at a mechanistic level. Computational models also provide the best paradigm for cognitive robotics and artificial intelligence that can get us closer to understand higher cognitive functions of the human brain in a fully situated environment, such as language processing and social interactions (Cangelosi, 2006; Cangelosi et al., 2010; Sporns, 2007; Stramandinoli et al., 2012b).

The starting point of neural network research goes back to 1943 with Warren McCulloch and Walter Pitts's paper '*A Logical Calculus of Ideas Immanent in Nervous Activity*' that showed for the first time how a neuron could be implemented as a simple mathematical operation. Inspired by biological neurons that either generate a spike (action potential) or remain silent, the activity of the artificial cells in the network was defined in binary units, with the state of 1 for active and 0 for inactive. Each neuron was constructed with an internal threshold that defined its state, which was subjected to the synaptic modification with other cells (McCulloch and Pitts, 1943). Around the same period, Donald Hebb (1949) developed a theory of cognition and learning that is best summed up by his phrase, 'Neurons that fire together wire together'. He proposed that the mechanism for how communication between neurons and the changing of synapses between them is based on a simple form of associative learning. Specifically, he wrote the following:

> ‘When an axon of cell A is near enough to excite B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased’ (Hebb, 1949).

This mechanism postulated by Hebb explains how, by means of correlated activity, repeatedly co-activated neurons wire together, forming the so-called cell assemblies (CA). The interconnected CA neurons have been assumed to represent the building blocks of cognitive functions, such as language and semantics, which is supported by extensive empirical studies positing the foundation of development, learning, and memory (e.g., Braitenberg, 1978; Palm et al., 2014; Pulvermüller, 1996).

McCulloch and Pitts’s contributions and Hebb’s notion of synaptic modification provided the groundwork for computational models, which in the last decades have provided valuable insight into ‘how’ and ‘why’ the brain processes language in a particular way (e.g. Chen et al., 2017; Dell, 1986; Elman, 1991; Farah and McClelland, 1991; Gaskell et al., 1995; Joanisse and Seidenberg, 1999; Norris, 1994; Plaut and Gonnerman, 2000; Plunkett and Marchman, 1996; Rumelhart et al., 1986; Seidenberg and McClelland, 1989; Ueno et al., 2011). Farah and McClelland (1991) implemented the first connectionist model for semantic processing, which consisted of two layers of functional and visual hidden units representing semantics as well as two layers of input and output nodes. The network was trained with living and non-living things, which were generated as random patterns of activity. After training, the model was lesioned to each semantic unit, producing category–specific impairments which were similarly documented in neurophysiological data (e.g., Basso et al., 1988; Miceli et al., 1988). Although this first semantic model replicated and interpreted the results of several studies, it was far from what we know from biology, hence, suffering from a lack of neuroscientific plausibility. Incorporating biological principles of the human brain have been defined to be essential to model and understand human-specific cognitive functions and their related mechanisms (Breakspear, 2017; Pezzulo et al., 2013). Recently computational approaches have indeed reached a higher degree of realism, incorporating fine microstructural and functional details of millions of neurons (Izhikevich and Edelman, 2008; Markram et al., 2011), but they have not yet addressed specific questions about the neurobiological basis of specific cognitive functions, such as semantic processing. Such an approach believe that solely in implementing realistic neural architecture will lead to the

spontaneous emergence of cognitive functions in the network. Other studies have addressed important aspects of language and semantic processing (e.g., Christiansen and Chater, 2001; Dell et al., 1999), but most did not attempt to mimic neurophysiological mechanisms, neuroanatomical structure, or realistic learning mechanisms. While recent simulation studies of semantic brain processes have included neuroanatomical information (Chen et al., 2017; Guenther et al., 2006; Ueno et al., 2011), they used learning mechanisms (i.e., backpropagation), which have been argued to be implausible for cortical networks (Mazzoni et al., 1991; O'Reilly, 1998). Hence, in order to advance research on how language and semantics are represented in the human brain, it is critical to implement precise mathematical models that mimic the properties and functions of the human brain based rigorously on biological principles.

Neurobiologically constrained neural network

In this study, we specifically designed a biologically constrained model of semantic brain areas, mimicking realistic properties of the human cortex to reconcile the diverging experimental evidence of semantic processing described above. These properties included connectivity, neurophysiology, and neuroanatomical functions and structures. At the micro level, the neural network implements physiologically artificial neurons. At the system level, there are 12 areas of relevance for language and semantic processing situated in the frontal, temporal, and occipital lobes. We mimicked the left perisylvian cortex involved in processing spoken words and their corresponding auditory-phonological signals (Fadiga et al., 2002; Pulvermüller, 1999; Pulvermüller and Fadiga, 2010; Zatorre et al., 1996).

- The 'auditory stream': the primary auditory cortex (A1), auditory belt (AB), and modality-general parabelt areas (PB).
- The 'articulatory stream': the inferior part of the primary motor cortex ($M1_i$), inferior premotor (PM_i) and multimodal prefrontal motor cortex (PF_i).

An additional six areas outside the perisylvian cortex (extrasylvian system) modelled referential meaning-related information about visual object identity (Ungerleider and Haxby, 1994), and about executable manual actions (Deiber et al., 1991; Dum and Strick, 2005, 2002; Lu et al., 1994):

- The 'ventral visual stream': the primary visual cortex (V1), temporo-occipital (TO), and anterior temporal (AT) areas.
- The 'dorsolateral motor stream': lateral primary motor ($M1_L$), premotor (PM_L), and prefrontal (PF_L) cortices.

The cortical connectivity structure between the areas closely reflects the existing neuroanatomical studies using diffusion tensor and diffusion-weighted imaging (DTI/DWI) in humans and non-human primates. Table 3.2 in Chapter 3 summarises the evidence of the studies reporting such connectivity between the areas modelled in the network. Below a summary of the neurobiological principles based on brain data with which the neural network model was constructed.

Principle 1: a biologically constrained model of the frontal, temporal and occipital lobes with distributed neurophysiological dynamics of pyramidal neurons that encode information and participate in multiple representations (Matthews, 2001);

Principle 2: synaptic modification by way of non-supervised Hebbian-type learning, including both long-term potentiation (LTP) and long-term depression (LTD, Artola and Singer, 1993) critical for shaping brain functions and experience-depended plasticity.

Principle 3: local lateral inhibition and area-specific global regulation mechanisms (local and global inhibition) (Braitenberg, 1978; Yuille and Geiger, 2003) relevant for regulating the activity among neurons and for the refinement of neural circuits as a result of learning.

Principle 4: a sparse, random, and initially weak connectivity implemented locally with a neighbourhood bias towards nearby links typically found pervasively in the cortex (Braitenberg and Schüz, 1998; Kaas, 1997);

Principle 5: between-area connectivity based on neurophysiological principles and motivated by neuroanatomical evidence using diffusion tensor and diffusion-weighted imaging (DTI/DWI) in humans and non-human primates (e.g., Rilling et al., 2011; Thiebaut de Schotten et al., 2012) that sets the basis for the spreading of activity between the cortical areas during learning;

Principle 6: constant presence of uniform, uncorrelated white noise in all neurons mimicking the spontaneous activity (baseline firing) of real neurons during all phases of learning and retrieval with additional noise added to the stimulus patterns to simulate realistic noisy input conditions during retrieval (Rolls and Deco, 2010). Noise in the brain has been defined a key component for neural communication (in particular for oscillatory fluctuation and neural variability) and in turn for the nervous system function (Faisal et al., 2008).

The aforementioned biological principles provided the basic set of constraints for the computational model that was used to simulate aspects of language acquisition and semantics in the action and perception system of the human brain under visually deprived and undeprived conditions. Here, we simulated associative word learning between object words and their referent objects present in the environment (Vouloumanos and Werker, 2009) as well as between action words and the performance of their semantically-related actions (Tomasello and Kruger, 1992). To induce CA formation through Hebbian learning, the network was stimulated via co-activation of specific sets of 'cells' (D'Esposito, 2007; Fuster, 2003) in the primary articulatory motor and auditory cortex, along with grounding referential-semantic neurons (primary visual area for object words and primary motor area for action words). The model also implements 'anti-Hebb learning', defined as 'cells out of sync, delink' (Artola and Singer, 1993; Bienenstock et al., 1982; O'Reilly, 1998). These two mechanisms, biologically described as long-term potentiation (LTP) and long-term depression (LTD), have been defined as fundamental principles for brain functions, which shape the brain and cognition throughout the entire life of a human being (e.g., Tsumoto, 1992). This dissertation describes two versions of the same neural network model: one model adopted a 'mean-field' approach by using graded-response neurons (Chapter 2), whereas the other implemented 'leaky integrate-and-fire neurons' simulating realistic spiking neurons (Chapter 3). The details of single-neuron properties, the synaptic plasticity rule, and single-area model structure are described in their respective chapters.

Overview of the present work

One of the controversial topics in the neuroscience of language is the functional role of the various cortical areas (semantic hubs and modality-preferential areas) that are active during meaning processing. Chapter 2 focuses on answering *why* the human brain uses such a complex system of hubs and modality-preferential areas for meaning processing, and *how* they emerge at the circuit level by means of a neurobiologically constrained neural network. Additionally, it provides a precise activation time course (simulating EEG/MEG activation) of the semantic brain areas modelled in the network by offering novel predictions for word meaning processing in the human brain.

Building upon previous mean-field neural-network models presented in Chapter 2, Chapter 3 describes a more sophisticated and realistic model of the cortex. Specifically, we added critical neurobiological constraints by introducing realistic spiking neurons, biologically plausible non-supervised learning mechanisms, and a more complex set of connectivity structures based on neuroanatomical studies. It is agreed upon neuroscientists that there is the need to build more detailed and realistic neurocomputational models, which are essential in obtaining a better understanding (based on biological principles) of brain functions, such as semantic and language processing in the human brain.

Relevant contributions to linguistic and semantic processing come from language studies of congenitally blind people, reporting neural organisation of the distributed lexico-semantic network in the visual regions compared to sighted individuals. Little is known about the hotly debated questions of *why* and *how* this functional reorganisation takes place at the cellular and synaptic level. Chapter 4 focuses on offering a novel explanation for these unresolved questions by simulating word meaning acquisition in visually deprived and in underpived control conditions.

The general aim of this study is to biological explain and reconcile the diverging evidence of semantic hubs and modality-preferential regions active during semantic knowledge processing in the human brain and to explore the factors underlying the neural organization in visual cortical regions caused by blindness. To this end, we simulated the acquisition and storage of different concepts used to speak about objects and actions on healthy and visually deprived populations by means of a biologically constrained neural network. This

work is built on the hypothesis that a set of neurobiological principles acting within specific neuroanatomical structures and functions are the key foundation for understanding the mechanisms underlying word meaning processing in the human brain.

2. Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex

This chapter has been published as:

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., (2017). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia* 98, 111–129. doi:http://doi.org/10.1016/j.neuropsychologia.2016.07.004

This version of the article may not exactly replicate the final version published in the Journal. It is not the version of record.

Abstract

Neuroimaging and patient studies show that different areas of cortex respectively specialize for general and selective, or category-specific, semantic processing. Why are there both semantic hubs and category-specificity, and how come that they emerge in different cortical regions? Can the activation time-course of these areas be predicted and explained by brain-like network models? In this present work, we extend a neurocomputational model of human cortical function to simulate the time-course of cortical processes of understanding meaningful concrete words. The model implements frontal and temporal cortical areas for language, perception, and action along with their connectivity. It uses Hebbian learning to semantically ground words in aspects of their referential object- and action-related meaning. Compared with earlier proposals, the present model incorporates additional neuroanatomical links supported by connectivity studies and downscaled synaptic weights in order to control for functional between-area differences purely due to the number of in- or output links of an area. We show that learning of semantic relationships between words and the objects and actions these symbols are used to speak about leads to the formation of distributed circuits, which all include neuronal material in connector hub areas bridging between sensory and motor cortical systems. Therefore, these connector hub areas acquire a role as semantic hubs. By differentially reaching into motor or visual areas, the cortical distributions of the emergent 'semantic circuits' reflect aspects of the represented symbols' meaning, thus explaining category-specificity. The improved connectivity structure of our model entails a degree of category-specificity even in the 'semantic hubs' of the model. The relative time-course of activation of these areas is typically fast and near-simultaneous, with semantic hubs central to the network structure activating before modality-preferential areas carrying semantic information.

Introduction

The human brain is able to acquire and store knowledge about people, facts, objects, actions, and culture through experiences in everyday life. Much of this knowledge comes in units, as 'conceptual' or 'semantic representations', and carries symbolic linguistic labels in language, whereby the relationships between word-forms and semantic meaning appears as arbitrary. When semantic functions are damaged, serious consequences in daily cognitive activity can arise, being manifest as impairments of language and verbal communication and in some cases extending to domains such as planning, object recognition, or goal directed action such as drinking a glass of water (Bak and Chandran, 2012; Damasio et al., 1996; Gainotti, 2010; Kemmerer et al., 2012; Pulvermüller and Fadiga, 2010). Given the centrality of semantics in human life, it is crucial to understand the neural mechanisms underlying the nature of semantic knowledge in the brain, which, despite decades of research, is still one of the most controversial issues among cognitive neuroscientists, who propose quite diverging perspectives on this issue.

One view puts forth that one or more area(s) is/are active during meaning processing in the brain, which appear to function as general convergence zones or semantic hubs and process the meaning of all types of signs and symbols. 'Semantic hubs' have been proposed to be situated in the frontal, temporal and parietal cortices, especially in the left language dominant hemisphere (Bookheimer, 2002; Patterson et al., 2007; Price, 2000; Pulvermüller, 2013). For example, evidence for a multimodal semantic hub in anterior-inferior temporal cortex comes from patients suffering from semantic dementia, because damage in this region seems to be the best predictor of their semantic deficit (Mion et al., 2010). Although there is strong evidence for semantic hub areas, that is, for cortical regions which are generally important for meaning processing, an explanation of why several regions seem to play a role as semantic hubs and, especially, why they are localised in their specific cortical areas, is necessary.

A second important observation is that some additional cortical areas contribute to semantic processing in a more selective fashion, being particularly relevant for specific semantic categories, such as words typically used to speak about animals, tools, or actions and their related concepts. Some evidence also indicates that when recognising a word such

as *run,* activity in motor cortex, and even more specifically in leg-motor cortex, emerges, whereas, when hearing an object- and visually-related word such as *sun*, activity in visual areas is relatively more pronounced (Boulenger et al., 2009; Damasio et al., 1996; Gainotti, 2010; Hauk et al., 2004; Pulvermüller et al., 2009). Support for category-specific semantic processes is provided by a number of neurocognitive empirical studies that have focused on the importance of the motor and premotor cortex during conceptual processing, demonstrating for example that perceiving action words and sentences evokes activity in motor and premotor cortices (Boulenger et al., 2009; Hauk et al., 2004, 2008; Hauk and Pulvermüller, 2004; Pulvermüller, 1999, 2001; Rüschemeyer et al., 2007; Shtyrov et al., 2004). Furthermore, activation in the premotor and motor cortex is so fine grained that we can differentiate semantic subcategories of action-related words somatotopically (Grisoni et al., 2016; Hauk et al., 2004; Hauk and Pulvermüller, 2004). Category-specific effects have also been seen in the visual areas, especially in the ventral temporal-occipital areas, when visually-related words are being processed (e.g. animal, colour or object-related words) (Chao et al., 1999; Kiefer, 2005; Sim and Kiefer, 2005). Importantly, category-specific semantic effects are also documented in the lesion literature, where sometimes rather small lesions in modality-preferential areas can selectively impair the processing of specific semantic categories (Dreyer et al., 2015; Hernández et al., 2008). A neurobiological explanation of category-specificity has been proposed, which relates the differential activation patterns and lesion signatures to the functional level of cortical circuits with different distributions across areas. Accordingly, widely distributed cortical circuits for word-forms carried by neuronal assemblies in the perisylvian language areas are linked with neuronal ensembles storing semantic information. These semantic circuits reach into modality-preferential motor and/or sensory areas depending on whether the perceptual or action-related information is relevant for grounding the meaning of the words (Barsalou, 2008; Martin, 2007; Pulvermüller and Fadiga, 2010; Pulvermüller et al., 2005; Pulvermüller, 2001). The different distribution of the semantic circuits across the cortex, therefore, explains aspects of category-specificity. Notably, some studies reported that both category-general and category-specific semantic activation in the brain has been found to emerge rather fast, i.e. within ~200 ms after a meaningful symbol can be recognized (Hoenig et al., 2008; Penolazzi et al., 2007; Pulvermüller et al., 2005b, 2004, 2000; Shtyrov et al., 2014). For example, Moseley et al. (2013) recorded brain signals using magnetoencephalography

(MEG) and found different responses for action-related, object-related and abstract written words already at 150 ms after their onset, with gradually stronger activations for the action/object items in motor/visual regions, respectively. An explanation of category-specificity has been offered in terms of neurobiological principles. However, in order to integrate theory and data about semantic hubs with established knowledge about category-specificity, it is necessary to develop formal models of cortical structure and function that explain the presence of both.

An effort towards such explanation was recently made by Garagnani and Pulvermüller (2016), who used a network implementation of cortical areas and their connectivity to mimic the function of the perisylvian language cortex, in particular inferior frontal and superior temporal cortex, along with general visual and motor areas function in order to simulate the binding of phonological/lexical and semantic information. Using Hebbian mechanisms for synaptic modification, this model was used to simulate the emergence of neuronal circuits that process information about word-forms and their related action- vs. object-related meanings. However, the model used a simplified connectivity structure, and was applied to make predictions about magnitude and topography of brain activation, but not its time course. Here, we improve on this earlier architecture by incorporating additional cortico-cortical connections documented by neuroanatomical studies. This neuroanatomically more appropriate model was used, as in the earlier version, to predict the cortical distribution of the memory circuits for words with object- and action-related meaning. However, this type of model can be used to predict not only *where* in the brain linguistic and semantic brain activity occurs, but also *when* these processes take place, i.e., the time course of such activation. Although the spatio-temporal dimension was already present in the previous network architecture (Garagnani and Pulvermüller, 2016), we provide here, for the first time, a precise activation time course analysis of different areas of the network. Furthermore, the previous model included connector hub areas, which exhibited increased numbers of links compared with other areas. To make sure that the specific activation signatures that we observed there – in particular, the generally strong activation seen in connector hub areas – were not just a result of an increased weighted sum of incoming and outgoing synaptic connections to and from neighbouring areas ('more and stronger links, more semantics'), an in-degree normalization across areas was used here

to balance the overall input across areas and emphasise the role of network topology (or connection structure) as a factor influencing circuit topographies (or cell assembly distributions).

To investigate word meaning processing in the human brain, we used a neural network model implementing realistic anatomical and physiological features of the human cortex. The model simulates primary and secondary sensorimotor areas in frontal, temporal and occipital cortex along with 'connector hub' areas interfacing between different sensory and motor systems (Garagnani et al., 2009, 2008, Garagnani and Pulvermüller, 2016, 2013, 2011; Pulvermüller and Garagnani, 2014). The short and long distance connections between model areas are based on existing neuroanatomical evidence. Functionally, the model takes advantage of realistic Hebbian learning mechanisms (Hebb, 1949). The network was trained with repeatedly presented specific sensorimotor patterns coding for the articulatory and acoustic phonological structure of single words and some of their action- or perception-related semantic features. As a result of learning and area/connectivity structure, distributed 'semantic circuits' emerged in the network, spanning different areas. Importantly, the topographies of these circuits showed similarities and differences between semantic types (action vs. object words), which can be related to the semantic information stored. We document circuit distributions and their dynamic activation and discuss the results in the context of specific model features, existing experimental evidence, and novel predictions for future research.

Materials and Methods

We applied a neurobiologically grounded computational model replicating structure and functional properties of the human cortex to investigate the neural mechanisms underlying word meaning acquisition and processing in the perception and action systems of the mind and brain. The model's architecture mimics the left perisylvian cortex involved in spoken word processing, corresponding to articulatory and acoustic phonological word-forms (Fadiga et al., 2002; Fry, 1966; Pulvermüller, 1992, 1999; Pulvermüller and Fadiga, 2010; Zatorre et al., 1996), areas outside the perisylvian cortex involved in processing visual object identity (Ungerleider and Haxby, 1994), and the execution of manual actions (Deiber et al., 1991; Dum and Strick, 2005, 2002; Lu et al., 1994). The model mimics a range of biologically realistic properties of the human cortex including the following features:

1. Area structure: 12 cortical areas were modelled, including modality-preferential sensory and motor ones as well as connector hub areas interlinking sensory and motor systems.
2. Between-area connectivity: different areas were linked based on neuroanatomical principles and data, realising sparse, random, initially weak and topographic connectivity.
3. Within-area connectivity: similarly sparse, random and initially weak connectivity was implemented locally, along with a neighbourhood bias towards local links (Braitenberg and Schüz, 1998; Kaas, 1997).
4. Local lateral inhibition and area-specific global regulation mechanisms (local and global inhibition) (Braitenberg, 1978; Palm et al., 2014; Yuille and Geiger, 2003).
5. Synaptic modification by way of Hebbian type learning, including both long-term potentiation and depression (LTP, LTD) (Artola and Singer, 1993).
6. Neurophysiological dynamics of single cells including temporal summation of inputs, nonlinear transformation of membrane potentials into neuronal outputs, and adaptation (Matthews, 2001).

7. Constant presence of uniform uncorrelated white noise in all neurons during all phases of learning and retrieval, and additional noise added to the stimulus patterns to mimic realistic noisy input conditions during retrieval (Rolls and Deco, 2010).

Word learning processes in the model are based entirely on mechanisms of Hebbian plasticity, often summarized by the phrase 'cells that fire together, wire together', although the learning rule applied (see above and Appendix 2A) implements 'anti-Hebb' learning too, colloquially described by the phrase 'cells out of sync delink' (for discussion, see Garagnani et al., 2009). Accordingly, within a network of interconnected neurons, repeatedly and consistently co-active sub-populations of cells strengthen their connections, forming the so called cell assemblies (CAs) (Hebb, 1949). According to Hebb (1949), assemblies can be considered functional units in the brain representing the building blocks of cognitive functions, including language (Braitenberg, 1978; Palm et al., 2014; Pulvermüller, 1996). In principle, the emerging neuronal assemblies can be *local*, that is, restricted to a small area or even cortical column of a fraction of a cubic millimetre or, alternatively, be spread out across wide cortical regions, and it is not clear a priori whether a given network and input pattern leads to the formation of local or distributed circuits. Different cortical distributions, or topographies, of cell assemblies have been postulated for symbols with different meaning. Standard postulates are that words related to actions include neurons in the motor cortex – which control the movements a word such as *run* is typically used to speak about – while words referring to objects (such as *sun*) will include neurons in areas along the ventral visual stream of object processing (Huyck and Passmore, 2013; Pulvermüller, 1999; Pulvermüller and Preissl, 1991). Previous simulation studies have already shown the formation of distributed neuronal assemblies exhibiting differential cortical distributions as a result of repeated concomitant presentation of activation patterns and Hebbian plasticity mechanism (Garagnani et al., 2009, 2008, Garagnani and Pulvermüller, 2016, 2013, 2011; Wennekers et al., 2006).

Model architecture

The model consists of 12 cortical areas of artificial neurons with area-intrinsic connections and mutual connections between them. In the left perisylvian language cortex, we identify six cortical areas divided into two sub-systems: auditory and articulatory systems (areas highlighted in blue and red in Fig. 1. A). The auditory system includes the primary auditory cortex (A1), auditory belt (AB), and parabelt areas (PB) - whereas the articulatory system includes the primary articulatory motor cortex (inferior part of primary motor cortex, $M1_i$), inferior premotor (PM_i) and prefrontal motor cortex (PF_i). Six additional areas outside the perisylvian cortex (which we call 'extrasylvian') were included to model the ventral visual stream and dorsolateral motor system (green and yellow highlighted areas). The ventral visual system is relevant for processing visual object identity and includes, apart from primary visual cortex (V1), temporo-occipital (TO) and anterior-temporal (AT) areas. Finally, the motor system which, for example, is relevant for the execution of manual actions, includes the dorsolateral fronto-central motor ($M1_L$), premotor (PM_L), and prefrontal cortices (PF_L). Each model area consists of two layers of 25 x 25 excitatory and inhibitory artificial neurons (*e*- and *i*-cells) (see Fig. 2.1. C). Each *e*-cell represents a cluster of excitatory pyramidal cells, and the underlying *i*-cell models represent the cluster of inhibitory interneurons, situated within the same cortical column (Eggert and van Hemmen, 2000; Wilson and Cowan, 1972). As it is typical for the mammalian cortex, the connectivity between and within model areas is sparse, patchy and topographic (Amir et al., 1993; Braitenberg and Schüz, 1998; Gilbert and Wiesel, 1983). To regulate and control activity in the network, local and area-specific inhibition is implemented (Bibbig et al., 1995; Palm, 1982; Wennekers et al., 2006). Details of the model functions and of the Hebbian learning mechanism (including LTD and LTP) are summarized in previous works (Garagnani et al., 2009, 2008, Garagnani and Pulvermüller, 2016, 2013, 2011). For completeness, we recapitulate them in Appendix 2A.

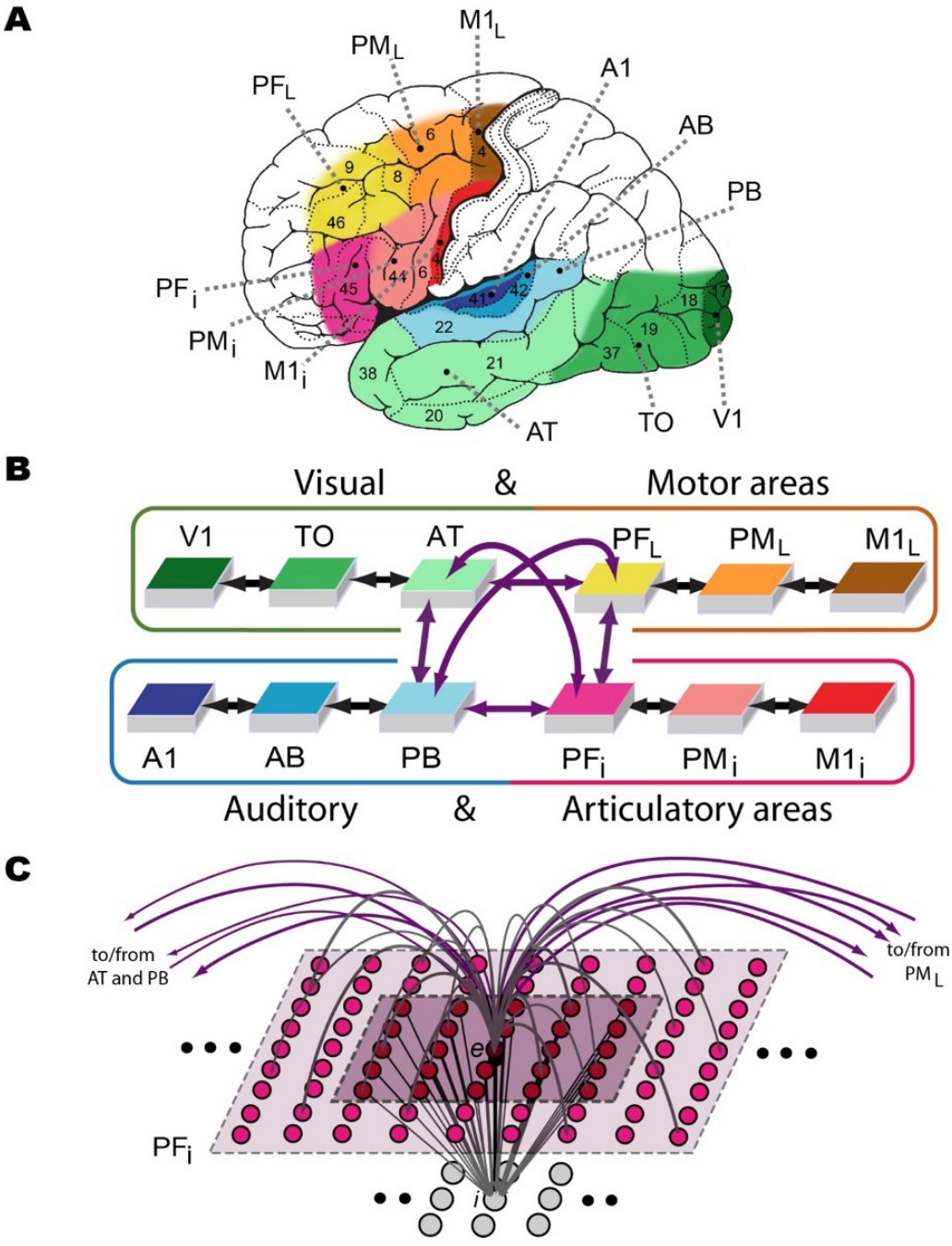
A
M1L
PML
A1
PFL
AB
PB
PFi
PMi
M1i
AT
TO
V1
B
Visual & Motor areas
V1
TO
AT
PFL
PML
M1L
A1
AB
PB
PFi
PMi
M1i
Auditory & Articulatory areas
C
to/from AT and PB
to/from PML
e
i
PFi

Figure 2.1. Model of lexical and semantic mechanisms: The 12 cortical areas modeled (A), their global connectivity architecture (B), and aspects of the micro-structure of their connectivity (C) are illustrated. **(A)** Six perisylvian (i) and six extrasylvian (ii) model areas are shown, each including a dorsolateral (frontal) and a ventral (temporal) part: (i) perisylvian cortex include an articulatory system (red colors), including inferior-prefrontal (PF_i), premotor (PM_i) and primary motor cortex ($M1_i$) and auditory system (areas in blue), including auditory parabelt (PB), auditory belt (AB) and primary auditory cortex (A1). These areas can store correlations between neuronal activations carrying articulatory-phonological and corresponding acoustic-phonological information, for example when phonemes, syllables and spoken word-forms are being articulated (activity in $M1_i$) and acoustic features of these spoken words are simultaneously perceived (stimulation of primary auditory cortex, A1). (ii) Extrasylvian areas include a motor system (yellow to brown), including dorsolateral prefrontal (PF_L), premotor (PM_L) and primary motor cortex ($M1_L$) and a 'what' visual stream of object processing (green), including anterior-temporal (AT), temporo-occipital (TO) and early visual areas (V1). Together with the perisylvian areas, these extrasylvian areas can store correlations between neuronal activations carrying semantic information, for example when words are used (activity in all perisylvian areas) to speak about objects present in the environment (activity in V1, TO, AT) or about actions that the individual engages in (activity in $M1_L$, PM_L, PF_L). Numbers indicate Brodmann Areas (BAs). **(B)** Schematic illustration of all 12 model areas and the known between-area connections implemented. The colours indicate correspondence between cortical and model areas. **(C)** Micro-connectivity structure of one of the 7,500 single excitatory neural elements modelled (labeled 'e'). Within-area excitatory links (in grey) to and from 'cell' e are limited to a local (19x19) neighbourhood of neural elements (light-grey area). Lateral inhibition between e and neighbouring excitatory elements is realized as follows: the underlying cell 'i' inhibits e in proportion to the total excitatory input it receives from the 5x5 neighbourhood (dark-purple shaded area); by means of analogous connections (not depicted), e inhibits all of its neighbours. Each pair (e,i) of model cells is taken to represent an entire cluster or column (grey matter under approximately 0.25 mm^2 of cortical surface) of pyramidal

Neuroanatomical and imaging studies have demonstrated the existence of next-neighbour between-area connectivity, which functionally binds adjacent cortical areas together (Pandya and Yeterian, 1985; Young et al., 1995, 1994). These functional links are modelled within each triple of areas forming the four domain-specific sub-systems in the model (see black arrows Fig. 2.1. B). In the perisylvian system, next-neighbour connections between locally adjacent areas are implemented within the auditory sub-system (A1, AB, PB) (Kaas and Hackett, 2000; Pandya, 1995; Rauschecker and Tian, 2000), as well as within the articulatory (PF_i, PM_i, $M1_i$) sub-system (Pandya and Yeterian, 1985; Young et al., 1995). Similarly, local next neighbour links are also realised in the extrasylvian system, between adjacent ventral visual (V1, TO, AT) (Bressler et al., 1993; Distler et al., 1993), and dorsolateral motor areas (PF_L, PM_L, $M1_L$) (Arikuni et al., 1988; Dum and Strick, 2005, 2002; Lu et al., 1994; Pandya and Yeterian, 1985; Rizzolatti, G. Luppino, 2001).

Long distance cortico-cortical links between sub-systems (see purple arrows Fig. 2.1. B) are realised between all pairs of multimodal hub areas (PB, PF_i, AT and PF_L). This is motivated by evidence for neuroanatomical connections between inferior prefrontal (PF_i) and auditory parabelt (PB) areas, carried by the arcuate and the uncinated fascicles (Catani et al., 2005; Makris and Pandya, 2009; Meyer et al., 1999; Parker et al., 2005; Paus et al., 2001; Rilling et al., 2008; Romanski et al., 1999b) and, in the extrasylvian system connections between anterior-temporal (AT) and lateral prefrontal (PF_L) areas, carried by the uncinate fascicle (Bauer and Jones, 1976; Chafee and Goldman-Rakic, 2000; Eacott and Gaffan, 1992; Fuster et al., 1985; Parker, 1998; Ungerleider et al., 1989; Webster et al., 1994). The peri- and extrasylvian systems are also linked by means of long distance cortico-cortical connections across the central hub areas; likewise parabelt (PB) and lateral prefrontal cortex (PF_L) are reciprocally connected (Pandya and Barnes, 1987; Romanski et al., 1999b, 1999a) as well as the anterior/middle-temporal (AT) and inferior prefrontal (PF_i) areas (Pandya and Barnes, 1987; Petrides and Pandya, 2009; Rilling, 2014; Romanski, 2007; Ungerleider et al., 1989; Webster et al., 1994). A recent simulation study adopting a similar network architecture did not implement connections between inferior and superior prefrontal or between auditory parabelt and anterior temporal cortex (Garagnani and Pulvermüller, 2016). We added both links because of the evidence for reciprocal connectivity between anterior-temporal (AT) and parabelt (PB) areas (Gierhan, 2013) and

between inferior and lateral prefrontal (PF_i, PF_L) areas (Yeterian et al., 2012). This also led to a more symmetric network structure. The asymmetries in the earlier network may account for some of its functional properties, which, as we discuss below, were not seen in the present network based on a (slightly) more realistic structure (see Discussion).

The previous study (Garagnani and Pulvermüller, 2016) found that semantic circuits included a massively enhanced number of neurons in connector hub areas compared with primary or secondary areas, which was seen as an explanation of semantic hub status. However, there are different mechanisms that could underlie the observation: One way to explain it is by way of topological network structure, especially the fact that 'connector hub' areas hold a central role in interlinking sub-systems. At the same time, and partly independent from their role as connector hubs, the same areas are also the targets and origins of an increased number of connections to other areas (i.e. a higher 'degree' of connectivity). In the case of our present model, 2 between-area connections exist for most areas (primary ones have input plus 1 connection), but connector hubs have 4 of them, thereby entailing larger amounts of activation reaching these areas when activity waves spread through the network from its different ends during learning. Any specific functional properties of hub areas, including their great involvement in carrying semantic circuit members, may therefore, result either from network topology, or from number of area input connections from other areas, or from both. If it is just the number of inputs to and thus amount of activation in an area – their 'in-degree' – that is relevant for an increased importance in semantics, the explanation of semantic hubs may trivially be based on the formula 'what activates most, is most relevant for cognition'. However, an explanation based on network topology and connectivity structure *per se* becomes plausible if general semantic relevance can be documented for hubs that have an overall input comparable to that of other areas. Therefore, we normalized the overall amount of input of all (equal-sized) areas by dividing the contribution of all long-distance connections (all links among the 'rich club' of connector hubs, central quadruplet in Fig. 2.1. B) by 3. After this *in-degree normalization* (which in the present symmetric architecture also implies out-degree normalization), each of the 12 areas receives two equal quantities of inputs (either 1*1 or 3*1/3), one from the left and one from the right side of the model. This procedure preserved differences in topology while normalising for amount of input activation per area.

Simulations

The simulations were carried out in two steps. After learning the semantic relationships between articulatory and acoustic information about the word-form (perisylvian activity patterns in $M1_i$ and A1) and 'grounding' action or object information (extrasylvian activity pattern either in $M1_L$ or in V1) (section 2.2.1), the network was used to simulate the neurophysiological correlates of word recognition and understanding (section 2.2.2).

Learning phase

The network architecture described above (Fig. 2.1. B) was initialized at random before the learning phase began (see appendix 2A): 12 different, randomly initialized networks were created, each with 12 different sets of sensorimotor patterns representing object- and action-related words. These 'word-learning patterns' represented six object-related and six action-related words. Each pattern consisted of a fixed set of 19 cells chosen at random from the 25 x 25 cells of an area (ca. 3% of the cells) which were simultaneously presented to the primary areas of the network. At the linguistic and semantic levels, the cells in $M1_i$ and A1 represented articulatory and acoustic phonetic features and their values (e.g., [+labial]) and those in $M1_L$ and V1 action-related and visually-related semantic features plus values of the words (e.g., [+LEG ACTION], [+ROUND SHAPE]). Each word in our training set was grounded in input to three of the four primary areas of the model: apart from perisylvian A1 and $M1_i$ activity, object-related words received concordant visual (V1) and action words lateral motor area ($M1_i$) grounding activity. This mimics a typical situation of object-related word learning, whereby the word is uttered while the referent object is present (Vouloumanos and Werker, 2009) or the relevant action is being performed (Tomasello and Kruger, 1992). Note that white noise was always present and overlaid all learning patterns (in addition to that already present in all areas of the network). This was implemented to account for variability of perceptions and actions of the same type. The model was set up to learn the correlation between word and referential semantic information; the critical question was which type of representations develops in the network as a consequence of learning.

Each word-learning pattern of 3 x 19 activated cells (57 cells in total) was simultaneously presented to the respective primary areas for 3000 times. Some trial-to-trial variability of

patterns was due to noise overlay (see below). The number of presentation was chosen on the basis of previous simulations (Garagnani and Pulvermüller, 2016). While three primary areas were directly activated by each learning pattern, the fourth non-relevant area ($M1_i$ for object- and V1 for action-related words) received additional variable noise input, i.e. a further pattern, consisting of 19 randomly chosen cells that changed inconsistently over learning episodes, was presented to the respective primary areas. This was done to make sure that the correlation of the word-form activity in the perisylvian cortex with that of the semantic information was high in one modality for action and object words in motor and visual systems, but low in the non-relevant one. A learning trial involved presentation of a word pattern for 16 time-steps, followed by a period during which no input (inter stimulus interval - ISIs) was given. The next stimulus was presented to the network only when the global inhibition of the PF_i and PB areas decreased below a specific fixed threshold; this allowed the activity in the network to return to a predefined baseline value, so as to minimize the possibility of one trial affecting the next one. During each ISIs, only the inherent baseline noise (simulating spontaneous neuronal firing) was present in the neural-network.

Cell assembly definition

During the learning phase, we noticed the gradual formation of cell assembly circuits with different assemblies responding to different input patterns. After 3000 presentations in which three of the four sub-systems were co-activated by stimulating specific neurons in their respective primary cortex, distributed neuronal circuits spontaneously emerged within the network areas, linking up word-form in the perisylvian language areas (auditory and articulatory sub-systems) with referential-semantic information in the sensorimotor areas (visual and motor sub-systems) (this is further explained in section 3.1).

To identify and quantify the neurons forming the 12 CA circuits across the network areas, we computed the average firing rate of each excitatory cell (7500 *e*-cells) over the 15 time-steps subsequent to a single presentation of the learned sensorimotor patterns (no semantic input was provided in the primary areas of the extrasylvian system). An *e*-cell was defined as a member of a given CA circuits, only if its time-averaged rate (output value or 'firing rate') reached a threshold ϑ which was area- and cell-assembly specific, and defined

as a fraction *γ* of the maximal single-cell's time-averaged response in that area to pattern *w.* More formally,

$$\vartheta = \vartheta_A(w) = \gamma \max_{x \in A} \overline{O(x,t)}_w$$

where $\overline{O(x,t)}_w$ is the estimated time-averaged response of cell *x* to word pattern *w* (see Eq. (A4.1) in Appendix 2A) and $\gamma \in [0,1]$ is a constant (we used $\gamma = 0.5$ on the basis of previous simulation results, see Garagnani et al., 2009, 2008). This was computed for each of the 12 trained networks and the number of CA cells per area was averaged over the six object- and six action-related words. CA distributions across areas were analysed statistically as described in section 2.3.

Neurophysiological word recognition simulations

After training, we used the network to simulate the process of perceiving, recognizing and understanding object- and action-related words and the neurophysiological mechanisms underlying these processes. To this end, each 'testing trial' started with primary auditory area (A1) stimulation using only the A1 component of the learning pattern of one learnt 'word'. Stimulation was for 2 time-steps, followed by 50 time-steps during which no input was provided and another 10 used as a baseline for the subsequent trial. To ensure that all testing trials started from analogous baselines, network activity was reset before the baseline. In order to obtain better signal-to-noise ratios, each of the auditory patterns was presented in 12 different testing trials. Results for each CA were obtained by averaging the 12 'trials' of its sensorimotor pattern presentation.

During word recognition, we recorded the area-specific 'within-cell assemblies (CA) activity' per simulation time-step during the 10 time-steps preceding the stimulus onset and the 50 time-steps following offset. The within-CA activity was computed as the sum of the output values (cumulative firing rates, CFRs) of the emerging CA cells in each area produced by stimulation of area A1 as a function of time. By 'CA cells', we mean here the cells forming the CA (as defined in Sec. 2.2.2 above); through Hebbian learning, these cells become strongly and reciprocally connected, forming the CA circuits. After this, we identified the 'peak amplitude' as the maximum value reached by the CA's cumulative firing rates during

the 50 post-stimulus time-steps, and the 'peak delay', the latency of the peak upon stimulation. These values were computed for each of the 12 learned networks, averaged over the two word-types and across network areas: results were submitted to statistical analysis as described below.

Statistical analysis

Statistics were performed on the six object- and six action-related words learnt by one network and across the 12 different network instances. To statistically test for the presence of significant differences in the topographical CA distribution and activation dynamics, we performed repeated-measures Analyses of Variance (ANOVAs). A 4-way ANOVA was run with factors WordType (two levels: *Object* vs. *Action*), PeriExtra (two levels: *Perisylvian* = {A1, AB, PB, $M1_i$, PM_i, PF_i}, *Extrasylvian* cortex = {V1, TO, AT, $M1_L$, PM_L, PF_L}), TemporalFrontal (TempFront) (2 levels: *temporal areas* = {A1, AB, PB, V1, TO, AT}, *frontal areas* ={$M1_L$, PM_L, PF_L, $M1_i$, PM_i, PF_i}) and Areas (three levels: *Primary* = {A1, V1, $M1_L$, $M1_i$}, *Secondary* = {TO, AB, PM_L, PM_i} and *Central* = {PB, AT, PF_L, PF_i} areas). We further performed a second statistical analysis on the data of the two systems separately, six perisylvian and six extrasylvian areas with factors 'WordType', 'TempFront', 'Areas', as described above. Analysis was performed on 3 different sets of data: (i) on CA cells distributions emerged from word acquisition, on the (ii) peak amplitudes, and (iii) peak delays during word recognition simulations. Finally, we performed Bonferroni-corrected planned comparison tests (24 comparisons, corrected critical $p < .0020$) to further explore the significant differences in CA cells distributions and peak delay data across the four sub-system areas.

Data analysis & Results

Learned CA topographies for object- and action-related words

Fig. 2.2 illustrates six of the twelve CA-cell distributions for object- (A) and action-related (B) words, as they spontaneously emerged during simulated word learning (the other CAs produced similar results). Each set of 12 squares is a snapshot of the CA distribution of a specific word across the network, and each white pixel in the squares represents a cell.

The emerging CA circuits are spread out to the same degree across the perisylvian language areas for object- and action-related words, whereas motor and visual areas of the extrasylvian cortex seem to exhibit different CA cell distributions. These distributions indeed appear to show a double dissociation. Object-related words extend more into the visual (V1, TO) areas, whereas they extend only minimally into the extrasylvian motor (PM_L, $M1_L$) areas; the reverse pattern emerges for the action-related words.

