Université Pierre et Marie Curie

Ecole doctorale Cerveau-Cognition-Comportement (ED 158)

Università degli Studi di Trento

Scuola di Dottorato in Cognitive and Brain Sciences

THE NEURO-COGNITIVE REPRESENTATION OF WORD MEANING RESOLVED IN SPACE AND TIME.

A dissertation submitted by

Valentina BORGHESANI

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Neuroscience

Publicly defended on the 28th of February 2017

Thesis committee:

Christophe PALLIER	INSERM-CEA Unicog	Advisor
Manuela PIAZZA	CIMeC, UNITN	Advisor
James HAXBY	Dartmouth College	Reviewer
Rik VANDENBERGHE	University Hospital Leuven	Reviewer
Laurent COHEN	UPMC	Examiner
Marius PEELEN	CIMeC, UNITN	Examiner

[intentionally left blank]

To those who know what they mean.

[intentionally left blank]

Abstract

One of the core human abilities is that of interpreting symbols. Prompted with a perceptual stimulus devoid of any intrinsic meaning, such as a written word, our brain can access a complex multidimensional representation, called semantic representation, which corresponds to its meaning. Notwithstanding decades of neuropsychological and neuroimaging work on the cognitive and neural substrate of semantic representations, many questions are left unanswered. The research in this dissertation attempts to unravel one of them: are the neural substrates of different components of concrete word meaning dissociated?

In the first part, I review the different theoretical positions and empirical findings on the cognitive and neural correlates of semantic representations. I highlight how recent methodological advances, namely the introduction of multivariate methods for the analysis of distributed patterns of brain activity, broaden the set of hypotheses that can be empirically tested. In particular, they allow the exploration of the representational geometries of different brain areas, which is instrumental to the understanding of where and when the various dimensions of the semantic space are activated in the brain. Crucially, I propose an operational distinction between motor-perceptual dimensions (i.e., those attributes of the objects referred to by the words that are perceived through the senses) and conceptual ones (i.e., the information that is built via a complex integration of multiple perceptual features).

In the second part, I present the results of the studies I conducted in order to investigate the automaticity of retrieval, topographical organization, and temporal dynamics of motor-perceptual and conceptual dimensions of word meaning. First, I show how the representational spaces retrieved with different behavioral and corpora-based methods (i.e., Semantic Distance Judgment, Semantic Feature Listing, WordNet) appear to be highly correlated and overall consistent within and across subjects. Second, I present the results of four priming experiments suggesting that perceptual dimensions of word meaning (such as implied real world size and sound) are recovered in an automatic but task-dependent way during reading. Third, thanks to a functional magnetic resonance imaging experiment, I show a representational shift along the ventral visual path: from perceptual features, preferentially encoded in primary visual areas, to conceptual ones, preferentially encoded in mid and anterior temporal areas. This result indicates that complementary dimensions of the semantic space are encoded in a distributed yet partially dissociated way across the cortex. Fourth, by means of a study conducted with magnetoencephalography, I present evidence of an early (around 200 ms after stimulus onset) simultaneous access to both motor-perceptual and conceptual dimensions of the semantic space thanks to different aspects of the signal: inter-trial phase coherence appears to be key for the encoding of perceptual while spectral power changes appear to support encoding of conceptual dimensions.

These observations suggest that the neural substrates of different components of symbol meaning can be dissociated in terms of localization and of the feature of the signal encoding them, while sharing a similar temporal evolution.

Keywords : Symbols, Semantic memory, Representational geometry, Neuropsychology, Neuroimaging, functional Magnetic Resonance Imaging (fMRI), magnetoencephalography (MEG)

Synopsis

INTRODUCTION:
ORGANIZATION AND CONTRIBUTIONS OF THE THESIS

The brain: a device through which we think we can think.

This introductory chapter frames the problem tackled during my PhD. I then briefly introduce the different chapters and the scientific publications stemmed from the experimental works conducted. Subsequently, as required by my dual PhD program, a summary in French and Italian is provided. Last, but surely not least, some due ackowledgements.

1. The symbols that made us what we are

"*We should start back, - Gared urged as the woods began to grow dark around them*". What I studied in this thesis is what is happening right now in your brain. I typed keys on a keyboard generating black lines on a white background. I have used those strokes (i.e., letters), to assemble symbols (i.e., words). Your brain, provided with information from your eyes, is translating them into meaningful mental representations. You can hear Gared talking and you know he is not alone. You can tell that it is dusk and you can see they are in a wood. You understand the meaning of those words and you use it to make sense of the situation. You can also push it further and begin to imagine what is not written: where were they headed? where will they get back to? how many are they? Above all, you might have recognized the piece of writing I typed: it is the incipit of "*A Song of Ice and Fire*", by George R.R. Martin.

You have been able to do so because you are equipped with a complex neuro-cognitive structure, the semantic system, that stores and processes various form of conceptual knowledge, including symbols meaning. The relevance of symbols understanding and manipulation in our lives cannot be overstated. We are constantly prompted by physical inputs (e.g., road signs, logos, spoken and written words), which we interpret as referring to more than what meets the eye. Throughout our life, we use symbols to evoke, communicate and reflect upon things that are not currently present to our senses. The term itself, symbol, derives from the ancient Greek *sumbolon*, fusion of the stem *ballein* (i.e., "*to throw*") and the

preposition sun (i.e., "*with*"), thus meaning literally *"that which is thrown or cast together*". Thanks to symbols, we are elevated from the reality of perception (dominated by the physical features of the stimuli) and gain access to the realm of semantic representations, where different features are cast together to assemble unitary concepts. Indeed, a simple concept such as "*cat*" includes information on both perceptual attributes, i.e., those features one experiences through the senses (e.g., cats are usually small, they have a soft fur, they meow), and conceptual features, i.e., those that emerge through combination of perceptual ones and/or one learns declaratively (e.g., cats belong to the felidae family).

Symbols, even if different in nature, can all be defined as pointers to concepts, sharing three key aspects: arbitrariness, culture-dependency, and unbounded combinatorial power. Contrary to signs, which have a natural affinity with their reference either iconiccally (e.g., a portrait – it relies on form similarity) or indexically (e.g., pointing – it depends on spatio-temporal contiguity), symbols are completely arbitrary. Their physical properties bearing no relation with the semantic content they provide access to. Moreover, their meaning is defined only within a given linguistic and cultural milieu (e.g., presented with /*burro*/ a Spanish speaker would think of a donkey, an Italian one of butter). Finally, they can be merged limitlessly, creating new symbols-concepts pairings (for instance consider the –relatively-recently introduced concept of *smartphone*), and interact endlessly: their reciprocal relationship can be analyzed in light of different context and goals, changing the corresponding representational geometry (e.g., according to the context, *cups, mugs, and glasses* can be used interchangeably or not). Given the relevance of symbols in our daily life, it is not surprising that some of the most outstanding questions tackled by cognitive neuroscience revolve around the neural correlates of symbols acquisition, storage, and processing.

Symbols are mentally represented at different levels of complexity. The first and most simple level sees symbols being processed as physical objects in the corresponding primary and secondary sensory cortices (i.e. visual cortex for written words, auditory cortex for spoke words). This stage corresponds to generic sensory (or motor) neural representations evoked by the presentation of any stimulus to a sensory organ. The moment one sees a flower, light is transduced by the eyes, information carried by the axons along the optical nerve, projected to the primary visual cortex, and subsequently a coherent representation of the visual features (e.g., shape and color) is reconstructed. Similar processes apply to the olfactory sensory representation (i.e., the smell of the flower entering the nostrils).

However, symbols also evoke higher order multifaceted representations, which we call cognitive semantic representations. They are rich internal states that reflect our knowledge of their meaning, including both motor-perceptual and conceptual dimensions. The semantic representation of flower is the summation of all the features of the concept, both motor-perceptual (e.g., a flower is usually something I can hold with my hand and has a pleasant smell) and conceptual (e.g., a flower is the reproductive structure of angiosperms) ones. We use the term neural semantic representations to refer to the neural activity automatically evoked by symbol meaning, which appears to be implemented in distributed neural networks spanning a large portion of the cortex

The main goal of the thesis is to investigate cognitive and neural aspect of the semantic space, exploiting cutting edge techniques for brain imaging data analysis which allow to test the mapping of given representational geometries onto neural pattern of activations.

2. Outline of the thesis

This manuscript consists of six chapters: the first two introduce the theory and the methods behind the experimental work undertaken, while the following three describe such endeavor. The last chapter summarizes the results, discussing the theoretical implications and the future perspectives.

Chapter 1 aims at explicitly define the field of inquiry (what am I going to talk about) and the operationalization of the variables at play (how am I going to do so). First of all, I define semantic representations in terms of their content, providing evidence of their relevance as a psychological and neurological reality. Second, I revise the hypotheses on the localization of their neural correlates in light of the experimental findings in the literature: are they distributed over a broad portion of the cortex or localized in pivotal areas? is the organization driven by evolutionary principles or anatomical constrains? Third, I describe the current results relative to the timing of activation within the semantic system at both short and long time scales (e.g., task requirements vs personal experiences). Finally, I highlight the relationship between the content of the representation, the format adopted (i.e. the operations

that can be performed), and the underlying implementation (i.e. the neural code). Notably, this chapter includes my theoretical contribution: an operational definition of word meaning that can foster both theoretical speculations and empirical research. The meaning of words is conceptualized as a multidimensional representation that includes both motor-perceptual (e.g., average size, prototypical color) and conceptual (e.g., taxonomic class) dimensions. This chapter, thus, not only offers a review of the literature on semantic representations, but also introduces the theoretical framework I adopted for the following experimental investigations.

Chapter 2 offers an overview of the methods used during my experimental work, illustrating how cognitive and neural representations can be investigated with behavioral tasks (Semantic Distance Judgment, Semantic Feature Listing, and Semantic Priming) as well as neuroimaging techniques (functional magnetic resonance imaging and magnetoencephalography). In particular, I focus on multivariate methods for the analyses of neuroimaging data. This chapter can be easily skipped (or read in diagonal) by those that are already familiar with the above mentioned techniques. However, its conclusion, and in particular the discussion on the application of multivariate methods to the investigation of neural substrate of cognitive representations, are crucial to the understanding of my perspective while navigating through the rest of the manuscript.

In Chapter 3, I describe the outcomes of our behavioral experiments. First, Semantic Distance Judgment and Semantic Features Listing experiments were conducted on two set of data. The goal was two-fold: the comparison of the semantic space the two methods give access to, and the validation of the stimuli to be used in the following neuroimaging experiments. The results indicate that, given a set of words, different measures converge in describing the same semantic space. Second, I conducted 4 priming experiments aiming at elucidating the automaticity of retrieval of different perceptual dimensions. The results suggest a delicate interaction between the task subjects are performing and whether the two words refer to objects that share (or not) the same visual (i.e., implied real world size) and auditory (i.e., prototypical sound) features.

Chapter 4 describes the functional magnetic resonance (fMRI) experiment I conducted and its results. We tested the hypothesis that perceptual and conceptual dimensions of word meaning are coded in different brain regions: perceptual dimension in unimodal perceptual areas, conceptual dimension in heteromodal association areas. We tested the presence of a

mapping between a perceptual dimension (implied real object size) and two conceptual dimensions (taxonomic categories at different levels of specificity), and the patterns of brain activity recorded in six areas along the ventral occipito–temporal cortical path. Combining multivariate pattern classification and representational similarity analysis, we found that the visual-perceptual dimension appears to be primarily encoded in early visual regions, while the conceptual dimension in more anterior temporal regions. This anteroposterior gradient of information content, from perceptual to conceptual, indicates that different areas along the ventral stream encode complementary dimensions of the semantic space.

In Chapter 5, I present the magnetoencephalography (MEG) experiment I conducted and its results. We investigated whether perceptual and conceptual dimensions of word meaning could be dissociated not only in their topography, but also in terms of their temporal dynamics. We compared one conceptual dimension (semantic category) and two perceptual dimensions (one concerning a visual feature - the implied real world size, and one concerning an auditory feature – prototypical sound). Results indicate an automatic, rapid (~200ms) and essentially simultaneous recovery of information along both perceptual and conceptual dimensions of word meaning, a results that speaks against popular theories in the field. However, the three different dimensions appear to dissociate in terms of the brain dynamics involved (changes in phase coherence vs spectral power) and the corresponding underlying sources.

In Chapter 6, I discuss the general implications of our findings and the future work that will be needed to deepen our understanding of the cognitive and neural substrate of semantic representations.

Finally, the Appendix includes all the supporting materials, including the analyses I used either as “sanity checks” of data quality or as complementary evidence to the main findings of the different studies.