Fig. 2.3 illustrates the distribution of the CA circuits, given as the number of CA cells per areas averaged across 12 trained networks, for object- (light grey) and action-related words (dark grey). The extrasylvian system involved in processing visual-object identity and motor action seems to exhibit a double dissociation between the two word types, as already noted above and in Fig. 2.2. The perisylvian language cortices seem to show no significant differences between the circuits for the two word types. Note also that there is a larger number of CA cells in the multimodal hub areas (PB, PF_i, AT, and PF_L) than in the secondary areas (AB, PM_i, TO, PM_L), where there are more cells than in primary areas (A1, $M1_i$, V1, $M1_L$). This appears independent of whether an object- or action-related word is represented.

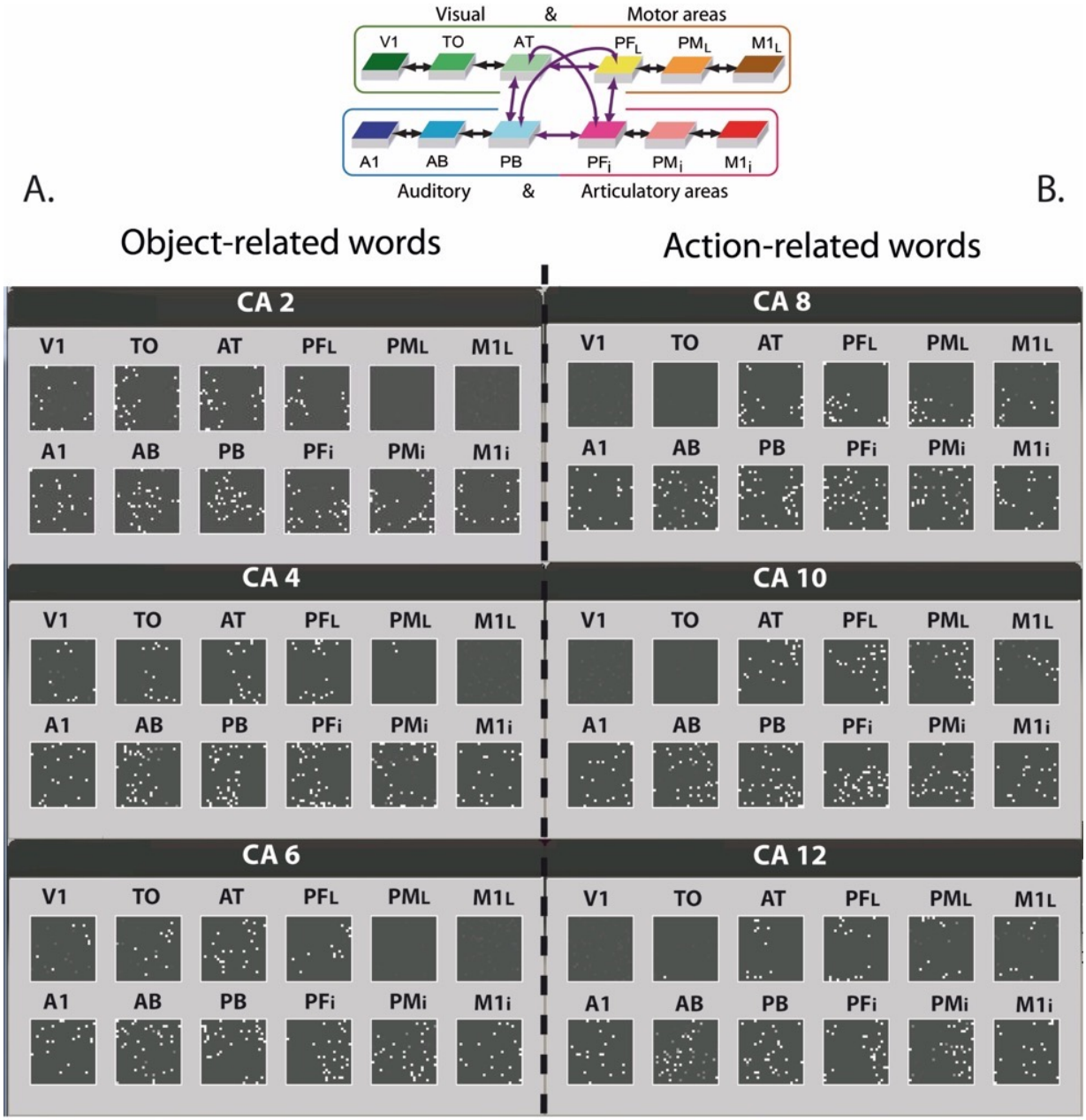

Figure 2.2. Distributions of cell-assemblies (CAs) emerging in the 12 area network during simulation of word learning in the semantic context of visual **(A)** and action **(B)** perceptions. Results of one typical instantiation of the model in Fig. 1 are shown, using the same area labels. Each set of 12 squares (in black) illustrates the distribution of 'cells' of one specific CA across the 12 network areas. Each white pixel in a square indexes one CA cell. CAs for object-related words extend into higher and primary visual cortex (V1, TO, but not $M1_L$), linking information about spoken word-forms (perisylvian pattern) with information from the visual modality (neural pattern in V1). Network correlates of action-related words extend into lateral motor cortex ($M1_L$, PM_L, but not V1), thus semantically grounding words in information about actions. For convenience, the area structure of the network is repeated at the top.

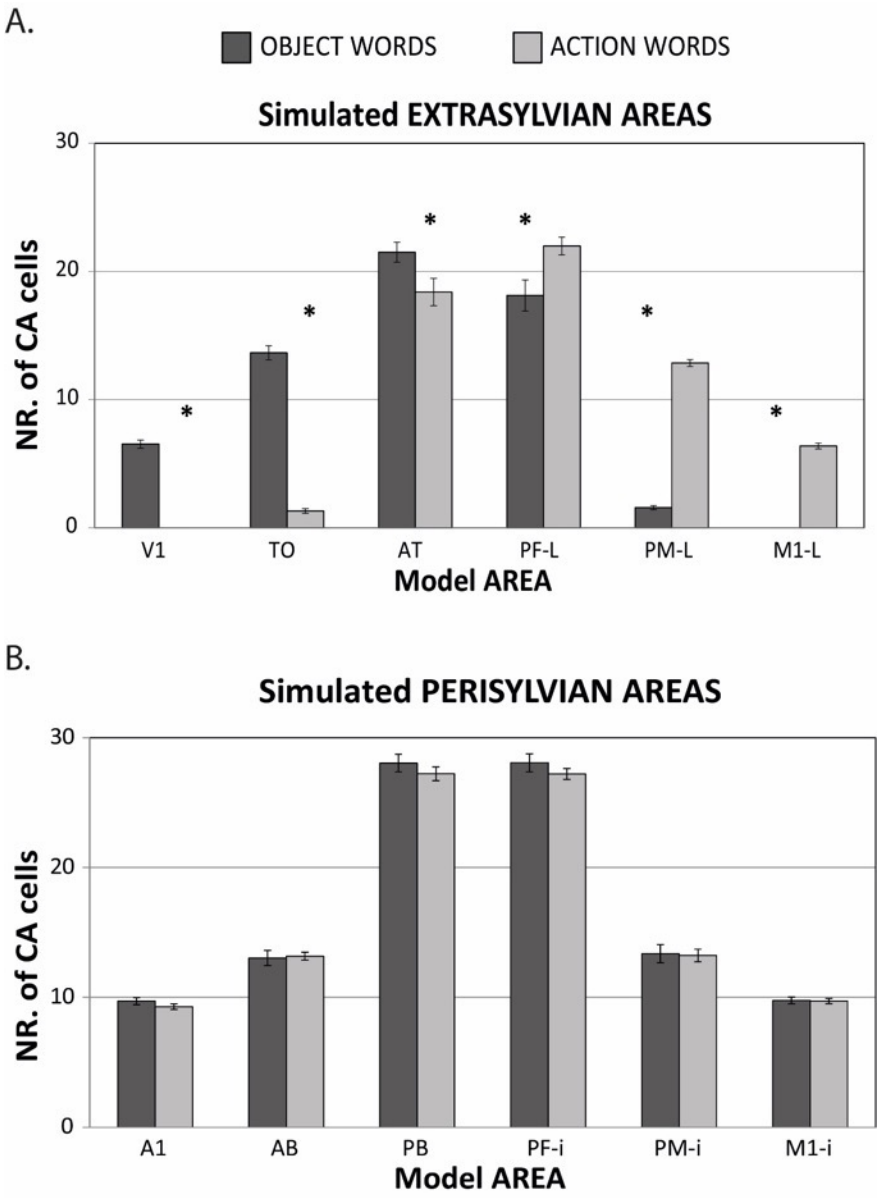

Figure 2.3. Average distributions of CAs emerging in 12 instantiations of the 12 area network architecture during simulation of word learning in the semantic context of actions and visual perceptions. Bars show average numbers of CA neurons per area for object- (dark grey) and action-related (light grey) word representations; error bars show standard errors over networks. **(A)** Data from the six perisylvian areas whose cells can be seen as circuit correlates of spoken word-forms do not show category-specific effects. **(B)** The extrasylvian areas whose cells can be seen as circuit correlates of word meaning show a double dissociation, with relatively more strongly developed CAs for object- than for action-related words in primary and secondary visual areas (V1, TO), but stronger CAs for action-related than for object-related words in dorsolateral primary motor and pre-motor cortices (PM_L, $M1_L$). Asterisks indicate that, within a given area, the number of CA neurons significantly differed between the circuits of action and object words (Bonferroni-corrected planned comparison tests, 24 comparisons; critical threshold $p < .0020$).

The observations described above were confirmed by the 4-way ANOVA. A main effect of Areas ($F_{2,24}$ = 1226.424, p < .0001) emerged, which confirms that the CA cell densities differed across areas, with CA cell densities being higher in hub than in secondary areas (p < .0001), and higher in secondary than in primary areas (p < .0001). In addition, we found a significant interaction between the factors WordType, PeriExtra, TempFront and Areas ($F_{2,24}$ = 130.795, p < .0001), indicating that the distributions of the two types of word-related CA circuits across the network differed. Because the interaction also demonstrates that CA-distribution differed between perisylvian and extrasylvian systems, we ran further statistical analyses on the data from the two systems separately, now using 3-way ANOVAs. We found a main effect of Areas for both perisylvian ($F_{2,24}$ = 2091.116, p < .0001) and extrasylvian systems ($F_{2,24}$ = 3959.92, p < .0001), as revealed by the 4-way ANOVA analysis. As expected, the perisylvian system did not show any significant differences between CA distributions of the two word types across the 6 areas ($F_{2,24}$ = 0.38, p = 0.68). In contrast, the extrasylvian system revealed a highly significant interaction of all three factors WordType, TempFront and Areas ($F_{2,24}$ = 156.555, p < .0001), confirming the word category differences in the CA topographies and local cell-density distributions across visual, motor and multimodal areas as suggested by Fig. 2.2 and 2.3. To further investigate the differences between CA types across the network, we ran Bonferroni-corrected planned comparison tests (24 comparisons, corrected critical p < .0020); these confirmed the presence of a larger number of CA cells in visual (V1, TO and AT) than in motor ($M1_L$, PM_L, and PF_L) areas for object- (p < .001), and the opposite for action-related words (p < .001). Post-hoc analysis of the data from the connector hubs (AT, PF_L) also showed a significant difference between the two word types there, i.e. stronger action-related word CA cell densities in PF_L compared to AT (p <.0001), and the opposite for object-related words (p <.001). Differences in CA-cell densities between word types and pairs of areas in the semantic systems were all significant (p < .002), as described in Fig. 2.2. In contrast, no significant differences emerged in the perisylvian system (p >.87).

Neurophysiological word recognition results

To obtain a simulation of spoken word recognition and comprehension processes, we analysed the time-course of the network's response to presentation of the learned auditory word-form patterns to area A1. To this end, we computed the sum of all CA cell activity values (quantified as the cumulative firing rates, CFRs, see Sec. 2.2.3) as a function of time across the entire network or for specific areas. Activation time courses showed an initial 'ignition' of CA circuits, a strong activation, which peaked at time-step ~ 16 and included a majority of the circuits' neurons (Fig. 2.4). Replicating, in part, the structural distributions of semantic circuits depicted in Fig. 2.3, both types of circuits were similarly spread out across all perisylvian areas of the model; by contrast, differences between semantic circuit types were present in extrasylvian cortex: object-related words (blue pixels) elicited activation in the visual system and less in the motor system, while the reverse happened for the action-related words (red pixels). Note also the low degree of overlap between CAs of the two different word types (yellow pixels) for these two specific CAs instances.

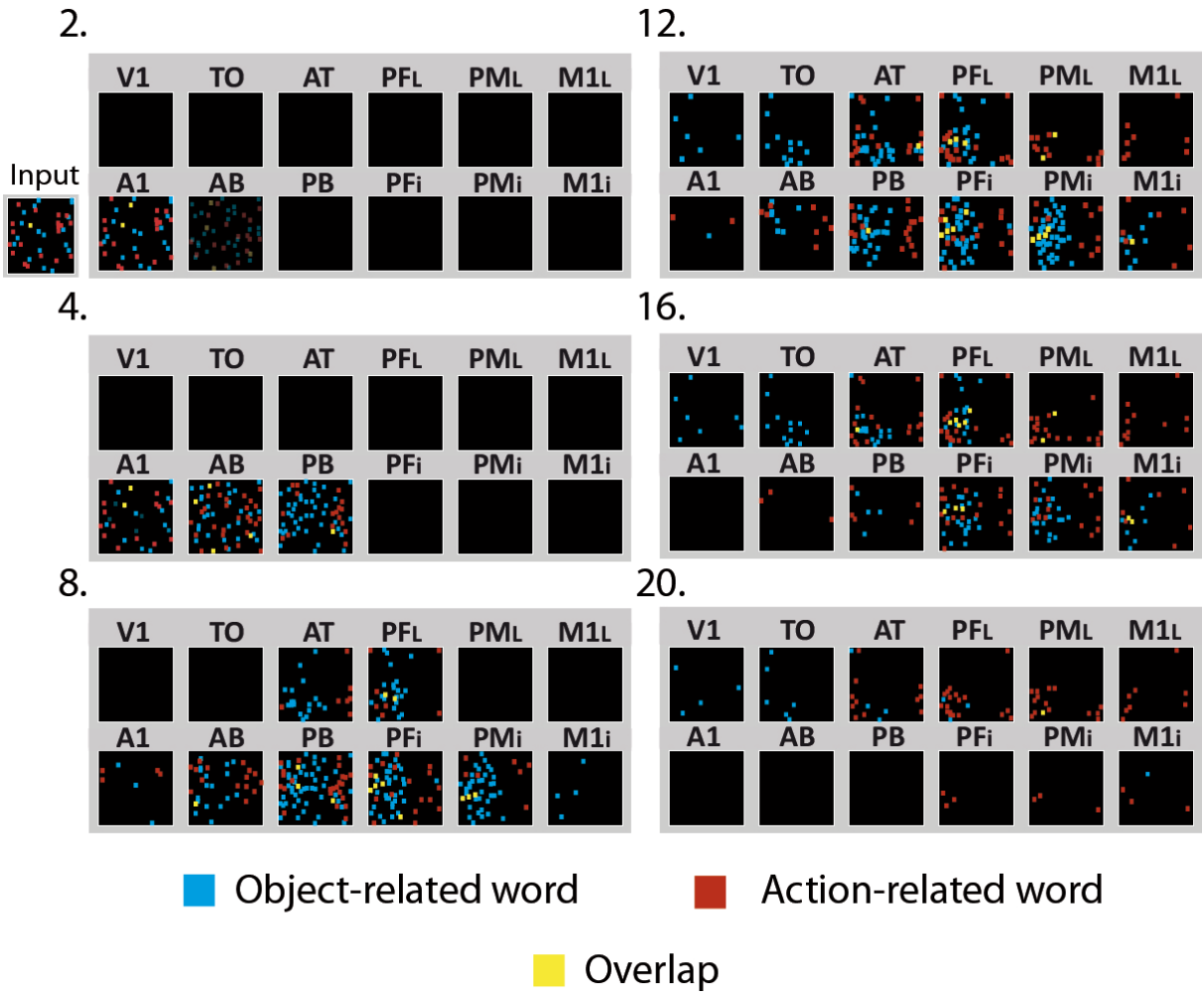

Figure 2.4. Activation spreading in the 12 area network showing the simulated recognition of object- (blue pixels) and action-related (red pixels) words. Yellow pixels illustrate the overlap between the two word-related CAs. Network responses to stimulation of A1 with the 'auditory' patterns of two of the learned words; each set of 12 'squares' depicts a selected snapshot of the entire network's activity (as in Fig. 2). Cell activity levels are indicated by brightness of pixels; snapshot numbers indicate simulation time-steps of the network output. See main text for details.

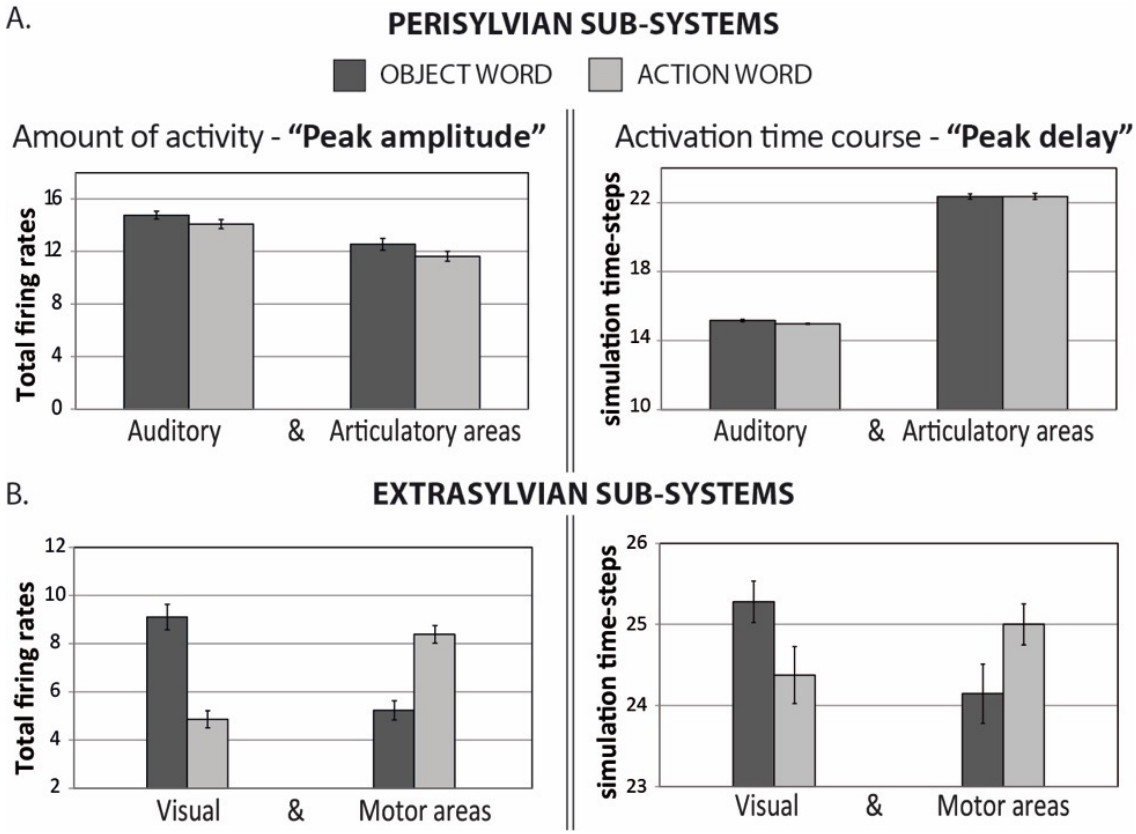

Figure 2.5. Bar plots illustrating the amount of activity - **'**peak amplitude' (left hand side) and the activation time- course – 'peak delay' (right hand side) of auditory and articulatory **(A)** and visual and motor **(B)** areas for object- and action-related words during auditory word recognition.

In extrasylvian areas, maximal area-specific activation levels significantly differed between the circuits carrying the two semantic word-types. A significant double dissociation showed that circuits for object-related words produced higher amplitude in the visual (cumulative firing rates (CFRs) = 9.10) sub-system than in the lateral (hand) motor system (CFRs = 5.23), and, vice versa, action-related words activated the lateral motor system (CFRs = 8.38) more strongly than the visual system (CFRs = 4.86, see Fig. 2.5. B – Bar plot left-hand side). As visual inspection indicates, the auditory and articulatory motor sub-systems (see Fig. 2.5. A – Bar plot left-hand side) did not show any differences in activity levels between semantic word types. Furthermore, comparing activity levels between areas of the network (see Fig. 2.6. A-B & 2.7.A-B), multimodal hub areas (AT, PF_L, PB, PF_i) seemed to show the strongest activation dynamics (CFRs ~ 15) in comparison with secondary (CFRs ~ 10) and primary areas (CFRs ~ 5).

The statistical analyses of the dynamic functional activation of the circuits confirmed these observations, which are in line with the CA-distribution results described in Section 3.1. In particular, the 4-way ANOVA performed on peak activation levels per area and word type revealed a main effect of Areas ($F2,22 = 630.246$, $p < .001$), again with maximal CA activation in 'central' connector hub areas. In addition, a significant interaction of factors WordType, PeriExtra, TempFront and Areas ($F2,22 = 137,433$, $p < .001$) emerged, confirming different activation levels between word type circuits across the network's areas. Because of the differences between the peri- and the extrasylvian systems, we also ran a second statistical analysis on each of the two systems separately. The 3-way ANOVA revealed a main effect of Areas on both perisylvian ($F2,22 = 667.146$, $p < .001$) and extrasylvian ($F2,22 = 268.1345$, $p < .001$) systems. Whereas the perisylvian areas did not show any significant differences in peak amplitude between the two circuit types ($F1,11 = 0. 98$, $p = .76$), the extrasylvian system revealed significant interactions of factors WordType, and TempFront ($F1,11 = 518.7315$, $p < .001$), and of WordType, TempFront and Areas ($F2,22 = 109.3367$, $p < .001$), showing different activation dynamics across the extrasylvian areas between the circuits of the two word categories (Fig. 2.5 – left-hand side).

Area-specific activation time-course – peak delay results

Fig. 2.6 and 2.7 delineate the area-specific activation time courses of semantic circuits of object- (A) and action-related words (B) across the network. The activation in different areas peaked at different times and showed different maximal amplitudes. The schematic brains at the top of each panel illustrate the area-specific peak delay and the boxplots indicate the latency of maximal activation together with their standard errors (boxes) and standard deviations (whiskers).

The activation time-courses in the perisylvian language areas exhibited a similar, cascade-like time-course for both object- and action-related CA circuits (see Fig. 2.6 A-B). Area A1 peaked at an early time (2 time-steps) after stimulus onset because it was driven by the sensorimotor pattern presented there. The auditory-belt (AB) area peaked at ~ 6 time-steps, and shortly followed by the parabelt (PB ~ 7) and inferior prefrontal (PF_i ~ 10) areas, and finally the premotor (PM_i ~ 12) and primary motor ($M1_i$ ~ 13) areas. This time-course was the same for both circuit types. By contrast, the extrasylvian semantic system (Fig. 2.7 A-B) seemed to exhibit different temporal activation patterns for the two types of semantic circuits. The extrasylvian connector hub areas (PF_L, AT) peak activated at similar latencies as the perisylvian hubs (PF_i, PB) central to the network structure (12-13 time-steps). Interestingly, the multimodal prefrontal area (PF_L) revealed a similar activation dynamics (~ 13 simulation time-steps) for both word types, whereas the anterior-temporal hub area (AT) peaked 1 time-step earlier for action-related words (~ 12) than for object-related ones (~ 13). Massive activation time-course differences were apparently present in non-central extrasylvian areas, i.e. in the primary and secondary visual and dorsolateral motor areas of the network. Object-related words activated their lateral premotor and temporo-occipital area shortly after the connector hubs (PM_L ~ 15, TO ~ 15), closely followed by the primary visual (V1 ~ 16) area. In contrast, the circuits underpinning action-related words first activated the lateral premotor (PM_L) area (~ 15), closely followed by temporo-occipital (TO) and lateral primary motor ($M1_L$) areas (~ 16). Both object- and action-related words activated the primary areas of the relevant system approximate ~ 15 time-steps after word onset and at the end of the activation cascade. As visible in Fig. 2.7. A-B, different activation dynamics can be observed for object- and action-related words in the secondary areas of the non-relevant system (PM_L for object-related words and TO for action-related words).

However, we note that the activation peaks were quite flat in these cases, thus leading to some variance in latencies.

Figure. 6 **Activation time course - PERISYLVIAN AREAS**

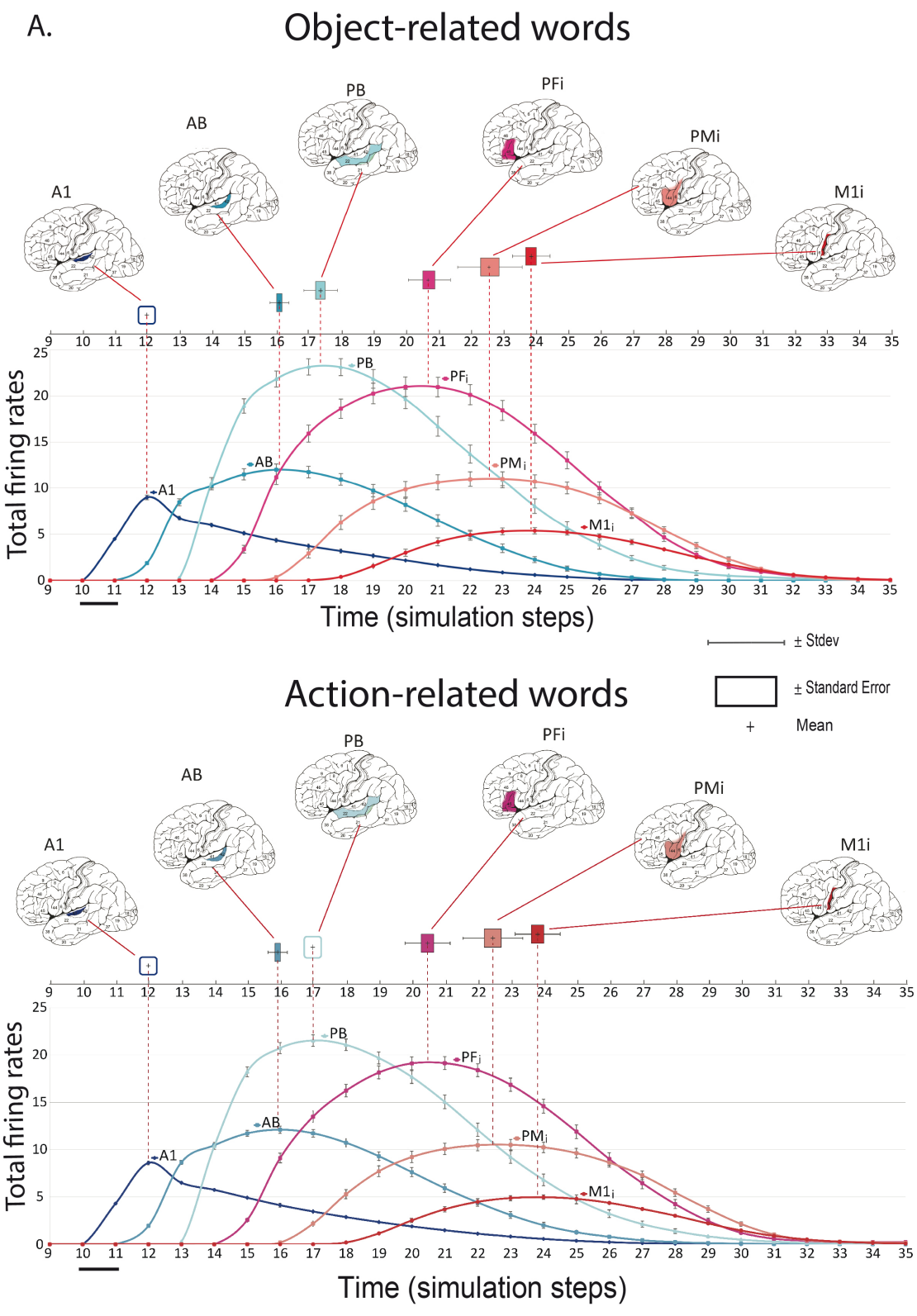

Figure. 7

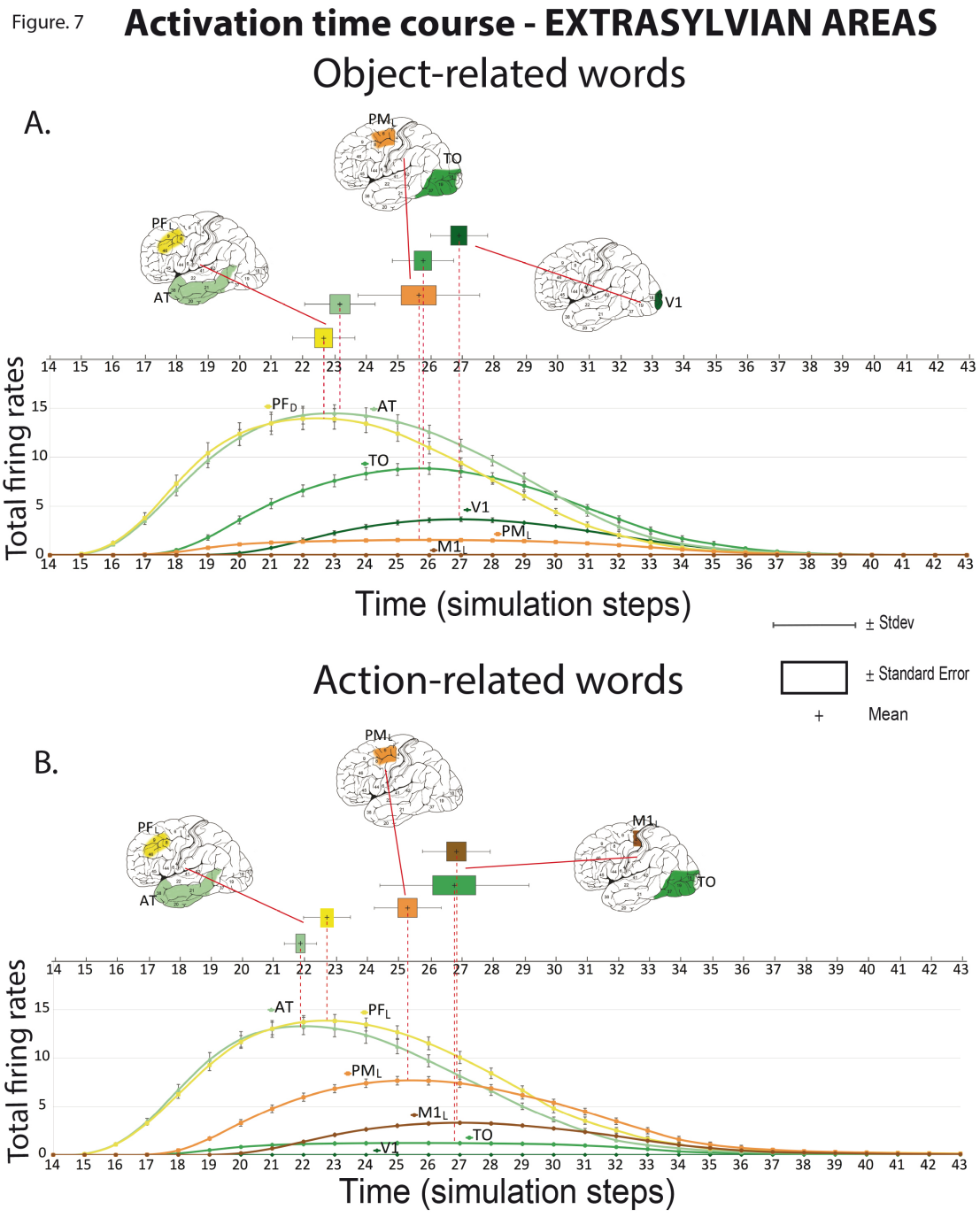

Figure 2.6 & 2.7 Spatio temporal activation patterns in the network: All curves (bottom part of each panel) illustrate area-specific activation dynamics plotted against time during the neurophysiological word recognition processes (time is in simulation time steps). Fig 2.6 shows simulation results for the six perisylvian model areas and Fig 2.7 shows them for the six extrasylvian areas. Brain schematics (at the top of each panel) highlight the cortical locations of the areas for each specific activation curve and peak. Two or more areas are plotted into the same brain schematic if there were no significant delay differences between their peak activations (Bonferroni-corrected for 24 comparisons; critical threshold $p < .0020$). The latency of maximal activation together with standard errors (boxes) and standard deviations (whiskers) are illustrated by a given boxplot. The small horizontal segment indicates stimulus onset and offset. Averages and statistics are calculated across 12 different

To confirm these observations about the activation time-course across areas for the different word-related CAs, we ran the same 4-way ANOVA as in the previous sections, but not using peak activation latencies. The statistical analysis revealed a significant interaction of factors WordType, PeriExtra, TempFront and Areas ($F2,22 = 3615.08$, $p < .0001$), which confirms the different area-specific activation time-courses between the two word type circuits. Once again, the perisylvian cortex showed no significant differences between circuit types across the six areas ($F2,22 = 0.4$, $p = .68$). The extrasylvian cortex revealed a significant interaction of the factors WordType, TempFront and Areas ($F2,22 = 4791.15$, $p < .0001$), which confirms a different activation time-course of the extrasylvian areas for object- and action-related words, as described above.

We further ran a Bonferroni-corrected planned-comparison test (24 comparisons, corrected $p < .0020$) to investigate the possible difference in temporal activation between the two word-types across the neural-network. Similar activation time-course for the two word types/circuits were found across the network areas, except for the temporo-occipital (TO, $p = 0.001$) and the anterior-temporal (AT, $p = 0.0002$) visual areas. Activation times for each word/circuit type showed no significant differences between the extrasylvian connector hub areas (AT, PF_L: $p > 0.0080$), which, however, activated significantly earlier than the modality-preferential ones ($p < 0.001$). Intriguingly, comparisons between modality-preferential cortices showed significant differences, expect between TO and PM_L ($p = 0.66$) for object-related word circuits and between TO and $M1_L$ ($p = 0.77$) for action-related ones. In the perisylvian language cortex, all comparisons between area-peak activation times showed significant differences ($p < 0.001$) (see Fig. 2.6 & 2.7, i.e. brain/boxplot).

For putative comparison of model data with experimental data (see also Discussion below), a further analysis of the activation dynamics was performed. Activation to both word types across sub-systems unfolded symmetrically in the perisylvian and extrasylvian cortex ('Motor'-then-'Visual' *vs*. 'Visual'-then-'Motor' – see Fig. 2.5., right-hand side). These observations were fully confirmed by the 2-way ANOVA run on the data of the two systems separately (i.e. peri- and extra-sylvian systems), with factors WordType (2 levels: *object* vs. *action*) and TempFront (2 levels: *temporal areas* vs. *frontal areas*). The statistical analysis showed a significant interaction of WordType and TempFront ($F1,11 = 24.52$, $p <$

.0004; action words, dorsal motor sub-system: 25 simulation time-steps, ventral visual sub-system: 24.37, object words, dorsal motor sub-system: 24.14, ventral visual sub-system: 25.27) in the extrasylvian systems, confirming the symmetrical time-course of activation of the two word types, with no differences in the perisylvian language cortex ($F1,11 = 0.6$, $p < 0.46$.). Notably, the significant interaction was due to slower average activation times in the relatively more relevant semantic system (dorsal action sub-system for action words, ventral visual sub-system for object words) compared with the less relevant sub-systems, a feature due to the absence of (slow) activation in the respective primary areas (see Fig. 2.5.).

Discussion

A neurocomputational model implementing a range of cortical areas in frontal, temporal and occipital lobes along with main features of their connectivity structure and neurophysiologically realistic learning mechanisms offers an explanation of known facts about the cortical basis of meaning processing, in particular, the fact that some areas serve a general role in semantic processing, whereas others primarily take a category-specific role. When the model was used to mimic semantic grounding of word-forms in action and perceptual information in motor and visual cortex, distributed neuronal assemblies developed, which functioned as 'semantic circuits' insofar as they interlinked information about word-form and meaning. Intriguingly, these semantic circuits showed different distributions across extrasylvian modality-preferential areas, as already found in a previous simulation study (Garagnani and Pulvermüller, 2016). This replicates the category-specificity of action and object words, which, in a range of neuroimaging studies, more strongly activated dorsolateral motor and ventral-stream visual areas, respectively. In contrast to the category-specific behaviour of modality-preferential areas outside the perisylvian domain, substantial amounts of neuronal machinery in connector hub areas in prefrontal and anterior temporal cortex were involved to similar degrees in both kinds of cell assemblies, consistent with a role of these connector hubs as 'semantic hubs'. As in-degree normalisation was used in the present simulations, we argue below that this functional segregation into general and category-specific semantic areas resulted from connectivity structure and especially the high 'degree' of connector hubs, rather than from overall strength of the input. In fact, in contrast to earlier work (Garagnani and Pulvermüller, 2016),

area function only gradually changed from category-specificity towards a category-general role, with even connector hubs exhibiting a degree of category-specificity, a feature which may be due, in part, to the inclusion of additional connections based on neuroanatomical evidence – we return to this issue below. Finally, the novel analysis of the time courses of activation indicated that in word recognition and comprehension, auditory areas are (trivially) activated first, closely followed by connector hub and modality-preferential frontal and temporal areas. Another intriguing observation was that the extrasylvian sub-systems carrying category-specific semantic information about a given word type (i.e., the dorsolateral motor sub-system for action words and the ventral visual sub-system for object words) showed a tendency toward delayed activation relative to the other areas. Moreover, a direct comparison of the activation dynamics of the model with real cortical activations observed during spoken word processing exhibit a degree of consistency (see section 4.2 Fig. 2.8). Below we discuss these findings in light of empirical data, previous neurocomputational work, and future research. It needs also to be emphasized that the present model testes, and demonstrates the validity of a neurobiological theory of language, which claims that semantic content is stored in the brain by distribution of the cell assembly circuits (CAs) spread out across cortical areas, and that the specific cortical distribution (topography) of these circuits across the network reflects semantic information, in particular, semantic category-specificity (see, for example, Pulvermüller, 1999). The semantic models most popular at present still stipulate semantic hubs as the main seat of conceptual and semantic processing without providing neurobiological explanations for such hubs, nor for their specific cortical locations. A purely verbal description of a distributed semantic circuits theory – in terms of 'what fires together must also bind together' – would already provide some plausibility, but one might still object that a working model of relevant cortical areas might give rise to entirely different mechanisms, for example to the emergence of local semantic processing in a single 'interface system' rather than distributed circuits that bind semantic information. Similarly, even if one is inclined to accept that distributed circuits reach into specific sensory and/or motor cortices, it would still be unclear – solely on the basis of a logical argument - whether such 'category-specific' distribution is restricted to primary areas, should include primary and secondary ones, or whether semantic specificity - as indicated by the present results - reaches the highest level

of connection hubs, which, as most models postulate, are category-general and relevant for all semantic categories to the same degree.

Semantic hubs vs. category-specificity in the human brain: explaining both by a neuromechanistic circuit-level model

Diverging theories of semantic representation have been proposed to explain the extensive empirical findings about the brain basis of meaning processing revealed by neuropsychological and neurophysiological/imaging studies in patients and healthy subjects. As mentioned in the introduction, cognitive neuroscience has posited the existence of several convergence areas or 'semantic hubs' that enable associating different aspects of conceptual and semantic knowledge. These areas have been located in the inferior and dorsolateral prefrontal, inferior parietal, superior temporal and anterior ventral temporal cortex, and postulated to equally process the meaning of all types of signs and symbols (Bookheimer, 2002; McCrory et al., 2000; Patterson et al., 2007; Pulvermüller, 2013). A complementary position emphasizes the importance of other cortical regions for semantic processing which are particularly relevant for specific word types related to specific semantic categories, such as animals, tools or actions. A range of relevant neuroimaging studies have shown the relevance of the motor cortex during conceptual processing of action-related words (Dreyer et al., 2015; Grisoni et al., 2016; Hauk et al., 2004; Hauk and Pulvermüller, 2004; Shtyrov et al., 2014) and of the sensory cortex during conceptual processing of visually related words (e.g. colours, animals or object-related words) (Damasio et al., 1996; Tranel et al., 1997). Furthermore, recent neurophysiological studies (EEG-MEG) show early (<200 ms) and automaticity brain activation reflecting semantic differences (e.g., Moseley et al., 2013; Pulvermüller et al., 2005b). This evidence, which we discussed extensively in the introduction above, is consistent with the claim that semantic processing is distributed across, and divided up between, category-general hubs and category-specific areas. The frequently emphasized need for an integrative explanation of both general and semantic areas along with their location (Binder and Desai, 2011; Pulvermüller, 2013) is now being answered by results from the network simulations we report here.

The explanation of hubs and category-specificity requires reference to an intermediate level of computational simulation of neuronal circuits which bind together specific word-

forms and their semantic, meaning-related features (Pulvermüller et al., 2014a). The formation of these semantic circuits results from (i) the correlation structure of 'grounding' sensorimotor semantic information and co-occurring word-forms, (ii) the neurobiologically realistic learning and therefore mapping of the correlations on neuronal connection strengths and (iii) the structural information immanent to the neuroanatomy of cortical areas and their connectivity. As these circuits map sensorimotor correlations, they bridge between those neurons in sensory and motor areas where information – and thus correlated activation – is present during learning. This leads to category-specificity of circuit topographies, with action words such as 'run' yielding cell assemblies reaching into motor systems and object words such as 'sun' being implemented as circuits strongly linking up with neurons in visual cortices (Kiefer et al., 2008; Pulvermüller, 2013). These distributed word-related CA circuits did not extend into the non-relevant sub-systems ($M1_L$ for object- and V1 for action-related words) because neural activity of these areas presented a low degree of correlation. This is because during training these areas were stimulated with random patterns that changed in every learning episode (see Materials and Methods). Consequently, following the correlation based learning rule, object-related CA circuits exhibited a larger density in the visual (V1, TO, AT) than in the motor areas ($M1_L$, PM_L, PF_L) and vice versa for action-related words (Fig. 2.3).

It should be clarified here that the presence of a random-noise pattern to the non-relevant sub-systems was necessary to prevent the extensions of the semantic circuits into motor areas for object-related and visual areas for action-related words. In fact, in an additional set of word learning simulations, network training without the random noise pattern being present in the non-relevant sub-systems failed to produce a category-specific distribution. This observation further documents the important function of neuronal noise in the brain and in brain-like networks (Doursat and Bienenstock, 2006), which prevents excessive CA growth. We conclude that noise in primary areas is critical for obtaining semantic cortical circuits with category-specific signatures. In essence, as it is important to learn that the word 'run' relates to certain motor patterns, it is likewise important to learn that variable visual inputs ('noise') typically occur during running so that specific visual features are de-correlated from the word-form. We note that under deprived conditions, for example in blind language learners, this type of de-correlating sensory-related noise is

missing in the deprived primary cortex. Resultant CA growth into the ventral stream may explain why blind individuals activate visual areas in linguistic and semantic processing (see Bedny et al., 2011; Neville and Bavelier, 2002).

In order to connect information about actions and perceptions available in the primary cortices, activity must run through connector hub areas. Therefore, neurons in multimodal cortices are included in all types of semantic circuits to a similar degree. This explains the existence and cortical location of semantic hubs in inferior and dorsolateral prefrontal cortex and in anterior and superior temporal cortex. Our model did not include areas of the parietal cortex, but if it did, it is foreseeable that the same localisation mechanisms will apply to the additional lobar system so that an additional 'semantic hub' in posterior parietal cortex (posterior supramarginal gyrus, intraparietal sulcus and angular gyri) might emerge. A new finding of the present work is the emergence of a degree of category-specificity also in extrasylvian hub areas. Earlier simulations by Garagnani and Pulvermüller (2016) had found no category differences in any of the hub areas. This may have been due, in part, to the reduced input to extrasylvian hub areas implicated by the absence of connections between ventral and dorsolateral prefrontal cortex and likewise between anterior inferior and posterior superior temporal cortex. As these connections have meanwhile been documented by anatomical studies (Gierhan, 2013; Yeterian et al., 2012), they were included in the present simulations and a small but significant degree of category-specificity in these hub areas was the consequence.