3. Author's publications

Contributions presented (partially or globally) in this thesis have been published (or are currently under review) in peer reviewed conference proceeding or journals:

Borghesani, V., Pedregosa, F., Eger, E., Buiatti, M., & Piazza, M. (2014). A perceptual-to-conceptual gradient of word coding along the ventral path. *International Workshop on Pattern Recognition in Neuroimaging*

Borghesani, V., Pedregosa, F., Buiatti, M., Alexis, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*

Borghesani, V., & Piazza, M. (under review). The neuro-cognitive representations of symbols: the case of concrete words. *Neuropsychologia*

Moreover, they have been presented at international conferences as posters and/or talks:

Borghesani, V., Pedregosa, F., Eger, E., Buiatti, M., & Piazza, M. (2014, Tubingen) A perceptual-to-conceptual gradient of word coding along the ventral path. *PRNI, International Workshop on Pattern Recognition in Neuroimaging* [selected for oral presentation]

Borghesani, V., Pedregosa, F., Eger, E., Buiatti, M., & Piazza, M. (2014, Rovereto) Word reading in the ventral stream: A perceptual to conceptual gradient of information coding. *CAOS, Concepts, Actions, Objects* [selected for oral presentation and winner of the best abstract award]

Borghesani, V., Eger, E., Buiatti, M., & Piazza, M. (2015, Chicago) Perceptual and conceptual semantic dimensions: where and when? *SNL, Society for the neurobiology of language*

Borghesani, V., Eger, E., Buiatti, M., & Piazza, M. (2015, Chicago) Can you see what I mean? Perceptual and conceptual semantics during word reading. *SfN, Society of Neuroscience*

Finally, during the three years of phd, other publications have stemmed from previously conducted research or current collaborations:

Borghesani*, V., de Hevia*, L., Viarouge*, A., Pinheiro Chagas, P., Eger, E., & Piazza, M. (2016). Comparing magnitudes across dimensions: a univariate and multivariate approach. *International Workshop on Pattern Recognition in Neuroimaging*

Borghesani*, V., de Hevia*, L., Viarouge*, A., Pinheiro Chagas, P., Eger, E., & Piazza, M. (2016, Rovereto). Few: Many = Short: Long? Coding Of Magnitude Across Different Quantitative Dimensions. *CAOS, Concepts, Actions, Objects*

Borghesani, V., Monti, A., Fortis, P., & Miceli , G. (sumitted). Monothematic spatial delusion: a review on reduplicative paramnesia for places and a case study.

(*) denotes joint first authorship as the authors contributed equally to this work

4. Résumé en français

L'une des capacités humaines fondamentales est la capacité d'interpréter des symboles. En présence d'un stimulus dépourvu de signification intrinsèque, comme un mot écrit, notre cerveau peut accéder à une représentation complexe et multidimensionnelle, appelée représentation sémantique. Malgré plusieurs décennies de travaux en neuropsychologique et neuroimagerie sur le substrat cognitif et neuronal des représentations sémantiques, de nombreuses questions restent sans réponse. Les présents travaux de thèse tentent de démêler l'un de ces mystères: les substrats neuronaux des différentes composantes du mot sont-ils dissociables?

Ce travail comporte deux composantes principales : l'une théorique et l'autre empirique. Dans la première partie, nous passons en revue les différentes positions théoriques concernant les corrélats cognitifs et neuraux des représentations sémantiques. Nous soulignons la façon dont les avancées méthodologiques récentes, notamment l'introduction de méthodes multivariées pour l'analyse de l'activité cérébrale, élargissent l'ensemble des hypothèses qui peuvent être testées empiriquement. Elles permettent notamment d'explorer les géométries représentationnelles des différentes zones du cerveau, ce qui est essentiel pour comprendre où et quand les différentes dimensions de l'espace sémantique sont activées dans le cerveau. De plus, nous proposons une distinction opérationnelle entre les dimensions moto-perceptives (c'est-à-dire les attributs des objets auxquels les mots se réfèrent perçus par les sens) et conceptuelles (c'est-à-dire l'information construite par l'intégration des multiples caractéristiques perceptives).

Dans la deuxième partie, nous présentons les résultats des études menées afin d'étudier l'automaticité de la récupération, l'organisation topographique et la dynamique temporelle des dimensions moto-perceptives et conceptuelles de la signification des mots. Tout d'abord, nous montrons comment les espaces représentationnels récupérés avec différentes méthodes comportementales et computationnelles (c'est-à-dire *Semantic Distance Judgment, Semantic Feature Listing, WordNet*) semblent être fortement corrélés et globalement cohérents entre les sujets. Ensuite, nous présentons les résultats de quatre expériences d'amorçage sémantique suggérant que les dimensions perceptives (telles que la taille et le son associées) sont récupérées d'une manière automatique mais dépendante de la tâche effectuée par les sujets au cours de la lecture. Puis, grâce à une expérience d'imagerie par résonance magnétique

fonctionnelle, nous montrons un gradient occipital-temporal le long de la voie visuelle ventrale: les caractéristiques perceptives sont préférentiellement encodées dans des zones visuelles primaires, tandis que les caractéristiques conceptuelles, dans les zones temporales médiane et antérieure. Ce résultat indique que des dimensions complémentaires de l'espace sémantique sont encodées d'une manière partiellement dissociée à travers le cortex cérébral. Enfin, au moyen d'une étude réalisée avec la magnétoencéphalographie, nous présentons des preuves d'un accès simultané et précoce (environ 200 ms après le stimulus) aux dimensions moto-perceptives et conceptuelles de l'espace sémantique grâce aux différents aspects du signal. La cohérence de phase semble être la clé pour le codage des aspects perceptifs, tandis que les changements de puissance spectrale semblent soutenir le codage des dimensions conceptuelles. Ces observations suggèrent que les substrats neuronaux de différentes composantes de la signification des symboles peuvent être dissociés en termes de localisation et également en termes de caractéristique du signal qui les encode, tout en partageant une évolution temporelle similaire.

Le manuscrit est constitué de six chapitres: les deux premiers introduisent la théorie et les méthodes du travail, tandis que les trois suivants décrivent les aspects expérimentaux. Le dernier chapitre résume les résultats, en discutant des implications théoriques et des perspectives futures. Enfin, l'annexe comprend tous les documents d'appui, y compris les analyses utilisées soit pour vérifier la qualité des données, soit comme preuves complémentaires aux conclusions principales des différentes études. Nous détaillons ci-dessous le contenu des cinq chapitres principaux.

Le **Chapitre 1** vise à définir explicitement le champ d'investigation (quel est le sujet abordé) et l'opérationnalisation des variables en jeu (comment nous allons l'aborder). Tout d'abord, nous définissons les représentations sémantiques en termes de contenu, en fournissant des preuves de leur pertinence comme une réalité psychologique et neurologique. Deuxièmement, nous révisons les hypothèses sur la localisation de leurs corrélats neuronaux à la lumière des résultats expérimentaux dans la littérature: sont-ils répartis sur une large portion du cortex ou localisés dans des zones-clés? L'organisation est-elle dirigée par des principes évolutifs ou des contraintes anatomiques? Troisièmement, nous décrivons les résultats relatifs à la dynamique temporelle du système sémantique à la fois à court terme (par exemple, les exigences de la tâche) et à long terme (par exemple, les expériences personnelles). Enfin, nous mettons en évidence la relation entre le contenu de la représentation, son format (c'est-à-dire les opérations pouvant y être exécutées) et

l'implémentation sous-jacente (c'est-à-dire le code neuronal qui le supporte). Ce chapitre inclut notamment notre contribution théorique : une définition opérationnelle du sens du mot qui peut favoriser à la fois les spéculations théoriques et la recherche empirique. La signification des mots est conceptualisée comme une représentation multidimensionnelle qui comprend des dimensions moto-perceptives (par exemple, la taille moyenne, le couleur prototypique) et conceptuelle (par exemple, la classe taxonomique). Ce chapitre propose odnc non seulement une revue de la littérature sur les représentations sémantiques, mais introduit également le cadre théorique adopté pour les recherches expérimentales suivantes.

Le **Chapitre 2** offre un aperçu des méthodes utilisées lors de notre travail expérimental, illustrant comment les représentations cognitives et neuronales peuvent être étudiées avec des tâches comportementales (*Semantic Distance Judgment, Semantic Feature Listing, Semantic Priming*) ainsi que des techniques de neuroimagerie (imagerie par résonance magnétique fonctionnelle et magnétoencéphalographie). En particulier, l'accent est mis sur les méthodes multivariées pour l'analyse des données de neuroimagerie qui utilisent des algorithmes de machine learning (une approche souvent appelée *decoding*) ou la corrélation entre activations neuronal (appelée *representational similarity analysis*, RSA).

Dans le **Chapitre 3**, nous décrivons les résultats des expériences comportementales. Premièrement, des expériences de *Semantic Distance Judgment et Semantic Feature Listing* ont été menées sur deux séries de données. L'objectif était double: la comparaison de l'espace sémantique auquel les deux méthodes donnent accès et la validation des stimuli à utiliser dans les expériences de neuroimagerie suivantes. Les résultats indiquent que différentes mesures convergent en décrivant le même espace sémantique. Deuxièmement, nous avons mené 4 expériences d'amorçage sémantique visant à élucider l'automaticité de la récupération de différentes dimensions perceptives. Les résultats suggèrent une interaction délicate entre la tâche réalisée par les sujets et l'effet d'amorçage en raison de la similitude des mots en termes de caractéristiques perceptives (par exemple, si deux mots se réfèrent à des objets qui partagent à peu près la même caractéristique visuelle ou auditive).

Le **Chapitre 4** décrit l'expérience de résonance magnétique fonctionnelle (IRMf) et ses résultats. Nous avons testé l'hypothèse selon laquelle les dimensions perceptives et conceptuelles de la signification des mots sont codées dans différentes régions du cerveau: la dimension perceptuelle dans les zones unimodales perceptuelles, la dimension conceptuelle dans les zones d'association hétéromodales. Nous avons testé la présence d'une correspondance entre une dimension perceptuelle (la taille implicite de l'objet réel) et deux dimensions conceptuelles (catégories taxonomiques à différents niveaux de spécificité) et les

patterns d'activité cérébrale enregistrés dans six zones le long de la voie ventrale occipito-temporale. En combinant les méthodes de *decoding* et *RSA*, nous avons constaté que la dimension perceptive (visuelle) semble être principalement codée dans la région visuelle primaire, tandis que la dimension conceptuelle est codée dans les régions temporales plus antérieures. Ce gradient antéro-postérieur du contenu informationnel, du conceptuel au perceptif, indique que différentes zones le long de la voie ventrale encodent des dimensions complémentaires de l'espace sémantique.

Dans le **Chapitre 5**, nous présentons l'expérience de magnétoencéphalographie (MEG) que effectuée et ses résultats. Nous nous sommes demandé si les dimensions perceptives et conceptuelles pouvaient être dissociées non seulement dans leur topographie, mais aussi dans leur dynamique temporelle. Nous avons comparé une dimension conceptuelle (catégorie sémantique) et deux dimensions perceptives (l'une concernant une caractéristique visuelle - la taille moyen - et l'autre concernant une caractéristique auditive – le son prototypique). Les résultats indiquent une récupération automatique, rapide (~200 ms) et essentiellement simultanée de l'information sur les trois. Cependant, les trois effets semblent se dissocier en ce qui concerne la dynamique cérébrale impliquée (changements dans la cohérence de la phase dans un cas, variations dans le spectre de puissance dans l'autre) et les sources cérébrales responsables.

Tout en contribuant à notre compréhension, encore partielle, de la manière dont le sens des mots est codé dans le cerveau et récupéré au cours du processus de lecture, les travaux présentés dans cette thèse ont des implications méthodologiques et théoriques importantes. En particulier, ils soulignent l'importance d'une intégration fructueuse entre les théories cognitives et les méthodes statistiques avancées afin d'éclairer les mystères entourant les représentations sémantiques.

5. Riassunto in italiano

Una delle capacità fondamentali degli esseri umani é quella di interpretare simboli. Posti davanti ad uno stimolo privo di significato intrinseco, ad esempio una parola scritta, il nostro cervello può accedere ad un'arbitraria, complessa e multidimensionale rappresentazione chiamata *rappresentazione semantica*. Nonostante decenni di indagine neuropsicologica e di studi di neuroimmagine sui correlati cognitivi e neurali delle rappresentazioni semantiche, molte domande sono, ad oggi, senza risposta. I lavori di ricerca presentati in questa dissertazione ambiscono a svelare uno di questi misteri: é possibile dissociare i substrati neurali delle diverse componenti del significato di una parola?

Il lavoro da me svolto si articola lungo due assi principali: uno teorico ed uno empirico. Nella prima parte, vengono riassunte le principali posizioni teoriche attualmente in auge relativamente ai correlati cognitivi e neurali delle rappresentazioni semantiche. Vengono inoltre evidenziati i recenti progressi metodologici, ovvero l'introduzione di metodi multivariati per l'analisi dei dati di neuroimmagine. Queste tecniche ampliano l'insieme di ipotesi che possono essere testate empiricamente, in particolare permettendo l'esplorazione (e la comparazione) delle geometrie rappresentazionali di diverse aree cerebrali. Tale passaggio é fondamentale ai fini di comprendere *dove* e *quando* le diverse dimensioni dello spazio semantico vengono attivate a livello cerebrale. Infine, propongo una distinzione euristica tra due tipologie diverse di dimensioni semantiche: da un lato quelle motorio-percettuali (vale a dire, gli attributi degli oggetti cui le parole si riferiscono che vengono percepiti attraverso i sensi), e dall'altro quelle concettuali (ad esempio, le informazioni frutto dell'integrazione di molteplici caratteristiche percettuali).

Nella seconda parte, vengono presentati i risultati degli studi che ho condotto al fine di indagare l'automaticità di recupero, l'organizzazione topografica, e le dinamiche temporali di diverse dimensioni motorio-percettuali e concettuali. Per prima cosa, mostro come gli spazi semantici ottenuti con diversi metodi comportamentali e computazionali (vale a dire, *Semantic Distance Judgment, Semantic Feature Listing, WordNet*) siano altamente riproducibili attraverso i soggetti e correlino tra loro. In secondo luogo, presento i risultati di quattro esperimenti di priming semantico che illustrano come le dimensioni percettuali (ovvero la dimensione fisica ed il suono emesso dall'oggetto cui la parola si riferisce) vengano recuperati in modo automatico durante la lettura, con importanti differenze a seconda

del compito svolto dai soggetti. Inoltre, grazie ai risultati di un esperimento di risonanza magnetica funzionale, illustro un gradiente occipito-temporale lungo la via visiva ventrale: le caratteristiche percettuali appaiono preferenzialmente codificate in aree visive primarie, quelle concettuali in aree associative temporali. Questo risultato indica che dimensioni complementari dello spazio semantico sono codificate in modo distribuito e parzialmente dissociato attraverso la corteccia cerebrale. Infine, mediante uno studio di magnetoencefalografia, dimostro come le diverse dimensioni dello spazio semantico possono essere recuperate in modo pressoché immediato (nei primi 200 ms dopo l'apparizione dello stimolo) e simultaneo, grazie però a diversi aspetti del segnale cerebrale. La coerenza di fase appare infatti fondamentale per la codifica delle dimensioni percettive, mentre le variazioni spettrali sembrano supportare la codifica delle dimensioni concettuali. Nel complesso, queste osservazioni suggeriscono che i substrati neurali delle diverse componenti del significato dei simboli, pur condividendo una simile evoluzione temporale, possono essere dissociate a livello della localizzazione cerebrale, e della caratteristica del segnale necessaria per codificarli.

Il presente manoscritto si compone di sei capitoli: i primi due introducono la teoria ed i metodi sfruttati per il lavoro sperimentale svolto, mentre i seguenti tre descrivono tale sforzo empirico. L'ultimo capitolo riassume i risultati, discutendone le implicazioni teoriche e proponendo possibili sviluppi. Infine, l'appendice include tutti i materiali di supporto, tra cui le analisi che usate come controllo della qualità dei dati o come supporto ai risultati principali illustrati nei capitoli precedenti. Segue dettaglio dei cinque capitoli principali.

Il **Capitolo 1** mira a definire esplicitamente il campo di indagine (quale argomento verrà affrontato) ed ad operazionalizzare le variabili in gioco (come verrà affrontato). Per prima cosa, definisco le rappresentazioni semantiche descrivendone i contenuti e fornendo prova della loro rilevanza come realtà psicologica e neurologica. In secondo luogo, riassumo le ipotesi sulla localizzazione dei loro correlati neurali alla luce dei risultati sperimentali publicati in letteratura. Le rappresentazioni semantiche sono distribuite su una vasta porzione della corteccia o localizzate in determinate aree chiave? La loro organizzazione è dettata da principi evolutivi o vincoli anatomici? In terzo luogo, descrivo i risultati relativi alla dinamica temporale con cui il sistema semantico viene attivato, considerando due scale temporali: a breve (es. obiettivo del compito svolto) ed a lungo termine (es. esperienze personali pregresse). Infine, evidenzio il rapporto tra il *contenuto* della rappresentazione, il *formato* adottato (ovvero le operazioni che possono essere eseguite), e l'*implementazione* sottostante

(ovvero il codice neurale). In particolare, questo capitolo comprende il mio contributo teorico: una definizione dello spazio semantico che, operazionalizzando due variabili fondamentali, può favorire sia le speculazioni teoriche, sia la ricerca empirica. Il significato delle parole viene concepito come una rappresentazione multidimensionale che comprende sia dimensioni motorio-percettuali (ad esempio, la dimensione media od il colore prototipo), che dimensioni concettuali (quali ad esempio la classe tassonomica). Questo capitolo, quindi, non solo offre una rassegna della letteratura sulle rappresentazioni semantiche, ma introduce anche il quadro teorico adottato per le seguenti indagini sperimentali.

Il **Capitolo 2** offre una panoramica dei metodi utilizzati durante il lavoro sperimentale, illustrando come le rappresentazioni cognitive e neurali possano essere studiate con compiti comportamentali (*Semantic Distance Judgment, Semantic Feature Listing, Semantic Priming*), nonché tecniche di neuroimmagine quali la risonanza magnetica funzionale e la magnetoencefalografia. In particolare, il focus é sui metodi multivariati per l'analisi dei dati di neuroimmagine che sfruttano algoritmi di *machine learning* (un approccio sovente chiamato *decoding*) o la correlazione tra pattern neuronali (chiamato *representational similarity analysis*, RSA).

Nel **Capitolo 3**, presento i risultati degli esperimenti comportamentali. Per prima cosa, ho condotto esperiementi di *Semantic Distance Judgment* e *Semantic Feature Listing* con il duplice intento di confrontare lo spazio semantico cui i due diversi metodi danno accesso, ed al contempo validare gli stimoli da utilizzare nei seguenti esperimenti di neuroimmagine. I risultati indicano che le diverse tecniche convergono nel descrivere il medesimo spazio semantico. In secondo luogo, ho condotto 4 esperimenti di priming semantico con l'obiettivo di chiarire il grado di automaticità con cui diverse dimensioni percettuali vengono recuperate durante la lettura. I risultati suggeriscono una delicata interazione tra il compito svolto dai soggetti e l'effetto di priming dovuto alla similarità tra parole in termini percettivi (ovvero se due parole si riferiscono a oggetti che condividono o meno la stessa caratteristica visiva od uditiva).

Il **Capitolo 4** ospita la descrizione dell'esperimento di risonanza magnetica funzionale condotto ed i suoi risultati. L'ipotesi testata era che le dimensioni percettuali e concettuali del significato di una parola siano codificate in differenti regioni del cervello: le dimensioni percettive in aree unimodali sensori-motorie, mentre le dimensioni concettuali in aree associative eteromodali. Ho cosi investigato la presenza di una mappatura tra una dimensione percettiva (la dimensione media nel mondo reale) e due dimensioni concettuali (due categorie tassonomiche ad un diverso livello di specificità), ed i pattern di attività cerebrale registrati in

sei aree lungo la via ventrale occipito-temporale. Grazie alla combinazione di tecniche di decoding ed RSA, ho evidenziato come la dimensione visivo-percettiva sembri essere codificata principalmente nelle regioni visive primarie (occipitali), mentre le dimensioni concettuali in regioni temporali più anteriori. Questo gradiente antero-posteriore, dal concettuale al percettuale, indica che diverse aree cerebrali codificano per dimensioni complementari dello spazio semantico.

Infine, nel **Capitolo 5**, presento l'esperimento realizzato mediante magnetoencefalografia ed i risultati cui ha condotto. L'ipotesi al banco di prova é che diverse dimensioni percettuali e concettuali possano essere dissociate non solo sulla base della loro topografia, ma anche in termini di dinamica temporale. Ho cosi confrontato una dimensione concettuale (la categoria semantica) e due dimensioni percettuali (una relativa ad un aspetto visivo, la dimensione media nel mondo reale, ed una relativa ad un aspetto uditivo, il suono prototipico). I risultati indicano un'automatico, rapido (~200 ms) ed essenzialmente simultaneo recupero delle informazioni lungo tutte e tre le dimensioni. Tuttavia, i tre diversi effetti sembrano dissociarsi in termini di quale dinamica cerebrale sia coinvolta (cambiamenti nella coerenza di fase in un caso, variazioni nello spettro di potenza nell'altro), e di quali sorgenti cerebrali ne siano responsabili.

Contribuendo alla nostra, ancora parziale, comprensione di come il significato delle parole sia codificato a livello cerebrale e recuperato durante il processo di lettura, i lavori presentati in questa tesi hanno importanti implicazioni metodologiche e teoretiche. In particolare, sottolineano l'importanza di una proficua integrazione tra teorie cognitive e metodiche statistiche avanzate, al fine di risolvere i misteri che circondano le rappresentazioni semantiche.

6. Acknowledgments

Specific acknowledgment for the experiments conducted (e.g., source of funding) can be found at the end of the relative chapter. Reading between the lines of this manuscript, one can grasp the most important part of these almost four years of graduate work: the learning process, the upheavals of my thoughts and abilities. I here express my sincere gratitude to all the people that witnessed it all (or part of it) from the frontline. Not to let all these people down is the honor and the burden I have to carry.

First of all, I'd like to thank my **advisors**. Manuela Piazza, who bravely bet on me and constantly pushed me further. Her "*allright, but what have learned about the brain?"* will always resonate in my ears. I can finally put pen to paper the anecdote I always tell about the first time I saw here. She was giving a talk on numerical cognition, I was a first year master student. I recall thinking "*this is the kind of (italian) (female) scientist I would like to become one day!*". I haven't changed my mind since. Christophe Pallier, for his incredible generosity and acumen. One of the very first time we interacted, as I arrived at Neurospin and started contributing to the lab wiki (fighting the embarrassment for my naïve codes), he told me I was "*playing a game he liked*". One of my goals is to keep doing so.

My **co-authors** deserve the most sincere and deep acknowledgment for sharing their expertise with me. Marco Buiatti, for teaching me everything I know about MEG, while putting up with (too frequent and rather silly) questions and the (overall very pessimistic) attitude I had. Evelyn Eger, my fMRI guru, for her intellectual honesty and great professionalism. She's been there for me, always, selfishlessly, and I cannot imagine how things would have gone otherwise. Fabian Pedregosa, for his priceless help diving into the world of statistical learning & python. He stroke a rare balance: never assuming I knew, never patronizing me. Finally, thanks to Lola de Hevia, Arnaud Viarouge, and Alexis Amadon.

Many thanks to the **jury members**, for the time and attention they devoted to my work. Not only I value their opinion and admire their research, but with hindsight I realize they became part of my committee the very first time I met them: James Haxby, who quite randomly attended my brown-bag meeting in Rovereto and immediately detected a commonality of interests; Rik Vandenberghe, who came at my poster at a big international conference and proved me that, as overawed as I was, I had something interesting to share; Marius Peelen, who, tirelessly organizing one CAOS edition after the other, awarded me with

the first prize of my graduate experience; Lauren Cohen, who followed my project since the first interim evaluation and paid me the best compliment ever: "*I wish I had done it!*".

I would also like to thank **Neurospin staff** and in particular all members (former and current) of **Parietal** and **Unicog** teams. I appreciate how much this environment has enriched me and forced me to grow. I learnt something from each and every one of you, mostly the hard way. Special mentions to Yann, Gabrielle, Marianne, Arthur, Benoit, Carole and the other co-organizers of the NeuroBreakfast (and the SpinDating(s)!) for their invaluable help thorugh the years. Same holds for Michael and Laetitia concerning the Unsupervised Decoding Metting. Many thanks also the community of Saclay Science at large for the incomparable experience that was the organization of News2016. Finally, thanks to INRIA for the great workplace it provides to *collaborators*. I wish I had had similar working condition all along. *Merci!*

Thanks to **CIMeC staff**, students and researchers for welcoming me when I was already overcooked, and for showing me that there is scientific hope in Italy. In particular, thanks to Paola Fortis, Serena Melison, Alessia Monti, Yuan Tao, and Simone Vigano. *Grazie mille!*

Keeping it honest and real, I'd like to thank also the **communities** behind *stackoverflow.com*, *scikit-learn.org*, *martinos.org/mne*, *neuroimage.usc.edu/brainstorm*, *pymvpa.org*, *fieldtriptoolbox.org*, and many scientific bloggers around the world. You have no idea how much you helped me, more than anyone else. If I have stood upon the shoulders of giants, it has been by reading blogposts, mailinglists and oftent just twits. If you ever contributed (even with few lines of code) to the internet-based knowledge about machine learning, programming, or neuroimaging: *kudos*!