A fruitful target for future research will be to investigate the possibility of category-specific semantic deficits after lesions in anterior temporal and dorsolateral prefrontal cortex. In this context, a closer look at patients in early stages of semantic dementia may be crucial, because these patients sometimes show lesions restricted to anterior and inferior temporal areas (Patterson et al., 2007). Some work in this field suggests no differences in processing different semantic categories (Lambon Ralph et al., 2007), but other studies have reported some differences, for example between colour- and form-related words (Gainotti, 2012; Pulvermüller et al., 2010). Stroke- and encephalitis-induced lesions of the multimodal parts of the left temporal lobe (corresponding to area AT in the network) have also been found to cause category-specific word processing deficits for animals, persons, and living things (Damasio et al., 1996; Gainotti, 2012; Hernández et al., 2008; Pulvermüller et al.,

2010). Thus, it seems that there is at least some evidence for category-specificity in the extrasylvian anterior-temporal connector hubs. Only future research can validate or falsify the model's prediction about a slight but significant category difference between object and action-related words after focal anterior-temporal and dorsolateral prefrontal damage.

There is quite a bit of debate about the prominence of different areas for semantic processing. Some approaches hold that true semantic processing is only present in the multimodal hubs, and modality-preferential areas only serve an optional, 'enriching' or 'colouring' function (Mahon and Caramazza, 2008). Although the network model we present here offers no justification for such a view – because all parts of the distributed semantic circuits contribute to their function and there is no basis for excluding circuit parts when it comes to function – the model offers an explanation of why some areas across which the circuits are distributed are functionally more important than others. Factors which come in here are the general location of an area's neurons with respect to the network's connectivity structure (topology) – with gradually more functional contributions of 'central' areas than 'peripheral' ones – and, importantly, the relative CA neuron density a circuit shows across areas. In this context, the generally observed main effects of the level of area, with relatively more CA neurons in secondary than primary and also much more neurons in connector hub areas than in secondary ones, is of critical importance. In the previous study (Garagnani and Pulvermüller, 2016), it was not entirely clear whether the relatively high number of CA neurons (and thus circuit neuron densities) in connector hub areas was due to the stronger input these areas generally received (higher 'in-degree') or to the network topology, or both. Here, we performed in-degree normalization (see Methods) and thus excluded the sheer amount of activity entering an area as explanatory factor. In spite of in-degree normalization across sub-systems, which ensured that all network areas received (on average) equal quantities of inputs, circuit cell density was still higher in the connector hub areas in the centre of the network architecture, where phonological and semantic word circuits converge. This result is consistent with the statement that network topology plays a major role in determining the prominence of connector hubs for general semantic processing. However, we note that larger circuit densities in the 'centre' of networks have also been observed with next neighbour between-area connections only, suggesting that,

apart from its 'degree' and resultant hub status as such, the 'centrality' of an area within the network is a relevant factor (Garagnani et al., 2008).

In sum, the present neural network simulations exhibit the spontaneous formation of semantic CA circuits distributed over modality-preferential and 'higher' multimodal convergence areas and mechanistically explain the emergence in the cortex of both category-specific and general semantic processes. In addition, the use of a more realistic architecture leads to the presence of moderate category-specificity in connector hub areas outside the perisylvian region. The spontaneous formation of these semantic circuits is based on, and explained by, well-documented learning mechanisms of Hebbian synaptic plasticity and cortical area and connectivity structure. These simulation results explain why modality-preferential areas are activated relatively more strongly by specific semantic categories and why the connector areas become semantic hubs and to a degree similarly great, relevance for processing all kinds of meanings.

Neurophysiological mechanisms underlying word recognition and understanding: simulating the time-course of semantic activation

The semantic circuits that had formed as a consequence of correlation learning were reactivated from the acoustic phonological end to simulate the area-specific cortical activation dynamics of spoken word understanding and to provide a functional estimate of category-general and -specific semantic activation strength, topography, and timing. Comparison of maximum circuit activity levels per area and word type revealed a dissociation similar to that found in the structural analysis of circuit topographies reported above. In particular, object-related words activated the visual system (V1, TO) more strongly than the motor system ($M1_L$, PM_L) and, for action-related words, motor system peak activation was relatively stronger (Fig. 2.5 B – left-hand side). As before, the perisylvian auditory and articulatory sub-systems did not show any significant difference in amplitude between word-types (see Fig. 2.5. A – left-hand side). Stronger activation in the connector hubs (AT, PF_L, PB, PF_i) than in secondary (TO, AB, PM_L, PM_i) areas, and stronger activation in secondary than in primary (A1, V1, $M1_L$, $M1_i$) areas, was found. The word-category dissociation and the different activity levels predicted during simulated word-recognition processes is a direct consequence of the distinct cortical topographies of object- and action-

related semantic circuits, which emerged in the model during learning, with more CA cells leading to correspondingly more activity during CA circuit ignition.

The area-specific activation time-course of the multi-area network illustrated in Fig. 2.6 and 2.7 (Brain and boxplot upper part) showed similar activation dynamics for object- and action-related words. For both word types, the perisylvian language system exhibited a cascade of activations whose peaks unfold (in a sequential manner) over a period of approximately 12 simulation time-steps. Activation was first present in the primary auditory areas A1, driven by the external stimulus, and then spread across the perisylvian areas, terminating in the primary articulatory areas ($M1_i$). In contrast, activation in the sensorimotor semantic areas is near-simultaneous, with all peaks concentrated within a period of just 5 simulation time-steps (hub areas activate first, regardless of word type). The 'near-simultaneous' effect of the CA cells activation processes in sensorimotor areas is caused by the rich neuroanatomical connections of the convergence hub areas, which link together the different modality-preferential cortices. Therefore, upon reaching the language hubs ($PB\text{-}PF_i$), activity leads to the simultaneous 'ignition' of the CA cells present in the anterior-temporal (AT) and dorsolateral prefrontal (PF_L) hub cortices, which, in turn, quickly activate the modality-preferential CAs. Thus, the inherent connectivity structure of the model leads to a near-simultaneous activation of the most richly connected hub areas as compared to the primary and secondary cortices. The multimodal hubs can be seen as a 'crossroad' where information from different modality-preferential systems converges; after full ignition, CA activity gradually disappears in the multi-area network (see Fig. 2.4), ending in the modality-preferential areas – i.e. primary hand-motor area ($M1_L$) for action-, and primary visual cortex (V1) for object-related words. In other words, the modality-preferential cortices (for object words V1 and TO areas and for action words $M1_L$ and PM_L areas) activate after all other areas. Hence, on the basis of the activation dynamics exhibited by the present model, we would predict that during semantic information retrieval, activation should spread in a cascade-like fashion across the perisylvian language areas; sensorimotor areas should then activate near-simultaneously, with semantic hubs activating before the modality-preferential areas, where additional semantic information is held.

In a recent simulation study (McNorgan et al., 2011), on the basis of within- and cross-modality feature- and concept-relatedness judgment data the authors argue that 'deep'

models of semantic grounding (i.e., which involve several processing steps between sensory, and between sensory and motor components) are necessary to explain their results. Because our model is neuroanatomically realistic and, as such, it incorporates indirect multi-step links between modality-preferential sensorimotor regions, it can be considered a neurobiologically motivated 'deep' semantic model in the sense of McNorgan et al. Therefore, we conjecture that it might also be compatible with their results, although, as our present focus was on modelling neurophysiological mechanisms, we have not attempted to replicate the outcome of their specific simulations.Experimental studies analysing the latency of semantic processes in language perception suggest that semantic information provided by words is already retrieved within ~200 ms after stimulus presentation (Brown and Lehmann, 1979; Hauk et al., 2008; Preissl et al., 1995; Pulvermüller, 1999). Moreover, recent MEG-EEG recordings have shown that different semantic categories (visually presented) activated different cortical areas within ~150 ms; at this point in time, action words activated mostly the motor system and object words activated the visual system (Moseley et al., 2013). However, these neuroimaging techniques with high temporal resolution (such as MEG and EEG) do not offer a sufficiently high spatial resolution to detect fine-grained differences between multimodal semantic hubs and modality-preferential areas implemented in the neural network (for example, between premotor and prefrontal areas). Therefore, we further investigated the activation dynamics of the four sub-systems, i.e. auditory, articulatory, visual and motor sub-systems implemented in the model, and compared their respective average activation time courses with each other and with real cortical activations observed during spoken word processing.

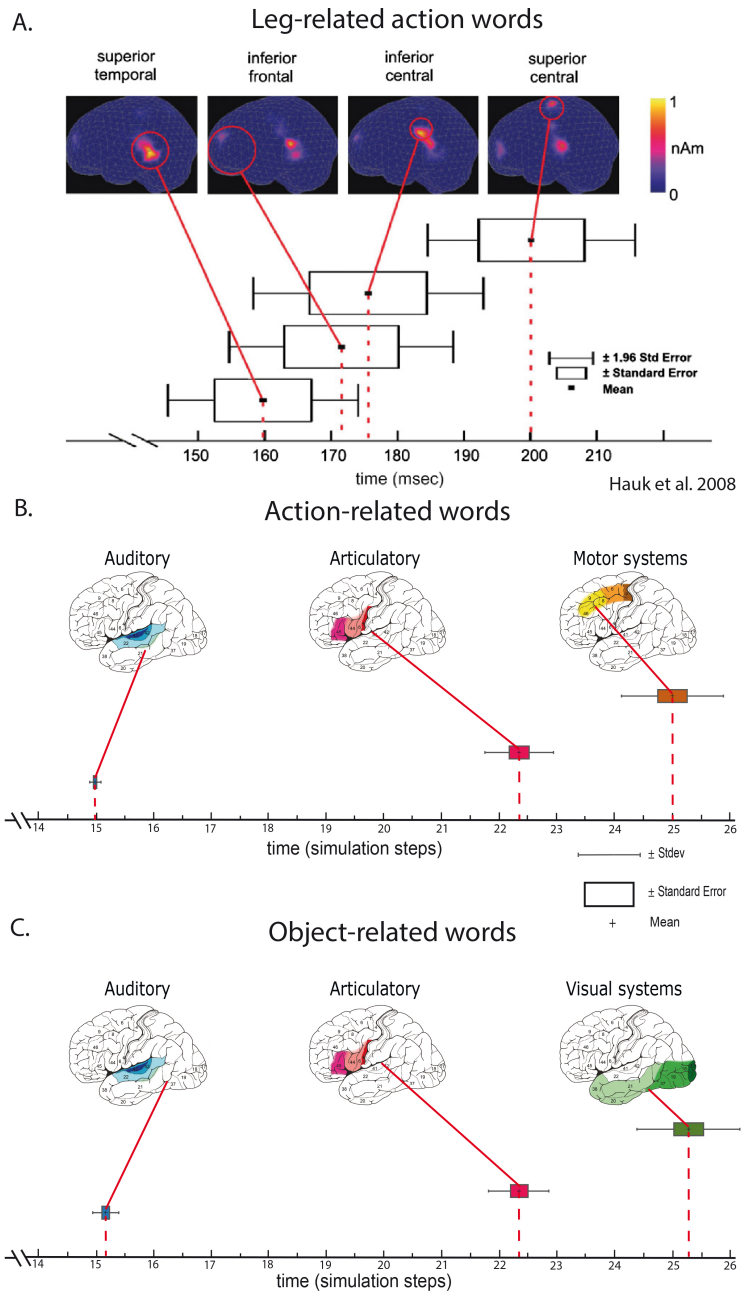

Figure 2.8. Comparison of real and simulated brain activations elicited by specific semantic word categories. **(A)** Time course of activation of cortical areas elicited by passive presentation of spoken action words and determined using magnetoencephalography (MEG) and distributed source localizations. Action words elicited sequential but near-simultaneous activations in left superior temporal, inferior frontal and superior central cortex. The average latency of maximal activation in the four ROIs is reported together with the standard errors (boxes; bars indicate 1.96 SE, data adapted from Pulvermüller et al., 2005b). The boxplots in panels B & C illustrate results from the corresponding simulated activation time-courses. The point in time at which stimulus-evoked activity is peaking in each of the modelled four sub-systems (auditory, articulatory, visual and motor systems) is plotted against time given in simulation time-steps. Boxes give standard errors and whiskers standard deviations. The average was computed across the 12 different networks and calculated separately for **(B)** Action and **(C)** Object-related words. Notice that the respective non-relevant sub-systems (Visual for action- and motor for object-related words) are not illustrated here, as the activation levels are relatively low.

Fig. 2.8 reports results from a Magnetoencephalography (MEG) study investigating the temporal activation dynamics evoked by action-related words (Pulvermüller et al., 2005b) and relates them to the activation time courses obtained from our model after stimulating area A1 with the 'acoustic patterns' of action- and object-related 'words'. Although the alignment of simulation time-steps and real time is always to a degree tentative, the near-simultaneous but still fast-cascading activation from superior temporal to inferior frontal and finally dorsal action-related areas exhibited by the cortical sources estimated from the MEG recordings is paralleled by the model results. Note, however, that the delay between superior temporal and inferior frontal activations is relatively longer in the simulations than in the MEG sources, thus also indicating a discrepancy. For relating simulation results more directly to empirical data, it might be advantageous to perform analogous semantic learning experiments in healthy subjects and then compare the brain and network responses of the processing of the learnt items (see also below).

In sum, the model shows a 'near-simultaneous' activation time-course of the semantic areas; the semantic hubs, anterior-temporal (AT) and dorsolateral prefrontal areas (PF_L), activate first, and are then followed by the modality-preferential areas carrying category-specific semantic information. The perisylvian language areas exhibited a cascade of activations, with no word type effects. Most of the empirical studies about semantic processing performed in the past used words from real natural language, making it impossible to control the way these words have been learned, or to isolate the relevant semantic features from the many other putatively confounding psycholinguistic and psychological features distinguishing the different lexical classes between each other (Kemmerer, 2014; Pulvermüller, 1999; Vigliocco et al., 2011). A well-designed word learning experiment employing neuroimaging methods with high spatial and temporal resolution (EEG/MEG and fMRI) is needed to test the validity of the present model's results and predictions, and identify *where* the neural correlates of novel object- and action-related words emerge in the brain, and at *which* point in time of the recognition process their activation occurs.

Summary and Conclusions

Current neurosemantic theories still diverge about the role of category-specific and category-general semantic mechanisms and about the contribution of modality-preferential and multimodal ('amodal') brain systems in semantic processing (Barsalou, 2008; Bookheimer, 2002; Devlin et al., 2003; Gallese and Lakoff, 2005; Martin and Chao, 2001; Patterson et al., 2007; Pulvermüller, 2005; Warrington and McCarthy, 1987). Here we applied a neural-network model replicating anatomical and physiological features of a range of cortical areas including sensorimotor, multimodal and language areas to investigate the neurobiological mechanisms underlying conceptual semantic grounding of words in action- and object-related information. The word learning simulations documented the spontaneous emergence of word/symbol-specific, tightly interconnected cell assemblies within the larger networks, each binding articulatory-acoustic word-forms to sensorimotor semantic information. Due to network structure, connectivity, and Hebbian associative learning, which maps neuronal correlations, the emerging 'semantic circuits' for object- and action-related words exhibited category-specificity primarily in modality-preferential areas; the 'higher' multimodal connector hub areas central to the network architecture showed only moderate category-specificity (Fig. 2.3 and 2.4). Due to their central position in the model architecture, connector hubs showed highest cell densities of both types of semantic circuits, therefore acting as 'semantic hubs'. Word category dissociations were confirmed by the reactivation of the cell assembly circuits during simulated word recognition and comprehension processes. The model's results, which can be compared with real experimental data (see Fig. 2.8), predict a symmetrical temporal activation for object- and action-related words, with the semantic hub areas activating first and modality-preferential ones slightly later (Fig. 2.6 and 2.7). Interestingly, extrasylvian systems relevant for semantic processing of a given word category activated with a delay upon the relevant system, whereby strong dorsal motor systems activation were preceded by weak ventral visual system activation to action words, while strong ventral visual activations to objects words were preceded by weak dorsal motor processes (Fig. 2.5). This observation (prediction) also calls for future experimental testing. The present simulations demonstrate that realistic neurocomputational models can elucidate aspects of semantic processing in the cortex and integrate findings from neuroimaging studies. In sum, the model illustrates the spontaneous

emergence of both category-specific and general semantic hub areas and, on the basis of well-established neuroscience principles, offers a mechanistic explanation of *where* and *when* meaning is processed in the brain.

Appendix 2A – Full model specification

Each of the 12 simulated areas (see Fig. 2.1.B) was implemented as two layers of artificial neuron-like elements ('cells'), 625 excitatory and 625 inhibitory, thus resulting in 15,000 cells in total. Each excitatory cell '*e*' can be considered the network equivalent of a local cluster, or column, of approximately 25,000 real excitatory cortical neurons, that is pyramidal cells, while its twin inhibitory cell '*i*' (see Fig. 2.1.C) models the cluster of inhibitory interneurons situated within the same cortical column (Eggert and van Hemmen, 2000; Wilson and Cowan, 1972). The activity state of each cell *e* is uniquely defined by its membrane potential $V(e,t)$, representing the average of the sum of all (excitatory and inhibitory) postsynaptic potentials acting upon neural pool (cluster) *e* at time *t*, and governed by the following equation:

$$\tau \cdot \frac{dV(e,t)}{dt} = -V(e,t) + k_1(V_{In}(e,t) + k_2\eta(e,t)) \qquad \text{(A1)}$$

where $V_{In}(e,t)$ is the net input to cell *e* at time *t* (sum of all inhibitory and excitatory postsynaptic potentials – I/EPSPs; inhibitory synapses are given a negative sign – plus a constant baseline value V_b), τ is the membrane's time constant, k_1, k_2 are scaling constants and $\eta(\cdot,t)$ is a white noise process with uniform distribution over [-0.5,0.5]. Note that noise is an inherent property of each model cell, intended to mimic the spontaneous activity (baseline firing) of real neurons. Therefore, noise was constantly present in all areas, in equal amounts (inhibitory cells have k_2=0, i.e., the noise is generated just by the excitatory cells, for simplicity).

Cells produce a graded response that represents the average firing rate of the neuronal cluster; in particular, the output (transformation function) of an excitatory cell *e* at time *t* is:

$$O(e,t) = \begin{cases} 0 & if\ V(e,t) \le \varphi \\ (V(e,t) - \varphi) & if\ 0 < (V(e,t) - \varphi) \le 1 \\ 1 & otherwise \end{cases} \qquad \text{(A2)}$$

O(*e,t*) represents the average (graded) firing rate (number of action potentials per time unit) of cluster *e* at time *t*; it is a piecewise-linear sigmoid function of the cell's membrane potential *V*(*e,t*), clipped into the range [0, 1] and with slope 1 between the lower and upper thresholds φ and $\varphi + 1$. The output O(*i,t*) of inhibitory cell *i* is 0 if *V*(*i,t*)<0, and *V*(*i,t*) otherwise. In excitatory cells, the value of the threshold φ in Eq. (A2) varies in time, tracking the recent mean activity of the cell so as to implement neuronal adaptation (Kandel et al., 2000). Thus, stronger activity leads to a higher threshold in subsequent time-steps. More precisely,

$$\varphi(e,t) = \alpha \cdot \omega(e,t) \qquad \text{(A3)}$$

where $\omega(e,t)$ is the time-average of cell *e*'s recent output and α is the 'adaptation strength' (see below for the exact parameter values used in the simulations). For an excitatory cell *e*, the approximate time-average $\omega(e,t)$ of its output O(*e,t*) is estimated by integrating the linear differential equation Eq. (A4.1) below with time constant τ_A, assuming initial average $\omega(e,0)=0$:

$$\tau_A \cdot \frac{d\omega(e,t)}{dt} = -\omega(e,t) + O(e,t) \qquad \text{(A4.1)}$$

Local (lateral) inhibitory connections (see Fig. 2.1.C) and area-specific inhibition are also implemented, realising, respectively, local and global competition mechanisms (Duncan, 2006, 1996) and preventing activation from falling into non-physiological states (Braitenberg and Schüz, 1998). More formally, in Eq. (A1) the input $V_{In}(e,t)$ to each excitatory cell of the same area includes an area-specific ('global') inhibition term $k_S \cdot \omega_S(e,t)$, which is subtracted from the total sum of the I/EPSPs postsynaptic potentials V_{In} in input to the cell, with $\omega_S(e,t)$ defined by:

$$\tau_S \cdot \frac{d\omega_s(e,t)}{dt} = -\omega_s(e,t) + \sum_{e \in area} O(e,t) \qquad \text{(A4.2)}$$

The low-pass dynamics of the cells (Eq. (A1), (A2), (A4.1-2)) are integrated using the Euler scheme with step size Δt, where Δt = 0.5 ms.

Excitatory links within and between (possibly non-adjacent) model areas are established at random and limited to a local (topographic) neighbourhood; weights are initialised at pattern, in the range [0, 0.1]. The probability of a synapse to be created between any two

cells falls off with their distance (Braitenberg and Schüz, 1998) according to a Gaussian function clipped to 0 outside the chosen neighbourhood (a square of size n =19 for excitatory and n=5 for inhibitory cell projections). This produces a sparse, patchy and topographic connectivity, as typically found in the mammalian cortex (Amir et al., 1993; Braitenberg and Schüz, 1998; Douglas and Martin, 2004; Kaas, 1997).

The Hebbian learning mechanism implemented simulates well-documented synaptic plasticity phenomena of long-term potentiation (LTP) and depression (LTD), as implemented by Artola, Bröcher and Singer (Artola et al., 1990; Artola and Singer, 1993). This rule, which covers both 'true' Hebbian co-occurrence ('what fires together wires together') as well as decorralative 'anti-Hebb' ('neurons out of sync delink') plasticity, provides a realistic approximation of known experience-dependent neuronal plasticity and learning (Finnie and Nader, 2012; Friedman and Donoghue, 2009; Malenka and Bear, 2004). In the model, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, $+\Delta w$ and $-\Delta w$ (with Δw<<1 and fixed). Following Artola *et al*., we defined as 'active' any link from an excitatory cell x such that the output O(x,t) of cell x at time t is larger than θ_{pre}, where $\theta_{pre} \in]0,1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP (or LTD) to occur. Thus, given any two cells x and y connected by a synaptic link with weight $w_t(x,y)$, the new weight $w_{t+1}(x,y)$ is calculated as follows:

$$w_{t+1}(x,y) = \begin{cases} w_t(x,y)+\Delta w & \textit{(LTP)} \quad \text{if } O(x,t) \geq \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y)-\Delta w & \textit{(LTD)} \quad \text{if } O(x,t) \geq \theta_{pre} \text{ and } \theta_- \leq V(y,t) < \theta_+ \\ w_t(x,y)-\Delta w & \textit{(LTD)} \quad \text{if } O(x,t) < \theta_{pre} \text{ and } V(y,t) \geq \theta_+ \\ w_t(x,y) & \textit{(no change)} \quad \text{otherwise} \end{cases} \qquad \text{(A5)}$$

Table 2.1. Parameter values used during the simulations

Eq. (A1)	Time constant (excitatory cells):	τ=2.5 (simulation time-steps)
	Time constant (inhibitory cells):	τ =5 (simulation time-steps)
	Scaling factor:	k_1=0.01
	Baseline potential	V_b=0
	Noise scaling factor	k_2=27·$\sqrt{48}$
	Global inhibition during training	k_S=95
	(during word recognition:	k_S=75)
Eq. (A3)	Adaptation:	α=0.01
Eq. (A4.1-2)	Time-average constant for CA definition:	τ_z =3 (simulation time-steps)
	(time constant for adaptation mechanism:	τ_A=15)
	Global inhibition time constant:	τ_S=12 (simulation time-steps)
Eq. (A5)	Postsynaptic potential thresholds for LTP/LTD:	
		θ_-=0.15
		θ_+=0.15

Presynaptic output activity required for any synaptic change:

θ_{pre}=0.05

3. A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity

---

This chapter has been published as:

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., (2018). A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity. *Front. Comput. Neurosci.* 12, 88. doi:https://doi.org/10.3389/fncom.2018.00088

Abstract

One of the most controversial debates in cognitive neuroscience concerns the cortical locus of semantic knowledge and processing in the human brain. Experimental data revealed the existence of various cortical regions relevant for meaning processing, ranging from semantic hubs to modality sensorimotor areas, involved in the processing of specific conceptual categories. *Why* and *how* the brain uses such complex organization for conceptualization can be investigated using biologically constrained neurocomputational models. Here, we improve pre-existing neurocomputational models of semantics by incorporating spiking neurons and a rich connectivity structure between the model 'areas' to mimic important features of the underlying neural substrate. Semantic learning and symbol grounding in action and perception were simulated by associative learning between co-activated neuron populations in frontal, temporal and occipital areas. As a result of Hebbian learning of the correlation structure of symbol, perception and action information, distributed cell assembly circuits emerged across various cortices of the network. These semantic circuits showed category-specific topographical distributions, reaching into motor and visual areas for action- and visually-related words, respectively. All types of semantic circuits included large numbers of neurons in multimodal connector hub areas, which is explained by cortical connectivity structure and the resultant convergence of phonological and semantic information on these zones. Importantly, these semantic hub areas exhibited some category-specificity, which was less pronounced than that observed in primary and secondary modality-preferential cortices. The present neurocomputational model integrates seemingly divergent experimental results about conceptualization and explains both semantic hubs and category-specific areas as an emergent process causally determined by two major factors: neuroanatomical connectivity structure and correlated neuronal activation during language learning.

Introduction

Although the brain mechanisms of meaning processing have been investigated for many years, cognitive neuroscientists have not reached a consensus about the function and the organizational principles of semantic knowledge. A range of neuroimaging and neuropsychological patient studies suggest a contribution of several cortical areas to semantic processing, but the precise role of each of them is still subject to debate. Cognitive and neuroscientists have suggested that the meaning of all words are equally processed and stored in a central 'symbolic system' cortically located in a 'semantic hub'. However, 'semantic hubs' have been proposed in different cortical regions, including the anterior-inferior-temporal lobe (Patterson et al., 2007; Ralph et al., 2017), the anterior-inferior-parietal (Binder et al., 2009; Binder and Desai, 2011) and the posterior-inferior-frontal cortex (Bookheimer, 2002; Carota et al., 2017; Posner and Pavese, 1998; Schomers and Pulvermüller, 2016; Tate et al., 2014). Whereas it is possible, in principle, that several semantic hubs co-exist, some researchers postulated the need for bringing together all semantic information into one focal area and consequently reject the existence of multiple semantic hubs (Patterson et al., 2007; Ralph et al., 2017). Furthermore, and over and above semantic hubs generally contributing to all types of semantics, the phenomenon of category-specific semantic processing has long been in focus (McCarthy and Warrington, 1988; Shallice, 1988): modality-preferential cortices, including visual, auditory, olfactory, gustatory, somatosensory and motor regions, have been shown to differentially activate when specific semantic types are processed, for example animal vs. tool nouns or verbs typically used to speak about different types of actions (Chao et al., 1999; Damasio et al., 1996; Grisoni et al., 2016; Hauk et al., 2004; Kemmerer et al., 2012; Vukovic et al., 2017). Also studies of patients with lesions in modality-specific regions revealed category-specific semantic deficits (Damasio et al., 1996; Dreyer et al., 2015; Gainotti, 2010; Neininger and Pulvermüller, 2003; Trumpp et al., 2013; Warrington and Mccarthy, 1983) which can not be explained by symbolic systems accounts presuming category-general semantic hubs. Likewise, these findings challenge proposals that see the semantic processing role of sensorimotor areas as optional, ancillary or epiphenomenal and deny them a genuine semantic conceptual function (Caramazza et al., 2014; Machery, 2007; Mahon and Caramazza, 2008). The evidence for multiple hubs and modality-specific areas for

conceptual-semantic knowledge is difficult to reconcile within most current neurobiological models of symbol processing.

To incorporate the diverging semantic theories and data from healthy and patient studies described above, it is necessary to build sophisticated models of relevant cortical areas that are biologically constrained by mimicking relevant features of brain function and connectivity. Ideally, such brain-constrained models may predict and offer mechanistic explanations for semantic processing in the human brain. Potentially, such modeling efforts can confirm a given theoretical framework, for example the existence of distributed semantic circuits spread out across several semantic hubs and modality-preferential areas or, as an alternative, the existence of a single focal 'semantic hub'. Based on previous integrative proposals (Damasio, 1989; Pulvermüller, 2013), we hypothesize that semantic category-specific and category-general behaviours of different cortical areas are a direct consequence of the neuroanatomical connectivity between the areas involved and learning experiences that are essential for grounding concepts in knowledge about objects and actions. Here, we attempt to address this theoretical hypothesis with a neurobiologically constrained spiking model of the cortex that in order to integrate data from healthy and patient studies described above.

Recent simulations of cortical function and learning incorporating fine microstructural and physiological details of millions of neurons (Izhikevich and Edelman, 2008; Markram et al., 2011) have not yet addressed specific questions about the neurobiological basis of specific cognitive functions such as semantic processing. Previous connectionist models have made significant progress in explaining language and semantic processing (Dell et al. 1999; Plaut and Gonnerman 2000; Christiansen and Chater 2001), but most of them do not attempt to replicate realistic properties of the human brain. Although recent simulation studies included neuroanatomical information to model semantic processing, they have used learning mechanism (i.e. back-propagation - Chen et al., 2017; Ueno et al., 2011), which were argued to be biologically implausible (Mazzoni et al., 1991; O'Reilly, 1998). Furthermore, these studies have incorporated just one semantic hub area in the anterior temporal lobe, whereas other evidence summarized above are not addressed. A recent modelling effort incorporates neuroanatomical structure and connectivity into models of semantic processing (Garagnani and Pulvermüller, 2016). By meticulously mimicking the

general parcellation of cortex into areas, their long-range cortico-cortical connections, features of local connectivity within cortical areas, local and global inhibitory mechanisms regulating cortical activity, and realistic neurobiological learning mechanisms, a stepwise approximation to response properties of real brain-internal networks could be achieved. Still, these previous implementation study has fallen short of implementing the complexity of cortico-cortical connectivity and the activation dynamics of spiking cortical neurons.

Building upon these previous efforts with graded-response neural-network models (Garagnani and Pulvermüller, 2016), we here set out to model the brain's semantic mechanisms using a mathematically precise model of multiple cortical areas, incorporating spiking neurons, biologically plausible non-supervised learning mechanisms and connectivity structure based on neuroanatomical studies. The network was used to simulate associative word learning by linking word-forms with their semantically-related object and action representations. The present biologically constrained model bridges the gap between neural mechanisms and conceptual brain functions, offering a biological account of how aspects of word meaning are acquired, stored, and processed in the brain.

Methods and Materials

General features of the model

We implemented a neurobiologically constrained model replicating cortical areas of fronto-temporo-occipital lobes and their connectivity to shed light on the mechanism underlying semantic processing grounded in action and perception. We created a neural architecture with 15,000 representative neurons for simulating activity in twelve cortical areas in the left language-dominant hemisphere (see Fig 3.1A). These 'areas' represented three levels of processing – primary, secondary and higher-association cortex – in four modality-systems: (motor) frontal superior-lateral hand-motor, (articulatory) inferior face-motor, (auditory) superior-temporal and (visual) inferior-temporo-occipital system. Two of these, the auditory and articulatory systems (areas highlighted in blue and red, Fig 3.1A) are in perisylvian language cortex and appear most relevant for language processing (Fadiga et al., 2002; Pulvermüller, 1999; Pulvermüller and Fadiga, 2010; Zatorre et al., 1996). The motor and visual system (yellow and green highlighted areas) are outside the perisylvian language

cortex (called 'extrasylvian' in the present work) and involved in processing visual object processing (Ungerleider and Haxby, 1994), and the execution of manual actions (Deiber et

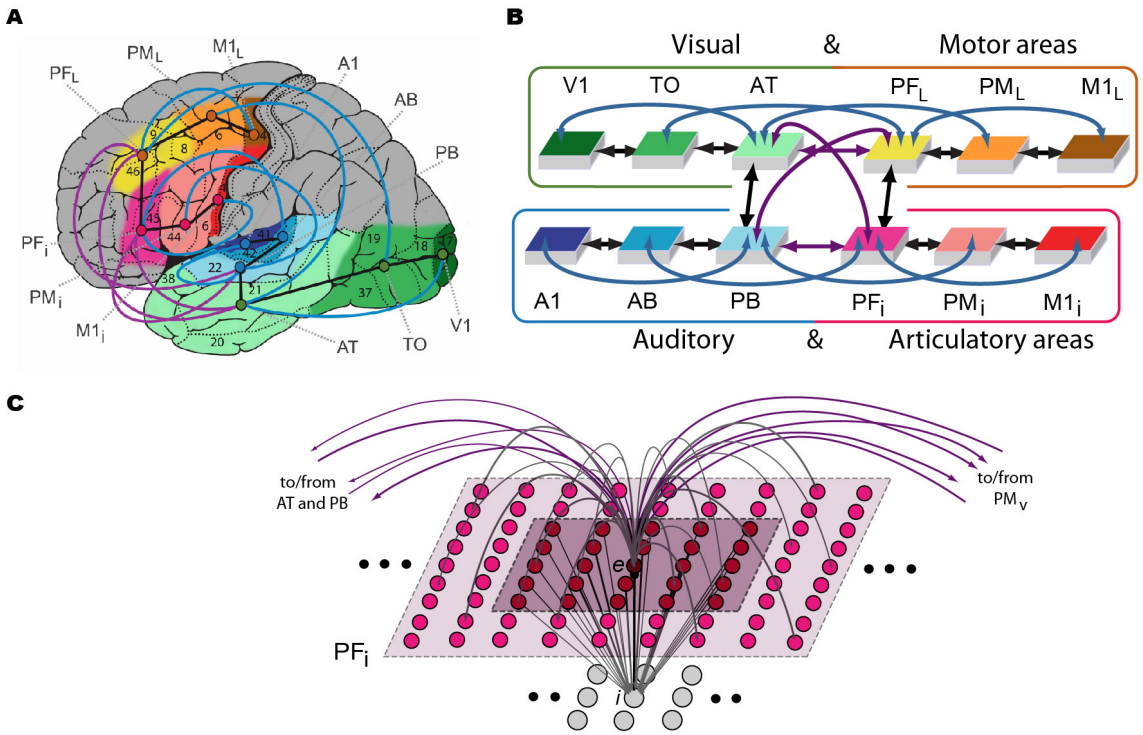

Figure 3.1. (A) Structure and connectivity of 12 frontal, temporal and occipital cortical areas relevant for learning the meaning of words related to actions. Perisylvian cortex comprises an inferior-frontal articulatory-phonological system (red colours), including primary motor cortex ($M1_i$), premotor (PM_i) and inferior-prefrontal (PF_i), and a superior-temporal acoustic-phonological system (areas in blue), including auditory parabelt (PB), auditory belt (AB) and primary auditory cortex (A1). Extrasylvian areas comprise a lateral dorsal hand-motor system (yellow to brown), including lateral prefrontal (PF_L), premotor (PM_L) and primary motor cortex ($M1_L$), and a visual 'what' stream of object processing (green), including anterior-temporal (AT), temporo-occipital (TO) and early visual areas (V1). When learning words in the context of perceived objects or to actions, both peri- and extrasylvian systems are involved. Numbers indicate Brodmann Areas (BAs) and the arrows (black, purple and blue) represent long distance cortico-cortical connections as documented by neuroanatomical studies. **(B)** Schematic global area and connectivity structure of the implemented model. The colours indicate correspondence between cortical and model areas. **(C)** Micro-connectivity structure of one of the 7,500 single excitatory neural elements modelled (labelled 'e'). Within-area excitatory links (in grey) to and from cell e are limited to a local (19x19) neighbourhood of neural elements (light-grey area). Lateral inhibition between e and neighbouring excitatory elements is realised as follows: the underlying cell i inhibits e in proportion to the total excitatory input it receives from the 5x5 neighbourhood (dark-purple shaded area); by means of analogous connections (not depicted), e inhibits all of its neighbours. Adapted from (Garagnani al., 1991; Dum and Strick, 2005, 2002; Lu et al., 1994).

The model replicates a range of important anatomical and physiological features of the human brain (e.g., Garagnani et al., 2017, 2008; Tomasello et al., 2017). As follow a summary of the six neurobiological principles incorporated in the neural network model:

(i) Neurophysiological dynamics of spiking pyramidal cells including temporal summation of inputs, threshold-based spiking, nonlinear transformation of membrane potentials into neuronal outputs, and adaptation (Connors et al., 1982; Matthews, 2001);

(ii) Synaptic modification by way of Hebbian-type learning, including the two biological mechanisms of long-term potentiation (LTP) and long-term depression (LTD) (Artola and Singer, 1993);

(iii) Area-specific global regulation mechanisms and local lateral inhibition (global and local inhibition) (Braitenberg, 1978; Yuille and Geiger, 2003);

(iv) within-area connectivity: a sparse, random and initially weak connectivity was implemented locally, along with a neighbourhood bias towards close-by links (Braitenberg and Schüz, 1998; Kaas, 1997);

(v) between-area connectivity based on neurophysiological principles and motivated by neuroanatomical evidence; and

(vi) uncorrelated white noise was constant present in all neurons during all stages of learning and retrieval with additional noise added to the stimulus patterns to mimic uncorrelated input conditions (Rolls and Deco, 2010).

Note that the connectivity structure implemented in the network reflects existing anatomical pathways between corresponding cortical areas of the cortex revealed by neuroanatomical studies using diffusion tensor and diffusion-weighted imaging (DTI/DWI) in humans and non-human primates (Table 3.2, Rilling et al., 2011; Thiebaut de Schotten et al., 2012). A detailed description of the single-neuron properties, synaptic plasticity rule, and single-area model structure is provided next, followed by details of the network anatomy and connectivity structure.

Structure and function of the spiking model

Each of the 12 simulated areas is implemented as two layers of artificial neuron-like elements ('cells'), 625 excitatory and 625 inhibitory, thus resulting in 15,000 cells in total (see Fig. 3.2b-c). Each excitatory cell '*e*' consists of a leaky integrate-and-fire neuron with adaptation and simulates a single pyramidal cell representative of excitatory spiking activity in a cortical micro-column, while its twin inhibitory cell '*i*' is a graded-response cell simulating the average inhibitory response of the cluster of interneurons situated in a local neighbourhood (Eggert and van Hemmen, 2000; Wilson and Cowan, 1972). The state of each cell *x* is uniquely defined by its membrane potential $V(x,t)$, specified by the following equation:

$$\tau \cdot \frac{dV(x,t)}{dt} = -V(x,t) + k_1(V_{In}(x,t) + k_2\eta(x,t)) \qquad \text{(B1)}$$

where $V_{In}(x,t)$ is the net input acting upon cell *x* at time *t* (sum of all inhibitory and excitatory postsynaptic potentials – I/EPSPs; inhibitory synapses are given a negative sign), τ is the membrane's time constant, k_1, k_2 are scaling values (see below for the specific parameter values used in the simulations) and $\eta(\cdot,t)$ is a white noise process with uniform distribution over [-0.5,0.5]. Note that noise is an inherent property of each model cell, intended to mimic the spontaneous activity (baseline firing) of real neurons. Therefore, noise was constantly present in all areas, in equal amounts (inhibitory cells have k_2=0, i.e., the noise is generated by the excitatory cells in the model for convenience).

The output (or transformation function) ϕ of an excitatory cell *e* is defined as follows:

$$\phi(e,t) = \begin{cases} 1 & if \;\; (V(e,t) - \alpha\,\omega(e,t)) > thresh \\ 0 & otherwise \end{cases} \qquad \text{(B2)}$$

Thus, an excitatory cell *e* spikes (=1) whenever its membrane potential $V(e,t)$ overcomes a fixed threshold *thresh* by the quantity $\alpha\omega(e,t)$ (where α is a constant and ω is defined below). Inhibitory cells are graded response neurons as they intend to represent the average impact of a cluster of local interneurons; the output $\phi(i,t)$ of an inhibitory neuron *i* is 0 if $V(i,t)<0$ and $V(i,t)$ otherwise.

To simulate neuronal adaptation(Kandel et al., 2000), function $\omega(\cdot,t)$ is defined so as to track the cell's most recent firing rate activity. More precisely, the amount of adaptation $\omega(e,t)$ of cell e at time t is defined by:

$$\tau_{ADAPT} \cdot \frac{d\omega(e,t)}{dt} = -\omega(e,t) + \phi(e,t) \qquad \text{(B3.1)}$$

where τ_{ADAPT} is the 'adaptation' time constant. The solution $\omega(e,t)$ of Eq. (B3.1) is the low-pass-filtered output ϕ of cell e, which provides an estimate of the cell's most recent firing-rate history. A cell's average firing activity is also used to specify the network's Hebbian plasticity rule (see Eq. (B4) below); in this context, the (estimated) instantaneous mean firing rate $\omega_E(e,t)$ of an excitatory neuron e is defined as:

$$\tau_{Favg} \cdot \frac{d\omega_E(e,t)}{dt} = -\omega_E(e,t) + \phi(e,t) \qquad \text{(B3.2)}$$

Local (lateral) inhibitory connections (see Fig. 3.2c) and area-specific inhibition are also implemented, realising, respectively, local and global competition mechanisms(Duncan, 2006, 1996). More precisely, in Eq. (B1) the input $V_{In}(x,t)$ to each excitatory cell of the same area includes an area-specific ('global') inhibition term $k_G\,\omega_G(e,t)$ (with k_G a constant and $\omega_G(e,t)$ defined below) subtracted from the total I/EPSPs postsynaptic potentials V_{In} in input to the cell; this regulatory mechanism ensures that area (and network) activity is maintained within physiological levels (Braitenberg and Schüz, 1998):

$$\tau_{GLOB} \cdot \frac{d\omega_G(e,t)}{dt} = -\omega_G(e,t) + \sum_{e \in area} \phi(e,t) \qquad \text{(B3.3)}$$

Excitatory links within and between (possibly non-adjacent) model areas are established at random and limited to a local (topographic) neighbourhood; weights are initialised at random, in the range [0, 0.1]. The probability of a synapse to be created between any two cells falls off with their distance (Braitenberg and Schüz, 1998) according to a Gaussian function clipped to 0 outside the chosen neighbourhood (a square of size n=19 for excitatory and n=5 for inhibitory cell projections). This produces a sparse, patchy and topographic connectivity, as typically found in the mammalian cortex (Amir et al., 1993; Braitenberg and Schüz, 1998; Douglas and Martin, 2004; Kaas, 1997).

The Hebbian learning mechanism implemented simulates well-documented synaptic plasticity phenomena of long-term potentiation (LTP) and depression (LTD), as implemented by Artola, Bröcher and Singer (Artola et al., 1990; Artola and Singer, 1993). This rule provides a realistic approximation of known experience-dependent neuronal plasticity and learning (Finnie and Nader, 2012; Malenka and Bear, 2004; Rioult-Pedotti et al., 2000), and includes both (homo- and hetero-synaptic, or associative) LTP, as well as homo- and hetero-synaptic LTD. In the model, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, +Δ and – Δ (with Δ<<1 and fixed). Following Artola *et al*., we defined as 'active' any (axonal) projection of excitatory cell *e* such that the estimated firing rate $\omega_E(e,t)$ of cell *e* at time *t* (see Eq. (B3.2)) is above ϑ_{pre}, where $\vartheta_{pre} \in [0,1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP to occur. Thus, given a pre-synaptic cell *i* making contact onto a post-synaptic cell *j*, the change $\Delta w(i,j)$ inefficacy of the (excitatory-to-excitatory) link from *i* to *j* is defined as follows:

$$\Delta w(i,j) = \begin{cases} +\Delta & if\ \omega_E(i,t) \geq \theta_{pre}\ and\ V(j,t) \geq \theta_+ \quad \textit{(LTP)} \\ -\Delta & if\ \omega_E(i,t) \geq \theta_{pre}\ and\ \theta_- \leq V(j,t) < \theta_+ \quad \textit{(homosynaptic LTD)} \\ -\Delta & if\ \omega_E(i,t) < \theta_{pre}\ and\ V(j,t) \geq \theta_+ \quad \textit{(heterosynaptic LTD)} \\ 0 & otherwise \end{cases} \tag{B4}$$

The values in table 3.1 describes the parameters used during word learning simulation in the network, which were chosen on the basis of previous simulations (e.g., Garagnani et al., 2009, 2007; Garagnani and Pulvermüller, 2011; Schomers et al., 2017; Tomasello et al., 2017).