Heartfelt thanks to all the **colleagues** that have also been good friends, putting up with me over and beyond what I deserved: Aina (kind and spontaneous as I wish I could be), Alex (and the way he cares), Ana Luisa (and her enviable proactive and positive attitude), Andres (for his company – and life-saving croissaints - during the thesis sprints), Benoit (my favorite French guy, ever), Darinka (a constant, unique, source of inspiration and support), Elisa (for her ability to smile – and make you smile - no matter what), Elodie (for sharing my clinical mindset, which I needed so badly), Gabriela (gentle companion, no matter whether in person or via emails), Kamalakar (the kindest nilearn guru), Martin (challenging and stimulating, just as I like it), Michael (my, perhaps unknowingly, friendly nemesis), Medhi (always thoughtful and sympathetic), Parvaneh (who examplifies the strength of candor and kindness), Pedro (for many things, not only the Matlab 101 crash course he asked me to mention), Ramon (infinite

source of positive thoughts), Rodrigo (and the science-bat-cave dream – which will never die), Solveig (the best conference roommate one can ask for), Thiago (and the late afternoon buses in the *Vallée de Chevreuse*), Yi-Chen (they don't make better friends & labmate!). Shout-outs also to those that I crossed only briefly and those that made my Parisian life outside the lab much better: André, Benedetta, Bianca, Emilie, Hernik, Idoia, Fosca, Marta, Mainak, Milad, Zafer. To all the ones I'm missing (because memory is terrible thing): I kept a journal, I'll be reminded of you eventually and thank you even more for your understanding!

Thanks (o better *merci* and *grazie*) to all the **volunteers** that participated in my experiments (more than 380!).

Grazie a Livia, l'unica che c'é sempre stata ed ha sempre capito. A Nicola. Ed a tutti gli **amici della *bassa***, che hanno aiutato il costante ricalibraggio delle prospettive: Arianna, Davide, Elena, Federica, Irene, Marcello, Marco, Marianna, Margherita, Massimiliano. Ai compagni d'università rimasti a portata d'abbraccio, foss'anche virtuale: Antonella, Beatrice, Claudia, Gaia, Ilaria, Karin, Maria, Martina, Sara, Valentina.

Il ringraziamento più sentito é quello verso la mia **famiglia**. Non importa quanti titoli mi daranno, quanti chilometri ci separeranno: per voi sarò sempre quella che cade anche da seduta (e voi sarete sempre li a ridere con me mentre mi fate rialzare). L'amore e supporto incondizionato di cui siete capaci sono la vera fonte della mia forza.

Por fin, gracias **Fabian**. I'm glad we met *before*, when I was an enthusiastic naïve girl. I'll be forever grateful you sticked around *during*, constantly repeating I was doing fine, I was going to make it, you would have loved me anyway. Deciding you were the right one to spend the *ever after* with, was overall an extremely easy decision. Trying to live up to you makes me a better scientist, but most importantly a better person. Nadie puede decir, en toda honestà, de merecer a friend, un compagno, un marido asi.

CHAPTER 1:
A MULTIDIMENSIONAL REVIEW OF THE LITERATURE

Could a machine think? The answer is, obviously, yes.
We are precisely such machines.
[Searle, 1980]

In this chapter, I explore the current state of the literature concerning the neuro-cognitive representations of semantic representations. First, I illustrate the role and properties of representations via examples stemming from sensory-motor systems. Then, I focus on the defining properties of semantic representations: their what (i.e., content and geometry), where (i.e., topographical organization), when (i.e., temporal dynamic) and how (i.e., format and implementation). The key findings from behavioral and neuroimaging experiments, as well as some of the key open questions, are presented. A subset of this chapter is currently under revision as a review paper:

Borghesani, V., & Piazza, M. (*under review*). The neuro-cognitive representations of symbols: the case of concrete words. *Neuropsychologia*

Highlights:

- Semantic knowledge is a complex cognitive and neurological reality, central to human nature.
- Concepts are represented across the neo-cortex in a distributed, yet specialized manner.
- Processing of semantic information is fast and automatic, yet not uniform.
- The question of the format of semantic representations is currently an ill-posed problem.

1. Neuro-Cognitive Introduction to Representations

As this thesis concerns cognitive and neural representations of semantic knowledge, I will begin by defining the concept of *representation* and, in particular, the properties of *neural representations*. This will be followed by the exploration of what we mean by *knowledge* and by *semantics*.

1.1 Cognitive Representations

A cognitive representation is a mental state, a mental information-bearing structure, that corresponds to an aspect of the external reality (e.g., a stimulus) or an internal state (e.g., being hungry). We can consider it the product of a function that maps the complexity of the external or internal world onto mental activity. Mental representations, banned from scientific psychology by behaviorists, who believed experimental efforts should be restricted towards what could be directly observed, were revived by cognitive psychologists and computer scientists. For them, representations are systems of symbols isomorphic to what is represented, such that conclusions drawn by processing symbols are valid inferences about the represented structure (Gallistel, 2001). Several aspects of representational systems have been problematized in the last three decades, with different perspective being taken (e.g., (Cummins, 1989). Introducing a topic developed later in the chapter, I here only briefly mention the crucial debate on the format of cognitive representations: is mental content stored in a symbolic, descriptive format or a depictive, pictorial one (see Fig. 1)? Intuitively, some concepts are better represented pictorially (e.g., *"red*"), some verbally (e.g., *"goalkeeper*"), others pose problems for both formats (e.g., *"justice*").

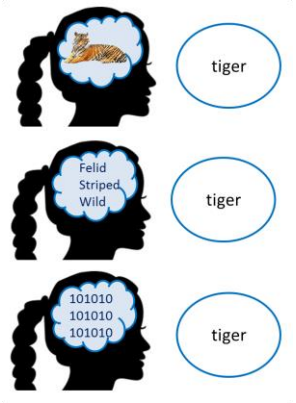


Figure 1 Tentative representation of the concept of representations. The same concept, for instance "tiger", can be stored via different representational systems: a pictorial depiction (upper), a verbal description (middle), or an abstract code (lower).

Considering representations as the codes that store information, we must distinguish cognitive representations from the cognitive processes that operate on them (i.e. that make use of that information). Cognitive neuroscience aims at describing the neural correlates of cognition in terms of both processes and representations (see Fig. 2). Indeed, authors working on mental representations, even if coming from different perspectives, agree on the necessity to analyze both sides of any representational system. The processes operating on a given representation appear to be an essential part of its definition (Marr, 1982), and thus any claim on that representation cannot be evaluated unless the processes operating on it are specified

as well (Anderson, 1978). Ultimately, it could be argued that representations and processes are indistinguishable, not only philosophically (i.e., would it make sense?), but also methodologically (i.e., can we really study one and not the other?).

In the neuropsychological literature, before the advent of neuroimaging, the problem of distinguishing between representations and processes was framed as the dissociation between a deficit "of access" (i.e., of the processes that give access to the representation) and a deficit "of storage" (i.e., of the representation itself) (see for instance Rapp and Caramazza, 1993). Concerning the object of the present thesis, semantics, this dissociation was supported by the description of two syndromes whose deficits selectively affect the level of processing (semantic aphasia) or of representations (semantic dementia) (for example see Corbett et al., 2009). More generally, comprehensive theories of semantic cognition attempt to explain both systems: that of semantic representations and that of semantic control (Lambon Ralph et al., 2016). In this thesis, I will focus only on the cognitive and neural correlates of semantic representations.

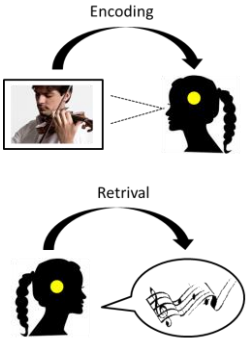


Figure 2 Representations vs processes. While listening to a piece of music, a process is in action: you are *encoding* (some of) the features of the melody. They are later available for *retrieval*, i.e. the process of accessing them in order to, for instance, repeat the tune to someone else. The first process created (or modified, if pre-existing) a representation of the melody, the second is reading it out to fulfill the task at hand.

1.2 Neural Representations

We commonly use the term *neural representations* to refer to the neural underpinnings of cognitive representations. They are the brain states product of a function that maps the external or internal world onto brain activity. Simplifying for the sake of clarity, these representations can be described answering the following, interrelated, questions:

a) *What* is the content of the representation and how is such content organized? Conceptualizing different entities as points in a multidimensional space, we can describe a representational geometry, i.e. the relationships (distances) between them.
b) *Where* is the representation stored in the brain? Answering this question requires the description of its topographical

organization, in terms of cortical and/or subcortical areas involved.

c) *When* does the content of the representation become available? Can we describe the temporal dynamics affecting the representation?

d) *How* is that content stored (i.e., what is the representational format), and how is it implemented (i.e., what is the underlying neural code)?

In some areas of cognition, neural representations have been described with great detail. As a prototypical example, let us consider the first cortical representation of the visual world as it is transduced by our eyes. The signal hitting our retina (*what*) is processed in primary visual areas - V1, calcarine cortex, occipital pole – (*where*), very fast – in the order of a few milliseconds, independently from the content (*when*), in a retinotopic fashion (Sereno et al., 1995) (macro-scale *how*), thanks to the firing of edges-detector neurons (Hubel and Wiesel, 1959) (micro-scale *how*) (see Fig. 3). Another prototypical example is the representation of motor and sensory information about our body parts in a somatotopic fashion (macro-scale *how*). Information about body movements (*what*) is encoded in primary motor area, M1 (*where*), with a constant rapid update (*when*). Likewise, the information about the state of our body (*what*) is rapidly (*when*) encoded in primary sensory area, S1 (*where*) (Penfield and Boldrey, 1938) (see Fig. 4). Neural representations of discrete sensory systems have been extensively studied and reviewed, not only in the case of sensory and motor representations cited above but also, for instance, of sounds (Rauschecker, 1998) and odors (Laurent, 1996).

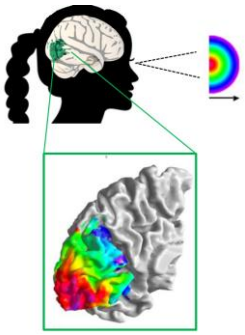


Figure 3 Retinotopy in primary visual areas. In V1, the information on the outside world is retinally mapped onto the cortex. [adapted from Dougherty et al. (2003) Journal of Vision]

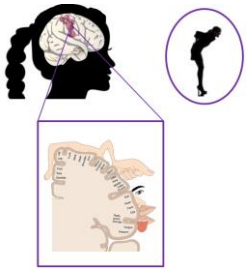


Figure 4 Somatotopy in primary sensory areas. In S1, each cortical area corresponds to a specific body part (motor homunculus) in a medial-to-lateral topographical mapping from the lower to upper body. Most sensitive areas (e.g., fingers) are overrepresented. [adapted from OpenStax College - Anatomy & Physiology http://cnx.org/content/col11496/1.6/]

However, as human beings, we do not only create an internal representation of the images hitting our retina, the sound waves reaching our ears, or the smells coming through our nostrils. We mentally represent very complex instances of both the external and the internal world. We can create a representation of everything that we experience in the world around us (e.g., objects, social roles, natural kinds), but also of things that we have never (and perhaps will never)

directly encounter, for instance things lacking a correspondence in the external world (e.g., unicorns, vampires, cylones). These different internal representations have been studied with varying degrees of depth. For instance, recently I have explored the concept of quantities (e.g., small magnitude, medium magnitude, big magnitude) across different dimensions (i.e., applied to numerosity – varying number of dots- and extension – lines of varying length) (Borghesani et al., 2016). To this end, we capitalized on some promising advances in neuroimaging data analyses that have the unprecedented potential of shedding light on the neural substrate of cognitive representations (see Chap. 2.4).

In this thesis, I am interested in one specific kind of representation, the semantic one, which has several exceptional properties. The first one is that it can be accessed by inputs coming from any sensory modality. Reading the letters /t i g e r/, hearing the sound /ˈtiːər/, seeing the picture of the stripped animal, are all means by which the concept of tiger would be triggered (see Fig. 5).

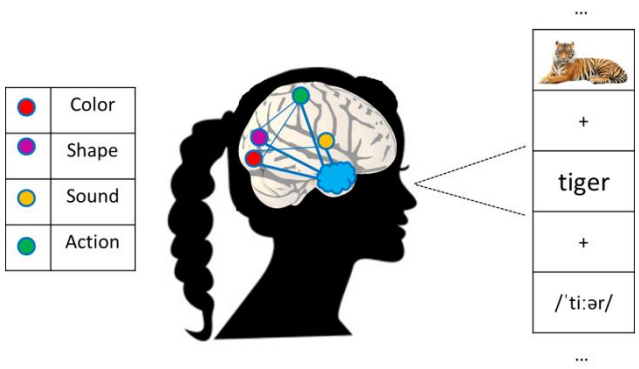


Figure 5 Semantic Representations. The concept of tiger can be accessed when prompted with stimuli of different nature, e.g., the picture of a tiger, the word /tiger/, and the sound /ˈtiːər/. It has been hypothesized that the different features that build up the multidimensional concept tiger are encoded in different brain regions located in proximity to primary motor-sensory regions. These different components of meaning are integrated by one (or more) hub(s) located in associative cortex.