Table 3.1 Parameter values used during simulations

Eq.(B1)	Time constant (excitatory cells)	τ = 2.5 (simulation time-steps)
	Time constant (inhibitory cells)	τ = 5 (simulation time-steps)
	Total input rescaling factor	k_1 = 0.01
	Noise amplitude	k_2= 5·√(24/Δt)
	Global inhibition strength	k_G= 0.60
Eq. (B2)	Spiking threshold	*thresh* = 0.18
	Adaptation strength	α= 7.0
Eq.(B3.1)	Adaptation time constant	τ_{ADAPT} = 10 (time steps)
Eq.(B3.2)	Rate-estimate time constant	τ_{Favg} = 30 (time steps)
Eq.(B3.3)	Global inhibition time constant	τ_{GLOB} = 12 (time steps)
Eq.(B4)	Postsynaptic membrane potential thresholds:	
		ϑ_+=0.15
		ϑ_-=0.14
	Presynaptic output activity required for LTP:	
		ϑ_{pre}=0.05
	Learning rate	Δ = 0.0008

Simulated brain areas and their connectivity structure

The spiking model mimics 12 different cortical areas with area-intrinsic connections and mutual connections between them. Six areas were modelled for the left-perisylvian language cortex including the primary auditory cortex (A1), auditory belt (AB), and modality-general parabelt areas (PB) constituting the auditory system, and the inferior part of primary motor cortex ($M1_i$), inferior premotor (PM_i) and multimodal prefrontal motor cortex (PF_i) representing the articulatory system (i.e. inferior face-motor areas). Additionally, six extrasylvian areas were modelled including the primary visual cortex (V1), temporo-occipital (TO) and anterior-temporal areas (AT) for the ventral visual system and the dorsolateral fronto-central motor ($M1_L$), premotor (PM_L), and prefrontal cortices (PF_L) for the motor system.

The network's connectivity structure reflects relevant features of cortical connectivity between corresponding areas of the cortex. These were modelled between neighbour cortical areas within each of the 4 'streams' (see black arrows Fig. 3.1 a-b) and between all pairs of multimodal areas (PB, PF_i, AT and PF_L) through the long distance cortico-cortical connections (purple arrows). Additionally, non-adjacent 'jumping' links were included within the superior or inferior temporal and superior or inferior frontal cortices (blue arrows). The neuroanatomical evidence motivated by studies using diffusion tensor and diffusion-weighted imaging (DTI/DWI) in humans and non-humans primates are reported in table 3.2 and described in previous study (Garagnani et al., 2017).

Table 3.2 Connectivity structure of the modelled cortical areas

Between-area connectivity (black arrows)	
Modelled Areas	**References**
Perisylvian system	
A1, AB, PB	(Kaas and Hackett, 2000; Pandya, 1995; Rauschecker and Tian, 2000)
PF_i, PM_i, $M1_i$	(Pandya and Yeterian, 1985; Young et al., 1995)
Extrasylvian system	
V1, TO, AT	(Bressler et al., 1993; Distler et al., 1993)
PF_L, PM_L, $M1_L$	(Arikuni et al., 1988; Dum and Strick, 2005, 2002; Lu et al., 1994; Pandya and Yeterian, 1985; Rizzolatti, G. Luppino, 2001)
Between system	
AT, PB	(Gierhan, 2013)
PF_i, PF_L	(Yeterian et al., 2012)
Long distance cortico-cortical connections (purple arrows)	
Perisylvian system	
PF_i , PB	(Catani et al., 2005; Makris and Pandya, 2009; Meyer et al., 1999; Parker et al., 2005; Paus et al., 2001; Rilling et al., 2008; Romanski et al., 1999b)
Extrasylvian system	
AT, PF_L	(Bauer and Jones, 1976; Chafee and Goldman-Rakic, 2000; Eacott and Gaffan, 1992; Fuster et al., 1985; Parker, 1998; Ungerleider et al.,

	1989; Webster et al., 1994)
Between system	
PB, PF_L	(Pandya and Barnes, 1987; Romanski et al., 1999b, 1999a)
AT, PF_i	(Pandya and Barnes, 1987; Petrides and Pandya, 2009; Rilling, 2014; Romanski, 2007; Ungerleider et al., 1989; Webster et al., 1994)
High-order 'jumping' links (blue arrows)	
Perisylvian system (Rilling et al., 2011, 2008; Rilling and Van Den Heuvel, 2018; Thiebaut de Schotten et al., 2012)	
A1, PB	(Pandya and Yeterian, 1985; Young et al., 1994)
PB, PMi	(Rilling et al., 2008; Saur et al., 2008)
AB, PFi	(Kaas and Hackett, 2000; Petrides and Pandya, 2009; Rauschecker and Scott, 2009; Romanski et al., 1999a)
PFi, M1i	(Deacon, 1992; Guye et al., 2003; Young et al., 1995)
Extrasylvian system (see also Thiebaut de Schotten et al., 2012)	
V1, AT	(Catani et al., 2003; Wakana et al., 2004)
AT, PML	(Bauer and Fuster, 1978; Chafee and Goldman-Rakic, 2000; Fuster et al., 1985; Pandya and Barnes, 1987; Seltzer and Pandya, 1989)
TO, PF_L	(Bauer and Jones, 1976; Fuster et al., 1985; Fuster and Jervey, 1981; Makris and Pandya, 2009; Seltzer and Pandya, 1989)
PF_L, $M1_L$	(Deacon, 1992; Guye et al., 2003; Young et al., 1995)

Simulating word acquisition

Prior to network training, all synaptic links (between- and within-areas) connecting single cells were established at random (see Methods section under '*Structure and function of the spiking model'*). Based on Hebbian (Hebb 1949) learning principles, word-meaning acquisition was simulated under the impact of repeated sensorimotor pattern presentations (D'Esposito, 2007; Fuster, 2003) to the primary areas of the network (see Fig 3.2), as follows: Each network instance used twelve distinct sets of sensorimotor neural patterns representing six action- and six object-related words. Each pattern consisted of a fixed set of 19 cells chosen at random within the 25 x 25 cells of an area (ca. 3% of the cells) and simultaneously activated in one of the primary areas of the network. The learning of object- and action-related words were grounded in sensorimotor information presented to the primary cortices of the model: besides perisylvian auditory A1 and articulatory $M1_i$ activity, object-related words received concordant visual (V1) and, similarly, action-related words received lateral motor area ($M1_L$) grounding activity. Note that white (so-called 'contextual') noise was continuously presented to all primary areas of the network, and thus superimposed on all learning patterns. This partly accounted for the variability of perceptions and actions of the same type. To sum up, the network was set up to learn correlations between word and referential semantic information in action and perception and to investigate which type of representations (i.e. cell assemblies) would develop in the model as a result of learning and cortical structure. Note that similar approaches to simulating spontaneous emergence of associations between articulatory and acoustic-phonetic neural patterns have been used in other computational studies (e.g., Guenther et al., 2006; Westermann and Miranda, 2004), although these previous works did not attempt to model semantic processes (i.e., word meaning acquisition).

Sensorimotor neural patterns in the arrangement of 3 x 19 cells, were presented for 3000 times to the relevant primary regions (this number was chosen on the basis of previous simulations obtained with a six area model, showing that no substantial change between 1000 and 2000 learning steps was revealed, Garagnani et al., 2009; Schomers et al., 2017). A word pattern was presented for 16 simulation time steps, followed by a period during which no input (interstimulus interval – ISI) was given. The next learning step (pattern presentation) occurred only when the global inhibition of PF_i and PB areas reduced

below a specific fixed threshold allowing the activity to return to a baseline value so that one trial is not affecting the next one. Only the inherent baseline noise (simulating spontaneous neuronal firing) and 'contextual' noise were present in the neural network during each ISI.

After learning, following a procedure which has become standard in our simulation studies (Garagnani et al., 2008, 2007; Garagnani and Pulvermüller, 2016; Schomers et al., 2017; Tomasello et al., 2017), we identified and quantified the neurons forming the 12 distributed CA circuits that emerged across the network areas during object and action word production. For simulating 'word production' in the network, the motor and auditory neurons of each word-form in areas M1 and A1 were activated together for 15 time-steps. Separate analyses were performed for object recognition and action execution, which was simulated by activating the corresponding stimulation pattern in visual or motor cortex (V1 or M1) thought to represent the object-related or action-related schemas semantically linked to the word-forms. During this period, we computed and displayed the average firing rate of each excitatory cell (7500 *e*-cells, cell's responses).

As an estimate of a cell's average firing-rate here we used the value $\omega_E(e,t)$ from Eq. (B3.2), integrated with time-constant τ_{Favg}= 5. An *e*-cell was then taken to be a member of a given CA circuit only if its time-averaged rate (output value or 'firing rate') reached a threshold ϑ which was area- and cell-assembly specific, and defined as a fraction γ of the maximal single-cell's time-averaged response in that area to pattern *w.* More formally,

$$\vartheta = \vartheta_A(w) = \gamma \max_{x \in A} \overline{O(x,t)_w}$$

where $\overline{O(x,t)_w}$ is the estimated time-averaged response of cell *x* to word pattern *w* (see in Method section under '*Structure and function of the spiking model*') and $\gamma \in [0,1]$ is a constant (we used γ = 0.5 on the basis of previous simulation results, see Garagnani et al. 2008; Garagnani et al. 2009; Tomasello et al. 2017). This was computed for each of the 12 trained network instances, averaging the number of CA cells per area over the 6 object- and 6 action-related words.

To statistically test for the presence of significant differences in the topographical CA distribution across the twelve network areas, for each network instance we performed a repeated-measures Analyses of Variance (ANOVA). A 4-way ANOVA was run with factors

WordType (two levels: *Object* vs. *Action*), PeriExtra (two levels: *Perisylvian* = {A1, AB, PB, $M1_i$, PM_i, PF_i}, *Extrasylvian* cortex = {V1, TO, AT, $M1_L$, PM_L, PF_L}), TemporalFrontal (TempFront) (2 levels: *temporal areas* = {A1, AB, PB, V1, TO, AT}, *frontal areas*={$M1_L$, PM_L, PF_L, $M1_i$, PM_i, PF_i}) and Areas (three levels: *Primary* = {A1, V1, $M1_L$, $M1_i$}, *Secondary* = {TO, AB, PM_L, PM_i} and *Central* = {PB, AT, PF_L, PF_i} areas). Finally, we further run a second statistical analysis on the data of the 6 perisylvian and 6 extrasylvian areas separately with factors 'WordType', 'TempFront', 'Areas', as described above.

Results

Word learning results

Twelve different instances of spiking networks were initialized at random having the same architecture as described above (Fig 3.1.B), providing analogues of 12 human subjects in a word learning experiment. Word-meaning acquisition was then simulated under the impact of repeated sensorimotor pattern presentations, in the 3 of the 4 sub-systems (see Fig 3.2), by co-activating specific neurons in their respective primary cortex. The cells activated in $M1_i$ and A1 represented articulatory and acoustic-phonetic features by which spoken words are typically characterized, while those presented to V1 and $M1_L$ simulated visually-related and action-related semantic features. This simulates associative learning of object-related word, whereby the word is uttered while the referent object is present (Vouloumanos and Werker, 2009) or the related action is being performed (Tomasello and Kruger, 1992). While each learning pattern directly activated three primary areas, the fourth unrelated area ($M1_i$ for object- and V1 for action-related words) received further uncorrelated noise pattern input that changed inconsistently over learning episodes. This aimed at ensuring that the correlation between word-form activity in perisylvian cortex and semantic information was high in one modality (for action /object words, in motor and visual systems respectively) but low in the non-relevant one.

Cell assemblies gradually emerged as a consequence of learning with different assemblies responding to different input patterns. These neural circuits spanned different areas, linking up word-forms in the auditory and articulatory sub-systems with referential-semantic information in the visual and motor sub-systems. Fig 3.2 illustrates 6 of the 12 CA-

distributions emerging across the novel spiking network along with the sensorimotor pattern presented as input during learning. Each set of 12 squares is a snapshot of a distributed word-related CA circuit across the network areas; 3 for object-related words (A) and 3 for action-related (B) words of one network instance (the other simulated networks exhibited similar results). Each white pixel in the squares represents an active cell of the CA.

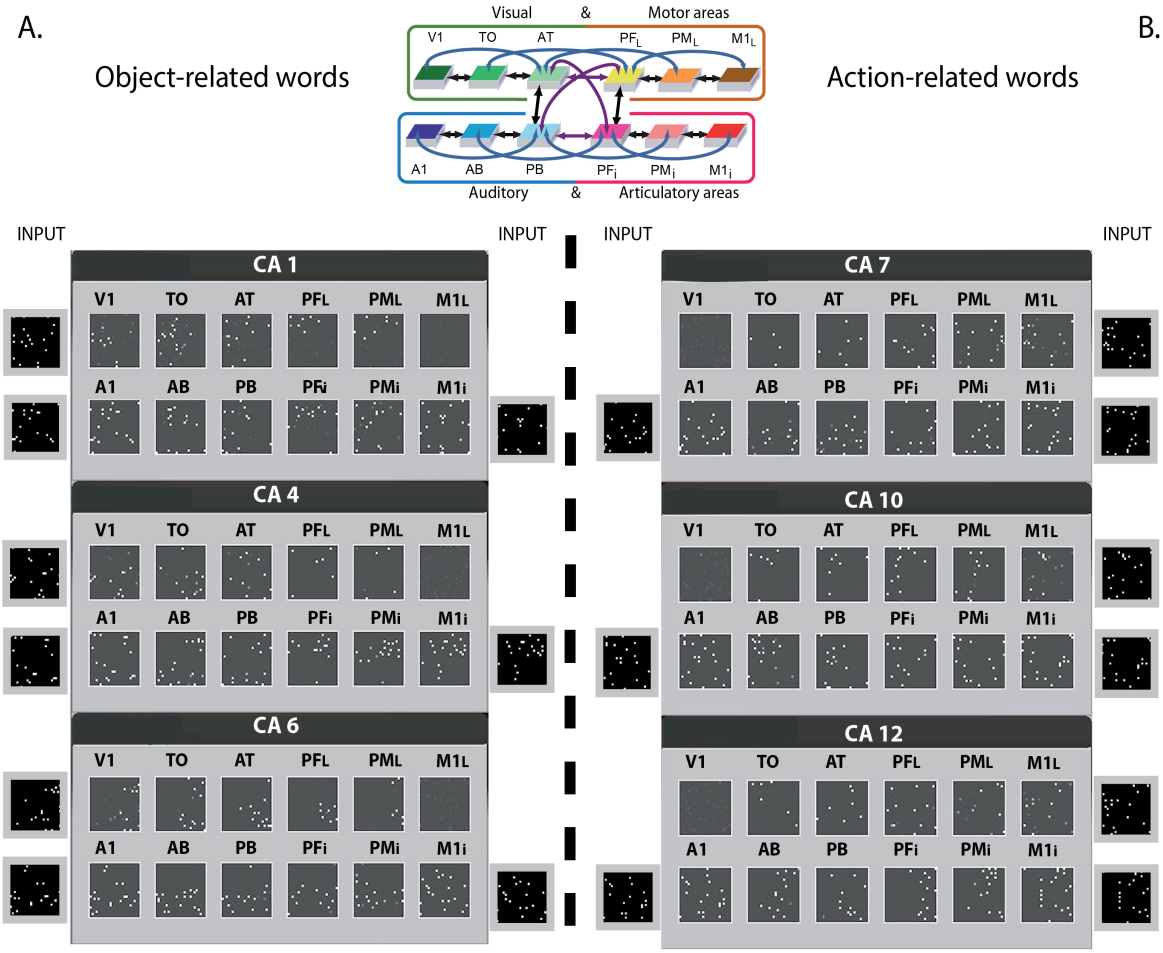

Figure 3.2. Distributions of cell-assemblies (CAs) emerging in the 12 area network during simulation of word learning in the semantic context of visual perception **(A)** and action execution **(B)**. Results of one typical instantiation of the model in Fig 1b are shown, using the same area labels. Each set of 12 squares (in black) illustrates one specific network area, with white dots indexing the distribution of CA neurons across the 12 network areas as a result of sensorimotor pattern presentation in 3 of the 4 primary areas. The perisylvian cortex was always stimulated, which mimics the learning of a spoken word form characterised by articulatory-acoustic features, while object words (A) received concordant stimulation to visual area (V1) and action words (B) to motor area ($M1_i$). Note that a random pattern simulating realistic noise input, changing in every learning phase, was presented to the non-relevant system (see Methods section). As a consequence of learning, CA circuits emerged in the network which extends into higher and primary visual cortex (V1, TO, but not $M1_L$) for object words. In contrast, network correlates of action-related words extend into lateral motor cortex ($M1_L$, PM_L, but not V1), thus semantically grounding words in information about actions. For convenience, the area structure of the network is repeated at the top.

The CA circuits in Fig 3.2 show roughly the same spread across the perisylvian areas for object and action-related words. By contrast, the visual and motor sub-systems of the extrasylvian cortex appear to show a different pattern of CA cell distribution, namely a double dissociation, i.e. object-related words seemed to extend more to the visual areas (V1, TO) and less to the motor areas (PM_L, $M1_L$) and vice versa for action-related words.

Figure 3.3 illustrates examples of CA circuit activation (i.e. each white pixel represents a spike) after the training has been undertaken. The network was confronted with the acoustic component (input pattern in primary auditory area) representing the auditory word-forms of the learned (A) object- and action-related (B) words, which in turn caused the 'ignition' of the whole CA circuit for that specific word-pattern. The snapshot numbers indicate simulation time-steps of the network activity. Similarly, as in the distribution of the emerging CA circuits illustrated in Fig 3.2, action- and object-related word recognition exhibited a semantic category-specific spreading of activity in the modality-preferential areas, which is near simultaneous (i.e. synchronous spikes) binding information from phonological (articulatory-acoustic) and semantic information. Interestingly, the re-activation of the word-related cell assemblies across the cortical areas exhibit the distinct consecutive neuronal and cognitive processes; the stimulation phase (time steps 1-2), which corresponds to word perception (orange pixel), the full activation or 'ignition' phase (time steps 5-8), the correlate of word comprehension (magenta pixel), and the reverberant maintenance of activity (time steps 12-14), which underpins verbal working memory (blue pixels).

The bar graph in Fig 3.4 reports the topographical distribution of the CA circuits across the network areas averaged over 12 networks. Different panels show results from the word production (A) and object and action recognition (B) 'experiments'. In each panel, average numbers of cell assembly neurons (plus standard errors) are shown for each area, with extrasylvian areas displayed at the top and perisylvian ones at the bottom. Intriguingly, the extrasylvian areas show a different CA distribution between the two word-type circuits, while the perisylvian language areas seem not to show any word-category differences.

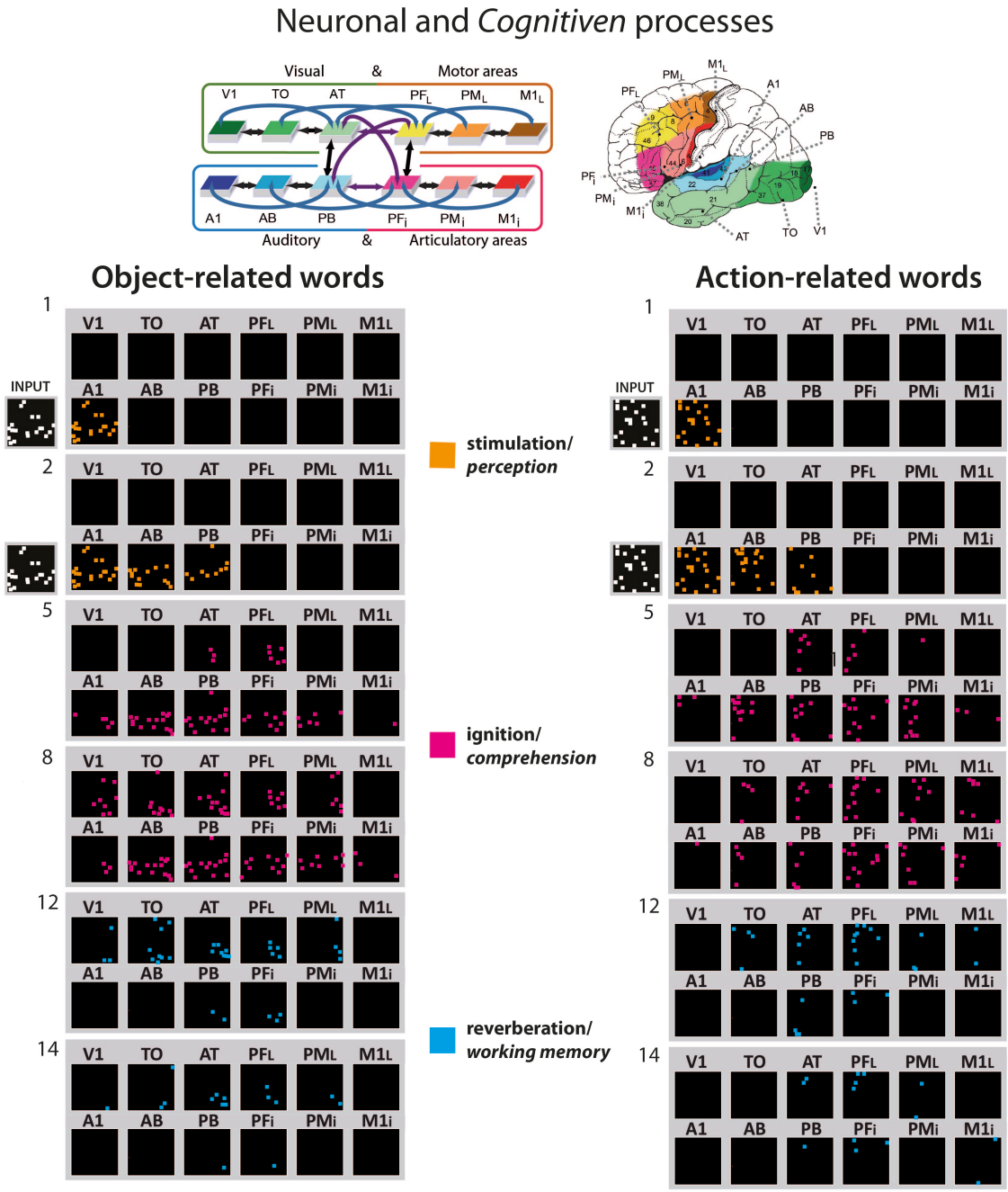

Figure 3.3. Activation spreading in the 12 area network showing the simulated recognition of object- and action-related words (see CA #6 and CA #10 in Fig 2, respectively). Network responses to stimulation of A1 with the 'auditory' patterns of two of the learned words; similar to Fig 2, the 12 network areas are represented as 12 squares, but, in this case, selected snapshots of network's activity are shown. The re-activation process comes in different consecutive neuronal and cognitive phases, the stimulation phase, which corresponds to word perception (orange pixel), the full activation or 'ignition' phase, the correlate of word comprehension (magenta pixel), and the reverberant maintenance of activity, which underpins verbal working memory (blue pixels). Each coloured pixel indicates one spike of the CA circuit at a given time step. At the top, the 12 brain areas modelled are shown.

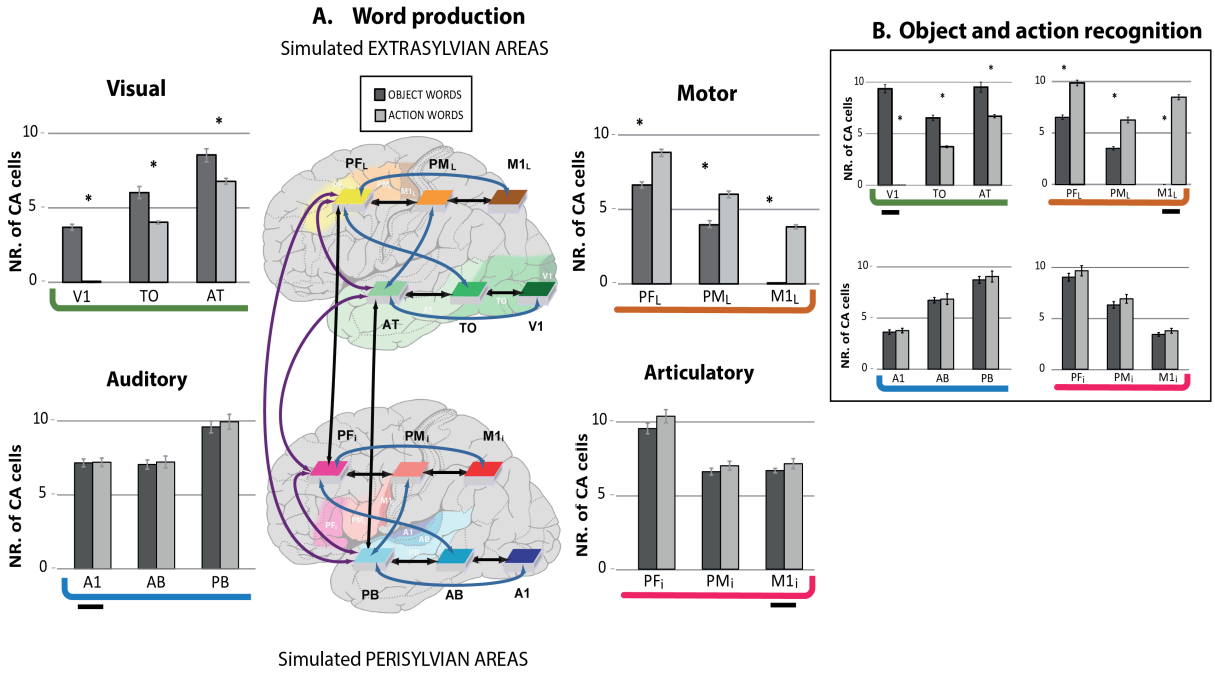

Figure 3.4. Mean numbers of cell assembly neurons in different model areas after simulating the learning of action- (light grey) and object-related words (dark grey) during word production **(A)** and object and action recognition **(B)**; error bars show standard errors over networks. **(A)** Simulated word production (simultaneous presentation of articulatory-auditory patterns in A1 and M1i areas) after word meaning acquisition. The extrasylvian areas (upper part) whose cells can be seen as circuit correlates of word meaning show a double dissociation, with relatively more strongly developed CAs for object- than for action-related words in primary and secondary visual areas (V1, TO), but stronger CAs for action-related than for object-related words in dorsolateral primary motor and pre-motor cortices (PM_L, $M1_L$). Also, the semantic hub areas (PFi, AT) showed a degree of dissociation between the two word types. Data from the perisylvian cortex (lower part), namely articulatory and auditory areas, whose cells can be seen as circuit correlates of spoken word-forms do not show category-specific effects. Brain areas and their connectivity structure are also illustrated. The shaded areas, but not the coloured boxes, indicate location in the cortex. **(B)** Simulated object and action recognition (alternated presentation of sensorimotor patterns in visual (for object) and in motor areas (for action words)). The present simulation exhibits similar results to the word production simulation. The small horizontal segment indicates the stimulus input presentation. Asterisks indicate that, within a given area, the number of CA cells significantly differed between the circuits of action and object words (Bonferroni-corrected planned comparison tests).

Furthermore, independently of whether an object or action-related word is represented, the word learning results showed higher density of CA cells in the connector hubs (PB, PF_i, AT and PF_L) than in the secondary (AB, PM_i, TO, PM_L) and primary areas (A1, $M1_i$, V1, $M1_L$). Similar results were revealed for both word production and action and object recognition, which is in line with the differential CA topographies already noted above and in Fig 3.2. However, there were minor differences in the estimated cell assembly topographies, as the relatively larger number of CA cells in the primary areas of the extrasylvian system were obtained for object and action recognition compared to word production, which was (trivially) due to the stimulus presentation there.

The 4-way repeated measurement ANOVA (with factors WordType, PeriExtra, TemporalFrontal and Areas) performed on the word production data from all of the 12 network areas fully confirmed the empirical and visual observation described above. A highly significant interaction emerged with factors WordType, PeriExtra, TempFront and Areas ($F_{2,22}$ = 14.012, p < .0002), revealing different CA circuits across the 12 area network between object- and action-related words. A main effect of Areas ($F_{2,22}$ = 265.721, p < .0001), indicating the different CA cell densities distributed across the network as noted above, namely higher CA cells in hubs than in secondary regions (p < .0001), and higher in secondary than in primary cortices (p < .0001). We separately ran a 3-way ANOVA on the data from the two systems, because of the significant interaction between peri- and extrasylvian areas. As expected, the extrasylvian system revealed a highly significant interaction of all 3 factors WordType, TempFront and Areas ($F_{2,22}$ = 53.11, p < .0001), confirming the word category dissociation in the CA topographies and local cell-density distributions across the extrasylvian regions as suggested by Figs 3.2 and 3.3. No significant differences between CA distributions of the 2 word types were found in the perisylvian areas ($F_{2,22}$ = 0.067, p = .93).

We further ran Bonferroni-corrected planned comparison tests (12 comparisons, corrected critical p < .0042) to investigate the differences between CA types that emerged after learning. Differences in CA-cell densities between word types and pairs of areas in the semantic systems were all significant (p < .0001), confirming the presence of a higher neuron-density in visual (V1, TO and AT) than in motor ($M1_L$, PM_L and PF_L) areas for object-related words (p < .0001), and the opposite for action-related words (p < .0001). Analysis of

the connector hubs (AT, PF_L) also showed a significant difference between the 2 word types there, i.e. stronger action-related word CA cell densities in PF_L compared to AT (p <.0001), and the opposite for object-related words (p <.0001). As observed above, no significant differences emerged in the perisylvian areas (p = .029) between the word types. We further run the same statistical analysis on the object and action recognition data, which revealed similar results as the word production simulation, i.e. double dissociation between action and object-related words in the extrasylvian system ($F_{2,22}$ = 467.321, p < .0001) with no significant difference in perisylvian cortex ($F_{2,22}$ = 0.060, p < .91).

Discussion

We investigated the neural mechanisms underlying word learning in a biologically constrained spiking model replicating connectivity and cortical features of the frontal, temporal and occipital areas to simulate aspects of semantic grounding in action and perception. The present neural-network showed

i. emergence of neuron circuits distributed across primary, secondary and multimodal areas, as a result of simulating the grounding of word-forms in their semantically-related objects and actions (Fig 3.2). We call these 'semantic circuits', because they interlink articulatory-acoustic word-from information with referential semantic representation coded in motor and visual areas;

ii. re-activation of the word-related circuits during word recognition exhibited the distinct consecutive neuronal and cognitive processes of word perception, word understanding and working memory (Fig 3.3);

iii. higher neuron densities of the semantic circuits and prolonged activity in the multimodal areas, where all semantic and phonological information first converges;

iv. pronounced semantic category-specificity primarily in the modality-preferential areas and moderate specificity also in multimodal areas for both word production and object and action recognition (Fig 3.4 A-B).

The present simulations offer a neurobiological explanation of a wide range of recent experimental results about word meaning processing and make critical predictions about the functional role of multimodal-association hubs, secondary and primary cortical regions

in language and semantic processing. Below, we provide a detailed discussion of the models and their results in light of previous empirical evidence, current semantic brain theories and its novel critical predictions.

Semantic brain processes: data and models

Accumulating evidence emphasises the relevance of several cortical regions for semantic processing, including inferior-frontal, superior- and anterior-temporal multimodal areas (Binder et al., 2009; Patterson et al., 2007; Pulvermüller, 2013), which are apparently to be relevant for all types of semantic processing, and modality-preferential areas, which seemingly take a category-specific role in semantics (Barsalou, 2008; Binder and Desai, 2011; Pulvermüller, 2013). Of great relevance in the current discussion about semantic grounding and 'embodiment' is the contribution of modality-preferential areas including primary and secondary cortices, for example, the motor and premotor cortex, or the primary and other 'early' visual area in semantic processing. These areas, which had classically been seen as 'perceptual' or 'motor' in their function, seem to partake in and contribute to semantic processing, as a range of previous experimental studies showed. The present results fit the postulate of semantic grounding (Harnad, 1990) that, in order to know the meaning of a symbol, it is necessary to relate it to real world entities, for example, the word 'grasp' to grasping actions and the word 'house' to the typical visual shape of houses. Grounding in this sense needs to be implemented in semantic representations that reach into motor and sensory systems. Our simulations applying brain constrained modelling at different levels demonstrate grounding in this very sense, hence fitting (and explaining) the experimental results mentioned above.

Some attempts to integrate both category-general and category-specific semantic mechanisms into one theoretical framework have been proposed. The 'hub-and-spoke' model postulates one single semantic hub in anterior-inferior-temporal lobe with category-specific spokes mainly in posterior brain areas (Ralph et al., 2017). This model explains crucial features of semantic dementia, but is inconsistent with hub-like properties of other multimodal areas (see introduction) and, in addition, does not address the motor system's role in category-specific processing (Vukovic et al., 2017), along with some fine-grained differences in the ability to process specific semantic categories which result from different

types of dementias (Shebani et al., 2017). Neurocomputational studies (Chen et al., 2017; Ueno et al., 2011) have investigated aspects of the hub-and-spoke model. However, as mentioned in the introduction, Chen and colleagues did not include all the brain areas for which experimental studies show a critical role in general semantic processing and they used learning mechanism (i.e. back-propagation - Chen et al., 2017; Ueno et al., 2011) which were criticized as implausible for cortical networks (Mazzoni et al., 1991; O'Reilly, 1998).

A claim about multiple semantic hubs has been made, in association with that about category-specific areas (Binder and Desai, 2011; Pulvermüller, 2013). However, formal neural-networks that could act as a foundation of a theory of semantic brain mechanisms did so far not reach the level of sophisticated neurobiologically constrained modelling with spiking neurons, realistic connectivity and learning. Earlier attempts were made using a preliminary version of the present architecture adopting non-spiking neurons (Garagnani and Pulvermüller, 2016; Tomasello et al., 2017). These previous models already suggest an explanation of category-general and category-specific semantic processing, but their conclusions were more limited by their less accurate modelling of neurophysiological and neuroanatomical features of the cortex.

Novel contribution: increased brain-constraints

Here, we added important neurobiological constraints, introducing leaky integrate-and-fire neurons that transform their summed input non-linearly into discrete output in the form of spikes. Similarly to biological neurons, functional interaction within the present model was based on discrete spikes, whereas previous mean-field networks used continuous activity functions (i.e. graded-response neurons), a less realistic implementation. Using graded-response neurons makes it easier to build distributed neural circuits across multiple areas as a result of action-perception learning since this type of neuron retains an increased firing rate for more extended periods. It was, therefore, crucial to investigate the possibility of distributed circuit formation with spiking neurons, which show an activation (action potential) for a short moment and then go silent again.

Compared with earlier studies, the present network included a more realistic set of cortico-cortical fibre tracts, adding second-next area connections or 'jumping links' (blue arrows Fig 3.1.A-B) indicated by DTI/DWI studies. A recent neurocomputational study

(Schomers et al., 2017) showed that these jumping links are instrumental for building verbal short-term memory, a capacity crucial for human language learning. Furthermore, previous exploratory implementation of 'jumping links' in an extended semantic network of mean-field (non-spiking/gradually active) neuronal elements suggested a degree of over-activation in case of implementation of the rich set of cortico-cortical connections, thus preventing precise simulation of more realistic connectivity. The use of spiking neuronal cells, whose action potentials only last for 1 simulation time-step and therefore produced less activity overall compared with the graded-neuron network, opened the possibility to include additional connection pathways documented by recent research without running into over-activation problems. On the other hand, spiking-neuron networks with just next neighbour connections between areas (thus omitting the 'jumping' links) ran into an under-activation problem, precisely because of the same feature (i.e. that spiking neurons lose their activity immediately). Thus, only the combined improvement of neuroanatomical (jumping connections) and neurophysiological (spiking) realism led to a functional network, which largely confirms conclusions formerly proposed on the basis of less realistic architectures. Incorporating significant biological detail into networks may be essential for obtaining a better understanding of the complex cortical mechanisms underlying semantic processing. Indeed, recent modelling results suggest that large-scale synchronous spiking within cell assembly circuits, also observed here, may be important for the binding of form to meaning during word learning and comprehension (Garagnani et al., 2017).

In summary, the comparison of less and more biologically constrained networks showed that improving the degree of realism does not always help. Moving from graded-response to spiking neurons alone renders an underactive network with little perspective on modelling semantic cognition, as the addition of a more detailed, elaborate and realistic connectivity structure on its own produces an overactive and thus, once again, dysfunctional networks. Only the parallel improvement on structural (anatomical) and functional (physiological) dimensions, that is, adding jumping links and spiking neurons, led to a functional network once again, which could confirm results from the earlier simulations obtained from the next-neighbour-connectivity and mean-field network, but provides a simulation at a more brain-constrained and therefore more realistic level.

Emergence of distributed symbolic circuits

The present model imitates elementary processes of semantic learning, where word-forms are presented in the context of object (Vouloumanos and Werker, 2009) or action information (Tomasello and Kruger, 1992). In our model, the co-occurrence of objects or actions with word-forms was implemented as correlated neuronal activation patterns in the model's primary articulatory ($M1_i$) and auditory (A1) along with either dorsolateral motor ($M1_L$) or visual cortex (V1). The first significant finding of this study is that such information about the semantic grounding of symbols can be mapped reliably onto biologically constrained associative networks. Each pattern representing the pairing of one specific symbol and one specific action or object led to the formation of a distributed circuit of spiking neurons spread out across several areas of the architecture. Each of these distributed circuits acted as a coherent functional unit, with its interlinked neurons in sensory, motor and multimodal areas activating together. The formation of each circuit required the spreading of activity across the network and the selective strengthening of a significant number of partaking neurons. Such strengthening was substantial enough so that, after learning, 'auditory input' was sufficient to revive the entire circuit, including its articulatory and semantic components. By comparing the mean-field next-neighbour model with the jumping-links spiking model, massive differences were revealed in the dynamics of cell assemblies activations during auditory word recognition (Fig. 3.3). Whereas the mean-field model showed cascade activation dynamics (with serial onset of activations and only partly overlapping activity of the hub areas AT, PF_L), the full-fledged three-phase dynamics with perception (activation of auditory areas), ignition (near-simultaneous activation of cell assembly neurons dispersed across wide cortical areas) and working memory (reverberation of activity in part of the cell assembly) was only present in the spiking and fully connected model. Intriguingly, after ignition, activity retreats from modality-preferential areas (time step 12, Fig. 3.3) to hub areas (time step 14), which predicts an 'anterior shift' from visual and motor areas to adjacent-anterior connector hub regions in temporal and prefrontal cortex during working memory (see also Fuster, 2009; Pulvermüller, 2018; Pulvermüller and Garagnani, 2014).

Although the formation of each circuit was driven by correlated information in sensory and motor areas, a widely distributed circuits with many neurons in multimodal

convergence zones got active. The involvement of neurons in multimodal areas is explained by long-distance connectivity structure, in particular by the absence of direct long-distance connections between sensory and motor areas; to bind information across modalities, activity must travel through connector hub areas (also called convergence zones, Damasio 1989), bridging between sensorimotor cortices. It is important to emphasize, however, that while the presence of connector hubs in the model is a (neuroanatomically motivated) structural feature, the result that the learned action and object word circuits reach *both* extrasylvian connector hubs AT and PF_L – hence forming semantic hubs – is not trivial, and could not be *a priori* predicted[1]. In other words, while the presence of *connector* hubs is a structural feature of the model, the formation of *semantic hubs* is not, and constitutes one of its crucial emergent properties.

The spontaneous formation of internal semantic circuits spanning the entire spiking neural network is a direct consequence of neurobiological principles modelled in the architecture that are known to govern the human brain. As discussed below, the activation of the learned distributed circuits explains relevant 'semantic area activations' seen in neuroimaging experiments (for further discussion, see Garagnani and Pulvermüller, 2016; Tomasello et al., 2017).

Explaining multiple semantic hubs

Not only did our model firmly bind neurons in multimodal areas to sensorimotor neurons involved in semantic processing, but, within each circuit, the proportion of these multimodal-area neurons was even greater than the percentage of circuit neurons in primary and secondary areas. On first view, this appears as surprising, because, during pattern presentation, sensory and motor neurons were directly stimulated together, whereas multimodal areas were activated only indirectly, by activity spreading from primary areas. However, the multimodal areas occupy a central location in the network topology because they bridge between sensory and motor areas, and therefore receive near-simultaneous convergent input from different (here, three) systems during learning. Such convergence also takes advantage of the higher 'degree' of connectivity characterising

[1] Note that the linkage of a perisylvian word circuit with semantic information coming from the visual (or motor) system does not necessarily have to go through connector hub PF_L (or AT).

[2] LTD and LTP are induced by the order and temporal interval between pre- and postsynaptic spikes (for more

multimodal areas and of their resultant role as 'connector hubs', for which a special role in cognition has previously been proposed (Van den Heuvel and Sporns, 2013). The cumulative effect of correlated inputs through several pathways converging on multimodal hubs accounts for their higher neuron-densities and their resultant major contribution to semantic circuit function. Thus, given that large fractions of the neurons of all semantic circuits were located in connector hubs, the model explains the prominent role of these connector regions in general semantic processing, which is due to the both well-known pre-existing neuroanatomical connectivity and the correlated neuronal activity during word learning.

Crucially, the model implicates and explains not only one, but at least four experimentally observed 'semantic hub' areas. One of these is in anterior-temporal lobe, providing a theoretical foundation for the critical postulate of the hub-and-spoke model (Patterson et al., 2007). Other semantic hubs are in superior-temporal-parabelt and in inferior- and dorsolateral-prefrontal cortex, where other models postulate sites of general semantic processing (Bookheimer, 2002; Carota et al., 2017; Posner and Pavese, 1998; Schomers and Pulvermüller, 2016; Tate et al., 2014). Our model, therefore, fits (and explains) data indicating the presence of frontal and temporal semantic hub areas, thus reconciling extant experimental evidence for a range of regions generally involved in conceptual processing (for reviews, see Kiefer and Pulvermüller, 2012; Pulvermüller, 2013).

Explaining category-specificity

We modelled the learning and processing of two different semantic categories: object- and action-related words. The formation of semantic circuits was driven by sensorimotor pattern information, involving visual cortex activity for object words and hand-motor cortex activity for action words. The respective other input system was activated with random noise to model the variable action output (visual input) in the context of specific visual objects (actions). Such uncorrelated noisy activity counters the spontaneous extension of neuron circuits towards inactive areas (Doursat and Bienenstock, 2006). Notably, as a consequence of the differential sensorimotor activation patterns, different circuit topographies developed across the areas for both word production and action or object recognition: circuits storing action-related information reached into the motor cortices

($M1_L$-PM_L) but not or less into visual areas (V1-TO), and vice versa for object words. Semantic circuits with different cortical topographies, which are a result of correlated neuronal activity in different sensorimotor areas during language learning, can therefore explain the emergence of category-specific semantic contributions of different cortical areas.

We take this observation as a proof-of-concept that the present type of spiking and jumping network is capable of spontaneously developing semantic-category specificity replicating a number of studies revealing neuroimaging and neuropsychological dissociations between action verbs and object nouns or between nouns sub-categories related to animals and tools (Damasio and Tranel, 1993; Kemmerer, 2015; Martin, 2007; Martin et al., 1996; Moseley and Pulvermüller, 2014). Interestingly, some category specificity was revealed in the semantic hubs, although it was less pronounced compared with primary and secondary areas. This area category-specific activation predicted by the model (Fig 3.4) seems to be of graded nature, with stronger category effect in the primary areas than in secondary areas and stronger in the secondary than in the hub areas and awaits experimental validations. The moderate category specificity predicted in the semantic hub areas is in line with recent evidence that semantic dementia patients due to anterior-temporal lesion show category-specific semantic impairments (Gainotti, 2012; Pulvermüller et al., 2010; Shebani et al., 2017), which sits less well with the suggested general-semantic function across all semantic types (Patterson et al., 2007).

It needs to be emphasized that most previous studies on semantics have investigated action and object words taken from natural languages, focusing mostly on the noun-verb distinction, which makes it difficult to control for all psycholinguistic proprieties and especially, when these words were acquired (e.g. Moseley and Pulvermüller 2014). If we take our present simulations as models of concrete action verb vs object noun processing, there is a good fit with the data, as these semantically and lexically different word types tend to differentially activate motor regions or ventral visual areas respectively (Damasio et al., 1996; Martin, 2007; Martin et al., 1996; Moseley et al., 2013; Pulvermüller et al., 2014b, 1999; Vigliocco et al., 2004). However, note that the ‘action’ and ‘object words’ simulated here capture the differential action- and object-relatedness of many verbs and nouns, but not the lack of such semantic differences seen between abstract verbs/nouns and certainly

not the combinatorial, or distributional differences between word categories, which result from their differential placements in specific grammatical contexts. Hence, for directly comparing the predictions of the present simulations to empirical data, it will be advantageous to perform analogous learning experiments and brain imaging studies to investigate *where* in the brain the neural signatures of novel object and action words first emerge. Nevertheless, the present simulation demonstrate the validity of a neurobiological theory of language processing (see Introduction, and Damasio, 1989; Pulvermüller, 2013), in which the mutual interaction of a set of neurobiological principles at work within anatomically-realistic structures and Hebbian learning are sufficient for explaining the emergence of semantic hubs and category specificity in the human brain.

It may be worthwhile to point to additional limitations of the present work along with possible extensons in the future. When an infant learns a new action word (e.g., 'grasp'), by hearing a novel word-form while performing the related action towards an object, concurrent activity might be present not just in the perisylvian language areas and motor cortices, but also in the visual occipital-parietal 'where' stream (Mishkin et al., 1983; Mishkin and Ungerleider, 1982), which was not implemented here. Therefore, an important extension of the present model would be to include parietal areas and the dorsal visual-where stream. Inclusion of left parietal areas would also be strongly motivated experimentally, as they are well known to play a role in general language processing (Pulvermüller and Fadiga, 2010) and also in category-specific processing of prepositions, number and tool words (Binder and Desai, 2011; Dehaene, 1995; Shebani et al., 2017; Tschentscher et al., 2012). Further model extensions should address other forms of language learning. Here we investigate but one aspect of word meaning acquisition, namely associative learning between a word and its referents, which represents only a very basic step of semantic learning. To capture other types of semantic learning, the emergence of semantic knowledge from variable contexts needs to be covered along with the semantic grounding of words learned from texts, where semantic links may be explained by co-activation of linguistic representations. Future work may address with realistic neuronal networks how, based on a kernel of early acquired words semantically grounded in referent object and action contexts, the co-occurrence of words in texts can lead to the formation of novel semantic circuits and semantic representations (Harnad, 2011; Stramandinoli et al.,

2012b). Furthermore, future simulations should extend the present work by investigating how combinatorial grammatical binding between pre-learnt and whole-form-stored lexical units emerges from correlated activity in co-activated neuronal circuits (see Pulvermüller, 2010).

Still, already in its current form, the present computational model makes critical predictions (some of which we spelled out in detail in discussion above) about how meaning is acquired, processed and stored in the human brain. Compared with earlier similar work, the spiking-and-jumping neural network developed in this work is based on a wider range of biological principles and features of the human brain, such as neurophysiological dynamics of spiking pyramidal cells, synaptic modification by way of Hebbian learning, local lateral inhibition and area-specific global regulation mechanisms, uncorrelated white noise present in all neurons during learning, brain-like connectivity structure based on neuroanatomical evidence. Therefore, the present model provides a sophisticated mechanistic explanation of the differential involvement of semantic cortical regions.

Conclusion

We used a biologically constrained neurocomputational model mimicking cortical features and connectivity of frontal, temporal and occipital cortices to simulate the brain mechanisms of word meaning acquisition. Extending our earlier work (Garagnani and Pulvermüller, 2016; Tomasello et al., 2017) by introducing, for the first time, spiking neuronal cells in a neuroanatomical constrained model with brain like connectivity, we show that Hebbian associative learning and connectivity together are sufficient to account for the emergence of general semantic areas ('semantic hubs'), as well as specific contributions of others modality-preferential ones to the processing of specific semantic categories. The present simulation results show that neurobiologically constrained networks can fruitfully contribute to bridging the gap between cellular-level mechanisms, behaviour and cognition by integrating brain theory with experimental data.

4. Visual cortex recruitment during language processing in blind individuals is explained by Hebbian learning

This chapter has been published as:

Tomasello, R., Wennekers, T., Garagnani, M., Pulvermüller, F., (2019). Recruitment of visual cortex for language processing in blind individuals is explained by Hebbian learning. *Scientific Reports* 9(1):3579.

doi: https://doi.org/10.1038/s41598-019-39864-1

This version of the article may not exactly replicate the final version published in the Journal. It is not the version of record.

Abstract

In blind people, the visual cortex takes on higher cognitive functions, including language. Why this functional organisation mechanistically emerges at the neuronal circuit level is still unclear. Here, we use a biologically constrained network model implementing features of anatomical structure, neurophysiological function and connectivity of fronto-temporal-occipital areas to simulate word-meaning acquisition in visually deprived and undeprived brains. We observed that, only under visual deprivation, distributed word-related neural circuits 'grew into' the deprived visual areas, which therefore adopted a linguistic-semantic role. Three factors are crucial for explaining this deprivation-related growth: changes in the network's activity balance brought about by the absence of uncorrelated sensory input, the connectivity structure of the network, and Hebbian correlation learning. In addition, the blind model revealed long-lasting spiking neural activity compared to the sighted model during word recognition, which is a neural correlate of enhanced verbal working memory. The present neurocomputational model offers a neurobiological account for neural changes followed by sensory deprivation, thus closing the gap between cellular-level mechanisms, system-level linguistic and semantic function.

Introduction

The classical model of the neurobiology of language, based on brain lesion data (Broca, 1861; Wernicke, 1874), proposed a left-lateralized linguistic network of the fronto-temporal regions located around the perisylvian fissure (Lichtheim, 1885). However, recent neuroimaging studies, as well as patient data, reported a more detailed cortical organization of the language areas, showing that brain areas outside the classical perisylvian cortex as well relevantly contribute to the processing of meaningful symbols and language (Binder and Desai, 2011; Pulvermüller, 2013; Pulvermüller and Fadiga, 2010). A range of cortical areas have been documented to be differentially involved, depending on the semantic type of symbols or larger meaningful constructions (Chao et al., 1999; Damasio et al., 1996; Dreyer et al., 2015; Grisoni et al., 2016; Hauk et al., 2004; Kemmerer, 2015; Moseley et al., 2013; Vukovic et al., 2017). For example, Moseley et al. (2013), reported enhanced neuromagnetic (MEG) responses for action words in the fronto-central areas, including motor regions, and for object-related words in the visual temporo-occipital areas, respectively. This and similar observations support neurobiological language models postulating that linguistic and semantic processes are carried by neuron circuits distributed across the perisylvian language regions as well as modality-preferential and multimodal areas in 'extra-sylvian' space (Garagnani and Pulvermüller, 2016; Pulvermüller, 1999; Pulvermüller and Fadiga, 2010; Tomasello et al., 2017, 2018).

A range of studies reported that the distributed language network shows striking capabilities to re-organize and adapt to focal lesions or sensory deprivation (Chen et al., 2002; Keck et al., 2008; Neville and Bavelier, 1998). Compared with healthy individuals, blind people's language processing in the so-called verb generation task leads to relatively stronger activation of visual areas in occipital cortex (Amedi et al., 2004, 2003, Burton, 2003, 2002; Raz et al., 2005; Struiksma et al., 2011). Several brain imaging studies showed activation of the primary visual (V1) and higher extra-striate visual cortices when congenitally blind individuals were required to generate semantically related verbs to heard nouns (Amedi et al., 2003; Burton, 2002; Struiksma et al., 2011) (see Fig. 4.1). In contrast, sighted subjects showed activation of the typical language regions (e.g., Broca's and Wernicke's areas) and motor areas, but no or significantly less visual area activation than blind individuals (Burton, 2002; Struiksma et al., 2011). Similar differences in V1 activation

have also been reported for single word (Burton, 2003; Burton et al., 2012) and sentence processing tasks (Bedny et al., 2011; Röder et al., 2002), which imply semantic understanding (Burton, 2003; Burton et al., 2012; Röder et al., 2002). Furthermore, congenitally blind people with relatively stronger V1 activity in the processing of meaningful language were reported to show better verbal working memory (Amedi et al., 2003) and generally enhanced verbal abilities compared to sighted individuals (Amedi et al., 2003; Occelli et al., 2017; Pasqualotto et al., 2013; Withagen et al., 2013). Although one might argue that visual responses in blind individuals are epiphenomenal with no functional relevance for language processing, a study inducing temporary virtual lesions of the primary visual area (V1) using transcranial magnetic stimulation (TMS) during a verb generation task has shown an increase in semantic (but not phonological) errors in blind individuals. In contrast, sighted control subjects showed a similar behavioural change only when TMS was applied to the left prefrontal cortex (lPFC) (Amedi et al., 2004). These results demonstrate that, in congenitally blind subjects, visual cortices respond in a similar way as classic language regions (Bedny et al., 2011) and are functionally relevant for language and semantic processing.

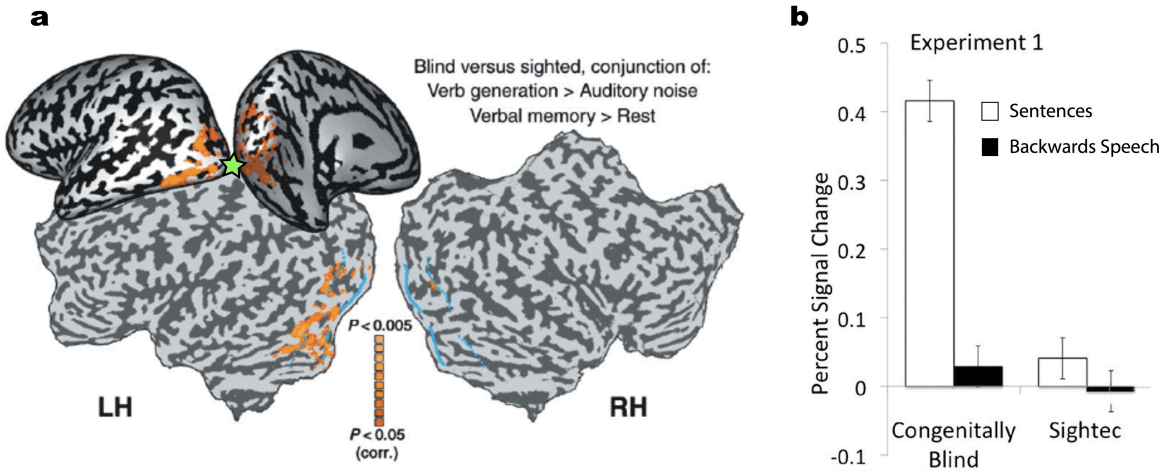

Figure 4.1. fMRI activation patterns between blind and sighted groups. (a) Activation of the primary and higher extra-striate visual areas (V1) when blind people recall words from memory or generate verbs from nouns compared to the sighted individuals (data adapted from Amedi et al., 2003). Green star indicates the stimulated cortical area (V1) delivered with rTMS causing substantial semantic errors during the verb generation task (data adapted from Amedi et al., 2004). **(b)** Perceptual signal change in the left primary visual area between blind and sighted control populations during meaningful sentence comprehension and backwards speech (data adapted from Bedny et al., 2011, this figure is not covered by the CC BY licence. [Credits to National Academy of Science]. All rights reserved, used with permission).

Undeprived healthy individuals may also activate their visual areas in language processing, but this is specific to words and sentences with a strong semantic relationship to visual information, for example words like 'cow' or 'tower', which have visually perceivable referents (Chao et al., 1999; Kiefer, 2005; Moseley et al., 2013; Pulvermüller et al., 1999; Sim and Kiefer, 2005). Associative learning can explain this category-specific semantic activation in the human brain: Because symbols with 'visual semantics' frequently co-occur with visually perceived referent objects during learning (Vouloumanos and Werker, 2009), the correlated neuronal activations are mapped at the neuronal level. However, such stimulus-driven correlation is obviously impossible in congenitally blind subjects. Therefore, the generally robust visual cortex activations during language processing and the associated relevance of visual areas for general language processing in the blind appear as a mystery.

Why is the visual cortex generally relevant in language processing in congenitally blind individuals, and why would a role of visual areas in sighted subjects, if present at all, be restricted to only specific semantic categories?

It is unlikely that congenitally blind and undeprived human subjects differ in the neuroanatomical connections interlinking visual areas and language regions, as diffusion tensor imaging (DTI) studies do not consistently demonstrate such differences (Noppeney et al., 2005; Shimony et al., 2005; Shu et al., 2009a, 2009b). However, at the functional level, there is evidence for relatively stronger functional connectivity (estimated from fMRI) between visual and frontoparietal language regions in blind people (Bedny et al., 2011; Burton et al., 2014; Butt et al., 2013; Striem-Amit et al., 2015). Therefore, the critical question to answer is how, given the absence of differences in anatomical long-range connectivity, it is possible that visual cortex function changes in congenitally blind people. It has been suggested that the lack of competing inputs to the deprived cortical areas during development may be critical; this would leave the blind's visual cortices available for recruitment for language processing (Bedny, 2017). However, the neural mechanisms determining such takeover remain to be specified. Here, we hypothesise that general neurobiological mechanisms and principles can explain the functional changes in the visual cortex, and we aim at isolating the factors that drive such plastic change.

To test this hypothesis, we applied a neurobiologically constrained model implementing properties of fronto-temporo-occipital areas and their connectivity in an attempt to

simulate features of language acquisition in undeprived (i.e. sighted) and deprived (i.e. congenitally blind) human subjects. The models were given information for learning the referential relationships between individual verbal symbols and the actions and objects they are typically used to communicate about. By comparing (congenitally) 'blind' and 'undeprived' models, we aimed to shed light on the neural language mechanisms consequent to sensory deprivation.

Results

General model architecture

At the micro level, the neural-network implements physiologically realistic spiking neurons, and at the system level, twelve areas of relevance for language and semantic processing situated in the frontal, the temporal and the occipital lobes (see Fig. 4.2a). The implemented area-intrinsic, as well as between-area, connectivity was guided by prior neuroscience evidence (Rilling et al., 2011; Thiebaut de Schotten et al., 2012). Six of the areas were in the left-perisylvian cortex [superior temporal Brodmann areas (BAs) 41, 42, 22 and inferior frontal areas, BAs 44, 45/6, 4], which is known to be most crucial for spoken language processing (Fadiga et al., 2002; Pulvermüller, 1999; Pulvermüller and Fadiga, 2010; Zatorre et al., 1996).

- The model's 'auditory stream' includes the primary auditory cortex (A1), auditory belt (AB), and modality-general parabelt areas (PB), and
- its 'articulatory stream' comprises the inferior part of primary motor cortex ($M1_i$), inferior premotor (PM_i) and multimodal prefrontal motor cortex (PF_i).

An additional six extrasylvian areas modelled referential meaning-related information about visual object identity (Ungerleider and Haxby, 1994), and about executable manual actions (Deiber et al., 1991; Dum and Strick, 2005, 2002; Lu et al., 1994).

- The 'ventral visual stream' includes the primary visual cortex (V1), temporo-occipital (TO) and anterior-temporal areas (AT) and
- the 'dorsolateral motor stream' the corresponding lateral primary motor ($M1_L$), premotor (PM_L), and prefrontal cortices (PF_L).

For clarity, we will mark area labels by an asterisk when speaking about model areas (e.g. *V1), whereas the conventional labels are used for the areas in the cortex (V1). Single-neuron properties, synaptic plasticity rule, and single-area model structure are specified in

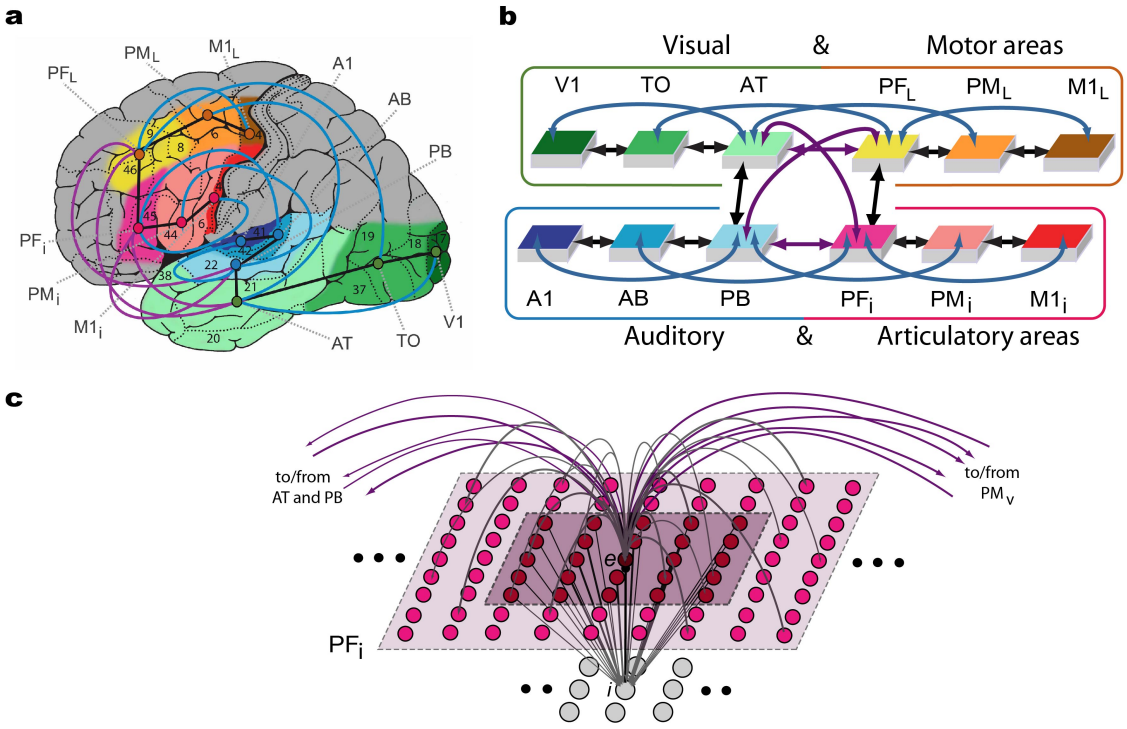

Figure 4.2. (a) Structure and connectivity of 12 frontal, temporal and occipital cortical areas relevant for learning the meaning of words related to actions. Perisylvian cortex comprises an inferior-frontal articulatory-phonological system (red colours), and the extrasylvian areas comprise a lateral dorsal hand-motor system (yellow to brown) and a visual "what" stream of object processing (green). Numbers indicate Brodmann Areas (BAs) and the arrows (black, purple and blue) represent long distance cortico-cortical connections as documented by neuroanatomical studies. **(b)** Schematic global area and connectivity structure of the implemented model. The colours indicate correspondence between cortical and model areas. **(c)** Micro-connectivity structure of one of the 7,500 single excitatory neural elements modelled (labelled '*e*'). Within-area excitatory links (in grey) to and from cell *e* are limited to a local (19x19) neighbourhood of neural elements (light-grey area). Lateral inhibition between *e* and neighbouring excitatory elements is realised as follows: the underlying cell *i* inhibits *e* in proportion to the total excitatory input it receives from the 5x5 neighbourhood (dark-purple

more detail in the Methods section under '*Structure and function of the spiking neuron model*' and in previous publications (Garagnani et al., 2017; Tomasello et al., 2018).

Briefly, the following biological, anatomical and physiological features of the cerebral cortex were replicated in the model:

(i) neurophysiological dynamics of spiking pyramidal cells including temporal summation of inputs, threshold-based spiking, and adaptation(Connors et al., 1982; Matthews, 2001);

(ii) synaptic modification by way of Hebbian-type learning, including both long-term potentiation and depression (LTP, LTD) (Artola and Singer, 1993);

(iii) local lateral inhibition and area-specific regulation mechanisms (called 'local and global control' below) (Braitenberg, 1978; Yuille and Geiger, 2003);

(iv) within-area connectivity: a sparse, random and initially weak connectivity was implemented locally, along with a neighborhood bias towards close-by links (Braitenberg and Schüz, 1998; Kaas, 1997);

(v) between-area connectivity based on neurophysiological principles and motivated by neuroanatomical evidence further explained below; and

(vi) presence of ongoing uniform uncorrelated white noise in all neurons during all phases of learning and retrieval (Rolls and Deco, 2010), and additional static noise added to the stimulus patterns to mimic realistic variability of input conditions during learning and retrieval.

The network´s connectivity structure reflects existing anatomical pathways revealed by neuroanatomical studies using diffusion tensor and diffusion-weighted imaging (DTI/DWI) (Rilling et al., 2011; Thiebaut de Schotten et al., 2012). These were modelled between adjacent cortical areas within each of the 4 'streams' (see black arrows Fig. 4.2 a-b) and between all pairs of multimodal areas (PB, PF_i, AT and PF_L) through the long distance cortico-cortical connections (purple arrows). Additionally, non-adjacent 'jumping' links were implemented within the superior or inferior temporal and superior or inferior frontal cortices (blue arrows). Detailed descriptions of the connectivity structure and the neuroanatomical evidence reporting such links are documented in the Methods section under '*The model's connectivity structure'.*

Word learning results

Thirteen different instances of 'sighted' and 'blind' model networks (in total 26 networks) were initialised having the same architecture as described above (Fig. 4.2b), but each with randomly generated synaptic connections. These model instances were used to simulate plastic changes in normal-sighted and congenitally blind humans during word learning. We mimic associative learning between word-forms used to speak about objects and their referent objects present in the environment as well as between action words and the performance of their semantically-related actions, as it is well-documented in the literature on language learning (Tomasello and Kruger, 1992; Vouloumanos and Werker, 2009). Although other forms of semantic learning (e.g., from texts or by definition) also play a role in meaning acquisition, we focus on the direct semantic grounding of words in object and action knowledge, because it is both prominent in early language learning and a precondition for other forms of semantic learning (Harnad, 1990; Vincent-Lamarre et al., 2016). In the sighted model simulations, object- and action-related word acquisition was grounded in sensorimotor information presented to the primary areas of the model: object-related word learning was driven by perisylvian activity in A1 and $M1_i$ and concordant visual (V1) activity patterns; similarly, action-related word learning was driven by semantic activity in the lateral motor area ($M1_L$) along with perisylvian activity (Fig. 4.3). The fourth non-relevant area ($M1_i$ for object- and V1 for action-related words) received an uncorrelated input pattern that was changing in every learning step. This aimed to mimic variable input patterns uncorrelated with word-form, reflecting, for example, the many different objects that can be grasped - and visually perceived - during the acquisition of the meaning of 'grasp', or the different motor inputs that might occur during the learning of novel concrete (object) words. In contrast, the congenitally blind models were trained with the same parameters but without any visual experience during the entire learning processes (i.e., no correlated *or* uncorrelated input to V1*).

Learning the association of word-forms in perisylvian language areas with the related referential semantic information in the extrasylvian system in sighted and congenitally blind models led to the formation of ensembles of strongly interconnected neurons, the so-called 'cell assemblies' (CA) once envisaged by Hebb (1949). These were scattered across several areas of the multi-area networks. After the learning had been completed, the CA neurons

were identified by simulating 'word production' processes by presenting the auditory-articulatory word-form patterns in the primary perisylvian areas (see Method section '*Data processing and statistical analysis*' for more details). Fig. 4.3 illustrates distributions for CAs underpinning 2 object- and 2 action-related words learned under undeprived (turquoise pixels) and deprived conditions (magenta pixels; other simulated networks led to similar topographies).

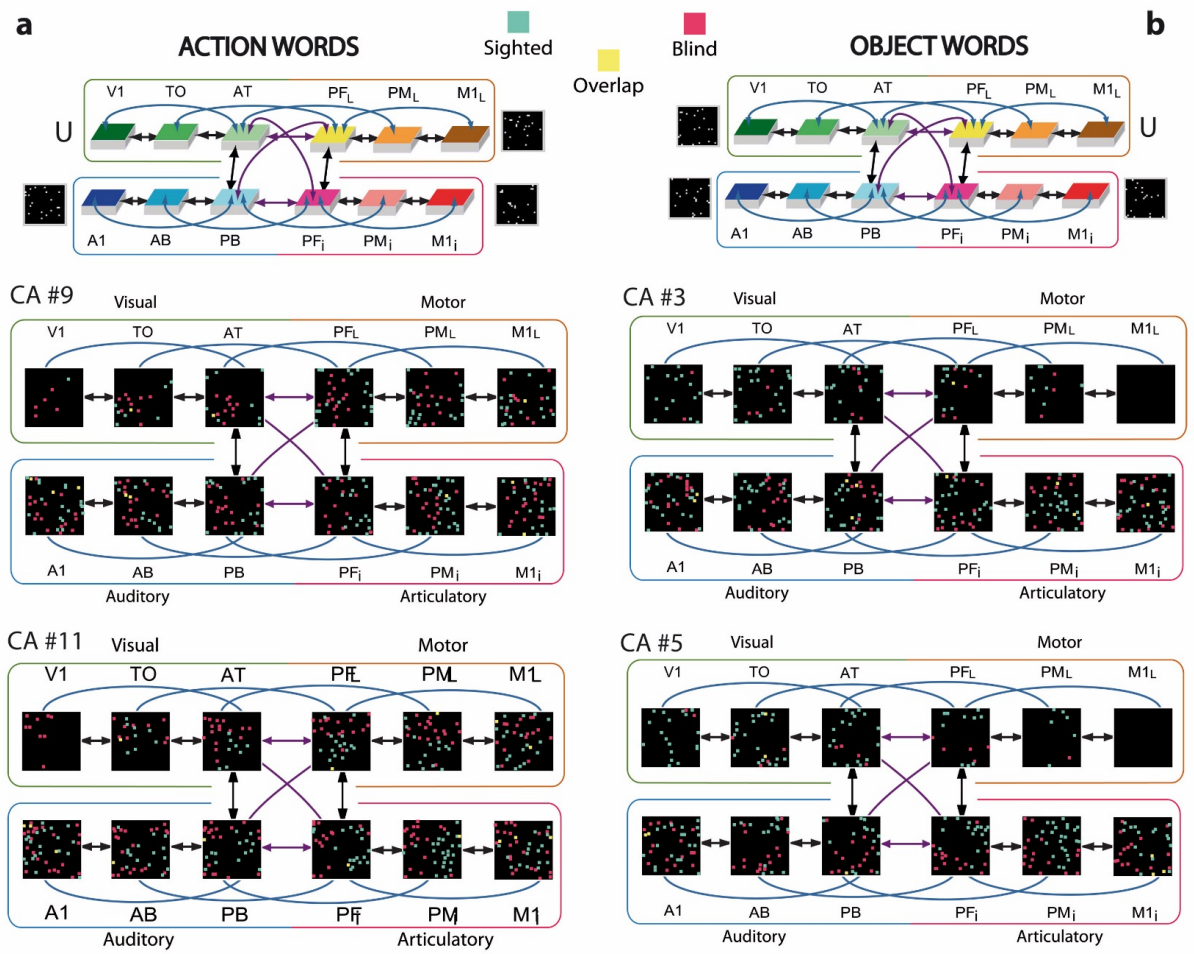

Figure 4.3 Distributions of cell-assemblies (CAs) emerging in the 12 areas network during simulation of action **(a)** and object **(b)** word learning under normal (sighted) and deprived conditions. Each set of 12 squares (in black) illustrates one specific network area, with coloured pixels indexing the distribution of CA neurons across the 12 network areas as a result of sensorimotor pattern presentations. The perisylvian cortex was always stimulated, which mimics the learning of a spoken word form characterised by articulatory-acoustic features, while action words received concordant stimulation to the motor area ($M1_i$), object words were grounded to visual areas (V1). The symbol "U" indicates the uncorrelated pattern presentation simulating variable sensory or motor input typically occurring during word learning (see Methods section for more detail). The blind model was trained in the same way, but without any visual input during the entire learning phase.

Visual inspection of the results suggested that the two types of word-related circuits did not differ in distribution across the perisylvian part of the networks. Likewise, sighted and blind model architectures produced similar perisylvian CA topographies (Fig. 3). This observation was confirmed by counts of CA neurons per area (see bar plots in Fig. 4) and by statistical results failing to support a difference in perisylvian CA distributions between word or network types. In contrast, the extrasylvian regions of the sighted model revealed a clear double dissociation between the two word types. CAs carrying object-related words seemed to extend more into the visual areas (V1, TO) and less into the motor areas (PM_L, $M1_L$), whereas action-related words showed the opposite pattern. Intriguingly, the CA circuits for action-related symbols in the blind model not only reached into the motor cortices (PM_L, $M1_L$) - to a similar degree as in the sighted model - but also extended into the visual areas, including higher order and primary visual regions (TO, V1). The blind model's object-word CA circuits also reached the visual system, although no (correlated or uncorrelated) visual input pattern had been presented during learning.

The bar plots in Fig. 4.4 shows the number of CA neurons of action- (a) and object-word (b) circuits situated in extrasylvian and perisylvian systems for sighted (turquoise) and blind (magenta) models. Visual illustration of the word-related CA circuits between sighted and blind models in the extrasylvian system (see bar plots in Fig. 4.4), shows a higher CA circuit density in the primary visual area (V1) for action related words under the deprived condition, which is consistent with the range of studies mentioned in the introduction about language processing in congenitally blind people. In contrast, object-related words seem to differ in all the areas of the extrasylvian system, i.e., they reveal a lower density of CA circuits under deprived condition.

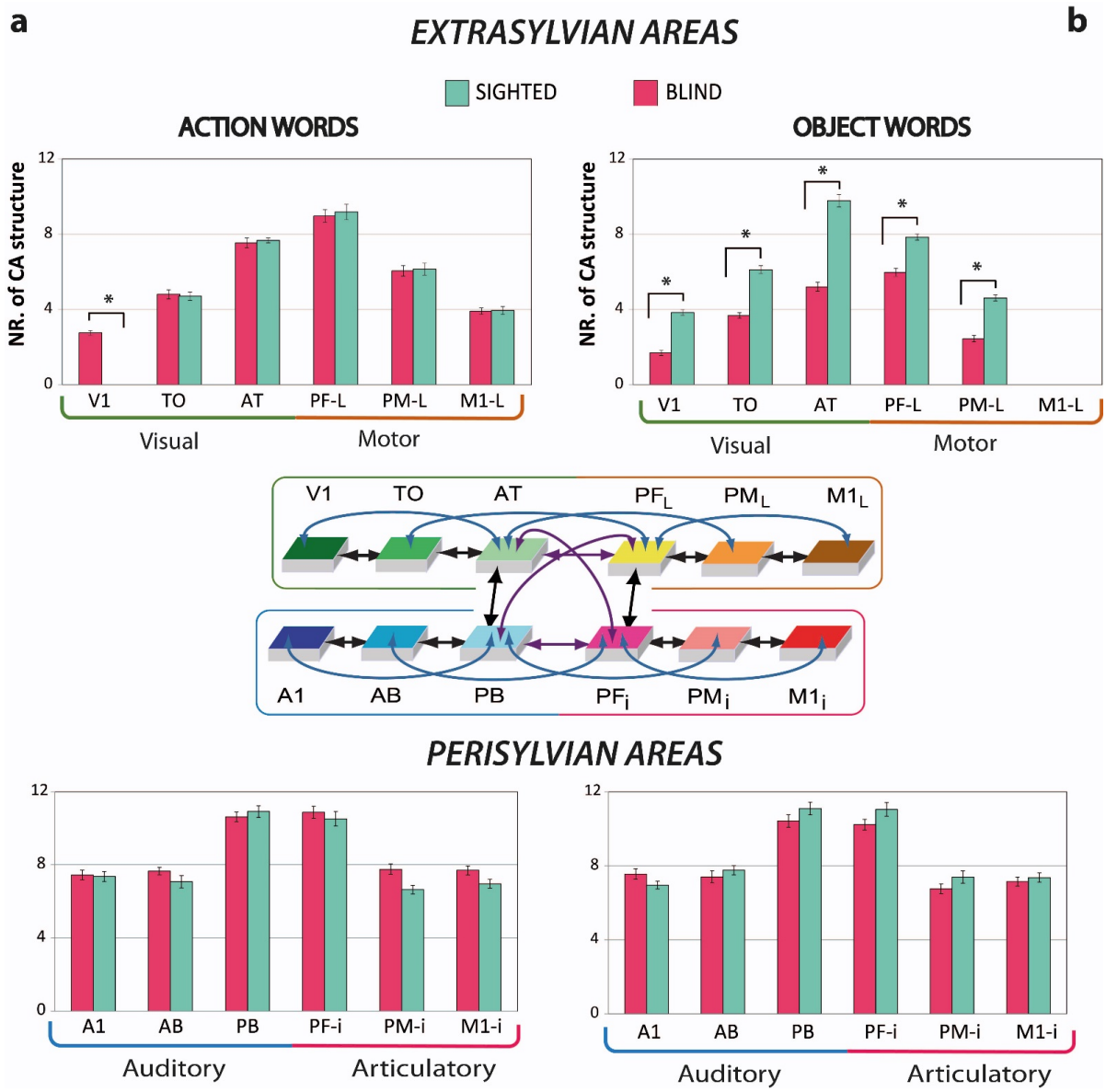

Figure 4.4 Mean numbers of cell assembly neurons in the different cortical areas of the sighted (turquoise bars) and blind models (magenta bars) after simulating the learning of action- **(a)** and object-related words **(b)** during word production; error bars show standard errors over networks. Data of the extrasylvian system are shown above and the one of the perisylvian cortex below. Asterisks indicate that, within a given area, the number of CA cells significantly differed between the sigthed and blind model for the two word types (Bonferroni-corrected planned comparison tests, 6 comparisons; critical threshold $p<.0083$).

Fig. 4.5 illustrates the correlates of action word recognition in sighted and blind models after training. The re-activation was simulated by presenting the auditory patterns of previously learned word-forms to the primary auditory area (*A1, Fig. 4.5). Similar to the CA structure illustrated in Fig. 4.3, action-related words in the blind model showed a higher density of CA cells in motor and deprived visual areas compared to the sighted one. Intriguingly, the blind model revealed a prolonged activation time course (CA Ignition) compared to the sighted model. In this particular example, the different neuronal and cognitive correlates of word *perception* (stimulation), word *understanding* (full ignition) and *memory* trace (reverberation) lasted more than 25 percent longer in the blind model as compared to the sighted one.

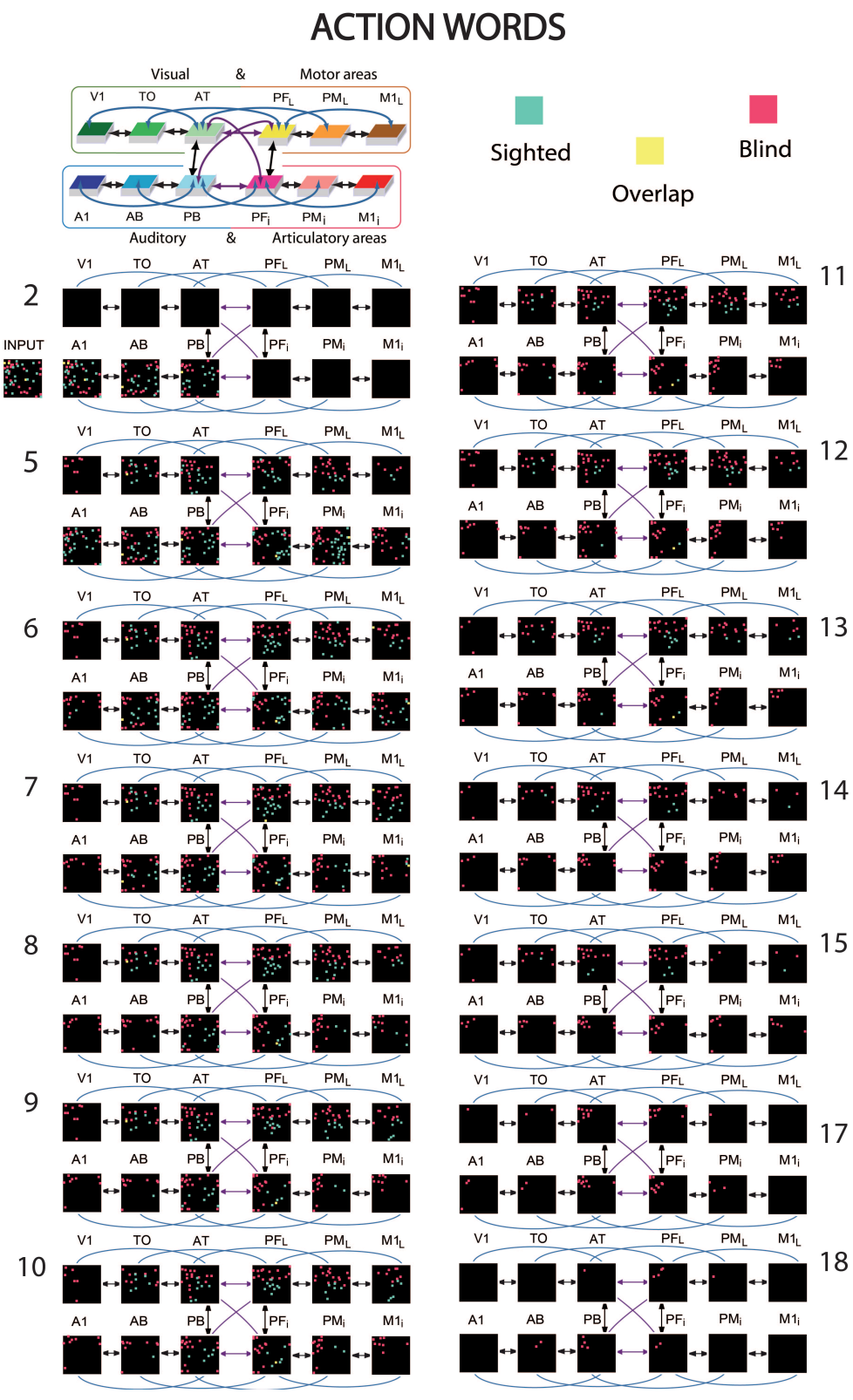

Figure 4.5. Activation spreading in the 12 area network during simulated action word recognition. Network responses to stimulation of A1 with the 'auditory' patterns of the learned words (CA #11 in Fig. 3, respectively); the 12 network areas are represented as 12 squares, but, in this case, selected snapshots of network's activity are shown (as in Fig. 3) with numbers indicating the simulation time-steps. Each pixel represents one spike of the CA circuit for sighted (turquoise pixel) and blind people (magenta pixels). Notice the prolonged spiking activation of the blind model compared to the sighted one. See main text for details.

The observations described above were confirmed by a 3-way repeated measurement ANOVA with the factors Model, WordType and Area, which revealed a main effect of Model ($F_{2,24}$ = 11.91, p = .0047, η_p^2 = .49) and a significant interaction between all three factors ($F_{2,24}$ = 13.32, ε = .43, p < .00001, η_p^2 = .52). Consistent results were revealed by the 5-way ANOVA breaking down the areas into cortical streams, which showed a significant 5-way interaction between Model, WordType, PeriExtra, TemporalFrontal and Area ($F_{2,24}$ = 7.45, ε = .83 p = .0054, η_p^2 = .38). To further investigate this complex effect, the interaction was broken down into component analyses (4- and 3-way ANOVAs), as specified below.

First, we performed separate ANOVAs on the peri- and extrasylvian systems. A significant interaction was found in the extrasylvian system involving the factors Model, WordType, TemporalFrontal and Area ($F_{2,24}$ = 21.46, ε = .82, p < .0001, η_p^2 = .65), while, as expected, no significant differences were revealed in the perisylvian system ($F_{2,24}$ = 0.389, p = .68). 3-way ANOVAs investigating performance on the two word categories separately showed significant interactions of the factors Model, TemporalFrontal and Area for both action ($F_{2,24}$ = 21.46, ε = .73, p < .0001, η_p^2 = .64) and object ($F_{2,24}$ = 14.99, ε = .80, p < .0001, η_p^2 = .55) words. The Bonferroni-corrected planned comparison tests (6 comparisons, corrected critical p < .0083) confirmed the observation of the higher density of action-related CA circuits in the blind compared to the sighted model in the primary visual area (V1, p < .0001), whereas, for object-related word CAs, a relatively lower neuron density was revealed in the primary visual (V1), temporo occipital (TO), anterior temporo (AT), lateral prefrontal (PFL) and lateral premotor (PML, p < .0001) areas (Fig. 4).

To contrast the different distributions of CA neurons across areas within each model separately, we ran another set of 4-way ANOVAs with the factors WordType, PeriExtra, TemporalFrontal and Area for the two models separately. The sighted model showed a significant interaction between WordType and Area ($F_{2,24}$ = 19.07, ε = .41, p <.001, η_p^2 = .72) and a significant interaction involving all four factors ($F_{2,24}$ = 19.07, ε = .41, p <.001, η_p^2 = .62), which confirms differences in CA distributions between the two word types. Additionally, a main effect of Area ($F_{2,22}$ = 747.838, ε = .98, p < .0001, η_p^2 = .98) was found, indicating the different CA cell densities distributed across the multi-area network, namely higher CA densities in hubs than in secondary areas (p < .0001), and in secondary than in primary areas (p < .0001). To determine whether differential CA distributions were present in peri- or extrasylvian systems, we separately ran further 3-way ANOVAs. The extrasylvian

system showed a highly significant interaction of the factors WordType, TemporalFrontal and Area ($F_{2,24}$ = 78.3, ε = .91, p < .0001, η_p^2 = .86), confirming the distinct word category distribution over the motor, visual and hub areas. The perisylvian regions did not show any significant distributional differences between the two word types ($F_{2,24}$ = 0.46, p = .63).

The blind model showed a 2-way interaction involving WordType and Area ($F_{2,24}$ = 19.07, ε = .43, p <.001, η_p^2 = .63), but the 4-way interaction of the factors WordType, PeriExtra, TemporalFrontal and Area was only marginally significant ($F_{2,22}$ = 3.47, ε = .95 p = .054). The additional statistical analysis performed separately on the two systems showed similar results as in the sighted model, supporting distributional differences of CA topographies in extrasylvian ($F_{2,24}$ = 13.0, ε = .88, p = .0003, η_p^2 = .51) but not perisylvian ($F_{2,24}$ = 0.14, p = .86) space. Bonferroni-corrected planned comparison tests assessed the presence of distributional differences between word types in the blind model area by area (6 comparisons, corrected critical p < .0083). This analysis revealed higher neuron densities for action- compared to object-related words in the dorsal motor stream, i.e. in lateral prefrontal (PF_L p < .0001), premotor (PM_L p < .0001) and primary motor cortex ($M1_L$ p < .0001), and, surprisingly, also in the ventral visual stream, anterior-temporal (AT, p < .0001), temporo-occipital (TO, p = .0027) and primary visual (V1, p = .0048) areas.

In summary, our neurobiologically constrained model of human cortex applied to simulate aspects of early word learning in congenitally blind and undeprived human individuals revealed the following results: Whereas in the undeprived case, contingencies between word-forms and actions or perceptions were mapped in the network by establishing tightly interconnected neuronal assemblies distributed across linguistic, ventral visual and dorsal motor streams, similar semantic mapping was only possible for action-related symbols in the blind model. Compared with the circuits for action-related words in the undeprived case, 'blind networks' showed an unexpected extension of these circuits into visual areas, with significantly higher neuron densities in primary (*V1) and higher visual cortices. Circuits of object-related words showed relatively reduced neuron densities in both extrasylvian streams.

Discussion

Activation of ventral stream visual cortex has been reported in healthy sighted subjects for the processing of object- and visually-related words specifically (Chao et al., 1999; Kiefer, 2005; Sim and Kiefer, 2005), but not or significantly less in action verb and tool word processing. In contrast, congenitally blind people were shown to activate visual areas, including the primary visual cortex, in semantic retrieval during verb generation (Amedi et al., 2004, 2003; Burton, 2002; Raz et al., 2005; Struiksma et al., 2011), single word comprehension (Burton, 2003; Burton et al., 2012) and sentence processing tasks (Bedny et al., 2011; Röder et al., 2002). Involvement of visual cortices in the healthy brain can be explained by their role in grounding symbolic meaning in visual perception of objects and their features (McCarthy and Warrington, 1988; Pulvermüller, 2001; Pulvermüller and Fadiga, 2010). However, under sensory deprivation, it is impossible that the correlation between visual and linguistic information leads to the strengthening of neuronal links into visual streams because blind people lack such modality-specific grounding information.