However, our understanding of this kind of representation is fuzzy. Its content is often vaguely defined: what do we mean exactly when we talk about semantic representations? What is the geometry of the semantic representational space? What are the neural underpinnings of semantic representations? Are they distributed in the

cortex, or are they stored in one comprehensive warehouse of concepts? Finally, what are the representational format, and the underlying neural code of semantic representations? In the following paragraphs, I will approach all these questions, one by one

2. The Content of Semantic Representations?

What do we mean by semantic knowledge? Different perspectives can be taken to answer this question, as the term *semantic* occupies a prominent role in many different (soft and hard) sciences. As I believe that they all contribute (or should contribute) to the current discourse on semantic representations, I will briefly introduce the main inputs from the relevant scientific fields.

2.1 Etymology and Philosophy

Being passionate about words often leads to a certain affection for etymology: that combination of letters, which means so much to you, how far did it travel? How did it get here? In ancient Greek, the verb *to know*, *οἶδᾰ́* [/ói̯.da/], was derived from the past perfect of the verb *to see, εἴδομαι* [/e͡ː.do.mai̯/] (indicative present: *ὁράω* [/ho.rá.ɔː/]). Similarly, the root of the English term *know* can be traced to the Old English *cnawan* (sharing roots with the Latin *gnoscere* and the Greek **gno*) and means "*perceive a thing to be identical with another*", "*perceive or understand as a fact or truth*" (as opposed to believe): again, perception is in the spotlight. Thus, etymologically speaking, at least for Indo-European languages, to know is to have seen: our knowledge is the outcome of our (visual) experiences. The term semantic, on the other hand, stems from the ancient Greek *σημαίνω* [/sɛː.mai̯.nɔː/] which means *to symbolize*, *to mean*. A *σῆμα* [/sɛ́ɛ̯ma/] is a mark, a token, a pointer. Therefore, semantic knowledge is the collection of all the tokens - and all the things the tokens refer to – that we have learned, that we have

experienced. The orthographic form ROME, and its phonological form /ˈroʊm/, are tokens referring to the capital city of Italy. A picture of the Colosseum would likely evoke the same general concept, perhaps highlighting the fact that Rome was the capital of the Roman Empire. If your personal knowledge of Rome also includes its traditional *cuisine*, those same images and words will additionally remind you, for instance, of a great lamb dish.

Can knowledge be reduced to our bodily experiences? The history of philosophy of science is studded with authors spreading over the continuum between *empiricism*, *idealism* and *rationalism*. According to the proponents of the first view (e.g., Thomas Hobbes, John Locke, David Hume), at birth our mind is a *tabula rasa*, ready to be filled with knowledge acquired through sensory-motor experiences: "*No man's knowledge here can go beyond his experience.*" in John Locke's words. Idealist authors (e.g., Plato, Kant) believe that we are born with innate ideas, core conceptual knowledge that does not require any learning processes. Finally, rationalists (e.g., Descartes, Spinoza, Leibniz) refute the identification of knowledge with perception, and state that the former can be derived from reason independently of any sensory data. Until recently, most of traditional Western philosophy has embraced Descartes' mind-body dualism: a clear-cut divide between mental and physical properties. Naturalists and pragmatists have paved the way for the so-called embodied cognition flow that has radically changed the way mind and body are thought to interact (Johnson, 2006): the mind is not a separate entity, but an emerging property of the interaction between the body and the environment.

Philosophically, there are thus two topics of discord: the relationship between mind and body, and the origin and nature of knowledge. The branch of philosophy concerned with the theory of knowledge, epistemology, sees a fourth position: that of *skepticism*. These authors (Socrates in primis) argue that a questioning attitude and the suspension of any judgment should be preferred, while

critically evaluating all the evidence. This is the perspective I will take throughout this thesis.

2.2 Linguistics and logic

In linguistics, semantics is defined as the study of the meaning of all linguistic expressions (i.e., morphemes, words, phrases and sentences) (Bréal, 1904). The link between words (symbols) and meaning (concepts) is far more complex than what can be superficially appreciated, complicated by phenomena such as polysemy (i.e., a sign has multiple meanings, related by contiguity within a given semantic field) and homonymy (i.e., a sign has multiple meanings, totally unconnected or unrelated) (see Fig. 6). Semantics should not be confused with pragmatics, the study of "speaker meaning", in other words the meaning of language in its context of use. This distinction parallels the one made in cognitive and clinical psychology between semantic representations (the information stored) and processes operating on them (which will determine changes according to the context/behavioral goals). While aware that pragmatic and contextual factors affect semantic representations, in this manuscript I focus on static representations of the meaning of single words

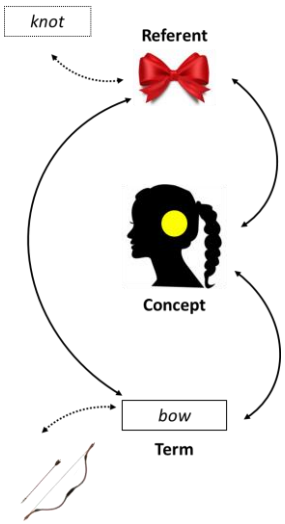


Figure 6 The link between words and meaning. A given concept can be expressed via a term (i.e., a word) and refers to a referent (i.e., an object in the real world). However, this linear relation is complicated by the observation that the same term might refer to different concepts (e.g., bow is "*a flexible strip of wood, bent by a string stretched between its ends, for shooting arrows*" but also a "*piece of looped, knotted, or shaped gathering of ribbon, cloth, paper, etc., used as a decoration*"). Moreover, the same referent might be accessed via different terms (e.g., a bow can be called knot).

The origin of modern semantics is usually traced back to the point of intersection between the logico-philosophical tradition and structural and generative approaches. Belonging to the first line of research, Frege (1982) distinguished between the *reference* (in German, *Bedeutung*) and the *sense* (in German, *Sinn*) of a concept. The first denotes a word extension (i.e., what it corresponds to in the world), the second its intension (i.e., what we know about its meaning, the way in which it refers to its referent). For instance, the sentences "*Bruce Wayne is Batman*" and "*Bruce Wayne is Bruce Wayne*" have the same referent/extension (i.e. the American billionaire owner of Wayne Enterprises), but rather different sense/intension (i.e. only the first one denotes knowledge of his secret identity). The second line of

research is best represented by Chomsky who stressed the dissociations (and interactions) between semantics and other aspect of language, such as syntax and grammar. As exemplified by its notorious sentence, "*colorless green ideas sleep furiously*", grammatically correct propositions can be completely meaningless.

Mirroring what pointed out in philosophy of knowledge, according to Buccino and colleagues (2016), we can talk about two streams in the current philosophy of language: *externalist* (i.e., the meaning of words resort to external entities - physical or social) and *internalist* (i.e., embodied experiences). As an example of the first perspective, consider Putnam's (1975) claim that the meaning of a word is given not only by the set of items it refers to (the *extension*), but also by the socially defined notion of its typical features (the *stereotype*). This kind of reasoning, along with notions such as the one adopted by Frege, implies that meaning does not follow from what speakers perceive or experience (their psychological state is irrelevant), but rather from some kinds of (physical or social) external entities like *senses* (Frege) or *stereotypes* (Putnam). The opposite perspective, the internalist positions, started with the observation by Russell (1910) that we can understand only those expressions we are "*acquainted with*". Sure, one can be taught of *hobbits* - short and fattish, with curly hair and a round jovial face -, however this description will be understood only by those that are familiar (i.e., had been exposed to) with the terms "short", "fattish", "curly", "hair", etc... We will see that this tension between internal and external sources of meaning expands to cognitive (neuro)science.

2.3 Computer Science and Artificial Intelligence

Insofar as to *know*, to have *knowledge*, is seen as tightly linked with being an intelligent agent, computer scientists have debated about the characteristics of semantic memory (perhaps unknowingly). As we will see later on (3.1), one of the first, pivotal, cognitive models of the organization of semantic knowledge stems from a hypothesis

generated by a computer scientist, Quillian (1967). However, the contributions I will review here focus on a deeper question: what is knowledge in the first place?

An important divide in the field of artificial intelligence (AI) should be mentioned. On one hand, the weak AI hypothesis states that a machine running a program will always be, even at its best, only capable of *simulating* real human behavior and consciousness. On the other hand, the strong AI hypothesis states that a machine running the proper (yet to be coded) program, would *be* a mind, thus positing no difference between a software emulating the actions of the brain, and the actions of a human being, including understanding and consciousness. The American philosopher John Searle (1980) responded to strong artificial intelligence advocates with what is now known as the Chinese room argument. The idea in vogue at the time was that intelligence in computers could be assessed with the so-called Turing test. Human beings are asked to have chat conversations with unknown interlocutors; if a machine, acting as interlocutor, can fool humans in thinking they are chatting with a conspecific, that machine can be considered intelligent. However, noticed Searle, pure symbols manipulation, in absence of any meaningful comprehension, cannot be considered knowledge, cannot be enough to call a system "intelligent". Provided with the right tools (e.g. a Chinese vocabulary and a textbook of Chinese grammar) one can manipulate Chinese symbols correctly, up to the point of fooling native speakers. However, it would just be a *simulation* of knowledge; there would not be any real *understanding*.

Ten years later, the Hungarian cognitive scientist Harnad formalized the core problem of semantic knowledge: symbols need to be grounded (1990). Grasping the meaning of something is the result of the capacity of picking out a referent in the outer world and of being conscious of such a process. Symbols need to be grounded, and this is not simply a computational property, it is a dynamic implementation-dependent property (i.e., it will depend on the sensory-motor states the system can experience) (Harnad, 2003). His

conclusion is thus that a complete separation between a central symbolic system and peripheral input/output systems is not sufficient to give rise to human-like intelligence (and knowledge). Some kind of interaction between hardware (i.e., the input and output systems) and software (i.e., the symbolic system) appears to be necessary. How this could be implemented is still an open question, and, as we will see later, parallels the open question on how such an interaction is carried out by our brains. While slightly changing its meaning over time, one term has been used to refer to the complex systems of inputs, outputs and symbolic operations characterizing biological life forms: *wetware*.

2.4 Psychology and Neuropsychology

The term *semantic memory* was coined for the first time by Quillian in his doctoral thesis (1966). In his seminal work, Tulving (1972) then formalized the distinction, within the declarative (i.e., consciously accessible) long-term memory system, between:

a) Episodic memory is tied to precise spatio-temporal coordinates, to unique personal events one *remembers*. For instance, I recall yesterday (*when*) I wrote one paragraph (*what*) while on the train back home (*where*). Episodic memory is thought to be dependent on medial-temporal lobe structures (MTL), while prefrontal cortex (PFC) seems to support strategic retrieval (Agosta et al., 2016).
b) Semantic memory is a mental thesaurus, containing all the general concepts one *knows*. As example, I know the most ancient recognized predecessor of the rail system was the rutway near Corith (the *Diolkos*). As described in the rest of chapter, semantic memory relies on a distributed network of cortical areas.

Semantic memory is thus defined as the general knowledge of facts (e.g., 25th August 1991, first public announcement of the existence of a Linux kernel), people (e.g., Tolkien, the author of "*The Lord of the Rings*") and objects (e.g., an astrolabe is an ancient inclinometer). Items are described both in terms of features (i.e., how things are – from the example before: typically made of brass) and functions (i.e.,

what things are for – from the example before: used for navigation and locating astronomical objects). Semantic memory is tightly linked with language, as it includes word meaning, and it is shared within a given cultural milieu (e.g., in Bologna, nobody understands what "*spaghetti alla bolognese*" means, it simply does not exist).

The early investigations on semantic memory were of neuropsychological nature. Cognitive neuropsychology seeks to understand the relationship between cognitive functions (as described by cognitive models) and brain areas and functions (as studied via the observation of patients with acquired brain damage). As I will develop later, single case studies of patients with memory deficits revealed important dissociations, which led to key inference on the structure of the semantic system (Caramazza, 1986). A single dissociation (i.e., a patient showing a deficit of semantic memory, not of episodic memory) can only demonstrate that the two constructs are somehow different, they cannot be reduced to one another. However, it does not provide any indication on what distinguishes them. For instance, if one of the two poses a higher demand on attention (or language, or any other cognitive function) this could explain away the difference. Nevertheless, if a double dissociation is observed (i.e., two patients, one showing a deficit of semantic memory and preserved episodic memory, another with the reverse pattern) then it is possible to conclude that the two constructs are functionally different (i.e., the differences cannot be reduced to, say, higher/lower attentional demands). Moreover, if the two patients are also showing different patterns of brain anomalies that can be linked to the cognitive deficits, then it is possible to conclude that semantic and episodic memory differ not only functionally, but also in their neural substrate.