Here, we show that a spiking neural network constrained by cortical neuroanatomy and function and obeying well-established neuroscience principles can simulate the known visual cortex recruitment in both sighted and blind individuals during word meaning acquisition. The neuromechanistic explanatory account that we wish to offer based on these network simulations builds upon two mechanisms:

First, CA circuits grow spontaneously: in a network with random connectivity between spontaneously active neurons, a neuron firing above the level of its connected neighbours will strengthen its links to some of these neighbours, therefore giving rise to the spontaneous emergence of a relatively more strongly connected set of neurons (Doursat and Bienenstock, 2006). We call this process, which is explained by correlation learning between co-active neurons, 'Doursat-Bienenstock expansion' or DB-expansion. If such expansion happens at the level of large neuronal assemblies, these circuits will 'grow into' adjacent and connected areas (Garagnani and Pulvermüller, 2016; Tomasello et al., 2018, 2017).

Second, noise suppresses spontaneous CA circuit growth: stimulus- and action-induced uncorrelated activity in the extrasylvian streams of the network is critical for preventing the expansion of CA circuits into these streams. In this sense, it is the variability of visual inputs

in processing action-related symbols that guarantees variable activation in the visual stream and therefore neural activity uncorrelated to these symbolic-linguistic activations. For instance, when learning the meaning of an action word such as 'run' while performing the corresponding action (Tomasello and Kruger, 1992), the variable sensory information perceived during running can be seen as variable uncorrelated input, which works against DB expansion into the ventral visual stream.

Our present simulations suggest that it is the *absence of uncorrelated input to the ventral visual stream in the blind network* and brain that is necessary for DB-expansion of action-word-related CA circuits. In essence, as observed in previous simulations (Garagnani and Pulvermüller, 2016; Tomasello et al., 2018, 2017), the uncorrelated visual input is crucial for preventing DB-expansion of action-word-related circuits into visual areas of the undeprived brain.

We propose that the strong activation of primary visual areas in language processing observed in congenitally blind people is explained by DB-expansion of CA circuits described above. The relatively weaker visual activation in language processing in healthy people is explained by noise-related CA growth suppression. As mentioned in the Introduction, neuroimaging studies documented relatively stronger activation of the primary visual area (fMRI activity in V1) in blind than in undeprived individuals when generating semantically related verbs to given nouns (Amedi et al., 2003; Burton, 2002; Struiksma et al., 2011). Consistently, a study employing transcranial magnetic stimulation (TMS) in the primary visual area reported impairments in the verb generation task in blind but not in sighted individuals (Amedi et al., 2004). The verb generation task implies the activation of multiple CA circuits for verbs, most of which are action-related (Moseley and Pulvermüller, 2014), and this engages the ventral visual system more in blind people than in undeprived control subjects. Stronger V1 activation in blind than in sighted people has also been reported during sentence processing (see Fig. 4.1), which likely included action-related words too (Bedny et al., 2011; Röder et al., 2002). Therefore, the aforementioned fMRI and TMS results are consistent with the predictions of the present simulations, in which the modelled primary visual area (*V1) becomes more actively involved in the processing of action-related meaningful symbols and complex utterances including such symbols (Fig. 4.1 & 4.3). These results represent a significant advance in the debate about the mechanisms underlying the neural changes in the visual cortex: evidence indicates that such cortical areas can take over

a particular function depending on input information received during the developmental period (Bedny, 2017); On the basis of our results, it is precisely the lack of informative input to visual cortex that drives the Hebbian synaptic modifications and consequent extension of linguistic representations into visual cortex seen in congenitally blind individuals. The underlying mechanisms are consistent with general neurobiological plasticity mechanisms documented in other deprived sensory systems (Buonomano and Merzenich, 1998; Merzenich et al., 1984) and, even though a higher cognitive function, language, is involved, the explanation rests on the same neuroscience principles.

Intriguingly, the present neurobiologically constrained 'blind' neural network was not only able to reproduce the visual cortex recruitment in the blind but also showed prolonged spiking neural activity for action-related words during word recognition simulations (Fig. 4.5). Sustained neural activity is a neural correlate of working memory (Baddeley and Hitch, 1974; Leavitt et al., 2017), which, in the present study, persisted longer in the blind compared to the sighted model. This phenomenon in the network is consistent with the observation of enhanced verbal working memory ability in congenitally blind individuals compared to control sighted population during working memory performance (Amedi et al., 2003; Occelli et al., 2017; Pasqualotto et al., 2013; Withagen et al., 2013). Note, furthermore, that during the reverberation phase, activity retreats from modality-specific to the modality general association cortices in frontal and temporal cortex (*AT, *PF) in both sighted (time steps 12-14) and blind models (time steps 17-19). This is consistent with, and provides an explanation for, the so-called 'anterior shift' of cortical activation from sensorimotor cortices to temporal and prefrontal connector hub regions during working memory (Fuster, 1998; Leavitt et al., 2017; Pulvermüller and Garagnani, 2014; Tomasello et al., 2018).

In the present simulation of undeprived referential-semantic learning, CA circuits emerged spontaneously across the fronto-temporo-occipital areas of the spiking neural network linking word-form in the perisylvian cortex with semantic information about referent objects and actions in the extrasylvian system. The learning of object- and action-related words was grounded in correlated sensorimotor information presented in the primary cortices of the architecture: besides perisylvian *A1 and *$M1_i$ activity, object-related words received concordant visual (*V1) and, similarly, action-related words received lateral motor area (*$M1_L$) grounding activity. Because of noise suppression of CA growth,

the fourth 'non-relevant' input area (*$M1_i$ for object- and *V1 for action-related words) was not left void of any sensory input, but instead processed uncorrelated ('suppressing') information and neuronal activation patterns. As reported by the present and previous simulations, noise-suppression of CA growth becomes relevant in the undeprived brain's formation of category-specificity of circuit topographies with action-related word circuits reaching into the motor cortices (*$M1_L$-*PM_L), but not or less into visual areas (*V1, *TO), and vice versa for object words (Garagnani and Pulvermüller, 2016; Tomasello et al., 2018, 2017). Here we replicated these previous results with a spiking neural network and went one step further by systematically investigating the consequences of *not* presenting such uncorrelated noise patterns to the model's primary visual cortex during action-word learning. This was meant to specifically simulate a learning situation in which the meaning of such action words is acquired in the absence of any visual input (i.e., in blindness).

The current observations and their possible explanation in terms of DB-expansion of CA circuits and noise-related suppression of such growth suggest that these mechanisms are more broadly applicable to cases of sensory deprivation. Similar to blind individuals, deaf individuals activate their deprived auditory cortex in processing visual stimuli (Finney et al., 2001) and in the processing of visually presented units of their native language, typically a manual signing system (Petitto et al., 2000). Some of these results had previously been used to strongly argue for an inborn mechanism linking abstract (but not acoustic or other sensory or motor) features of language to specific brain parts. Our present work offers an alternative explanation based on established neurobiological mechanisms (see Results, points (i) – (v) – (vi)).

For object-related words, simulation results indicate a generally reduced relevance of extrasylvian areas in blind people – both compared with action words in the same population and compared with the same word type in the healthy undeprived (see Fig. 5.4). This suggests reduced grounded semantic knowledge in blind people, at least for some specific word types requiring visual knowledge for complete acquisition of their related concepts. For the semantics of colour terms, such partially deficient semantic knowledge in the blind has been supported by experimental studies (Connolly et al., 2007; Shepard and Cooper, 1992), although other work reported comparable semantic similarity ratings (Marmor, 1978). However, for other object-related words, it is less plausible that substantial differences in semantic knowledge are present between congenitally blind and sighted

infants. It is known that, when blind people learn words for objects, they naturally draw more on manual exploration and touch than undeprived individuals. In her seminal studies, Gleitman noted, for example, that, when a blindfolded undeprived child is advised to 'look up', it would raise its head, whereas a blind one would explore the space above its head with the hands (Gleitman, 1990). This and similar observations suggest that, for a range of words typically grounded in visual experience, congenitally blind individuals use tactile and motor knowledge in the semantic grounding process. This difference in stimulation modality implies a degree of similarity between semantic grounding processes of object and action words in the blind. On the other hand, this difference in modality also implies that congenitally blind people can use similar grounding information for object words as healthy subjects, although this same (or very similar) information is provided through a different channel. This is particularly the case if information about the form or shape of referent objects is acquired through vision or tactile exploration. Future experimental works and simulation studies are still needed to explore more closely the learning of different subtypes of visually-related words in blind brains and networks taking into account, in particular, information in the tactile modality. Instead of aiming at capturing such fine-grained differences in semantic grounding, our present study specifically addressed the effect of sensory deprivation and the consequent conquering of visual cortex by linguistic and semantic processes.

We wish to conclude by pointing to further obvious limitations of the present work. First, we simulated semantic learning in a 'grounding' context, where words are co-present with actions and objects. Useful next steps in the modelling effort shall focus on the acquisition of novel word meaning in the context of already grounded meaningful words (Pulvermüller, 2010; Pulvermüller and Knoblauch, 2009) and on the learning of word sequences and whole constructions along with their semantics. With regard to blind individuals, we have restricted our scope to congenitally blind subjects, because they provide the clearest case of deprivation. The more complex situation of later deprivation, where normal learning takes place first and deprivation kicks in at a later stage, may also provide a basis for fruitful future simulations. We note that there are some important differences in reorganisation processes between congenitally, early and late blind persons (Burton, 2002; Kujala et al., 1997; Voss et al., 2008), which may be attributed to altered learning histories and possibly also to altered neural substrates and plasticity at different

developmental stages. In spite of its focus on only one type of semantic learning and only the most typical type of visual sensory deprivation, our model offers a novel neurobiological explanation of the linguistic takeover of visual cortex.

In sum, the present study aimed to simulate the effect of visual deprivation on the neuronal mechanisms of semantic and language processing in sighted and congenitally blind people by means of a neurobiological constrained neural network of the frontal, temporal and occipital lobes. Specifically, we focus on the mechanisms responsible for the activation of the deprived areas during semantic processing consistently reported by a number of experimental studies described above, and show that the interaction of three main factors may lead to the takeover of visual cortex for linguistic and semantic processing: (i) the changes in the balance of activity related to the absence of uncorrelated sensory input, (ii) constrained neuroanatomical connectivity and (iii) Hebbian correlation learning. Mechanisms of DB-expansion (resulting from (ii)-(iii)) are crucial for visual cortex recruitment in the blind, and those of 'noise'-related prevention of such expansion for the category-specific nature of semantic circuits in healthy individuals. The present architecture explains action-related word processing in both dorsal motor and deprived ventral visual streams. Here we bridge the gap between neural mechanisms and conceptual brain functions, offering a biological account of visual cortex reorganization following sensory loss from birth and its functional recruitment for language and semantic processing.

Methods

Structure and function of the spiking neuron model

Each of the 12 simulated areas is implemented as two layers of artificial neuron-like elements ('cells'), 625 excitatory and 625 inhibitory, thus resulting in 15,000 cells in total (see Fig. 4.2b-c). Each excitatory cell '*e*' consists of a leaky integrate-and-fire neuron with adaptation and simulates a single pyramidal cell representative of excitatory spiking activity in a cortical micro-column, while its twin inhibitory cell '*i*' is a graded-response cell simulating the average inhibitory response of the cluster of interneurons situated in a local neighbourhood (Eggert and van Hemmen, 2000; Wilson and Cowan, 1972). The state of each cell *x* is uniquely defined by its membrane potential $V(x,t)$, specified by the following equation:

$$\tau \cdot \frac{dV(x,t)}{dt} = -V(x,t) + k_1(V_{In}(x,t) + k_2\eta(x,t)) \qquad (1)$$

where $V_{In}(x,t)$ is the net input acting upon cell *x* at time *t* (sum of all inhibitory and excitatory postsynaptic potentials – I/EPSPs; inhibitory synapses are given a negative sign), τ is the membrane's time constant, k_1, k_2 are scaling values (see below for the specific parameter values used in the simulations) and $\eta(\cdot,t)$ is a white noise process with uniform distribution over [-0.5,0.5]. Note that noise is an inherent property of each model cell, intended to mimic the spontaneous activity (baseline firing) of real neurons. Therefore, noise was constantly present in all areas, in equal amounts (inhibitory cells have k_2=0, i.e., the noise is generated by the excitatory cells in the model for convenience).

The output (or transformation function) ϕ of an excitatory cell *e* is defined as follows:

$$\phi(e,t) = \begin{cases} 1 & if \quad (V(e,t) - \alpha\,\omega(e,t)) > thresh \\ 0 & otherwise \end{cases} \qquad (2)$$

Thus, an excitatory cell *e* spikes (=1) whenever its membrane potential $V(e,t)$ overcomes a fixed threshold *thresh* by the quantity $\alpha\omega(e,t)$ (where α is a constant and ω is defined below). Inhibitory cells are graded response neurons as they intend to represent the

average impact of a cluster of local interneurons; the output $\phi(i,t)$ of an inhibitory neuron i is 0 if $V(i,t)<0$ and $V(i,t)$ otherwise.

To simulate neuronal adaptation (Kandel et al., 2000), function $\omega(\cdot,t)$ is defined so as to track the cell's most recent firing rate activity. More precisely, the amount of adaptation $\omega(e,t)$ of cell e at time t is defined by:

$$\tau_{ADAPT} \cdot \frac{d\omega(e,t)}{dt} = -\omega(e,t) + \phi(e,t) \qquad (3.1)$$

where τ_{ADAPT} is the 'adaptation' time constant. The solution $\omega(e,t)$ of Eq. (3.1) is the low-pass-filtered output ϕ of cell e, which provides an estimate of the cell's most recent firing-rate history. A cell's average firing activity is also used to specify the network's Hebbian plasticity rule (see Eq. (B4) below); in this context, the (estimated) instantaneous mean firing rate $\omega_E(e,t)$ of an excitatory neuron e is defined as:

$$\tau_{Favg} \cdot \frac{d\omega_E(e,t)}{dt} = -\omega_E(e,t) + \phi(e,t) \qquad (3.2)$$

Local (lateral) inhibitory connections (see Fig. 4.2c) and area-specific inhibition are also implemented, realising, respectively, local and global competition mechanisms(Duncan, 2006, 1996). More precisely, in Eq. (1) the input $V_{In}(x,t)$ to each excitatory cell of the same area includes an area-specific ('global') inhibition term $k_G\,\omega_G(e,t)$ (with k_G a constant and $\omega_G(e,t)$ defined below) subtracted from the total I/EPSPs postsynaptic potentials V_{In} in input to the cell; this regulatory mechanism ensures that area (and network) activity is maintained within physiological levels(Braitenberg and Schüz, 1998):

$$\tau_{GLOB} \cdot \frac{d\omega_G(e,t)}{dt} = -\omega_G(e,t) + \sum_{e \in area} \phi(e,t) \qquad (3.3)$$

Excitatory links within and between (possibly non-adjacent) model areas are established at random and limited to a local (topographic) neighbourhood; weights are initialised at random, in the range [0, 0.1]. The probability of a synapse to be created between any two cells falls off with their distance (Braitenberg and Schüz, 1998) according to a Gaussian function clipped to 0 outside the chosen neighbourhood (a square of size n=19 for excitatory and n=5 for inhibitory cell projections). This produces a sparse, patchy and topographic

connectivity, as typically found in the mammalian cortex (Amir et al., 1993; Braitenberg and Schüz, 1998; Douglas and Martin, 2004; Kaas, 1997).

The Hebbian learning mechanism implemented simulates well-documented synaptic plasticity phenomena of long-term potentiation (LTP) and depression (LTD), as implemented by Artola, Bröcher and Singer (Artola et al., 1990; Artola and Singer, 1993). This rule provides a realistic approximation of known experience-dependent neuronal plasticity and learning (Finnie and Nader, 2012; Malenka and Bear, 2004; Rioult-Pedotti et al., 2000), and includes both (homo- and hetero-synaptic, or associative) LTP, as well as homo- and hetero-synaptic LTD. In the model, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, +Δ and – Δ (with Δ<<1 and fixed). Following Artola *et al*., we defined as 'active' any (axonal) projection of excitatory cell *e* such that the estimated firing rate $\omega_E(e,t)$ of cell *e* at time *t* (see Eq. (3.2)) is above ϑ_{pre}, where $\vartheta_{pre} \in [0,1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP to occur. Thus, given a pre-synaptic cell *i* making contact onto a post-synaptic cell *j*, the change $\Delta w(i,j)$ inefficacy of the (excitatory-to-excitatory) link from *i* to *j* is defined as follows:

$$\Delta w(i,j) = \begin{cases} +\Delta & \text{if } \omega_E(i,t) \geq \theta_{pre} \text{ and } V(j,t) \geq \theta_+ \quad \textit{(LTP)} \\ -\Delta & \text{if } \omega_E(i,t) \geq \theta_{pre} \text{ and } \theta_- \leq V(j,t) < \theta_+ \quad \textit{(homosynaptic LTD)} \\ -\Delta & \text{if } \omega_E(i,t) < \theta_{pre} \text{ and } V(j,t) \geq \theta_+ \quad \textit{(heterosynaptic LTD)} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Table 4.1 Parameter values used during simulations

Eq.(B1)	Time constant (excitatory cells)	τ = 2.5 (simulation time-steps)
	Time constant (inhibitory cells)	τ = 5 (simulation time-steps)
	Total input rescaling factor	k_1 = 0.01
	Noise amplitude	$k_2 = 1 \cdot \sqrt{(24/\Delta t)}$
	Global inhibition strength	k_G = 0.60
Eq. (B2)	Spiking threshold	*thresh* = 0.18
	Adaptation strength	α = 7.0
Eq.(B3.1)	Adaptation time constant	τ_{ADAPT} = 10 (time steps)
Eq.(B3.2)	Rate-estimate time constant	τ_{Favg} = 30 (time steps)
Eq.(B3.3)	Global inhibition time constant	τ_{GLOB} = 12 (time steps)
Eq.(B4)	Postsynaptic membrane potential thresholds:	
		ϑ_+=0.15
		ϑ_-=0.14
	Presynaptic output activity required for LTP:	
		ϑ_{pre}=0.05
	Learning rate	Δ = 0.0008

The model's connectivity structure

The **between-area connectivity** binds adjacent cortical areas together (Pandya and Yeterian, 1985; Young et al., 1995, 1994). In the perisylvian system, next-neighbour connections between cortically adjacent areas are implemented within the auditory (A1, AB, PB) (Kaas and Hackett, 2000; Pandya, 1995; Rauschecker and Tian, 2000), as well as within the articulatory (PF_i, PM_i, $M1_i$) sub-systems (Pandya and Yeterian, 1985; Young et al., 1995). Similarly, local next neighbour links are also realised in the extrasylvian system, between adjacent ventral visual (V1, TO, AT) (Bressler et al., 1993; Distler et al., 1993), and dorsolateral motor areas (PF_L, PM_L, $M1_L$) (Arikuni et al., 1988; Dum and Strick, 2005, 2002; Lu et al., 1994; Pandya and Yeterian, 1985; Rizzolatti, G. Luppino, 2001). Furthermore, reciprocal links also exist between anterior-temporal (AT) and parabelt (PB) areas (Gierhan, 2013) and inferior and lateral prefrontal (PF_i, PF_L) areas (Yeterian et al., 2012).

The **long distance cortico-cortical connections** implemented reciprocally link all pairs of multimodal hub areas (PB, PF_i, AT and PF_L) of the four sub-systems, modelling documented anatomical connections between inferior pre-frontal (PF_i) and auditory parabelt (PB) (Catani et al., 2005; Makris and Pandya, 2009; Meyer et al., 1999; Parker et al., 2005; Paus et al., 2001; Rilling et al., 2008; Romanski et al., 1999b) and between anterior-temporal (AT) and lateral prefrontal (PF_L) areas, realised by the arcuate and the uncinated fascicles (Bauer and Jones, 1976; Chafee and Goldman-Rakic, 2000; Eacott and Gaffan, 1992; Fuster et al., 1985; Parker, 1998; Ungerleider et al., 1989; Webster et al., 1994). The peri- and extrasylvian systems are also linked by means of long distance cortico-cortical connections across the central hub areas; likewise parabelt (PB) and lateral prefrontal cortex (PF_L) are reciprocally connected (Pandya and Barnes, 1987; Romanski et al., 1999b, 1999a) as well as the anterior/middle-temporal (AT) and inferior prefrontal (PF_i) areas (Pandya and Barnes, 1987; Petrides and Pandya, 2009; Rilling, 2014; Romanski, 2007; Ungerleider et al., 1989; Webster et al., 1994).

The present neural spiking network implemented additional **high-order 'jumping' links,** which skip one intermediate area (blue arrows Fig. 4.2b), documented by a range of recent neuroanatomical and diffusion tensor and diffusion-weighted imaging (DTI/DWI) studies in humans and non-human primates. These links exist within (auditory) superior temporal and

(articulatory) inferior frontal cortex of the perisylvian cortex, that is amongst: primary auditory (A1) - parabelt (PB) areas (Pandya and Yeterian, 1985; Young et al., 1994), parabelt (PB) - inferior premotor (PMi) areas (Saur et al., 2008), auditory belt (AB) - inferior prefrontal (PFi) (Kaas and Hackett, 2000; Rauschecker and Scott, 2009; Romanski et al., 1999a) and as well inferior prefrontal (PFi) - primary motor (M1i) areas (Deacon, 1992; Guye et al., 2003; Young et al., 1995). Additional evidence for the presence of high-order jumping links within the perisylvian system are well-documented also in DTI/DWI studies in humans (Rilling et al., 2011; Thiebaut de Schotten et al., 2012). The ventral visual and the dorsolateral motor sub-systems of the extrasylvian cortex were also endowed with jumping links, similarly to the perisylvian cortices listed above. In particular, primary visual (V1) area is reciprocally linked to anterior-temporo (AT) area (Catani et al., 2003; Wakana et al., 2004), as well as anterior-temporo (AT) and dorsolateral premotor (PML) area, as documented by both anatomical (Pandya and Barnes, 1987; Seltzer and Pandya, 1989) and monkey studies (Bauer and Fuster, 1978; Chafee and Goldman-Rakic, 2000; Fuster et al., 1985). Additional jumping links were implemented between temporo-occipital (TO) and dorsolateral prefrontal areas (PF_L), as supported by evidence from anatomical studies in humans (Makris and Pandya, 2009) and monkeys (Bauer and Jones, 1976; Fuster et al., 1985; Fuster and Jervey, 1981; Seltzer and Pandya, 1989), and between dorsolateral prefrontal (PF_L) and dorsolateral premotor ($M1_L$) areas (Deacon, 1992; Guye et al., 2003; Young et al., 1995). Further neuroanatomical DTI studies also showed connections within the extrasylvian system as described above (Thiebaut de Schotten et al., 2012). Notice that the connectivity structure of both sighted and blind models was kept the same, as a number of DTI studies have shown similar anatomical connectivity structure between sighted and blind populations (Noppeney et al., 2005; Shimony et al., 2005; Shu et al., 2009a, 2009b).

Simulating word learning

Prior to the training, each network was initialised with all the synaptic links (between- and within-areas) connecting single cells established at random (see Methods section under '*Structure and function of the spiking neuron model*'). Similar to previous simulation studies (Garagnani et al., 2017; Garagnani and Pulvermüller, 2016; Tomasello et al., 2018, 2017), word-meaning acquisition was then simulated under the impact of repeated sensorimotor

pattern presentations to the primary areas of the network. Each network instance used 12 different sets of sensorimotor word patterns representing six object- and six action-related words. Each pattern consisted of a fixed set of 19 cells chosen at random within the 25 x 25 cells of an area (ca. 3% of the cells). Note that additional white (so-called 'contextual') noise was continuously presented to all primary areas of the network, and thus superimposed on all learning patterns. This partly accounted for a degree of variability during word meaning acquisition of the two word-types.

Word-related sensorimotor patterns were presented 3000 times (previous simulations using a six area model showed no substantial change in the relevant primary areas between 1000 and 10000 learning steps (Garagnani et al., 2009; Schomers et al., 2017)) as described above. A trial started with a word pattern presentation for 16 simulation time steps, followed by a period during which no input (interstimulus interval – ISI) was given. The next word pattern (learning step) was presented to the network only when the global inhibition of the PF_i and PB areas decreased below a specific fixed threshold; this allowed the activity to return to a baseline value, so as to minimise the possibility of one trial affecting the next one. Only the inherent baseline noise (simulating spontaneous neuronal firing) and 'contextual' noise were present in the neural-network during each ISI.

Data processing and statistical analysis

Cell assemblies, which are strongly interconnected networks of neurons, spontaneously emerged during word learning simulation. After learning, the word-form neurons in the primary perisylvian auditory-articulatory areas (A1, $M1_i$) simulating the 'word production' were activated for 15 simulation time-steps to identify and quantify the neurons forming the 12 distributed CA circuits that emerged across the network areas. During this period, we computed and displayed the average firing rate of each excitatory cell (7500 *e*-cells, cell's responses).

As an estimate of a cell's average firing-rate here we used the value $\omega_E(e,t)$ from Eq. (B3.2), integrated with time-constant τ_{Favg}= 5. An *e*-cell was then taken to be a member of a given CA circuit only if its time-averaged rate (output value or 'firing rate') reached a threshold ϑ which was area- and input-pattern specific, and defined as a fraction *γ* of the maximal single-cell's time-averaged response in that area to pattern *w.* More formally,

$$\vartheta = \vartheta_A(w) = \gamma \max_{x \in A} \overline{O(x,t)_w}$$

where $\overline{O(x,t)_w}$ is the estimated time-averaged response of cell *x* to word pattern *w* (see Eq. [B3.3] in Methods section under '*Structure and function of the spiking model'*) and $\gamma \in [0,1]$ is a constant (we used γ = 0.5 on the basis of previous simulation results Garagnani et al., 2009, 2008; Tomasello et al., 2017). This was computed for each of the 13 trained network instances, averaging the number of CA cells per area over the 6 object- and 6 action-related words.

To investigate the presence of significant statistical differences between sighted and blind neural network models, we performed an initial statistical analysis including both neural network models. To this end, a 3-way ANOVA was run with factors Model (two levels: *Sighted* vs. *Blind*), WordType (two levels: *Object* vs. *Action*) and Area (12 levels: *Perisylvian* = {A1, AB, PB, $M1_i$, PM_i, PF_i}, *Extrasylvian* cortex = {V1, TO, AT, $M1_L$, PM_L, PF_L}). Additionally, to further investigate difference of the modelled cortical regions between the two models a 5-way ANOVA was run with factors Model (two levels: *Sighted* vs. *Blind*), WordType (two levels: *Object* vs. *Action*), PeriExtra (two levels: *Perisylvian* = {A1, AB, PB, $M1_i$, PM_i, PF_i}, *Extrasylvian* cortex = {V1, TO, AT, $M1_L$, PM_L, PF_L}), TemporalFrontal (TempFront) (2 levels: *temporal areas* = {A1, AB, PB, V1, TO, AT}, *frontal areas*={$M1_L$, PM_L, PF_L, $M1_i$, PM_i, PF_i}) and Area (three levels: *Primary* = {A1, V1, $M1_L$, $M1_i$}, *Secondary* = {TO, AB, PM_L, PM_i} and *Central* = {PB, AT, PF_L, PF_i} areas). Subsequently, each system, 6 peri- and 6 extrasylvian areas, were investigated separately with factors 'Model', 'WordType', 'TempFront' and 'Area'. The same statistical analysis, but this time omitting 'WordType' as a factor was additionally performed to disentangle the different CA distribution of action- and object-related words between the two models.

A second level of analysis was run on each Model (blind and sighted) separately, first with a 2-way ANOVA with factors 'WordType' and 'Area' and a 4-way ANOVA with factors 'WordType', 'PeriExtra', 'TempFront' and 'Area' and subsequently, with 3-way ANOVA on each system within the sighted and blind model, peri- and extrasylvian systems, separately. Corrected p-values along with epsilon (ε) values are reported throughout. Partial eta-square (η_p^2) values are also stated, which is defined as an index of effect size (0.01-0.06 small, 0.06-0.14 medium and >0.14 large, Cohen, 1988)

5. General Discussion

Summary of results and main contributions

Chapter 2

Current semantic theories still diverge on the nature of semantic processing in the brain. *Why* do specific parts of the brain specialise in processing meaning, and *why* are some of the brain's semantic processes category-specific to semantics types such as animals, tools or actions? We applied a neurocomputational model replicating anatomical and physiological features of a range of cortical areas relevant for language and semantic processing to simulate (i) the learning of semantic relationships between word-forms and specific object perceptions and motor movements of the own body and (ii) the neurophysiological responses to perception of learned object and action words.

The model showed spontaneous emergence of stimulus-specific, tightly interlinked CAs, connecting the processing of word-form information with that of sensorimotor semantic information. These simulations (i) explain the presence of category-specificity in the cortical distribution of word-related circuits, with highly-connected hub areas exhibiting an only moderate category specificity, and (ii) predict a symmetric activation time-course in the sensorimotor systems for both object- and action-related word recognition, with analogous temporal dynamics in the hub areas. These results account for the presence of both category-specific and general semantic hub areas in the human brain, which is a direct consequence of the mutual interaction of network structure, connectivity, and Hebbian plasticity. Here we offer a mechanistic account of *when* and *where* semantic knowledge is processed in the human brain by reconciling the diverging neurocognitive empirical studies.

Chapter 3

Previous computational models of semantic processing have failed to implement precise mathematical neural architectures of the human brain, mostly using basic neuron (non-spiking) model and incorporating only a set of connectivity structure revealed by neuroanatomical studies. Here, we improved the realism of the previous computational mean-field architecture presented in Chapter 2, by adding important biological constraints;

(i) adaptation-based spiking cells (versions of leaky integrate-and-fire neurons), each thought to represent a single pyramidal neuron, and

(ii) a more realistic connectivity structure based on prior neuroanatomical evidence.

The spiking neural network confirmed the conclusions of previous simulations achieved with basic mean-field model by exhibiting a category-specificity in the cortical distribution of word-related circuits. The high-degree connection hub areas that bind information across different modalities showed the involvement in all types of semantic processing. After training, the network was re-activated with the learnt auditory pattern to simulate word recognition processes exhibiting the different cognitive processes of word perception, comprehension, and verbal working memory. The spiking model showed an 'anterior shift' from sensorimotor areas to frontal, temporal regions during working memory (reverberation time) that were not present in previous basic mean-field models. The present findings demonstrate that biologically computational models at different level of details can consistently explain semantic processing in the human brain.

Chapter 4

Studies investigating language processing in congenitally blind people have shown that the deprived visual areas are functionally recruited by other modalities, such as language and semantic processing. *What* are the consequences of the neural circuits changes representing language processing in the visual system deprived from visual input? *How* does it emerge at the cellular/synaptic level? Here we applied a spiking neural network of the fronto-temporal-occipital lobes to simulate word meaning acquisition in sighted and blind population, in order to move forward the debate of neurobiological factors underlying the functional changes in the visual cortex.

Equipped with correlation-based Hebbian learning, both sighted and blind models showed the spontaneous emergence of CAs across the network, binding word-form information to that of sensorimotor semantic information. Nevertheless, we observed that only under visual deprivation, distributed word-related neural circuits extended into the deprived visual areas, which therefore adopted a semantic and linguistic role. Two mechanisms are the consequences of the visual area's recruitment under sensory deprivation: (i) the 'Doursat-Bienenstock' expansion, which gives rise to the formation of

strongly connected assemblies of cells extending into adjacent regions and (ii) the changes in the network's activity balance brought about by the absence of uncorrelated sensory input. The mutual interaction of these two mechanisms offers an explanation for the neural plastic changes in the blind brain for word meaning processing.

General aims & scope

The present work aims to investigate the neurobiological mechanisms underlying language and semantic processing in the human brain. One of the primary objectives was to shed light on the functional cortical organisation of the semantic system and, in turn, to reconcile the diverging semantic theories and experimental evidence of the various cortical contributions (semantic hubs vs modality-preferential areas) for meaning processing. Notably, we seek to answer the hotly debated questions of *why* specific parts of the brain specialise in processing meaning and *why* some of the brain's semantic processes are category-specific to semantic types such as animals, tools, or actions. Additionally, we investigated the precise activation time course of semantic processing by seeking an answer at *which* point in time the activation of the modelled semantic brain areas first emerges and compare it with MEG activation responses. A second central objective of the present work was to investigate the mechanisms of *why* and *how* the distributed language network adapts to and reorganises itself by visual deprivation, as documented in neurocognitive studies of language processing in congenitally blind people.

We show how a set of biological mechanisms acting within specific neuroanatomical structures is sufficient to provide a direct and straightforward explanation for the unresolved questions mentioned above. The present work applied a neurobiologically constrained model of the human cortical function at different levels of detail to investigate the components of the semantic knowledge system in the human brain. The network replicates the structure and connectivity of frontal, temporal and occipital areas to simulate the emergence of neural circuits underpinning information about object- and action-related words under deprived and undeprived conditions. The neural network was capable to replicate and reconcile the diverging neuroimaging data of semantic processing and to mimic the processes of neural plastic change in the visual system caused by blindness for linguistic and semantic processing. This research work offers a neurobiological explanation for the diverging neurocognitive evidence of multiple semantic hubs and modality-

preferential cortical regions for meaning processing of different linguistic word types and how semantic processing is instantiated in visually deprived populations. The simulation results described in detail in the present dissertation are in line with modal grounding approach for word meaning processing, in particular showing that words are represented by CA distributed in multimodal semantic hubs and modality-preferential sensorimotor regions of the human brain, speaking against a pure amodal symbolic system for semantic knowledge processing. Here, we bridge the gap between cellular-level mechanisms and system-level language function in sighted and blind people.

Semantic knowledge: Semantic theories, experimental data and computational models

It is widely accepted that the main neurobiological mechanism for learning depends on Hebbian plasticity mechanisms (Hebb, 1949), which, along with correlation learning, have been defined the basic neuroscience principles for language acquisition (see in Introduction section under '*How biological computational models can help').* During learning, the flow of activity within a network circuit in the brain leads to microscopic chemical changes in synapses between neurons (i.e., strengthening, LTP and weakening, LTD, the connection points), which sets the foundation for how words, symbols, and knowledge of the outside world are encoded and stored in the human brain. The computational neural network presented here is governed by these biological learning mechanisms, which were used to simulate associative word learning, a foundational mechanism for lexical acquisition. In particular, the network simulates referential association between a word-form and its referent in the context of object information, (e.g., saying the word 'dog' while the dog is physically present, Vouloumanos and Werker, 2009) or action information (e.g., saying the word 'grasp' while performing the related action movement, Tomasello and Kruger, 1992). Although the model is only 'fed' in with information, the Hebbian correlation principle implemented in the network is not only biologically plausible but also ecologically valid, in the sense that it relies solely on the input given to the network. Neither supervision nor specific tasks are needed for learning, in contrast to the backpropagation learning rule (a supervised learning mechanism) used by recent computational semantic models (Chen et al., 2017; Ueno et al., 2011), defined as biologically implausible (Mazzoni et al., 1991;

O'Reilly, 1998). Hence, the impact of repeated sensorimotor pattern presentations in the primary articulatory (M1i) and auditory (A1) perisylvian systems along with either dorsolateral motor ($M1_L$) or visual (V1) cortices (simulating semantic grounding) leads to the spontaneous emergence of word-related circuits of concrete object or action words linking auditory-articulatory information (i.e., word-form) with its semantic referential information in action and perception systems.

The cortico-cortical connectivity structure implemented between the brain regions sets the basis for such learning by enabling the spreading of activity through the network areas and, in turn, the strengthening of synapses between cells and the formation of assemblies of neurons. Here we closely replicated the existing between-area, long distance, and high-order jumping pathways between the corresponding areas of the cortex (Chapter 3, Table 3.1). The mutual interaction of the correlated Hebbian learning mechanism and the connectivity structure between the cortical areas led to the formation of neural circuits, binding word-form and semantic information together. Each of these distributed CA circuits acted as a single functional unit, distributed across primary, secondary, and multimodal hub areas of the fronto-temporal-occipital lobes.

The emerged neural circuits exhibited a category-specific topographical distribution, reaching into motor and visual areas for action- and visually-related words, respectively. The formation of such topographical distribution was not only due to the mutual interaction of the learning mechanism and the connectivity structure but also for the presence of uncorrelated input patterns in the fourth non-relevant areas (V1 for action words and $M1_L$ for object words). This was intended to simulate variable sensory or motor information, typically occurring during the learning of object or action words—for example, think about the different ways an object can be grasped or the variable sensory input during an action performance. In this way, the correlation mapping of word-forms in the perisylvian language areas with that of semantic information (in visual and action systems) was higher in one modality (for object or action words) but lower in the non-relevant regions. The presence of an uncorrelated pattern played an essential role in preventing the excessive CA extension (Doursat and Bienenstock, 2006) in the non-relevant areas and, crucially, for the emergence of category-specific CA distribution. In other words, constantly activated CAs during learning cause the strengthening of some synapses and the weakening of others; in turn, some cells

are more strongly linked to a CA than to others. If the network is continuously stimulated, CAs spontaneously grow into connected adjacent cortical areas taking over the entire network, unless uncorrelated variable patterns block their extension. This self-organized growth principle has been postulated as an important basis of learning and neural development (Doursat and Bienenstock, 2006). Hence, the joint interaction of connectivity structure with the biological mechanisms of LTP and LTD in strengthening and weakening the connections between cells (i.e., Hebbian plasticity) and the aforementioned CA growth principle with the presence of uncorrelated input lead to the formation of category-specific semantic activation, consistent with a wealth of brain data from neurophysiology, patient and neuroimaging studies (Damasio et al., 1996; Martin, 2007; Martin et al., 1996; Moseley et al., 2013; Pulvermüller et al., 2014b, 1999; Vigliocco et al., 2004).

The multimodal connector hubs central in the neural architecture showed a higher density of CA circuits compared to primary and secondary regions of the network. This is due to their higher degree of connections and their strategic position between sensory and motor areas, which are thought to play a role in linking multiple pieces of information across different modalities (Braitenberg and Schüz, 1998; Damasio, 1989; Sporns et al., 2007; Van den Heuvel and Sporns, 2013). Intriguingly, these hub regions become the loci for general semantic processing, functioning as a semantic hub, as the same degree of CA cells for both semantic categories (action and object words) emerged there. This is mainly due to their role in linking the distinct modality-preferential regions together, in which during word learning, correlated neural activity needs to flow through the connectors to reach the modality-preferential areas. In contrast to the hub-and-spoke model (Ralph et al., 2017) mentioned in the introduction, which explains the presence of a single hub located in the anterior temporal lobe (ATL), the present simulations are able to explain the spontaneous emergence of four different semantic hub areas reported by experimental data on semantic processing: the anterior inferior temporal lobe, which many neuroscientists still believe to be the only brain locus for meaning processing (Patterson et al., 2007; Ralph et al., 2017), the superior temporal parabelt, the inferior and lateral prefrontal cortex (Bookheimer, 2002; Carota et al., 2017; Posner and Pavese, 1998; Tate et al., 2014). Intriguingly, the model predicts a moderate category-specificity in the semantic hub areas (AT and PF_L) compared to primary and secondary areas. This is in line with recent patient studies that

revealed category-specific impairments after lesions in the mentioned semantic hub areas (Gainotti, 2012; Pulvermüller et al., 2010; Shebani et al., 2017; Silveri et al., 2018), which is in contrast to the general semantic impairment previously postulated (Patterson et al., 2007). Specifically, lesions in multimodal parts of the left temporal lobe (corresponding to area AT in the network) have been found to cause category-specific word processing deficits for animals, people, and other living things (Damasio et al., 1996; Gainotti, 2012; Hernández et al., 2008; Pulvermüller et al., 2010). However, in order to properly test the predictions made by the model, especially for the category-specificity of object- and action-related words in the anterior-temporal and dorsolateral prefrontal hub areas, a well-designed experiment employing Transcranial Magnetic Stimulation (TMS), producing virtual lesions in healthy subjects, could further investigate the possibility of category-specific semantic deficits in AT and PFL areas. Note that further semantic hubs have been also reported in the parietal lobe, in particular, in the anterior inferior parietal area and inferior frontal gyrus (Binder and Desai, 2011) for general meaning processing (Pulvermüller and Fadiga, 2010), and also for category-specific activation of numbers and propositions (Dehaene, 1995; Shebani et al., 2017; Tschentscher et al., 2012), which were not implemented in the present model. Hence, an important step will be to incorporate peri- and extrasylvian parietal areas, extending the architecture to a total of 18 modelled cortical areas. Although the extension of the present architecture to such a sophisticated model will lead to additional insights, the possibility exists that the parietal cortex duplicates some of the action-related processes in the frontal lobe.

The present model demonstrates that while modality-preferential areas are involved in and of functional relevance for the processing of word meaning, the connector hub areas, which exhibited a similar degree of neurons for all semantic categories, acquired the function of a semantic hub for general meaning processing (Damasio and Damasio, 1994; Pulvermüller, 2013). Hence, the present simulation results demonstrate that conceptual knowledge is not processed uniquely in an amodal symbolic system (e.g., Ellis and Young, 1988; Fodor, 1983), but in part grounded in the real world, which is a necessary requirement for establishing the semantic link between word-form and its referent and the formation of distributed CA (see also, Harnad, 1990; Pulvermüller, 1999b). Cortical regions functioning as domain general semantic processing are also an indispensable part of the semantic system

having a specific functional role, namely to link the different modality-preferential sensorimotor information together. Once words based on the described mechanisms are acquired, they posit the basis for learning the meaning of novel words through combinatorial principles (e.g., from textbooks, see for more details Harnad, 2011; Stramandinoli et al., 2012, and conclusion section). Intriguingly, as predicted by the present simulations, the brain loci where combinatorial mechanisms might act, seems to be in the semantic hub regions, since these regions exhibited higher neuronal material than primary and secondary areas.

In summary, the present simulation work shows that word meaning is represented in the brain by cell assemblies spread out in a wide number of cortical areas, showing different functional roles for semantic processing as described above, speaking against for a purely amodal semantic representation in the human brain. Furthermore, the formation of distributed CA is the result of the semantic mapping between words and the objects and actions (referential semantic information) that are used to speak about by means of correlation learning and Hebbian synaptic plasticity. The present computational model offers a biological explanation of language and semantic processing in the human brain based on three neurobiological mechanisms acting together: network structure, neuroanatomical connectivity, and Hebbian associative learning.

Improved biological constrains

Neural networks have become useful tools that have been applied successfully in a broad range of higher cognitive processes, have provided a better understanding of neural functions and predicted the outcome of new evidence that was impossible to investigate with conventional techniques. Recently, neuroscientists have successfully managed to build large-scale neural networks in conjunction with neuroscience techniques in order to combine experimental results, brain theories and neurocomputational predictions (e.g. linking neural circuits with functional systems) as well as deepen their understanding of language learning processing in the human cortex. However, as mentioned in the introduction, most of these neural networks did not attempt to replicate the neuroanatomy of the regions in question (M H Christiansen and Chater, 2001; Dell et al., 1999; Elman, 1996; Plaut and Gonnerman, 2000; Plunkett, 1997), and if some biological constraints where

included (i.e., connectivity structure, Chen et al., 2017; Guenther et al., 2006; Husain et al., 2004; Ueno et al., 2011), they have not applied learning mechanisms that mimic well-documented neurophysiological phenomena (Braitenberg and Schüz, 1998; Mazzoni et al., 1991; O'Reilly, 1998). Although Chen et al. (2017) and Ueno et al. (2011) have implemented a semantic model based on the hub-and-spoke theoretical approach incorporating brain constrained connectivity structure, they included just one area, the anterior temporal lobe (ATL), of the multiple semantic hubs revealed by recent experimental studies (Binder and Desai, 2011; Pulvermüller, 2013), as already mentioned previously. Moreover, Ueno et al. (2011) have set the cortical locus of semantics in the ATL area a priori, rather than explaining it based on neuroscientific principles. '… we implemented the vATL semantic system alone. Specifically, it was set to generate semantic outputs for comprehension and provided the semantic input for speaking/naming' (p. 393, Ueno et al., 2011). Such an approach does not provide any insights into the brain loci of word meaning processing in the brain. Hence, it is essential to building neurobiologically realistic models, which closely mimic neuroanatomical structures and neurophysiological characteristics of the cortex, to investigate the biological principles that governs the neural system of the human brain (i.e., Hebbian learning, neural plasticity, associative learning, adaptations, self-organization behaviour), and importantly that the functional cortical role of the different linguistic functions (phonological, syntactical and semantic) are not set a priori (e.g., Dell et al., 1999; Ueno et al., 2011).

This was successfully applied in the present research in the domain of language processing and word meaning acquisition in the brain by implementing two variants of the same neuronal architecture, simulating the same set of fronto-temporal-occipital cortical areas at two different levels of biological detail.