Thanks to neuropsychological investigations, since the mid-seventies it is acknowledged that episodic and semantic memories constitute two dissociable cognitive and neurological realities. In 1975, Warrington described three patients showing a "*selective impairment of semantic memory*". Episodic memory was preserved. Ten years later, Mesulam (1982) described six more patients showing

this peculiar deficit of memory not imputable to Alzheimer's Disease and coined the term progressive fluent aphasia (Mesulam, 1987). From these first reports stemmed two interconnected and prolific lines of research: on one hand, the study of categorical dissociation within semantic memory (Warrington and Shallice, 1984), on the other hand, the definition of a syndrome called semantic dementia (SD) as one of the forms of progressive fluent aphasia (Snowden et al., 1989; Hodges et al., 1992; Neary et al., 1998). We will later see how the bulk of evidence stemming from this neuropsychological-oriented research has contributed to the investigation of the neural substrate of semantic memory. For one of the first examples of the mirror dissociation (i.e., spared semantic memory and impaired episodic one), see (Vargha-Khadem et al., 1997)

2.5 Dimensions and Geometries

Semantic representations lie in a complex multidimensional space described by the intersection of numerous features. We have recently proposed a novel way to conceptualize the mental representation of the meaning of concrete words, which we think could be a useful heuristic to foster theoretical speculations as well as empirical research (Borghesani and Piazza, under review). Considering the way they are learned/acquired, we can distinguish (at least) two kinds of features:

- *Motor-perceptual ones.* The umbrella term motor-perceptual features includes all features of the objects referred to by the words that can be (and typically are) perceived through the senses. These features, under normal circumstances, are apprehended through direct physical interaction with the items. It comprises modality-specific features, for instance aspects solely apprehended through vision such as color (e.g., a tomato is typically red), purely gustatory such as taste (e.g., a tomato is a particular combination of acid and sugar flavor), purely auditory such as sound (e.g., a tomato is not associated with

any specific sound). Moreover, it encompasses features that can be equally resolved via multiple sensory systems, such as the average size or shape (e.g., the average size and shape, which can be sensed both through vision and through touch). Finally, some classes of concrete nouns (for example tools) are also defined through action descriptors, hence the reference to a combination of motor-perceptual features. As these kinds of features are constrained by the physical laws of the world we live in, and they quite often correlate among each other (e.g., small objects tend to produce high pitch sounds; green food tends to be acidic, thin small objects can be grasped with precision grip). According to our proposal, however, they can and should be considered separately when attempting to describe the neural substrate of word meaning.

- *Conceptual ones.* These higher-order descriptors constitute another key dimension of the semantic space, and are either (1) derived from the integration of multiple motor-perceptual features, and thus refer to multimodal aspects of item (e.g., I know a tomato is a fruit, which is a largely cultural label I learned to attach to things that are edible and have seeds) or (2) learned explicitly in a declarative fashion, as they bear no direct link with any motor-perceptual feature (e.g., I know tomatoes were not cultivated in Europe before the discovery of the Americas). Whether this latter case, in which a given feature cannot be entirely resolved by the integration of motor-perceptual ones, should be classified separately (and which term should be used to refer to it) is currently an open question. Future work should also attempt to investigate how the integration of unimodal motor-perceptual features (e.g. *yellow* + *acidic* + *small* + *round* = *lemon*) is implemented, and how it differs from the integration of symbols referring to two or more integrated features (e.g. *lemon* + *Italian* + *liqueur* = *limoncello*).

One clarification with respect to the concept of modality is needed. When talking about semantic knowledge, it is important to distinguish between:

- input modality of the stimulus (e.g., the picture of a tomato, the smell of lavender), determined by the sensory organ that transduces the information, and
- content modality which is the modality specific component of the representation (e.g., the color of a tomato is red)

As we have previously seen with the example of Rome, the content of semantic knowledge can be accessed via stimuli of any modality: visual (e.g., a picture, a written name), auditory (e.g., a spoken name, a sound), olfactory (e.g., a smell), etc… Particularly interesting is the case of words, arbitrary symbols whose physical properties (i.e. strokes on paper or vibrations of the air) greatly differ from the semantic content we have access to. As a matter of fact, the written (orthographic) and spoken (phonological) surface form of words carry meaning only thanks to cultural conventions. Given the heterogeneity of ways by which semantic knowledge can be accessed, when assessing semantic memory one needs to exploit a rich set of tests. A neuropsychologist's aim may be to reveal a core semantic deficit (i.e., deficit of the semantic representations) as opposed to, for instance, an impairment preventing the access to the information or the production of the response (i.e., processes acting upon the representations). Thus, neuropsychologists use tests relying on both visual and auditory inputs (verbal and non-verbal), some of which ask for complex answers (requiring good motor skills or good verbal skills) while others probe a simple yes/no answer or a binary choice (see Fig.7).

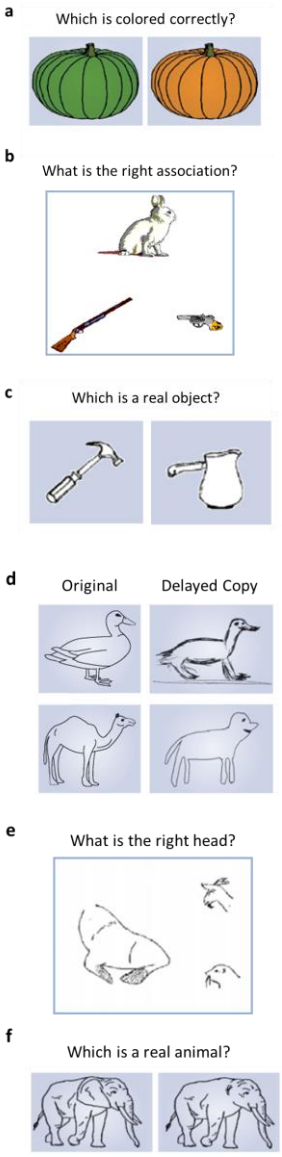


Figure 7 Testing semantic knowledge. Subjects' semantic memory for artificial as well as natural kinds can be tested by asking: (a) to select the image which presents the correct prototypical color; (b) to choose which elements are linked by a semantic association (e.g., based on functional links); (c) to select the plausible item, one of the two being the unrealistic merge of two different common objects, e.g., a knife and a kettle); (d) to draw a delayed copy of simple animals (notice the absence of any key feature that would allow identification of the animals); (e) to choose the correct missing piece; (f) to select the right image.

To sum up, *semantics* is the branch of linguistics studying the meaning of the different linguistic expressions and *semantic knowledge* (or *semantic memory*) is the memory system dedicated to store information on meaning of words and, more generally, our knowledge of the world. Philosophy, experimental psychology,

cognitive science and computer science have greatly contributed to the debate on the origin (i.e., how much is innate/learned?), structure (i.e., which concepts cluster together?), and implementation (i.e., how to achieve the needed interaction between symbols and input/output systems?) of semantic knowledge. With these questions in mind, we are now ready to explore its neural substrate: neural semantic representations.

3. Organization and Localization of Semantic Representations

I have stressed the tight link between language and semantic memory. Decades before any formal definition of semantic memory, conceptual knowledge was already included in the first models attempting to describe the language system. It all started with the pivotal descriptions of patients with selective deficits of production (Broca, 1861; Broca, 1865) and understanding (Wernicke, 1874/1977) of language. Following these accounts, traditional models of language (Lichtheim, 1885), have posed that a center for speech production (so called Broca's area) and a center for speech comprehension (so called Wernicke's area) are connected to a *concepts center/ideation center*, where meaning is stored (see Fig. 8 and 9). No attempt was made to precisely localize this center.

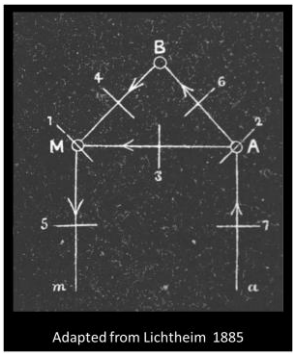


Figure 8 Lichtheim's diagram. The center of auditory images (A) and the center of motor images (M) are connected both by a direct pathway and by an indirect one, going through the center of concepts (B).

During the XX and XXI centuries, alongside the progress of studies on language comprehension and production, much work has been conducted in the attempt to localize the *concepts center(s)*, the neural substrate of semantic knowledge. Neuropsychological studies of patients manifesting semantic deficits have been pivotal in shedding light on the possible cognitive and neural dissociations. They helped develop most of the cognitive theories later tested with neuroimaging methods. I here review the milestones of both the neuropsychological and the neuroimaging perspective, after a brief excursus on cognitive

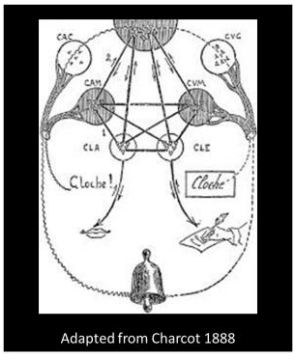


Figure 9 Charcot's bell diagram. The auditory center for words (CAM) and the visual center for words (CVM) are connected respectively with the common auditory center (CAC) and the common visual center (CVC), both of which lead to the ideation center (ID).

and computational models of semantic memory. These models, while not necessarily detailing the possible neural implementations, have paved the way for many of the following approaches.

3.1 Cognitive and Computational Models

At the end of the sixties, Quillian proposed a model of how denotative, factual information can be stored in a computer (and in the human mind) through a semantic network (Quillian, 1967). In this model, concepts are stored as series of nodes and associative links between those nodes. Links usually go both ways between concepts, but with different *criterialities* (i.e. they can be more or less essential): for example, it is highly criterial for the concept of ukulele that it is a musical instrument, and not very criterial for the concept of musical instrument that one kind is ukulele. The computational model's predictions were later tested behaviorally in collaboration with Collins (Collins and Quillian, 1969). Retrieving properties from a node and moving up in the hierarchy of links requires time, thus comparing the processing time of different words/sentences permits the understanding of how they are organized (one relative to the other) in the semantic network (see Fig. 10).

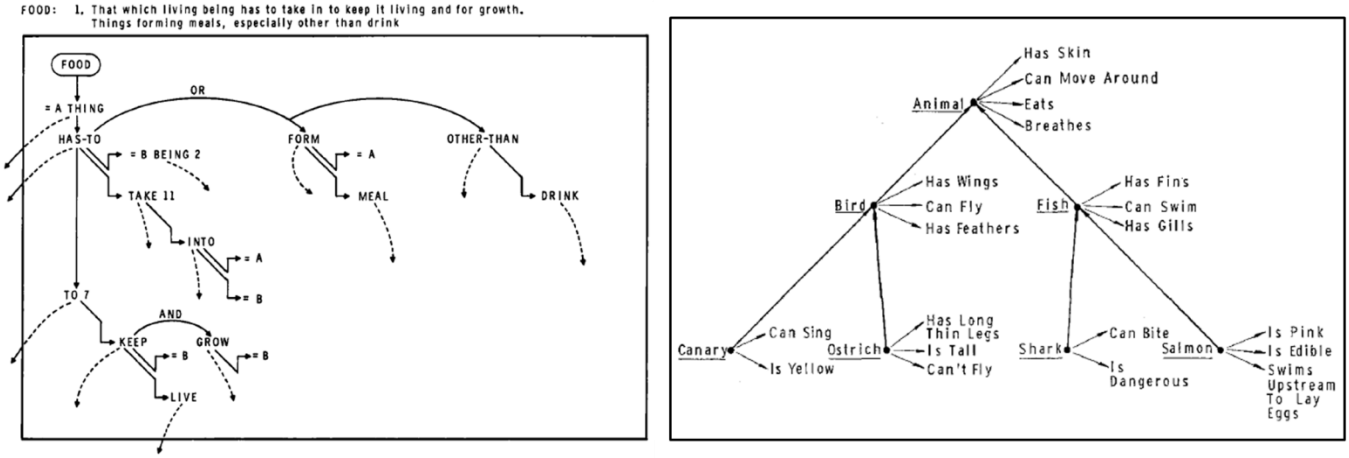


Figure 10 Representation of the semantic network proposed by Quillian. Left: original computational model characterized by type node ("food"), token nodes (e.g., "form","drink") and semantic links – of which 5 types where defined: e.g. conjunctive, disjunctive, subordinate (Quillian, 1967). Right: Example of the the hierarchical structures of nodes and connections: distances (in terms of number of nodes and links to be travelled) determined the speed at which properties are retrieved. For instance, assessing whether it is true or not that a Canary can sing, takes less time than assessing whether it can fly (one needs to retrieve first the knowledge of the fact that it is a bird, than of the fact that birds can fly). From (Collins and Quillian, 1969)

Collins further developed the model, defining the Spreading of Activation Theory (Collins and Loftus, 1975). The name comes from the assumption that, when a word is processed at the semantic level, the corresponding activation spreads out along all connected paths in the network. Such activation progressively decreases, in a way that is proportional to the accessibility or strength of the links. The longer a word is processed, the longer activation is released, while activation decreases over time or if another activity interferes. Between two given words, it is thus possible to compute a semantic distance value (i.e., the distance along the shortest path), as well as a semantic similarity value (i.e., an aggregate measure of all possible the paths) (see Fig. 11).

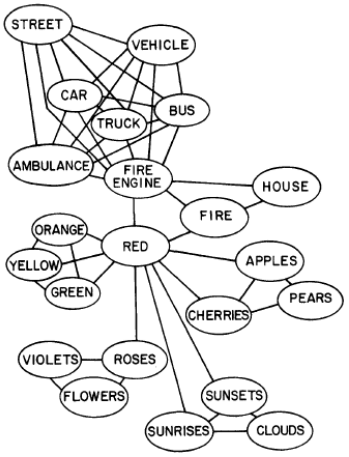


Figure 11 Example of a Spreading of Activation Theory graph. Given two concepts, we can compare their semantic distance (e.g., the path from roses to cherries is shorter than the one from violets and cherries) and their semantic similarity (e.g., between roses and cherries the way through red is the only possible path, they are not very similar; on the contrary, ambulance and vehicle are very similar as they are connected by many paths of different length). [figure adapted from (Collins and Loftus, 1975)]

The early seventies saw the development of antagonist featural models, such as the one proposed by Smith and colleagues (1974). In these kinds of models, a concept is not an unanalyzable unit: it is represented as a set of semantic features. Critical is the distinction between essential aspects of word meaning, called defining features (e.g., for birds: being a biped, having wings) and other accidental, characteristic features (e.g., for birds: flying, perching in trees). This observation led the authors to the definition of *typicality*: an instance of a category will be highly typical if it possesses most of the characteristic features (while, by definition, all instances manifest defining features). For instance, a *canary* is a more typical exemplar of the category *birds* than a *penguin*. They also suggested that differences in typicality ratings can be used as a measure of semantic distance, which in turn can be displayed in a low dimensional space thanks to techniques such as multidimensional scaling. One can then attempt to interpret the different dimensions as reflecting underlying characteristic features of the category (see Fig. 12). Analyses of behaviorally collected semantic data allowed researchers to notice how different domains (i.e., living vs non-living things) present substantial variance on factors such as feature correlations and distinguishing features (McRae and Cree, 2002).

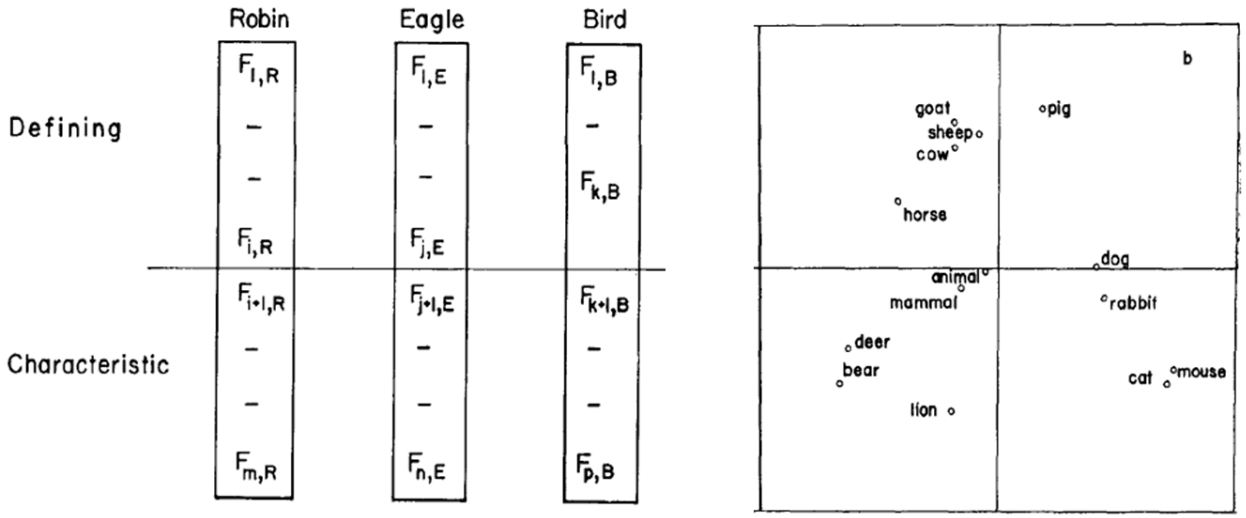


Figure 12 Featural model by Smith and colleagues. Left: each concept is described in terms of defining (i.e., necessary and sufficient) and characteristic (i.e., additional, optional) features. Right: multidimensional scaling of the category "mammals". Notice how some animals tend to cluster together (e.g., sheep, cow and goat), while others appear very distant (e.g., lion and pig). The dimension lying on the X axis can be interpreted as size (i.e., from big to small animals), while the one on the Y axis as predacy (i.e., from dangerous to harmless animals). [figures adapted from (Smith et al., 1974)]

The attempts to implement semantic networks with computer programs led to the development of connectionist models. The key feature of this family of approaches is that they aim at modeling not only how semantic concepts are stored, but also how they are learned, acquired. Generally speaking, knowledge is represented in terms of a set of units interconnected via weighted connections. Learning (i.e., adjusting the weights) can be accomplished either in a supervised or in an unsupervised fashion. Different architectures have been proposed, mostly involving a series of input, output and hidden units (i.e. the ones intervening between the different layers). The term feed-forward networks is associated with models where activation flows from input units to hidden units to output units. An illustrative example is the model proposed by Rumelhart and Todd (1993) and later developed by Rogers and McClelland (2004). This kind of model permits the observation of how concepts are re-arranged according to the semantic context (see Fig. 13). Models whose architecture involves feedback, bidirectional or recurrent connectivity as well are called dynamic models. In general, these kinds of models have two important consequences: they support the idea of a distributed semantic system (as already proposed by featural models); and they highlight the importance of simulations and correlation with behavioral evidence in

order to address complex phenomena such as semantic priming (Masson, 1995). Usually, the representational geometries that these models describe are organized around interpretable elements, encoding specific properties of the items the concepts refer to (e.g., the color, the shape, the size), hence the alternative name: attribute-based models.

The assumption that concepts are not represented in unitary nodes (as suggested by Quillian), but instead in a distributed fashion, is at the core of another family of models, that of distributional ones, also referred to as co-occurrence, or corpus-based models. Different alternative structures have been proposed, but they all share the hypothesis that semantics is learned via statistical extrapolation of relations among symbols during direct encounters in the linguistic environment. Modeling in these cases involves studying large text corpora, varying the kind of learning mechanisms to be used: from Hebbian learning to probabilistic inference. Examples include Latent Semantic Analyses (LAS, Landauer and Dumais, 1997) and Hyperspace Analogue to Language (HAL, Lund and Burgess, 1996). The first one, LSA, is based on the assumption that words that are close in meaning will co-occur in similar texts. As a first step, a document-term matrix is computed, describing how many times each concept appears in each text. Then, a low-rank approximation of such a matrix is computed, which can be used to assess similarities and relations between words, and to compare documents. The second approach, HAL, considers that words (e.g., "horse" and "donkey") are semantically related if they frequently appear in the same context (i.e., with the same words, e.g., "barn"), even if they never actually co-occur (e.g., "jugs" and "butter", mediated by the food context). The HAL matrix representing how all the words in its lexicon are associated is computed, for instance, over a 10-word reading frame moving through a corpus of text: whenever two words are simultaneously in the frame, the association between them is increased (inversely with their distance in the frame). In these *company-based*

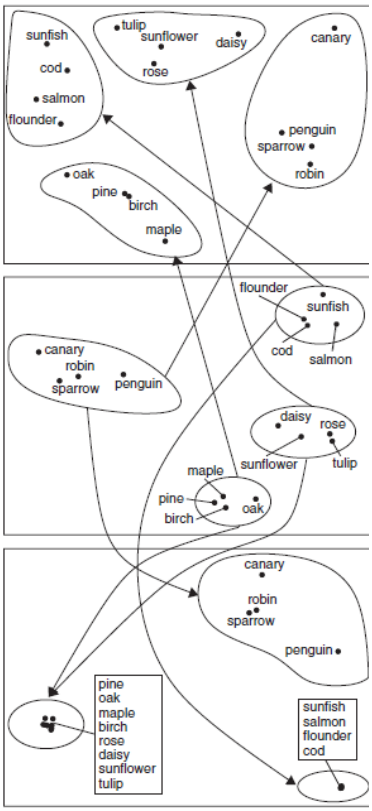


Figure 13 Multidimensional scaling of the similarities represented by the model by Rogers and McClelland. The middle panel illustrates the similarities among items at the level of the Representation. The upper and lower panels illustrate the similarities at the level of Hidden Units when different relational context are activated: *is* and *can* respectively. Note, for instance, how different trees are well spread out in the *is* context (they are all different instances), while the *can* context collapses differences among the plants (all they can do is to grow). [figure adapted from (Rogers and McClelland, 2004)]

models, the representations (i.e., vector-spaces) express conceptual structure, but are otherwise devoid of content, and thus of difficult psychological interpretation.

This, perhaps simplistic, dichotomy between cognitive (attribute-based, theory-driven) and computational (company-based, data-driven) models here presented, illustrate the divide still present in the current literature. On one hand, the representations characterizing attribute-based models are built of interpretable elements, encoding specific properties such as color, shape, size, etc… We will see that such an approach culminates with recent studies pursuing the identification of the neural substrate of those features (Binder et al., 2016). On the other hand, there are those aiming at resolving semantic content in fully distributed models where the interpretation of the different emerging dimensions is rarely helpful in clarifying their content, while being good in predicting behavioral performance (for a review on the success, shortcoming and future direction of this approach see Pereira et al., 2016).

3.2 Clinical Evidence

In the previous section, I have mentioned one of the core concepts in neuropsychology, that of dissociations (in particular double dissociations) and the inferential power they carry. Two other important points should be mentioned before reviewing the clinical evidence on the neural substrate of semantic knowledge. As the inferences in neuropsychology are drawn by observing a link between a given cognitive impairment and a given brain damage, they can only be as accurate as the neuropsychological assessment conducted and the brain imaging results obtained. Concerning the neuropsychological evaluation, before ascribing the performance in one given test to a semantic deficit, one needs to conduct a differential diagnosis with respect to modality specific access deficits including (but not limited to):

- visual agnosia (i.e., an impairment in recognition of visually presented items not due to visual defects; recognition is spared if items are presented in another modality, for instance if allowed to touch them or hear the sound they produce) (Farah, 2004);
- tactile agnosia (i.e., an impairment in recognition of items when they can only be explored by touch) (Reed, 1996);
- auditory agnosia (i.e., defective recognition of sounds that can be observed in different pure forms: only for speech (i.e., word-deafness), only for music (i.e., amusia) or only for non-verbal sounds) (Goldstein, 1974);

It is harder to frame semantic deficits with respect to disorders that affect production and/or comprehension of language, so called aphasias. Different aphasic syndromes have been described and only some of those include semantic deficits among their prominent symptoms. Before ascribing a given behavioral performance to a semantic deficit, it is important to verify that the difficulties are not limited to verbal material. Regarding the brain damage analyses, one has to pay attention to the different etiology:

- focal vascular damage, usually follows ischemic strokes (decreased or absent circulation of blood due to a thrombus or embolus) and more rarely intracerebral hemorrhage (rupture or leak of a blood vessel). These events are more frequent close to big brain vessels such as the Middle Cerebral Artery (for a map of the distribution of MCA infarcts (Phan et al., 2005). The onset of the symptoms is abrupt and recovery of the affected cognitive function will depend, among other factors, on the extension of the resulting brain damage (once all possible medical procedures have been applied).
- progressive degenerations, due to viruses (Whitley and Gnann, 2002) or proteopathies (Walker and LeVine, 2000), tend to develop from specific locations (made vulnerable by particular anatomical and genetic factors) and then spread to neighboring

regions. As the degeneration progresses gradually, possible compensatory mechanisms can come into play at the neural level, as well as at the behavioral one. Timing of the assessment is crucial: early signs can be missed while patients in advanced stages can be too compromised to be tested.

To expand further the interplay of neuropsychology and neuroimaging is beyond the scope of the present work, however for a review of the current challenges faced by clinical and cognitive neuropsychology, please see recent review by Price and colleagues (2016).