(i) One version adopted a mean-field approach, in which the activity of each graded-response neuron represented the average activity of a cluster of cortical neurons (i.e, communication between neurons are based on continues firing rate value, Eggert and van Hemmen, 2000) and included only a subset of the cortico-cortical connections known to exist between the modelled areas (Fig. 5.1A).

(ii) A second variant building upon the mean-field model used adaptation-based spiking cells (versions of leaky integrate-and-fire neurons) each thought to represent a single

pyramidal neuron (Matthews, 2001) and a more complex brain-like connectivity structure based on neuroanatomical evidence (Fig. 5.1B).

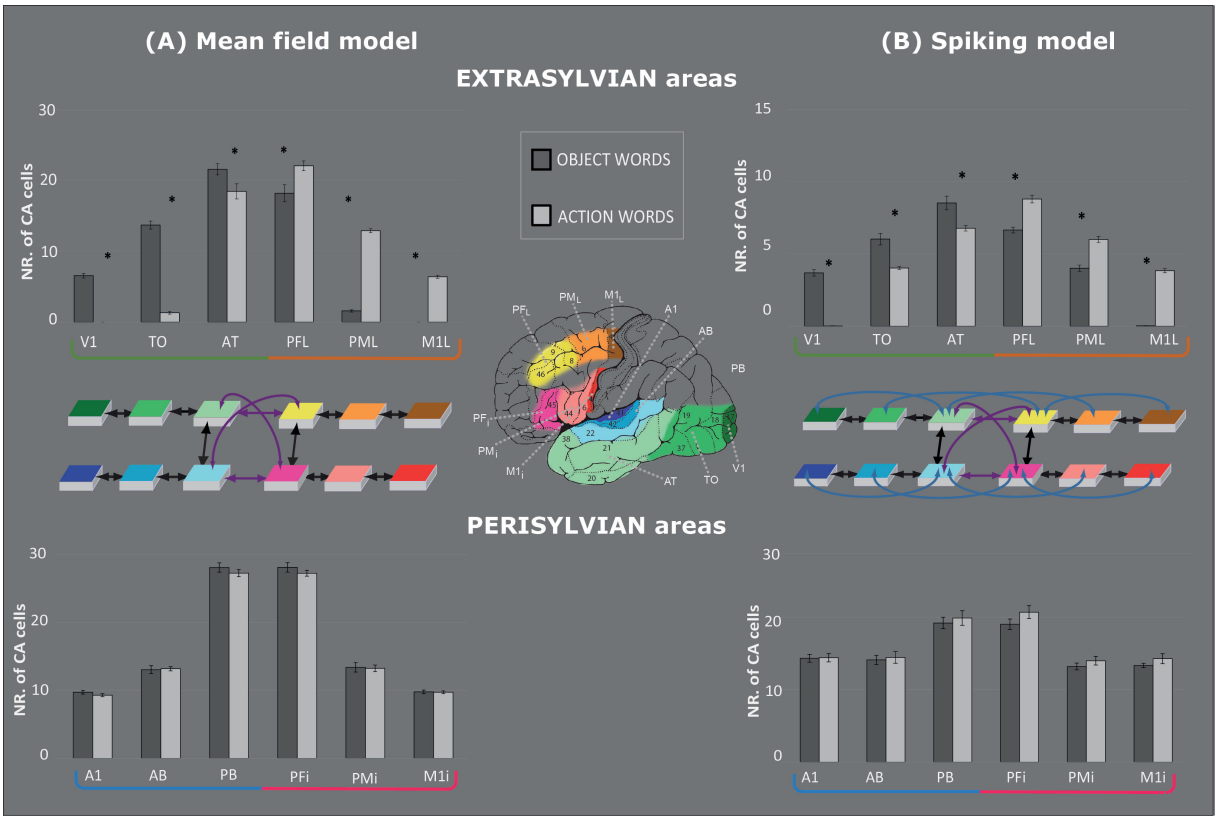

Figure 5.1. Average distributions of CAs emerging in both mean-field model (A) and spiking model (B) during simulation of word learning in the semantic context of actions and visual perceptions. Bars show average numbers of CA neurons per area. Both models show category-specificity in the cortical distribution of word-related circuits, and high-degree connection hub areas central to the network architecture exhibit all types of semantic processing with moderate category specificity. Asterisks indicate that, within a given area, the number of CA cells significantly differed between the circuits of action and object words (Bonferroni-corrected planned comparison tests).

Improving the biological realism of the neural model was motivated by the fact that some of the conclusions revealed by the simple mean-field network might be due to the basic neural behaviour or/and the simplified connectivity structure. For instance, the formation of neural circuits within the mean-field neural network is facilitated by the higher and continuous firing rate of the neurons during learning, which in the case of spiking neurons is much lower. Secondly, by implementing the full set of connectivity know to exist between the modelled cortical areas in the basic mean-field model lead to an over-activation problem, hence limiting the possibility to create more sophisticated cortex model. Therefore, it was essential for obtaining a better understanding of higher cognitive functions, such language and semantic processing, to implement more sophisticated and realistic neural architecture, which is agreed upon (Breakspear, 2017; Pezzulo et al., 2013).

Neurons within the improved neural architecture interact now primarily with action potential (i.e., spikes) to encode neuronal information, similarly documented in realistic biological neurons in the nervous systems of humans. A recent simulation work done by us has shown the importance of synchronous oscillatory spiking activity within cell assembly circuits for the binding of phonological and semantic information (Garagnani et al., 2017). Besides, the model incorporates now high-order 'jumping' links within perisylvian and extrasylvian systems, which have been defined essential for the formation of verbal working memory (Schomers et al., 2017). Interestingly, the improved spiking model revealed similar results as the mean-field architecture on the distribution of CA cells (see Fig. 5.1). Both models consistently show that verbal utterances (word-forms) encoded in the perisylvian language areas are thus linked with the semantically-related action and object information manifested in motor and visual cortices, with multimodal hub regions processing of all types of words. The lexico-semantic circuits distributed across the network regions showing different functional role emerged spontaneously as a result of learning, in contrast to previous connectionist models that have set the cortical locus of linguistic functions a priory (e.g., Dell et al., 1999; Ueno et al., 2011). Here we explain by means of a biologically constrained model at different levels of detail *why* and *how* some areas are more specialised in general semantic meaning processing, and others are more sensitive to specific semantic categories.

Apart from reproducing similar CA distributions in the spiking model, several interesting differences have been observed between the two neural models. For instance, the presence of the jumping links in the spiking model leads to a higher density of CA cells for object- and action-related words to the secondary non-relevant areas (TO for action and PML for object words) compared to the mean-field model (Fig. 5.1). Note also the CA size (the average number of CAs across the network areas as a result of learning) is approximately 50% less in the spiking neural architecture (CA=~75 cells on average) than in the mean-field model (CA=~163 cells on average). This suggests a better memory performance of the spiking model, as more cells are available for acquiring a larger number of lexicons. Interestingly, by visual observation, the spiking model shows an explosion-like activation during the initial learning phase (not present in the mean-field model), which seems to be related to a first period of synaptic elimination (or the so-called 'pruning' phase) that subsequently facilitates the rendering of the remaining synaptic circuits that are frequently activated. This experience-dependent plasticity, similarly reported during early stages of the developmental period in infants, has been defined crucial for the fine-tuning of functional networks, such as for language and general brain development (e.g., Blakemore and Choudhury, 2006). Further simulations could explore more closely the Hebbian plasticity behaviours and their synaptic modifications (LTP and LTD mechanisms) between pre- and postsynaptic spiking cell during learning, which might show a synaptic transmission similarly induced by the novel spike-timing-dependent plasticity[2] paradigma. Apart from this, intriguing differences have also been revealed during the neurophysiological word recognition and comprehension processes described in detail below. Importantly, it needs to be emphasized that only the combined improvement of neuroanatomical and neurophysiological (spiking) realism lead to a functional neural network with the spontaneous emergence of word-related circuits during learning. Additional simulations by omitting the jumping links from the network lead to an under-activation problem (i.e., low neural firing). Hence, the jumping links in the spiking model played an important role for the spreading of spiking activity during word learning, in turn, for the formation of cell assemblies.

[2] LTD and LTP are induced by the order and temporal interval between pre- and postsynaptic spikes (for more details see Dan and Poo, 2004).

Neurophysiological responses underlying word recognition

The neural network model not only firmly replicated the diverging experimental data and offered a unified explanatory account for the emergence of both category-specific and general semantic processing, but it also made precise and crucial predictions on the activation time course of the implemented cortical areas of the model. In other words, the model was able to predict not only *where* in the brain semantic processing emerge, but also *when* in time these processes take place. Here we applied the model to simulate EEG/MEG responses of the learnt object and action words during auditory word recognition. The primary auditory (A1) area of the model was stimulated with the learned acoustic component of the word-related CA circuits, which in turn lead to the full CA ignition. Apart from reproducing the topographical distribution of words with object- and action-related meaning in the action and perception system, similarly documented in a recent MEG study (Moseley et al., 2013), the neural network predicted a serial activation of the perisylvian cortices with overlapping activation of the hub regions (AT and PF_L), followed by the modality-preferential areas. By directly comparing the neurophysiological simulation responses with brain data (Fig. 2.8, Pulvermüller et al., 2005), we observed a great degree of consistency with the time course activation from superior temporal to inferior frontal areas and finally dorsal action- or visually-related regions.

Notice that a systematic activation time course analysis was performed only with the mean-field model approach (Chapter 2). Further simulation studies should investigate the neurophysiological responses underlying word recognition of the more biological constrained spiking model with brain-like connectivity (Chapter 3). Nevertheless, by visual observation of the CA dynamics between the simple mean-field model and the fully connected spiking model (Fig. 2.4 and Fig. 3.3), massive differences in the activation time courses were identified. As mentioned above, the mean-field model exhibited more a serial activation dynamics in the primary and secondary areas with overlapping hub regions activation, while the spiking model seems to shows a more simultaneous/cascade activation of the whole set of semantic brain areas implemented; ~ 15 simulation time-steps from perception to reverberation instead of the ~ 35 time-steps exhibited by the mean-field model. Interestingly, the spiking model exhibited the presence of the three-phase cognitive processes of perception, full ignition, and working memory. After full ignition, reverberant

activity is maintained in the semantic hub areas of the model, which is a consequence of the higher degree of CA cells emerged there during learning. These neural/cognitive processes of word recognition, defined as the primary stages of human memory processing, predicts an 'anterior shift' from sensorimotor areas to adjacent semantic hub regions contributing to working memory (e.g., Fuster, 2009; Pulvermüller, 2018; Pulvermüller and Garagnani, 2014) that was not present in the mean-field architecture.

In summary, the present simulation offers a mechanistic explanation of the current dispute over the different semantic theories of amodal symbolic system or grounded approach for meaning processing and in particular on the cortical locus of the semantic system in the brain. In particular, based on correlation learning, neuroanatomical and connectivity structure, the neural network model showed the spontaneous emergence of neural circuits in primary, secondary, and multimodal hub areas— regions observed to be active in a range of experimental studies about semantic processing. These results were consistently provided by a basic computational model using a mean-field approach and simple connectivity structures (Chapter 2) as well as a more realistic model of the cortex using adaptation-based spiking cells and brain-like connectivity (Chapter 3). Furthermore, the reactivation of word-related circuits provided novel predictions on the temporal dynamics of the cortical areas of the model, which could be used to guide future experimental work in the field.

Visual system recruitment for language processing following sensory loss

The intrinsic interaction of realistic neurophysiological learning mechanisms, connectivity and neuroanatomy structure are able not only to explain conceptual encoding and processing in the human brain (described above) but can also reproduce and explain the mechanisms underlying neural plastic change of the language system that takes place in congenitally blind people. A substantial number of neurocognitive studies in blind individuals have shown the recruitment of the visual system for semantic processing, in particular for verb generation tasks (Amedi et al., 2004, 2003, Burton, 2003, 2002; Raz et al., 2005; Struiksma et al., 2011), single word (Burton, 2003; Burton et al., 2012) and sentence processing (Bedny et al., 2011; Röder et al., 2002).

In this research work, we applied a neurobiologically constrained model of spiking neurons in human cortical function (introduced in Chapter 3) to describe the putative neural mechanisms underlying word learning at the cellular/synaptic level under visually deprived condition. As in previous simulations, object- and action-related word learning were simulated under the impact of repeated sensorimotor patterns in the primary areas of the model but without sensory experience. This was meant to simulate learning situations of word meaning acquisition in the absence of any visual input (i.e., blindness). As a consequence of Hebbian plasticity, distributed CA circuits spontaneously emerged across the network areas linking word-form data with semantic information. Intriguingly, by comparing blind and sighted models, only in the deprived architecture, neurons of the deprived visual areas (V1) were recruited for linguistic and semantic processing. In particular, the blind models produced word-related neuronal circuits extending into the visual cortex for all semantic categories (more action- than object-related circuits). In particular, the visual cortices (to which no input was given during the entire learning phase), exhibited a similar dissociation between the two word types, as in the motor system, with higher density of CAs for action compared to object words in the primary (V1), secondary (TO) and central (AT) areas of the visual cortex. Whereas in the undeprived simulations, only words denoting visual entities grew into the visual domain (see Fig 5.2). These simulation results are in line with the range of neuroimaging and TMS studies on blind people that have documented the functional involvement of the visual cortex (including V1) during semantic retrieval in a verb generation task (Amedi et al., 2004, 2003, Burton, 2003, 2002; Raz et al., 2005; Struiksma et al., 2011).

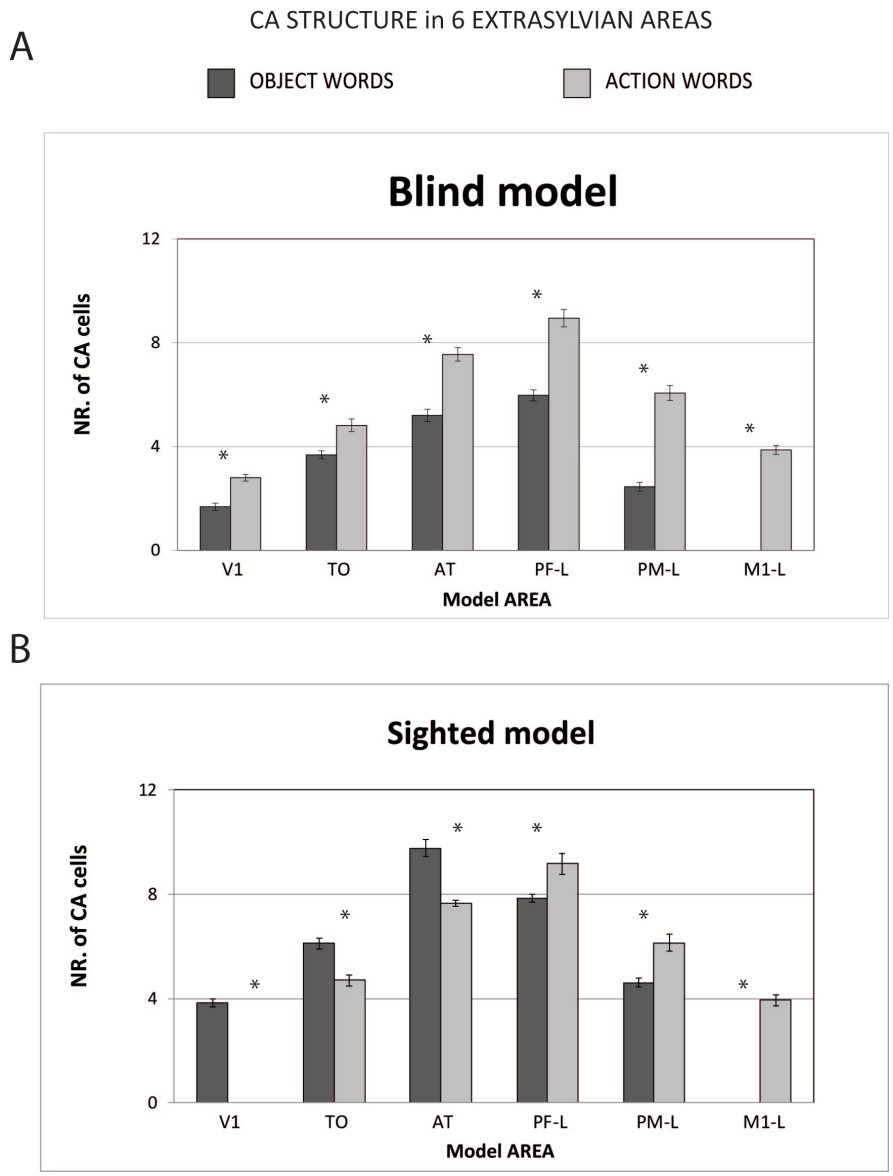

Figure 5.2 Mean numbers of cell assembly neurons in the extrasylvian areas after simulating the learning of action- (light grey) and object-related words (dark grey) during word production in blind (A) and sighted (B) models; error bars show standard errors over networks. Simulated word production (simultaneous presentation of articulatory-auditory patterns in A1 and M1i areas) after word meaning acquisition. The blind model (A) shows higher density of CA cells for both word types (especially for action words) in the primary visual system (V1), which was never stimulated during learning, while higher number of CA cells in the visual regions were found only for object related words in the sighted model (B). Asterisks indicate that, within a given area, the number of CA cells significantly differed between the circuits of action and object words (Bonferroni-corrected planned comparison tests, 6 comparisons; critical threshold $p<.0084$).

The first significant finding of this work is that the biologically constrained spiking network is able to replicate the experimental evidence for the visual area's recruitment under sensory deprivation for language and semantic processing. However, as often mentioned in this dissertation, the advantage of applying neural networks is the ability not only to replicate experimental data but also to understand the mechanisms and their interactions behind complex cognitive functions. Specifically, for the present work, it allows us to consider *why* and *how* neural organisation emerges in the brain as a consequence of blindness. Apart from the spontaneous emergence of CA circuits across the network areas based on the mutual interaction of Hebbian plasticity and connectivity structure, the observed 'CA overgrowth' into the deprived visual areas is the direct consequence of two biological principles/mechanisms acting together:

(i) The 'Doursat-Bienenstock´ expansion: A neurobiological mechanisms that give rise to the formation of strongly connected assemblies of cells extending into adjacent/connected cortical areas (Doursat and Bienenstock, 2006). In other words, neurons repeatedly and constantly activated tend to strengthen their connections, forming the so-called CAs by means of Hebbian learning mechanism (Hebb, 1949), and if continuously stimulated, they tend to extend into linked cortical regions of the brain, which has been defined as a principle of self-organization.

(ii) The absence of uncorrelated neural input to the deprived regions, which under healthy conditions is critical for blocking the excessive neural extensions and importantly, for the formation of semantic neural circuits with category-specific signatures.

Here, we propose that by means of these two aforementioned mechanisms acting within specific neuroanatomical structures can explain the relatively stronger activation of the visual system in blind individuals. In particular, the changes in activity balance is due to the absence of uncorrelated input in areas typically receiving sensorimotor information, which enables the spontaneous extension of CAs into the deprived areas and in turn, to the functional recruitment for language and semantic processing. Interestingly, the additional neural recruitment in the deprived visual system in the blind model exhibited a prolonged neural activation during auditory word recognition processes compared to the sighted (control) model (Fig. 4.5). Sustained activity has been often related to neural correlates of

working memory (Baddeley and Hitch, 1974; Leavitt et al., 2017), in which the longer spiking neural activation of the blind model can be seen as a sign for better working memory compared to the sighted populations, consistent with a number of neurocognitive studies (Amedi et al., 2003; Occelli et al., 2017; Pasqualotto et al., 2013; Withagen et al., 2013). Also, the present word recognition simulations of blind individuals revealed an anterior shift during reverberation activity from sensorimotor to frontal, temporal hub regions (Fuster, 1998; Leavitt et al., 2017; Pulvermüller and Garagnani, 2014) as already documented in previous simulations (Tomasello et al., 2018).

In contrast, under normal (i.e., sighted) conditions, uncorrelated input plays a vital role in preventing CA growth into such areas (Doursat and Bienenstock, 2006) and is an essential element for the formation of CA circuits with category-specific distributions (see Chapters 2 & 3). The present computational work offers a novel biological explanation for the neural changes following visual deprivation reported by numerous empirical studies, and it makes critical predictions on the role of the primary visual areas (V1) during semantic processing in blind people. The present results go one step further in the debate (Amedi et al., 2017; Bedny, 2017; Heimler et al., 2015) about the mechanisms behind the neural changes in the visual cortex, in which cortical areas can take over a particular function depending on information inputs received during the developmental period and/or because of the lack of competing inputs in deprived cortices. Here, we add that it is exactly the absence of informative (uncorrelated) input to the visual cortex that drives the Hebbian synaptic competitions (the strengthening or the weakening between connected cells) and consequent of the Doursat-Bienenstock CA extension of linguistic representations into the blind visual cortices.

Conclusions, limitations and future perspectives

The present thesis applied a neurobiologically constrained model with anatomical structure, neurophysiological function, and connectivity of the fronto-tempo-occipital lobes in order to investigate how word meaning is acquired, stored, and processed in sighted and blind populations. The findings of the present work support the hypothesis that neural correlates of semantic word types are represented in the brain by distributed CAs across both multimodal hub areas for general semantic processing and modality-preferential regions for category-specific semantic processing. This study shows how by means of a single

neurobiologically constrained neural model can elucidate *how*, *when* and *where* semantic knowledge is acquired, processed, and stored in the human brain and *how* semantic meaning is implemented at the cellular/synaptic level under deprived conditions. Below is a summary of the original contributions derived from the present research work:

(i) Two variants of a neurobiological model mimicking different cortical areas of the human brain to simulate word meaning acquisition in action and perception system. Chapter 2 introduces a mean-field model with a simple cortical connectivity structure, and Chapter 3 describes a fully connected network with realistic spiking neurons.

(ii) Based on neuroanatomical principles and Hebbian plasticity, both models consistently provide the same explanation for the nature of semantic processing, offering a solution to the debate on the functional role of semantic hubs and category-specific cortical regions during meaning processing.

(iii) A precise time course activation is offered (simulating EEG/MEG activation, Chapter 2) during auditory word recognition, which predicts a near simultaneous semantic cortical activation of the two learnt word types, comparable with experimental data.

(iv) A neuromechanistic explanation at the neural level of *how* and *why* the visual cortex is functionally recruited for linguistic and semantic processing in congenitally blind people is offered (Chapter 4);

It is important to emphasize that the experimental studies mentioned in the present work used natural language focusing on the action verb and object noun distinctions, which makes it difficult to control how and when these words have been acquired and to match for all the different psycholinguistic components between words of different lexical classes (for discussion see, Moseley and Pulvermüller 2014). Hence, to test the predictions and the validity of the neural network architecture, it might be essential to perform analogous learning experiments using fMRI or EEG/MEG techniques with high spatial and temporal resolutions. However, in overall the present computational model (i) offers a mechanistic explanation of how word meaning is acquired, stored and processed in the human brain (i.e., the formation of word circuits in the brain) by reconciling the diverging neurocognitive empirical evidence; (ii) provides a number of strong predictions, which can be easily tested

(for instance, the anterior shift in frontal and temporal lobes for working memory, the amount of activity expected to be observed in sensorimotor cortices during word comprehension is larger in hubs/secondary areas than in primary areas, a prolonged neural activity during word recognition in the blind compared to the sighted model); and (iii) is based on a small number of assumptions, all strongly grounded in well-documented neurophysiological principles and existing knowledge of brain neuroanatomy.

However, like any other neural networks, the model is simplified in a number of ways. For instance, at the microstructural level, the model does not include different ion channels, glial cells or neurotransmitters that are present in the cortex; at the macrostructural level, distinct connectivity density of the different cortical links between the network areas were not incorporated. As already mentioned above, cortical regions of the parietal lobe relevant for semantic processing (e.g., Binder and Desai, 2011) are missing in the present model, a valuable addition will be to include those areas in the model to further investigate the brain loci of meaning processing. Another simplification is that the learning of a word in the model requires hundreds of neural pattern stimulations. It is well known that humans are able to learn novel words by being exposed only one or a few times (so-called one-shot learning). It also exhibits the learning of a small number of words; humans are able to store more than 3,000 words.

Additionally, the present research work focused only on one type of word learning, in which the meaning of words are directly grounded in action and perception systems of the human brain. However, as already mentioned in the present work, many words are learned from textbooks or sentence context (i.e., acquired from indirect referential meaning). This type of word learning could be simulated by the simultaneous co-activation of previous emerged CA circuits of word-form and its referent by means of combinatorial mechanisms (see e.g., Harnad, 2011; Stramandinoli et al., 2012), which plays an important role in lexical acquisition. Similarly, also sentence processing could be simulated, in particular, once the model has learned the meaning of an initial lexicon, as described in this dissertation, associative links between already-learned linguistic representations may develop, due to co-activation of existing word circuits (e.g., Pulvermüller, 2010; Pulvermüller and Knoblauch, 2009). Because all words include numerous cells in the perisylvian hub regions (showing also prolonged activation), it might be that the binding between word-related circuits is best

mapped in these regions. In this way, combinatorial mechanisms, for instance, of adjectives-nouns, verb-nouns or phrases that are more complex, could be simulated and investigated in the different cortical areas of the model. Moreover, an important direction for further works would be to simulate referential learning of different categories of object or action-related words (e.g., Hauk et al., 2004). For instance, in order to simulate the learning of an action word, such as 'lick', whose referential meaning is related to articulators, firstly the model should be modified so that the patterns of motor semantics can also occur within the inferior, articulatory, motor areas. Within the mouth representation space of the model, for example, the cell groups controlling the 'licking' motion and those controlling the articulation of the word-form 'lick' should lie close to each other (and possibly overlap) in the same inferior motor areas. In this way, the brain loci of different types of action or object words can be further investigated.

A challenge of the present simulations work would be to explain how the learning of abstract words and their brain loci could be investigated within the computational model. While amodal symbolic system does not give a special status to abstract meaning, modal grounded approaches face the difficulties to explain how this type of meanings are grounded in sensorimotor regions of the brain (e.g., Mahon and Caramazza, 2008). The problem resides on the fact that while concrete words refer to a narrow set of objects, scene or actions in the real world, the mapping of abstract words with their sensorimotor information is not so straightforward. For instance, the word 'beauty' can refer to many different entities, having a 1-to-many mapping with different physical elements of the external world. Exactly such variability creates a problem in explaining the semantic link of abstract words. However, a solution to this problem has been proposed based on the so-called *family resemblance* theory (Wittgenstein, 1953), in which the meaning of abstract words might be constitute by partially overlapping neural sets shared among concrete words in the sensorimotor regions (Pulvermüller, 2018b, see Fig. 2). For example, the neural representation of the word 'beauty' would partially share its neuronal representation with all the concrete instances that have the characteristic of beauty, for example, blue eyes, panorama, or sunset. These mechanisms behind abstract word meaning acquisitions could be simulated by indirectly grounding abstract words in action and perception systems by associating them (with a degree of neural overlap) to previous sets of learned concrete

words (Della Rosa et al., 2010; Stramandinoli et al., 2012a). Hence these mechanisms would be sufficient enough for the formation of cell assemblies linking word-form to its semantic meaning in overlapping neural representations in modality-preferential cortical areas.

Although the neural network model can be improved in a number of ways, and further interesting simulation work can be conceived, the neural architecture in the present form is fully biologically constrained and it was sufficient enough to make critical predictions on the cortical locus of semantic processing by reconciling the diverging semantic theories and experimental data by means of a single neurocomputational model. The present research work provides a first step towards a better understanding of the biological mechanisms underlying language and semantic processing at the cortical-circuit level of the human brain under deprived and undeprived conditions. We believe that such an approach using biologically constrained computational models which follow precise neurobiological principles can also be used to provide neurobiological explanations for distinct cognitive functions apart from language and semantics.

References

Ackrill, J.L., 1963. Aristotle's Categories and De Interpretatione. Oxford Univ. Press.

Amedi, A., Floel, A., Knecht, S., Zohary, E., Cohen, L.G., 2004. Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. Nat. Neurosci. 7, 1266.

Amedi, A., Hofstetter, S., Maidenbaum, S., Heimler, B., 2017. Task selectivity as a comprehensive principle for brain organization. Trends Cogn. Sci. 21, 307–310.

Amedi, A., Raz, N., Pianka, P., Malach, R., Zohary, E., 2003. Early "visual" cortex activation correlates with superior verbal memory performance in the blind. Nat. Neurosci. 6, 758–766. doi:10.1038/nn1072

Amir, Y., Harel, M., Malach, R., 1993. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. J. Comp. Neurol. 334, 19–46. doi:10.1002/cne.903340103

Arikuni, T., Watanabe, K., Kubota, K., 1988. Connections of area 8 with area 6 in the brain of the macaque monkey. J. Comp. Neurol. 277, 21–40. doi:10.1002/cne.902770103

Artola, A., Bröcher, S., Singer, W., 1990. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature 347, 69–72. doi:10.1038/347069a0

Artola, A., Singer, W., 1993. Long-Term depression of excitatory synaptic transmission and its relationship to long-term potentiation. Trends Neurosci. 16, 480–487.

Baddeley, A., Hitch, G., 1974. Working memory. In Psychology of learning and motivation. Acad. Press 8, 47–89.

Bak, T.H., Chandran, S., 2012. What wires together dies together: Verbs, actions and neurodegeneration in motor neuron disease. Cortex 48, 936–944. doi:10.1016/j.cortex.2011.07.008

Barsalou, L.W., 2017. What does semantic tiling of the cortex tell us about semantics? Neuropsychologia 105, 18–38.

Barsalou, L.W., 2008. Grounded cognition. Annu. Rev. Psychol. 59, 617–645.

doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L.W., 1999. Perceptual symbol systems. Behav. Brain Sci. 22, 577-609; discussion 610–60. doi:10.1017/S0140525X99252144

Barsalou, L.W., Simmons, W.K., Barbey, A.K., Wilson, C.D., 2003. Grounding conceptual knowledge in modality-specific systems. Trends Cogn. Sci. 7, 84–91. doi:10.1016/S1364-6613(02)00029-3

Basso, a, Capitani, E., Laiacona, M., 1988. Progressive language impairment without dementia: a case with isolated category specific semantic defect. J. Neurol. Neurosurg. Psychiatry 51, 1201–1207. doi:10.1136/jnnp.51.9.1201

Bauer, R.H., Fuster, J.M., 1978. The effect of ambient illumination on delayed-matching and delayed-response deficits from cooling dorsolateral prefrontal cortex. Behav. Biol. 22, 60–66. doi:10.1016/S0091-6773(78)92019-9

Bauer, R.H., Jones, C.N., 1976. Feedback training of 36-45 Hz EEG activity in the visual cortex and hippocampus of cats: evidence for sensory and motor involvement. Physiol. Behav. 17, 885–890.

Bedny, M., 2017. Evidence from blindness for a cognitively pluripotent cortex. Trends Cogn. Sci. 21, 637–648.

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., Saxe, R., 2011. Language processing in the occipital cortex of congenitally blind adults. Proc. Natl. Acad. Sci. 108, 4429–4434.

Bibbig, A., Wennekers, T., Palm, G., 1995. A neural network model of the cortico-hippocampal interplay and the representation of contexts. Behav. Brain Res. 66, 169–175.

Bienenstock, E.L., Cooper, L.N., Munro, P.W., 1982. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J. Neurosci. 2, 32–48. doi:10.1371/journal.ppat.0020109

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends Cogn. Sci. 15, 527–536. doi:10.1016/j.tics.2011.10.001

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19, 2767–2796. doi:10.1093/cercor/bhp055

Blakemore, S.-J., Choudhury, S., 2006. Development of the adolescent brain: implications for executive function and social cognition. J. Child Psychol. Psychiatry 47, 296–312. doi:10.1111/j.1469-7610.2006.01611.x

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci 25, 151–88. doi:10.1146/annurev.neuro.25.112701.142946

Boulenger, V., Hauk, O., Pulvermüller, F., 2009. Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. Cereb. Cortex 19, 1905–1914. doi:10.1093/cercor/bhn217

Braitenberg, V., 1978. Cell assemblies in the cerebral cortex, in: Heim, R., Palm, G. (Eds.), Theoretical Approaches to Complex Systems. Springer, Berlin, pp. 171–188.

Braitenberg, V., Schüz, A., 1998. Cortex: Statistics and Geometry of Neuronal Connectivity. Springer, Berlin. doi:10.1007/978-3-662-03733-1_27

Breakspear, M., 2017. Dynamic models of large-scale brain activity. Nat. Neurosci. 20, 340.

Bressler, S., Kelso, J. a S., 2001. Cortical coordination dynamics. Trends Cogn. Sci. 5, 26–36.

Bressler, S.L., 2002. Understanding cognition through large-scale cortical networks. Curr. Dir. Psychol. Sci. 11, 58–61.

Bressler, S.L., Coppola, R., Nakamura, R., 1993. Episodic multiregional cortical coherence at multiple frequencies during visual task performance. Nature 366, 153–156. doi:10.1038/366153a0

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290. doi:10.1016/j.tics.2010.04.004

Broca, P., 1861. Remarques sur la siège de la faculté de la parole articulée, suivies d'une observation d'aphémie (perte de parole). Bull. la Société d'Anatomie 36, 330–357.

Brown, W.S., Lehmann, D., 1979. Verb and noun meaning of homophone words activate

different cortical generators: a topographic study of evoked potential fields. Exp. Brain Res. 2, 159–168.

Buonomano, D. V, Merzenich, M.M., 1998. Cortical plasticity: from synapses to maps. Annu. Rev. Neurosci. 21, 149–186. doi:10.1146/annurev.neuro.21.1.149

Burton, H., 2003. Dissociating Cortical Regions Activated by Semantic and Phonological Tasks: A fMRI Study in Blind and Sighted People. J. Neurophysiol. 90, 1965–1982. doi:10.1152/jn.00279.2003

Burton, H., 2002. Adaptive Changes in Early and Late Blind: A fMRI Study of Verb Generation to Heard Nouns. J. Neurophysiol. 88, 3359–3371. doi:10.1152/jn.00129.2002

Burton, H., Sinclair, R.J., Agato, A., 2012. Recognition memory for Braille or spoken words: an fMRI study in early blind. Brain Res. 1438, 22–34.

Burton, H., Snyder, A.Z., Raichle, M.E., 2014. Resting state functional connectivity in early blind humans. Front. Syst. Neurosci. 8, 51.

Butt, O.H., Benson, N.C., Datta, R., Aguirre, G.K., 2013. The fine-scale functional correlation of striate cortex in sighted and blind people. J. Neurosci. 33, 16209–16219.

Cangelosi, A., 2006. The grounding and sharing of symbols. Pragmat. Cogn. 14, 275–285.

Cangelosi, A., Greco, A., Harnad, S., 2002. Symbol grounding and the symbolic theft hypothesis, in: Simulating the Evolution of Language. Springer, pp. 191–210.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., Tani, J., Belpaeme, T., Sandini, G., Nori, F., 2010. Integration of action and language knowledge: A roadmap for developmental robotics. IEEE Trans. Auton. Ment. Dev. 2, 167–195.

Caramazza, A., Anzellotti, S., Strnad, L., Lingnau, A., 2014. Embodied cognition and mirror neurons: a critical assessment. Annu. Rev. Neurosci. 37, 1–15.

Carota, F., Kriegeskorte, N., Nili, H., Pulvermüller, F., 2017. Representational Similarity Mapping of Distributional Semantics in Left Inferior Frontal, Middle Temporal, and Motor Cortex. Cereb. Cortex. doi:10.1093/cercor/bhw379

Catani, M., Jones, D.K., Donato, R., Ffytche, D.H., 2003. Occipito-temporal connections in the human brain. Brain 126, 2093–2107. doi:10.1093/brain/awg203

Catani, M., Jones, D.K., Ffytche, D.H., 2005. Perisylvian language networks of the human brain. Ann. Neurol. 57, 8–16. doi:10.1002/ana.20319

Chafee, M. V, Goldman-Rakic, P.S., 2000. Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J. Neurophysiol. 83, 1550–1566.

Chao, L.L., Haxby, J. V, Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2, 913. doi:10.1038/13217

Chen, L., Lambon Ralph, M.A., Rogers, T.T., 2017. A unified model of human semantic knowledge and its disorders. Nat. Hum. Behav. 1, 0039. doi:10.1038/s41562-016-0039

Chen, R., Cohen, L.G., Hallett, M., 2002. Nervous system reorganization following injury. Neuroscience 111, 761–773.

Chomsky, N., 1965. Aspects of the theory of syntax. MIT Press. Cambridge, MA.

Christiansen, M.H., Chater, N., 2001. Connectionist psycholinguistics: Capturing the empirical data. Trends Cogn. Sci. 5, 82–88.

Christiansen, M.H., Chater, N., 2001. Connectionist psycholinguistics: The very idea, Connectionist psycholinguistics. Greenwood Publishing.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences, 2nd ed, Statistical power analysis for the behavioral sciences (rev. ed.). Erlbaum, Hillsdale. doi:10.1234/12345678

Collins, A.M., Loftus, E.F., 1975. Spreading-activation theory of semantic memory. Psychol. Rev. 82, 407–428. doi:10.1037/0033-295X.82.6.407

Connolly, A.C., Gleitman, L.R., Thompson-Schill, S.L., 2007. Effect of congenital blindness on the semantic representation of some everyday concepts. Proc. Natl. Acad. Sci. 104, 8241 LP-8246.

Connors, B.W., Gutnick, M.J., Prince, D.A., 1982. Electrophysiological properties of neocortical neurons in vitro. J. Neurophysiol. 48, 1302–1320.

D'Esposito, M., 2007. From cognitive to neural models of working memory. Proc. R. Soc.

London B Biol. Sci. 362, 761–772. doi:10.1098/rstb.2007.2086

Damasio, A.R., 1989. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. Cognition 33, 25–62. doi:10.1016/0010-0277(89)90005-X

Damasio, A.R., Damasio, H., 1994. Cortical systems for retrieval of concrete knowledge: The convergence zone framework, in: Koch, C., Davis, J.L. (Eds.), Large-Scale Neuronal Theories of the Brain. MIT Press, Cambridge, MA, pp. 61–74.

Damasio, A.R., Tranel, D., 1993. Nouns and verbs are retrieved with differently distributed neural systems. Proc. Natl. Acad. Sci. U. S. A. 90, 4957–60. doi:10.1073/pnas.90.11.4957

Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D., Damasio, A.R., 1996. A neural basis for lexical retrieval. Nature 380, 499–505. doi:10.1038/380499a0

Dan, Y., Poo, M., 2004. Spike timing-dependent plasticity of neural circuits. Neuron 44, 23–30.

Deacon, T.W., 1992. Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. Brain Res. 573, 8–26. doi:10.1016/0006-8993(92)90109-M

Dehaene, S., 1995. Electrophysiological evidence for category-specific word processing in the normal human brain. Neuroreport 6, 2153–7. doi:10.1097/00001756-199511000-00014

Deiber, M.P., Passingham, R.E., Colebatch, J.G., Friston, K.J., Nixon, P.D., Frackowiak, R.S.J., 1991. Cortical areas and the selection of movement: a study with positron emission tomography. Exp. Brain Res. 84, 393–402. doi:10.1007/BF00231461

Dell, G.S., 1986. A spreading-activation theory of retreival in sentence production. Psychol. Rev. 93, 283–321.

Dell, G.S., Chang, F., Griffiths, Z.M., 1999. Connectionist models of language production: lexical access and grammatical encoding. Cogn. Sci. 23, 517–542. doi:10.1016/S0364-0213(99)00014-2

Della Rosa, P.A., Catricalà, E., Vigliocco, G., Cappa, S.F., 2010. Beyond the abstract—concrete

dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. Behav. Res. Methods 42, 1042–1048.

Devlin, J.T., Matthews, P.M., Rushworth, M.F., 2003. Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. J. Cogn. Neurosci. 15, 71–84. doi:10.1162/089892903321107837

Distler, C., Boussaoud, D., Desimone, R., Ungerleider, L.G., 1993. Cortical connections of inferior temporal area TEO in macaque monkeys. J. Comp. Neurol. 334, 125–150. doi:10.1002/cne.903340111

Douglas, R.J., Martin, K.A.C., 2004. Neuronal Circuits of the Neocortex. Annu. Rev. Neurosci. 27, 419–451. doi:10.1146/annurev.neuro.27.070203.144152

Doursat, R., Bienenstock, E., 2006. Neocortical self-structuration as a basis for learning. 5th International conference on development and learning (ICDL) 2006. Indiana University, Bloomington, Indiana.

Dreyer, F.R., Frey, D., Arana, S., Saldern, S. von, Picht, T., Vajkoczy, P., Pulvermüller, F., 2015. Is the motor system necessary for processing action and abstract emotion words? Evidence from focal brain lesions. Front. Psychol. 6, 242. doi:10.3389/fpsyg.2015.01661

Dum, R.P., Strick, P.L., 2005. Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. J. Neurosci. 25, 1375–1386. doi:10.1523/JNEUROSCI.3902-04.2005

Dum, R.P., Strick, P.L., 2002. Motor areas in the frontal lobe of the primate. Physiol. Behav. 77, 677–682.

Duncan, J., 2006. EPS Mid-Career Award 2004: brain mechanisms of attention. Q. J. Exp. Psychol. (Hove). 59, 2–27. doi:10.1080/17470210500260674

Duncan, J., 1996. Competitive brain systems in selective attention. Int. J. Psychol. 31, 3343.

Eacott, M.J., Gaffan, D., 1992. Inferotemporal-frontal disconnection: The uncinate fascicle and visual associative learning in monkeys. Eur. J. Neurosci. 4, 1320–1332.

doi:10.1111/j.1460-9568.1992.tb00157.x

Eggert, J., van Hemmen, J.L., 2000. Unifying framework for neuronal assembly dynamics. Phys. Rev. E. Stat. Phys. Plasmas. Fluids. Relat. Interdiscip. Topics 61, 1855–1874. doi:10.1103/PhysRevE.61.1855

Ellis, A.W., Young, A.W., 1988. Human Cognitive Neuropsychology. Lawrence Erlbaum Associates Ltd., Hove, UK.

Elman, J., 1991. Distributed representation, simple recurrent networks, and grammatical structure. Mach. Learn. 7, 195–225.

Elman, J.L., 1996. Rethinking Innateness : A Connectionist Perspective On Development Neural Network Modeling and Connectionism. MIT Press, Cambridge, MA.

Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. Eur. J. Neurosci. 15, 399–402. doi:10.1046/j.0953-816x.2001.01874.x

Faisal, A.A., Selen, L.P.J., Wolpert, D.M., 2008. Noise in the nervous system. Nat. Rev. Neurosci. 9, 292–303. doi:10.1038/nrn2258

Farah, M.J., McClelland, J.L., 1991. A computational model of semantic memory impairment: modality specificity and emergent category specificity. J. Exp. Psychol. Gen. 120, 339–357. doi:10.1037/0096-3445.120.4.339

Fernandino, L., Humphries, C.J., Seidenberg, M.S., Gross, W.L., Conant, L.L., Binder, J.R., 2015. Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia 76, 17–26.

Finney, E.M., Fine, I., Dobkins, K.R., 2001. Visual stimuli activate auditory cortex in the deaf. Nat. Neurosci. 4, 1171.

Finnie, P.S.B., Nader, K., 2012. The role of metaplasticity mechanisms in regulating memory destabilization and reconsolidation. Neurosci. Biobehav. Rev. 36, 1667–1707. doi:10.1016/j.neubiorev.2012.03.008

Fodor, J.A., 1983. The modulatity of mind. MIT Press, Cambridge, MA.

Fodor, J.A., 1975. The Language of Thought. Harvard University Press. doi:10.1111/j.1468-

0149.1977.tb01728.x"

Friedman, D., Donoghue, J.P., 2009. Learning-Induced LTP in Neocortex. Science (80-.). 533, 533–536. doi:10.1126/science.290.5491.533

Fry, D., 1966. The development of the phonological system in the normal and deaf child, in: Smith, F., Miller, G.A. (Eds.), The Genesis of Language. MIT Press, Cambridge, MA, pp. 187–206.

Fuster, J.M., 2009. Cortex and memory: emergence of a new paradigm. J. Cogn. Neurosci. 21, 2047–2072. doi:10.1162/jocn.2009.21280

Fuster, J.M., 2003. Cortex and mind: Unifying cognition. Oxford university press.

Fuster, J.M., 1998. Distributed memory for both short and long term. Neurobiol. Learn. Mem. 70, 268–74. doi:10.1006/nlme.1998.3852

Fuster, J.M., Bauer, R.H., Jervey, J.P., 1985. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. Brain Res. 330, 299–307.

Fuster, J.M., Jervey, J.P., 1981. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. Science 212, 952–955. doi:10.1126/science.7233192

Gainotti, G., 2012. The format of conceptual representations disrupted in semantic dementia: A position paper. Cortex 48, 521–529. doi:http://dx.doi.org/10.1016/j.cortex.2011.06.019

Gainotti, G., 2010. The influence of anatomical locus of lesion and of gender-related familiarity factors in category-specific semantic disorders for animals, fruits and vegetables: A review of single-case studies. Cortex 46, 1072–1087. doi:10.1016/j.cortex.2010.04.002

Gallese, V., Lakoff, G., 2005. The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. Cogn. Neuropsychol. 22, 455–79. doi:10.1080/02643290442000310

Garagnani, M., Lucchese, G., Tomasello, R., Wennekers, T., Pulvermüller, F., 2017. A Spiking Neurocomputational Model of High-Frequency Oscillatory Brain Responses to Words and Pseudowords. Front. Comput. Neurosci. 10, 1–19. doi:10.3389/fncom.2016.00145

Garagnani, M., Pulvermüller, F., 2016. Conceptual grounding of language in action and perception: A neurocomputational model of the emergence of category specificity and semantic hubs. Eur. J. Neurosci. 43, 721–737. doi:10.1111/ejn.13145

Garagnani, M., Pulvermüller, F., 2013. Neuronal correlates of decisions to speak and act: Spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas. Brain Lang. 127, 75–85. doi:10.1016/j.bandl.2013.02.001

Garagnani, M., Pulvermüller, F., 2011. From sounds to words: A neurocomputational model of adaptation, inhibition and memory processes in auditory change detection. Neuroimage 54, 170–181. doi:10.1016/j.neuroimage.2010.08.031

Garagnani, M., Wennekers, T., Pulvermüller, F., 2009. Recruitment and consolidation of cell assemblies for words by way of hebbian learning and competition in a multi-layer neural network. Cognit. Comput. 1, 160–176. doi:10.1007/s12559-009-9011-1

Garagnani, M., Wennekers, T., Pulvermüller, F., 2008. A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. Eur. J. Neurosci. 27, 492–513. doi:10.1111/j.1460-9568.2008.06015.x

Garagnani, M., Wennekers, T., Pulvermüller, F., 2007. A neuronal model of the language cortex. Neurocomputing 70, 1914–1919. doi:10.1016/j.neucom.2006.10.076

Gaskell, M.G., Hare, M., Marslen-Wilson, W.D., 1995. A connectionist model of phonological representation in speech perception. Cogn. Sci. A Multidiscip. J. 19, 407–439.

Gierhan, S.M.E., 2013. Connections for auditory language in the human brain. Brain Lang. 127, 205–221. doi:10.1016/j.bandl.2012.11.002

Gilbert, C.D., Wiesel, T.N., 1983. Clustered intrinsic connections in cat visual cortex. J. Neurosci. 3, 1116–1133.

Gleitman, L., 1990. The structural sources of verb meanings. Lang. Acquis. 1, 3–55.

Grisoni, L., Dreyer, F.R., Pulvermüller, F., 2016. Somatotopic Semantic Priming and Prediction in the Motor System. Cereb. Cortex 26, 2353–2366. doi:10.1093/cercor/bhw026

Grisoni, L., Miller, T.M., Pulvermüller, F., 2017. Neural Correlates of Semantic Prediction and

Resolution in Sentence Processing. J. Neurosci. 37, 4848–4858. doi:10.1523/JNEUROSCI.2800-16.2017

Guenther, F.H., Ghosh, S.S., Tourville, J.A., 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 96, 280–301. doi:10.1016/j.bandl.2005.06.001

Guye, M., Parker, G.J.M., Symms, M., Boulby, P., Wheeler-Kingshott, C.A.M., Salek-Haddadi, A., Barker, G.J., Duncan, J.S., 2003. Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. Neuroimage 19, 1349–1360. doi:10.1016/S1053-8119(03)00165-4

Harnad, S., 2012. From sensorimotor catefories and pantomime to grounded symbols and propositions. Oxford Handb. Lang. Evol. 387–392.

Harnad, S., 2011. From sensorimotor categories and pantomime to grounded symbols and propositions.

Harnad, S., 1990. The symbol grounding problem. Phys. D 42, 335–346.

Hauk, O., Johnsrude, I., Pulvermüller, F., 2004. Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. Neuron 41, 301–307. doi:10.1016/S0896-6273(03)00838-9

Hauk, O., Pulvermüller, F., 2004. Neurophysiological Distinction of Action Words in the Fronto-Central Cortex. Hum. Brain Mapp. 21, 191–201. doi:10.1002/hbm.10157

Hauk, O., Shtyrov, Y., Pulvermüller, F., 2008. The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. J. Physiol. Paris 102, 50–58. doi:10.1016/j.jphysparis.2008.03.013

Hebb, D.O., 1949. The organization of behavior, John Wiley. New York.

Heimler, B., Striem-Amit, E., Amedi, A., 2015. Origins of task-specific sensory-independent organization in the visual and auditory brain: neuroscience evidence, open questions and clinical implications. Curr. Opin. Neurobiol. 35, 169–177. doi:https://doi.org/10.1016/j.conb.2015.09.001

Hernández, M., Costa, A., Juncadella, M., Sebastián-Gallés, N., Reñé, R., 2008. Category-

specific semantic deficits in Alzheimer's disease: A semantic priming study. Neuropsychologia 46, 935–946. doi:10.1016/j.neuropsychologia.2007.11.018

Hodges, J.R., Davies, R.R., Patterson, K., 2009. Semantic dementia. Behaivoral Neurol. Dement. 115, 264–278. doi:10.1017/CBO9780511581410.018

Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B., Kiefer, M., 2008. Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. J. Cogn. Neurosci. 20, 1799–1814. doi:10.1162/jocn.2008.20123

Husain, F.T., Tagamets, M.A., Fromm, S.J., Braun, A.R., Horwitz, B., 2004. Relating neuronal dynamics for auditory object processing to neuroimaging activity: A computational modeling and an fMRI study. Neuroimage 21, 1701–1720. doi:10.1016/j.neuroimage.2003.11.012

Huth, A.G., Heer, W.A. De, Griffiths, T.L., Theunissen, F.E., Jack, L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. doi:10.1038/nature17637

Huyck, C.R., Passmore, P.J., 2013. A review of cell assemblies. Biol. Cybern. 107, 263–288. doi:10.1007/s00422-013-0555-5

Izhikevich, E.M., Edelman, G., 2008. Large-scale model of mammalian thalamocortical systems. Proc. Natl. Acad. Sci. USA 105, 3593–3598. doi:10.1073/pnas.0712231105

Jackendoff, R., 1983. Semantics and cognition. MIT press.

Jirsa, V.K., 2004. Connectivity and dynamics of neural information processing. Neuroinformatics 2, 183–204. doi:10.1385/NI:2:2:183

Joanisse, M.F., Seidenberg, M.S., 1999. Impairments in verb morphology after brain injury: a connectionist model. Proc. Natl. Acad. Sci. U. S. A. 96, 7592–7597. doi:10.1073/pnas.96.13.7592

Kaas, J.H., 1997. Topographic Maps are Fundamental to Sensory Processing. Brain Res. Bull. 44, 107–112. doi:http://dx.doi.org/10.1016/S0361-9230(97)00094-4

Kaas, J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. U. S. A. 97, 11793–9. doi:10.1073/pnas.97.22.11793

Kandel, E.R., Schwartz, J.H., Jessell, T.M., 2000. Principles of neural science. McGraw-Hill New York.

Katz, J.J., Fodor, J.A., 1963. The structure of a semantic theory. Language (Baltim). 39, 170–210.

Keck, T., Mrsic-Flogel, T.D., Afonso, M.V., Eysel, U.T., Bonhoeffer, T., Hübener, M., 2008. Massive restructuring of neuronal circuits during functional reorganization of adult visual cortex. Nat. Neurosci. 11, 1162.

Kemmerer, D., 2015. Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. Psychon. Bull. Rev. 22, 1068–1075. doi:10.3758/s13423-014-0784-1

Kemmerer, D., 2014. Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. Cortex 58, 27–51. doi:http://dx.doi.org/10.1016/j.cortex.2014.05.004

Kemmerer, D., Rudrauf, D., Manzel, K., Tranel, D., 2012. Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. Cortex 48, 826–848. doi:10.1016/j.cortex.2010.11.001

Kiefer, M., 2005. Repetition-priming modulates category-related effects on event-related potentials: further evidence for multiple cortical semantic systems. J. Cogn. Neurosci. 17, 199–211. doi:10.1162/0898929053124938

Kiefer, M., Pulvermüller, F., 2012. Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. Cortex 48, 805–825. doi:10.1016/j.cortex.2011.04.006

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., Hoenig, K., 2008. The sound of concepts: four markers for a link between auditory and conceptual brain systems. J. Neurosci. 28, 12224–12230. doi:10.1523/JNEUROSCI.3579-08.2008

Kujala, T., Alho, K., Huotilainen, M., Ilmoniemi, R.J., Lehtokoski, A., Leinonen, A., Rinne, T., Salonen, O., Sinkkonen, J., Standertskj??ld-Nordenstam, C.G., N????t??nen, R., 1997. Electrophysiological evidence for cross-modal plasticity in humans with early- and late-

onset blindness. Psychophysiology 34, 213–216. doi:10.1111/j.1469-8986.1997.tb02134.x

Lakoff, G., 1988. Cognitive semantics. Mean. Ment. Represent. 119, 154.

Lambon Ralph, M.A., Lowe, C., Rogers, T.T., 2007. Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. Brain 130, 1127–1137.

Leavitt, M.L., Mendoza-Halliday, D., Martinez-Trujillo, J.C., 2017. Sustained Activity Encoding Working Memories: Not Fully Distributed. Trends Neurosci. 40, 328–346. doi:https://doi.org/10.1016/j.tins.2017.04.004

Leshinskaya, A., Caramazza, A., 2016. For a cognitive neuroscience of concepts: Moving beyond the grounding issue. Psychon. Bull. Rev. 23, 991–1001.

Lichtheim, L., 1885. On aphasia. Brain 7, 433–484.

Lu, M.T., Preston, J.B., Strick, P.L., 1994. Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J. Comp. Neurol. 341, 375–392. doi:10.1002/cne.903410308

Machery, E., 2007. Concept empiricism: A methodological critique. Cognition 104, 19–46.

Mahon, B.Z., Caramazza, A., 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J. Physiol. Paris 102, 59–70. doi:10.1016/j.jphysparis.2008.03.004

Makris, N., Pandya, D.N., 2009. The extreme capsule in humans and rethinking of the language circuitry. Brain Struct. Funct. 213, 343–358. doi:10.1007/s00429-008-0199-8

Malenka, R.C., Bear, M.F., 2004. LTP and LTD: An embarrassment of riches. Neuron 44, 5–21.

Markram, H., Meier, K., Lippert, T., Grillner, S., Frackowiak, R., Dehaene, S., Knoll, A., Sompolinsky, H., Verstreken, K., DeFelipe, J., Grant, S., Changeux, J.P., Sariam, A., 2011. Introducing the Human Brain Project. Procedia Comput. Sci. 7, 39–42. doi:10.1016/j.procs.2011.12.015

Marmor, G.S., 1978. Age at onset of blindness and the development of the semantics of color names. J. Exp. Child Psychol. 25, 267–278.

Martin, A., 2016. GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. Psychon. Bull. Rev. 23, 979–990.

Martin, A., 2007. The Representation of Object Concepts in the Brain. Annu. Rev. Psychol. 58, 25–45. doi:10.1146/annurev.psych.57.102904.190143

Martin, A., Chao, L.L., 2001. Semantic memory and the brain: Structure and processes. Curr. Opin. Neurobiol. 11, 194–201. doi:10.1016/S0959-4388(00)00196-3

Martin, A., Wiggs, C.L., Ungerleider, L.G., Haxby, J.V., 1996. Neural correlates of category-specific knowledge. Nature 379, 649–652. doi:10.1038/379649a0

Matthews, G.G., 2001. Neurobiology: Molecules, Cells, and Systems, Blackwell Science.

Mazzoni, P., Andersen, R.A., Jordan, M.I., 1991. A more biologically plausible learning rule for neural networks. Proc. Natl. Acad. Sci. 88, 4433–4437.

McCarthy, R.A., Warrington, E.K., 1988. Evidence for modality-specific meaning systems in the brain. Nature 334, 428–430. doi:10.1038/334428a0

McCrory, E., Frith, U., Brunswick, N., Price, C., 2000. Abnormal functional activation during a simple word repetition task: A PET study of adult dyslexics. J. Cogn. Neurosci. 12, 753–62. doi:10.1162/089892900562570

McCulloch, W.S., Pitts, W., 1943. A Logical Calculus of the Idea Immanent in Nervous Activity. Bull. Math. Biophys. 5, 115–133. doi:10.1007/BF02478259

McIntosh, A.R., 2000. Towards a network theory of cognition. Neural Networks 13, 861–870. doi:10.1016/S0893-6080(00)00059-9

McNorgan, C., Reid, J., McRae, K., 2011. Integrating conceptual knowledge within and across representational modalities. Cognition 118, 211–233. doi:10.1016/j.cognition.2010.10.017

Merzenich, M.M., Nelson, R.J., Stryker, M.P., Cynader, M.S., Schoppmann, A., Zook, J.M., 1984. Somatosensory cortical map changes following digit amputation in adult monkeys. J. Comp. Neurol. 224, 591–605. doi:10.1002/cne.902240408

Meyer, J.W., Makris, N., Bates, J.F., Caviness, V.S., Kennedy, D.N., 1999. MRI-Based

topographic parcellation of human cerebral white matter. Neuroimage 9, 1–17. doi:S1053811998903834 [pii]

Miceli, G., Silveri, M.C., Nocentini, U., Caramazza, A., 1988. Patterns of dissociation in comprehension and production of nouns and verbs. Aphasiology 2, 351–358. doi:10.1080/02687038808248937

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y.T., Fryer, T.D., Williams, G.B., Hodges, J.R., Nestor, P.J., 2010. What the left and right anterior fusiform gyri tell us about semantic memory. Brain 133, 3256–3268. doi:10.1093/brain/awq272

Mishkin, M., Ungerleider, L.G., 1982. Contribution of striate input to the visuospatial functions of parieto-preoccipital cortx in monkeys . Behav. Brain Res. 6, 57–77.

Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: Two central pathways. Trends Neurosci. 6, 414–417. doi:Doi 10.1016/0166-2236(83)90190-X

Moseley, R.L., Pulvermüller, F., 2014. Nouns, verbs, objects, actions, and abstractions: Local fMRI activity indexes semantics, not lexical categories. Brain Lang. 132, 28–42. doi:10.1016/j.bandl.2014.03.001

Moseley, R.L., Pulvermüller, F., Shtyrov, Y., 2013. Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. Sci. Rep. 3, 1928. doi:10.1038/srep01928

Neininger, B., Pulvermüller, F., 2003. Word-category specific deficits after lesions in the right hemisphere. Neuropsychologia 41, 53–70. doi:10.1016/S0028-3932(02)00126-4

Neville, H., Bavelier, D., 2002. Human brain plasticity: Evidence from sensory deprivation and altered language experience. Prog. Brain Res. 138, 177–188. doi:10.1016/S0079-6123(02)38078-6

Neville, H.J., Bavelier, D., 1998. Neural organization and plasticity of language. Curr. Opin. Neurobiol. 8, 254–258. doi:10.1016/S0959-4388(98)80148-7

Noppeney, U., Friston, K.J., Ashburner, J., Frackowiak, R., Price, C.J., 2005. Early visual

deprivation induces structural plasticity in gray and white matter. Curr. Biol. 15, R488–R490.

Norris, D., 1994. Shortlist - a Connectionist Model of Continuous Speech Recognition. Cognition 52, 189–234. doi:10.1016/0010-0277(94)90043-4

O'Reilly, R.C., 1998. Six principles for biologically based computational models of cortical cognition. Trends Cogn. Sci. 2, 455–462. doi:10.1016/S1364-6613(98)01241-8

Occelli, V., Lacey, S., Stephens, C., Merabet, L.B., Sathian, K., 2017. Enhanced verbal abilities in the congenitally blind. Exp. Brain Res. 235, 1709–1718. doi:10.1007/s00221-017-4931-6

Ogden, C.K., Richards, I.A., 1923. The meaning of meaning. Harvest Book.

Palm, G., 1982. Neural Assemblies. An Alternative Approach to Artificial Intelligence. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Palm, G., Knoblauch, A., Hauser, F., Schüz, A., 2014. Cell assemblies in the cerebral cortex. Biol. Cybern. 108, 559–572. doi:10.1007/s00422-014-0596-4

Pandya, D.N., 1995. Anatomy of the auditory cortex. Rev. Neurol. (Paris). 151, 486–494. doi:10.1016/B978-0-323-05283-2.00129-4

Pandya, D.N., Barnes, C.L., 1987. Architecture and connections of the frontal lobe, in: Perecman, E. (Ed.), The Frontal Lobes Revisited. The IRBN Press, New York, pp. 41–72.

Pandya, D.N., Yeterian, E.H., 1985. Architecture and Connections of Cortical Association Areas, in: Peters, A., Jones, E. (Eds.), Association and Auditory Cortices SE - 1, Cerebral Cortex. Springer US, pp. 3–61. doi:10.1007/978-1-4757-9619-3_1

Parker, A., 1998. Interaction of frontal and perirhinal cortices in visual object recognition memory in monkeys. Eur. J. Neurosci. 10, 3044–3057. doi:10.1046/j.1460-9568.1998.00306.x

Parker, G.J.M., Luzzi, S., Alexander, D.C., Wheeler-Kingshott, C.A.M., Ciccarelli, O., Lambon Ralph, M.A., 2005. Lateralization of ventral and dorsal auditory-language pathways in the human brain. Neuroimage 24, 656–666. doi:10.1016/j.neuroimage.2004.08.047

Pasqualotto, A., Lam, J.S.Y., Proulx, M.J., 2013. Congenital blindness improves semantic and

episodic memory. Behav. Brain Res. 244, 162–165.

Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–87. doi:10.1038/nrn2277

Paus, T., Castro-Alamancos, M.A., Petrides, M., 2001. Cortico-cortical connectivity of the human mid-dorsolateral frontal cortex and its modulation by repetitive transcranial magnetic stimulation. Eur. J. Neurosci. 14, 1405–1411. doi:10.1046/j.0953-816X.2001.01757.x

Penolazzi, B., Hauk, O., Pulvermüller, F., 2007. Early semantic context integration and lexical access as revealed by event-related brain potentials. Biol. Psychol. 74, 374–388. doi:10.1016/j.biopsycho.2006.09.008

Petitto, L.A., Zatorre, R.J., Gauna, K., Nikelski, E.J., Dostie, D., Evans, A.C., 2000. Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. Proc. Natl. Acad. Sci. U. S. A. 97, 13961–6. doi:10.1073/pnas.97.25.13961

Petrides, M., Pandya, D.N., 2009. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. PLoS Biol. 7, e1000170. doi:10.1371/journal.pbio.1000170

Pezzulo, G., Barsalou, L.W., Cangelosi, A., Fischer, M.H., McRae, K., Spivey, M., 2013. Computational grounded cognition: a new alliance between grounded cognition and computational modeling. Front. Psychol. 3, 612.

Plaut, D.C., Gonnerman, L.M., 2000. Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? Lang. Cogn. Process. 15, 445–485. doi:10.1080/01690960050119661

Plunkett, K., 1997. Theories of early language acquisition. Trends Cogn. Sci. 1, 146–153. doi:10.1016/S1364-6613(97)01039-5

Plunkett, K., Marchman, V.A., 1996. Learning from a connectionist model of the acquisition of the English past tense. Cognition 61, 299–308. doi:10.1016/S0010-0277(96)00721-4

Posner, M.I., Pavese, a, 1998. Anatomy of word and sentence meaning. Proc. Natl. Acad. Sci. U. S. A. 95, 899–905. doi:10.1073/pnas.95.3.899

Preissl, H., Pulvermüller, F., Lutzenberger, W., Birbaumer, N., 1995. Evoked potentials distinguish between nouns and verbs. Neurosci. Lett. 197, 81–83. doi:10.1016/0304-3940(95)11892-Z

Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J. Anat. 197 Pt 3, 335–359. doi:10.1046/j.1469-7580.2000.19730335.x

Pulvermüller, F., 2018a. Neural reuse of action perception circuits for language, concepts and communication. Prog. Neurobiol. 160, 1–44.

Pulvermüller, F., 2018b. The case of CAUSE: neurobiological mechanisms for grounding an abstract concept. Philos. Trans. R. Soc. B Biol. Sci. 373, 20170129.

Pulvermüller, F., 2013. How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. Trends Cogn. Sci. 17, 458–470. doi:10.1016/j.tics.2013.06.004

Pulvermüller, F., 2010. Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. Brain Lang. 112, 167–179. doi:10.1016/j.bandl.2009.08.002

Pulvermüller, F., 2005. Brain mechanisms linking language and action. Nat. Rev. Neurosci. 6, 576–582. doi:10.1038/nrn1706

Pulvermüller, F., 2002. The neuroscience of language: On brain circuits of words and serial order. Cambridge University Press.

Pulvermüller, F., 2001. Brain reflections of words and their meaning. Trends Cogn. Sci. 5, 517–524. doi:10.1016/S1364-6613(00)01803-9

Pulvermüller, F., 1999. Words in the brain's language. Behav. Brain Sci. 22, 253–336. doi:10.1017/S0140525X9900182X

Pulvermüller, F., 1996. Hebb's concept of cell assemblies and the psychophysiology of word processing. Psychophysiology. doi:10.1111/j.1469-8986.1996.tb01057.x

Pulvermüller, F., 1992. Brain function and syntactic mechanisms: is there a connection?

Pulvermüller, F., Cooper-Pye, E., Dine, C., Hauk, O., Nestor, P.J., Patterson, K., 2010. The word processing deficit in semantic dementia: all categories are equal, but some categories are more equal than others. J. Cogn. Neurosci. 22, 2027–2041. doi:10.1162/jocn.2009.21339

Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat. Rev. Neurosci. 11, 351–360. doi:10.1038/nrn2811

Pulvermüller, F., Garagnani, M., 2014. From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment. Cortex 57, 1–21. doi:10.1016/j.cortex.2014.02.015

Pulvermüller, F., Garagnani, M., Wennekers, T., 2014a. Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. Biol. Cybern. 108, 573–593. doi:10.1007/s00422-014-0603-9

Pulvermüller, F., Härle, M., Hummel, F., 2000. Neurophysiological distinction of verb categories. Neuroreport 11, 2789–2793. doi:10.1097/00001756-200008210-00036

Pulvermüller, F., Hauk, O., Nikulin, V. V., Ilmoniemi, R.J., 2005a. Functional links between motor and language systems. Eur. J. Neurosci. 21, 793–797. doi:10.1111/j.1460-9568.2005.03900.x

Pulvermüller, F., Kherif, F., Hauk, O., Mohr, B., Nimmo-Smith, I., 2009. Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. Hum. Brain Mapp. 30, 3837–3850. doi:10.1002/hbm.20811

Pulvermüller, F., Knoblauch, A., 2009. Discrete combinatorial circuits emerging in neural networks: A mechanism for rules of grammar in the human brain? Neural Networks 22, 161–172. doi:10.1016/j.neunet.2009.01.009

Pulvermüller, F., Lutzenberger, W., Preissl, H., 1999. Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. Cereb. Cortex 9, 497–506. doi:10.1093/cercor/9.5.497

Pulvermüller, F., Moseley, R.L., Egorova, N., Shebani, Z., Boulenger, V., 2014b. Motor cognition-motor semantics: Action perception theory of cognition and communication. Neuropsychologia 55, 71–84. doi:10.1016/j.neuropsychologia.2013.12.002

Pulvermüller, F., Preissl, H., 1991. A neuronal model of aphasic syndromes, in: Elsner, N., Penzlin, H. (Eds.), Synapse - Transmission - Modulation. Contributions to the 19th Göttingen Neurobiology Conference. Thieme, Stuttgart, New York, p. 548.

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R., 2005b. Brain signatures of meaning access in action word recognition. J. Cogn. Neurosci. 17, 884–892. doi:10.1162/0898929054021111

Pulvermüller, F., Shtyrov, Y., Kujala, T., Näätänen, R., 2004. Word-specific cortical activity as revealed by the mismatch negativity. Psychophysiology 41, 106–112. doi:10.1111/j.1469-8986.2003.00135.x

Quillian, M.R., 1969. The teachable language comprehender: A simulation program and theory of language. Commun. ACM 12, 459–476.

Ralph, M.A.L., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. Nat Rev Neurosci 18, 42–55.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–24. doi:10.1038/nn.2331

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–6. doi:10.1073/pnas.97.22.11800

Raz, N., Amedi, A., Zohary, E., 2005. V1 activation in congenitally blind humans is associated with episodic retrieval. Cereb. Cortex 15, 1459–1468.

Rilling, J.K., 2014. Comparative primate neuroimaging: Insights into human brain evolution. Trends Cogn. Sci. 18, 46–55. doi:10.1016/j.tics.2013.09.013

Rilling, J.K., Glasser, M.F., Jbabdi, S., Andersson, J., Preuss, T.M., 2011. Continuity, divergence, and the evolution of brain language pathways. Front. Evol. Neurosci. 3, 11. doi:10.3389/fnevo.2011.00011

Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X., Behrens, T.E.J., 2008. The evolution of the arcuate fasciculus revealed with comparative DTI. Nat. Neurosci. 11,

426–428. doi:10.1038/nn2072

Rilling, J.K., Van Den Heuvel, M.P., 2018. Comparative primate connectomics. Brain. Behav. Evol. 91, 170–179. doi:10.1159/000488886

Rioult-Pedotti, M.-S., Friedman, D., Donoghue, J.P., 2000. Learning-Induced LTP in Neocortex. Science (80-.). 290, 533 LP-536.

Rizzolatti, G. Luppino, G., 2001. The cortical motor system. Neuron 31, 889–901.

Röder, B., Stock, O., Bien, S., Neville, H., Rösler, F., 2002. Speech processing activates visual cortex in congenitally blind humans. Eur. J. Neurosci. 16, 930–936. doi:10.1046/j.1460-9568.2002.02147.x

Rolls, E.T., Deco, G., 2010. The Noisy Brain: Stochastic Dynamics as a Princple of Brain Function. Oxford University Press, Oxford.

Romanski, L.M., 2007. Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. Cereb. Cortex 17, i61–i69. doi:10.1093/cercor/bhm099

Romanski, L.M., Bates, J.F., Goldman-Rakic, P.S., 1999a. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 403, 141–157. doi:10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 1999b. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2, 1131–1136. doi:10.1038/16056

Rumelhart, D.E., Hinton, G., Williams, R., 1986. Learning internal representations by backpropagation, in: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Mircrostructure of Cognition. MIT Press, Cambridge, MA.

Runes, D.D., 1984. Dictionary of philosophy: Revised and enlarged.

Rüschemeyer, S.-A., Brass, M., Friederici, A.D., 2007. Comprehending Prehending: Neural Correlates of Processing Verbs with Motor Stems. J. Cogn. Neurosci. 19, 855–865. doi:10.1162/jocn.2007.19.5.855

Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. Proc. Natl. Acad. Sci. U. S. A. 105, 18035–40. doi:10.1073/pnas.0805234105

Schomers, M.R., Garagnani, M., Pulvermüller, F., 2017. Neurocomputational Consequences of Evolutionary Connectivity Changes in Perisylvian Language Cortex. J. Neurosci. 37, 3045–3055. doi:10.1523/JNEUROSCI.2693-16.2017

Schomers, M.R., Pulvermüller, F., 2016. Is the Sensorimotor Cortex Relevant for Speech Perception and Understanding? An Integrative Review. Front. Hum. Neurosci. 10. doi:10.3389/fnhum.2016.00435

Searle, J.R., 1980. Minds, braind and Programs. Behav. Brain Sci. 3, 1–19.

Sedley, D., 2003. Plato's Cratylus. Cambridge University Press.

Seidenberg, M.S., McClelland, J.L., 1989. A distributed, developmental model of word recognition and naming. Psychol. Rev. 96, 523–68. doi:10.1037/0033-295X.96.4.523

Seltzer, B., Pandya, D.N., 1989. Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey. J. Comp. Neurol. 290, 451–471. doi:10.1002/cne.902900402

Shallice, T., 1988. From neuropsychology to mental structure. New York: Cambridge University Press. Cambridge University Press, New York.

Shebani, Z., Patterson, K., Nestor, P.J., Diaz-de-Grenu, L.Z., Dawson, K., Pulvermüller, F., 2017. Semantic Word Category Processing in Semantic Dementia and Posterior Cortical Atrophy. Cortex in press. doi:10.1016/j.cortex.2017.04.016

Shepard, R.N., Cooper, L.A., 1992. Representation of colors in the blind, color-blind, and normally sighted. Psychol. Sci. 3, 97–104.

Shimony, J.S., Burton, H., Epstein, A.A., McLaren, D.G., Sun, S.W., Snyder, A.Z., 2005. Diffusion tensor imaging reveals white matter reorganization in early blind humans. Cereb. Cortex 16, 1653–1661.

Shtyrov, Y., Butorina, A., Nikolaeva, A., Stroganova, T., 2014. Automatic ultrarapid activation

and inhibition of cortical motor systems in spoken word comprehension. Proc. Natl. Acad. Sci. U. S. A. 111, E1918--23. doi:10.1073/pnas.1323158111

Shtyrov, Y., Hauk, O., Pulvermüller, F., 2004. Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. Eur. J. Neurosci. 19, 1083–1092. doi:10.1111/j.0953-816X.2004.03126.x

Shu, N., Li, J., Li, K., Yu, C., Jiang, T., 2009a. Abnormal diffusion of cerebral white matter in early blindness. Hum. Brain Mapp. 30, 220–227.

Shu, N., Liu, Y., Li, J., Li, Y., Yu, C., Jiang, T., 2009b. Altered anatomical network in early blindness revealed by diffusion tensor tractography. PLoS One 4, e7228.

Silveri, M.C., Brita, A.C., Liperoti, R., Piludu, F., Colosimo, C., 2018. What is semantic in semantic dementia? The decay of knowledge of physical entities but not of verbs, numbers and body parts. Aphasiology 32, 989–1009. doi:10.1080/02687038.2017.1387227

Sim, E.J., Kiefer, M., 2005. Category-related brain activity to natural categories is associated with the retrieval of visual features: Evidence from repetition effects during visual and functional judgments. Cogn. Brain Res. 24, 260–273. doi:10.1016/j.cogbrainres.2005.02.006

Simmons, W.K., Martin, A., Barsalou, L.W., 2005. Pictures of appetizing foods activate gustatory cortices for taste and reward. Cereb. Cortex 15, 1602–1608. doi:10.1093/cercor/bhi038

Sporns, O., 2007. What neuro-robotic models can teach us about neural and cognitive development. Neuroconstructivism Perspect. Prospect. 2, 179–204.

Sporns, O., Chialvo, D.R., Kaiser, M., Hilgetag, C.C., 2004. Organization, development and function of complex brain networks. Trends Cogn. Sci. 8, 418–425. doi:10.1016/j.tics.2004.07.008

Sporns, O., Honey, C.J., Kötter, R., 2007. Identification and classification of hubs in brain networks. PLoS One 2, e1049.

Stramandinoli, F., Marocco, D., Cangelosi, A., 2012a. The grounding of higher order concepts

in action and language: a cognitive robotics model. Neural Networks 32, 165–173.

Stramandinoli, F., Marocco, D., Cangelosi, A., 2012b. A Neuro-Robotics Model for the Acquisition of Higher Order Concepts in Action and Language. Proc. Annu. Meet. Cogn. Sci. Soc. 32, 2868.

Striem-Amit, E., Ovadia-Caro, S., Caramazza, A., Margulies, D.S., Villringer, A., Amedi, A., 2015. Functional connectivity of visual cortex in the blind follows retinotopic organization principles. Brain 138, 1679–1695.

Struiksma, M.E., Noordzij, M.L., Neggers, S.F.W., Bosker, W.M., Postma, A., 2011. Spatial language processing in the blind: evidence for a supramodal representation and cortical reorganization. PLoS One 6, e24253.

Tate, M.C., Herbet, G., Moritz-Gasser, S., Tate, J.E., Duffau, H., 2014. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. Brain 137, 2773–2782.

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., Catani, M., 2012. Monkey to human comparative anatomy of the frontal lobe association tracts. Cortex 48, 82–96. doi:10.1016/j.cortex.2011.10.001

Tomasello, M., Kruger, A.C., 1992. Joint attention on actions: acquiring verbs in ostensive and non-ostensive contexts. J. Child Lang. 19, 311–333. doi:10.1017/S0305000900011430

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., 2018. A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity. Front. Comput. Neurosci. 12, 88. doi:10.3389/fncom.2018.00088

Tomasello, R., Garagnani, M., Wennekers, T., Pulvermüller, F., 2017. Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. Neuropsychologia 98, 111–129. doi:http://doi.org/10.1016/j.neuropsychologia.2016.07.004.

Tranel, D., Damasio, H., Damasio, A.R., 1997. A neural basis for the retrieval of conceptual knowledge. Neuropsychologia 35, 1319–1327. doi:10.1016/S0028-3932(97)00085-7

Trumpp, N.M., Kliese, D., Hoenig, K., Haarmeier, T., Kiefer, M., 2013. Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. Cortex 49, 474–486. doi:10.1016/j.cortex.2012.02.002

Tschentscher, N., Hauk, O., Fischer, M.H., Pulvermüller, F., 2012. You can count on the motor cortex: Finger counting habits modulate motor cortex activation evoked by numbers. Neuroimage 59, 3139–3148. doi:10.1016/j.neuroimage.2011.11.037

Tsumoto, T., 1992. Long-term potentiation and long-term depression in the neocortex. Prog. Neurobiol. 39, 209–228. doi:10.1016/0301-0082(92)90011-3

Ueno, T., Saito, S., Rogers, T.T., Lambon Ralph, M.A., 2011. Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. Neuron 72, 385–396. doi:10.1016/j.neuron.2011.09.013

Ungerleider, L.G., Gaffan, D., Pelak, V.S., 1989. Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. Exp. Brain Res. 76, 473–484. doi:10.1007/BF00248903

Ungerleider, L.G., Haxby, J. V., 1994. “What” and “where” in the human brain. Curr. Opin. Neurobiol. 4, 157–165. doi:10.1016/0959-4388(94)90066-3

Van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. Trends Cogn. Sci. 17, 683–696.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., Cappa, S.F., 2011. Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci. Biobehav. Rev. 35, 407–426. doi:http://dx.doi.org/10.1016/j.neubiorev.2010.04.007

Vigliocco, G., Vinson, D.P., Lewis, W., Garrett, M.F., 2004. Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. Cogn. Psychol. 48, 422–488. doi:10.1016/j.cogpsych.2003.09.001

Vincent-Lamarre, P., Massé, A.B., Lopes, M., Lord, M., Marcotte, O., Harnad, S., 2016. The latent structure of dictionaries. Top. Cogn. Sci. 8, 625–659.

Voss, P., Gougoux, F., Zatorre, R.J., Lassonde, M., Lepore, F., 2008. Differential occipital

responses in early-and late-blind individuals during a sound-source discrimination task. Neuroimage 40, 746–758.

Vouloumanos, A., Werker, J.F., 2009. Infants' learning of novel words in a stochastic environment. Dev. Psychol. 45, 1611–7. doi:10.1037/a0016134

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., Shtyrov, Y., 2017. Primary motor cortex functionally contributes to language comprehension: An online rTMS study. Neuropsychologia 96, 222–229. doi:https://doi.org/10.1016/j.neuropsychologia.2017.01.025

Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M., Mori, S., 2004. Fiber tract-based atlas of human white matter anatomy. Radiology 230, 77–87. doi:10.1148/radiol.2301021640

Warrington, E.K., Mccarthy, R., 1983. Category specific access dysphasia. Brain 106, 859–878. doi:10.1093/brain/106.4.859

Warrington, E.K., McCarthy, R.A., 1987. Categories of knowledge: further fractionations and an attempted integration, Brain. doi:10.1093/brain/110.5.1273

Webster, M.J., Bachevalier, J., Ungerleider, L.G., 1994. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. Cereb. cortex 4, 470–483. doi:10.1093/cercor/4.5.470

Wennekers, T., Garagnani, M., Pulvermüller, F., 2006. Language models based on Hebbian cell assemblies. J. Physiol. Paris 100, 16–30. doi:10.1016/j.jphysparis.2006.09.007

Wernicke, C., 1874. Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis, Wernicke's work on aphasia. Kohn und Weigert, Breslau. doi:10.1007/978-3-642-65950-8

Westermann, G., Miranda, E.R., 2004. A New Model of Sesorimotor Coupling in the Development of Speech. Brain Lang. 89, 393–400.

Willshaw, D.J., Buneman, O.P., Longuet-Higgins, H.C., 1969. Non-holographic associative memory. Nature 222, 960–962. doi:10.1038/222960a0

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized

populations of model neurons. Biophys. J. 12, 1–24. doi:10.1016/S0006-3495(72)86068-5

Withagen, A., Kappers, A.M.L., Vervloed, M.P.J., Knoors, H., Verhoeven, L., 2013. Short term memory and working memory in blind versus sighted children. Res. Dev. Disabil. 34, 2161–2172. doi:https://doi.org/10.1016/j.ridd.2013.03.028

Yeterian, E.H., Pandya, D.N., Tomaiuolo, F., Petrides, M., 2012. The cortical connectivity of the prefrontal cortex in the monkey brain. Cortex 48, 68–81. doi:10.1016/j.cortex.2011.03.004

Young, M.P., Scannell, J.W., Burns, G., 1995. The analysis of cortical connectivity. Springer, Heidelberg.

Young, M.P., Scannell, J.W., Burns, G.A.P.C., Blakemore, C., 1994. Analysis of connectivity: Neural systems in the cerebral cortex. Rev. Neurosci. 5, 227–250. doi:10.1515/REVNEURO.1994.5.3.227

Yuille, A.L., Geiger, D., 2003. Winner-Take-All Mechanisms, in: Arbib, M. (Ed.), The Handbook of Brain Theory and Neural Networks. MIT Press, Cambridge, MA, pp. 1056–1060.

Zatorre, R.J., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET Studies of Phonetic Processing of Speech: Review, Replication, and Reanalysis. Cereb. Cortex 6, 21–30. doi:10.1093/cercor/6.1.21

Appendix A

Curriculum vitae

Rosario Tomasello

Education

April 2015 – May 2019 **PhD student at the Berlin School of Mind and Brain - Humboldt University and Brain Language Laboratory - FUB**

Supervisors: Prof. Dr. Dr. Friedemann Pulvermüller & Prof. Angelo Cangelosi

WS 2011 – WS 2014 **Master of Arts in European languages – structure and use**

Free University of Berlin, – M.A. Thesis: 'Neurophysiological evidence for the Whorfian effects: The case of Italian Blues'.

WS 2006 - WS 2009 **Bachelor in Science for International Communication**

University of Catania (IT) – B.A. Thesis: **'***Linguistic and civil awareness among Turkish immigrants in Germany'*

1999/2000 - 2004/05 **High school graduations in foreign languages (Abitur)**

Institute 'G. Verga' Adrano (IT) – German, English and France, Literature, Philosophies etc.

Semester abroad

WS 2011/12 **M.A. Portuguese as a foreign language**
University of Coimbra – LPP/Erasmus, Interdisciplinary research of Portuguese comparing to other Romanic languages, pragmatic and semantic studies.

SS 2009 **B.A. English and German studies**

University of Hanover – LPP/ Erasmus
General linguistic in German and English, incl. Phonetic, syntax, morphology, semantic, pragmatic.

Professional Activities

02.2019 – currently **Post-Doctoral scholarship - Berlin School Mind & Brain at the Humboldt University Berlin**

07.2018 – 02.2019 **Research fellow 'Xprag' BraiSiCo project - Brain Language Laboratory**

Free University Berlin, project title: '*Brain Signatures of communication'*

06.2016 – 06.2018 **Doctoral stipend - Berlin School Mind & Brain at the Humboldt University Berlin**

01.2014 – 06.2016 **Research fellow - Plymouth University (UK), based at the Freie Universitaet Berlin - BABEL project:** *'Neurocomputational Modelling of Language Learning'*

04.2013 –01.2014 **Student Assistant - Brain Language Laboratory**

Free University Berlin, Supporting Prof. Dr. Pulvermüller for research and teaching, execution and evaluation of EEG and behavioral experiments, It-support for the lab.

Teaching Experience

WS2018/19 **Pragmatik: Sprache und Handeln**

Co-teaching a seminar for BA and MA Programs, Freie Universität Berlin, Germany

SS2017 & SS2018 **Language and the Brain course**

Lecture on neuropragmatics for MA students of the Berlin school of Mind & Brain, Humboldt University

WS2017/18 **Sprachfunktion: Experimentelle Pragmatik**

Seminar for BA and MA Programs, Freie Universität Berlin, Germany

Publications

- **Tomasello, R.,** Wennekers, T., Garagnani, M., Pulvermüller, F., (2019). Recruitment of visual cortex for language processing in blind individuals is explained by Hebbian learning. *Scientific Reports* 9(1):3579.
- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 2018 A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity, *Front. Comput. Neurosci.*, vol. 12, p. 88
- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 2017 Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*. 98.
- Garagnani M, Lucchese G, **Tomasello R**, Wennekers T, Pulvermüller F, 2017 A spiking neurocomputational model of high-frequency oscillatory brain responses to words and pseudowords. *Front. Comput. Neurosci.* 10.
- **Tomasello R,** Kim C, Dreyer FR., Grisoni L, Pulvermüller F. (Submitted) Neurophysiological Evidence For Early Interplay of Linguistic and Gestural Information in Understanding Communicative Actions.
- Shebani Z, Carota F, Hauk O, Rowe J B, Barsalou L B, **Tomasello R,** and Pulvermuller F, (Under Review) Brain correlates of action word memory, *bioRxiv*, p. 412676, Jan. 2018.

Talk & Article review

- **Talk - Tomasello R**, Neuromechanistics Perspectives on Language Processing - *Brain Dynamics on Multiple Scales Conference* at the Max Planck Institute for the Physics of Complex Systems, Dresden, DE, 06.2017
- **Lighting Talk - Tomasello R,** Neural correlates of basic gestures in communication, Graz Austria, Xprag Annual Meeting, 06.2018
- Acted as a peer reviewer for the *Language, Cognition and Neuroscience* Journal

Poster

- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 03.2019 Recruitment of visual cortex for language processing in blind individuals is explained by Hebbian learning, CNS, San Francisco, USA
- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 12.2018 A neurobiologically-constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity, Mind & Brain poster session Humboldt University, Berlin DE
- **Tomasello R,** Pulvermüller F, 06.2018, Neural correlates of basic gestures in communication, Graz Austria, Xprag Annual Meeting, 2018
- **Tomasello R,** & Pulvermüller F, 09.2017 Neural correlates of speech act processing in gestural context, Xprag.de, Köln, DE
- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 06.2017 A neurobiologically-constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity, Brain Dynamics on Multiple Scales at Max Planck Institute for the Physics of Complex Systems, Dresden, DE
- **Tomasello R**, Garagnani M, Wennekers T, Pulvermüller F, 03.2017 A neurobiologically-constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity, CNS, San Francisco, USA
- Garagnani M, Lucchese G, **Tomasello R,** Wennekers T, Pulvermüller F, 04.2016. Stable cell assembly formation and maintenance via spike-driven, non-homeostatic Hebbian synaptic plasticity. A Royal Society Meeting, London, UK.
- Pulvermüller F, **Tomasello R,** Kim C, Egorova N, 2016 Brain indexes of communicative function and interactive prediction. ESCAN, Porto, PT.
- **Tomasello R,** Garagnani M, and Pulvermüller F, 2015 Category specificity, hubs, and time course of semantic brain activation: A neurocomputational model. SNL. Chicago, USA.
- Kim C, **Tomasello R,** and Pulvermüller F, 2014 Neurophysiological and behavioral evidence for language influence on colors perception: SNL, Amsterdam.

Appendix B – Erklärung

Hiermit versichere ich, dass ich die vorgelegene Arbeit selbständig verfasst habe und keine anderen an die angegebenen Hilfsmittel verwendet habe. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin,

Rosario Tomasello, Berlin 2019