I have already mentioned that a specific neurodegenerative disorder, semantic dementia (SD), has provided researchers with crucial evidence on the neural substrate of semantic memory. SD is a member of a family of degenerative disorders called Fronto-Temporal Lobar Degeneration (FTLD, Agosta et al., 2015) that has three clinical manifestations affecting motility (includes: motor neuron disease, corticobasal degeneration, and progressive supra-nuclear palsy), behavior (called behavioral variant of frontotemporal dementia, bvFTD) or language (called primary progressive aphasia, PPA). This latter case includes three variants dissociated not only at the clinical level, but also at the anatomical one (Gorno-Tempini et al., 2004; Gorno-Tempini et al., 2011; Vandenberghe, 2016) (see Fig. 14): a nonfluent variant (nfvPPA), characterized by apraxia of speech (i.e., motor speech disorder) and deficits in processing complex syntax; a logopenic variant (lPPA), showing slow speech and impaired syntactic comprehension and naming; and a semantic one (SD). SD is considered a presenile disorder - i.e., the patients are relatively young at the onset, typically between the ages of 50 and 70 years. About two-third of the cases are associated with ubiquitin pathology (Grossman, 2010).

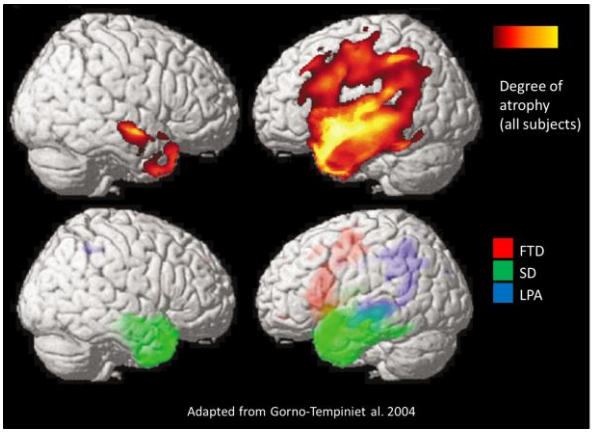


Figure 14 Pattern of atrophy in the three variant of PPA. The SD variant shows atrophy spreading posteriorly from the anterior temporal pole (in green). The nfvPPA atrophy appears to be confined to the lower and posterior part of the frontal lobe (in red). The lPPA is associated with atrophy in superior and posterior portions of the temporal lobe and inferior anterior portions of the parietal lobe (in blue).

Clinical manifestations include fluent speech (it might be very difficult for caregivers to realize they are witnessing a language disorder) in presence of semantic memory deficits. The key initial feature is a reduction of expressive and receptive vocabulary, often manifested by anomia embedded in sentences with normal phonological, grammatical and syntactical features (see Fig. 15). The semantic nature of the deficit is highlighted by the fact that conceptual knowledge appears compromised even in tasks that do not require verbal communication, for instance simple object use (Hodges et al., 2000), and item identification based on smell (Luzzi et al., 2007), sound (Bozeat et al., 2000) or taste (Piwnica-Worms et al., 2010). The deficits not only involve all modalities, but also all concepts, with the exception of basic numerical ones (Cappelletti et al., 2001). Three aspects of stimuli and task influence SD patients' performance: familiarity with a given item (the more, the better), typicality of such an item within a domain (e.g., for the category of wind instruments, *flute* would be more resistant to damage than *ocarine*), and specificity (performance decreases when a high level of specificity is required, e.g., distinguishing between *comté* and *beaufort* – both french cheese) (Lambon Ralph et al., 2016).

a

Item	Sept '91	Mar '92	Sept '92	Mar '93
Bird	+	+	+	Animal
Chicken	+	+	Bird	Animal
Duck	+	Bird	Bird	Dog
Swan	+	Bird	Bird	Animal
Eagle	Duck	Bird	Bird	Horse
Ostrich	Swan	Bird	Cat	Animal
Peacock	Duck	Bird	Cat	Vehicle
Penguin	Duck	Bird	Cat	Part of animal
Rooster	Chicken	Chicken	Bird	Dog

b

Item	P1	P2
Lion	Is it an animal? It has little legs and big ears, sleeps a lot	An animal, quite tall.
Deer	Is it an animal?	An animal, gives milk, like sheep.
Violin	A music thing, can't think.	Is it an instrument? I think it's made of metal
Guitar	Play music with it, can't remember, it's big, you play with an arrow.	It's what you do music with, put on your shoulder and put a bit across-like that…

Figure 15 Examples of verbal testing of SD patients. The results of a naming task (i.e., patients are presented with printed pictures and asked to name the item in it) exemplify the progressive loss of conceptual knowledge (a). Similarly, when asked to define a given concept, patients can produce grammatically correct sentences, but are unable to provide a proper description (b).

In vivo anatomical imaging has revealed fronto-temporal atrophy starting from the anterior temporal lobe and then progressively spreading posteriorly towards the parietal lobe (Galton et al., 2001; Rosen et al., 2002; Davies et al., 2006; Brambati et al., 2009), confirmed by post-mortem pathological findings (Davies et al., 2005) (see Fig. 16). Converging findings come from the analyses of white matter abnormalities (Agosta et al., 2009; Galantucci et al., 2011). SD patients show alteration, as compared to a control group, in all metrics (i.e., mean fractional anisotropy, axial, radial and mean diffusivities). In particular, they present a dysfunction of the ventral language system (i.e., a severe involvement of the uncinate fasciculus and of the inferior longitudinal fasciculus, especially the anterior portion bilaterally and the left middle section), with relative sparing of the dorsal network (i.e. the parietofrontal components of the superior longitudinal fasciculus are relatively spared). All tracts encompassing the temporal lobe are vastly damaged, including the left arcuate. Along the inferior longitudinal fasciculus, DTI changes decrease in severity from anterior to posterior regions. Finally, the atrophy observed with MRI is also supported by the evidence of anterior temporal hypometabolism as observed with positron emission tomography (PET) (Diehl et al., 2004; Nestor et al., 2006; Desgranges et al., 2007). Overall, this evidence suggests a major role of the anterior temporal lobe in the processing and storage of semantic knowledge.

Figure 16 Semantic Dementia. The atrophy of the temporal pole involves both hemispheres, but is prevalent on the left side.

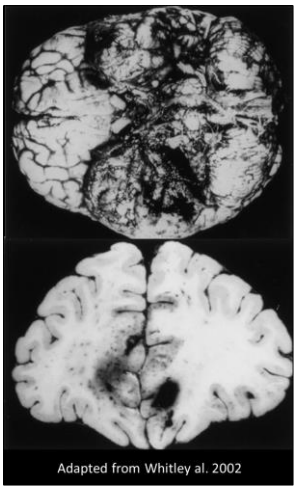


Figure 17 Herpes Symplex Enchephalities. Bilateral atrophy of the temporal poles is clearly visible.

Another set of patients has greatly contributed to the study of semantic memory: those suffering from herpes simplex virus encephalitis (or HSVE), a viral infection of the central nervous system with a predilection for temporal lobe involvement (Whitley and Gnann, 2002) (see Fig. 17). It is commonly associated with severe amnesia, naming difficulties and disexecutive symptoms (Kapur, 1994), however, it is important to point out that the diagnosis is based on positive virology irrespective of the cognitive profile. HSVE patients often show category-specific semantic deficits: performance appears to be disrupted for living things, spared for non-living items

or artefacts (Warrington and Shallice, 1984; Pietrini et al., 1988; Sartori et al., 1993; Laiacona et al., 2003). Thus, although pathology in both SD and HSVE is centered on the anterior temporal lobes, differences in cognitive profile and anatomical changes have been highlighted (see Fig. 18). This has led authors to suggest that the antero-medial temporal cortex (extensively damaged in HSVE patients) may be important for processing such as living things, whereas the inferolateral temporal cortex (where SD abnormalities predominate) may play a more general role within the semantic system (Noppeney et al., 2007). Others authors, combining neuropsychological data and computational simulations, have emphasized how the different neuropsychological profiles of SD and HSVE patients can be explained not solely by the location of damage, but also by the kind of impairment (Lambon Ralph et al., 2007). A generalized semantic impairment is found when the computational model is damaged by removing randomly selected connections entering, leaving or intrinsic to a central hub where all information converges (role assigned to the ATL, thus simulating SD patients' lesions). Conversely, when damage is achieved by changing the value of the weights of those connections (impairment thought to simulate HSVE's lesions), a category specific impairment emerges. It should be noted that HSVE is an acute disease: following treatment partial recovery and some degree of relearning are possible, leaving the subjects with a functioning yet less "semantically acute" semantic system (Lambon Ralph et al., 2016).

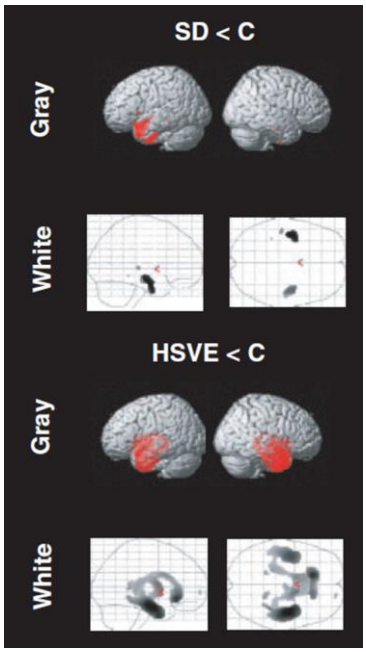


Figure 18 White and gray matter changes in SD and HSVE patients. Pathology is centered on the anterior temporal lobe in both cases, but the disease spreads in very different ways. Moreover, SD and HSVE are due to very different pathological processes, thus even an identical volume loss may not be functionally equivalent [from (Noppeney et al., 2007)].

Following the theoretical swing towards an embodied account of semantic knowledge (which I will develop later on in the chapter), growing attention has been given to neurological disorders affecting somato-sensory and motor systems: do they impact semantic as well? First, interesting semantic symptoms have been investigated in patients presenting one of the motor variants of FTLD. Motor neuron disease (MND) is an umbrella term for a group of neurological disorders that destroy upper and/or lower motor neurons

(amyotrophic lateral sclerosis, primary lateral sclerosis, progressive muscular atrophy, progressive bulbar palsy and pseudobulbar palsy) (Leigh and Ray-Chaudhuri, 1994). MND is usually associated with neuronal loss in the anterior horn of the spine and bulbar nuclei plus a widespread cortical atrophy, mainly frontotemporal (see Fig. 19). Selective deficits for verb processing have been associated with pathological changes in Brodmann areas 44 and 45 in MND patients (Bak et al., 2001; Bak and Hodges, 2004; Grossman, 2008; Bak and Chandran, 2012). Other motor variants of FTLD that have caught researchers' attention are progressive supra-nuclear palsy (or PSP) and corticobasal degeneration (or CBD). PSP is a neurodegenerative disorder whose diagnosis is purely clinical and based on the observation of symptoms such as supranuclear gaze dysfunction, extrapyramidal symptoms and cognitive dysfunction (Steele et al., 1964). As for CBD, degeneration involves both the cerebral cortex and the basal ganglia (Lee et al., 2011). Contrasting healthy subjects with SD, PSP and CBD patients, it was possible to observe that only the two groups of patients with motor variants of FTLD (i.e., PSP and CBD) were significantly more impaired in naming actions compared to objects (Cotelli et al., 2006). Moreover, verb deficits in lexico-semantic tasks have been reported in PSP patients (Daniele et al., 1994; Bak et al., 2006; Daniele et al., 2013) and CBD ones (Spatt et al., 2002; Silveri and Ciccarelli, 2007).

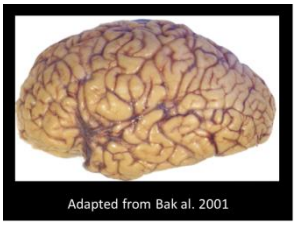


Figure 19 Motor Neuron Disease. It is possible to observe mild atrophy of the left temporal pole.

Second, patients with Parkinson's Disease (or PD), have been studied. PD's main pathological characteristic is cell death in basal ganglia – in particular in the substantia nigra (Davie, 2008). PD patients have been shown to be significantly more impaired in action than in object naming (as compared to healthy controls) (Pignatti et al., 2006; Cotelli et al., 2007; Rodriguez-Ferreiro et al., 2009). Moreover, priming for actions verbs appears to be affected by dopaminergic treatment: it is recovered (reaching a level comparable to those for concrete nouns and similar to that of healthy participants) only following Levodopa intake (Boulenger et al., 2008b).

Finally, patients with lesions in the right hemisphere have been compared with healthy controls, observing a dissociation between right frontal lobe lesions (affecting motor performance) and right temporo-occipital lesions (sparing motor performance). The first group showed worse performance when processing action verbs, whereas the second when processing visually-related nouns (Neininger and Pulvermüller, 2003). Furthermore, patients with ideomotor apraxia (i.e., deficits in producing actions or using tools following stroke affecting left primary motor cortex or the left inferior frontal and/or parietal lobe), can manifest impairments in retrieving conceptual knowledge related with actions/tools, and even at the single subject level many dissociations are observed (Buxbaum and Saffran, 2002; Negri et al., 2007; Pazzaglia et al., 2008a; Pazzaglia et al., 2008b; Papeo et al., 2010). Overall, these findings suggest that processing lexico-semantic information about action words might depend on the integrity of the cortical (and subcortical) motor system.

I have briefly mentioned that the second line of research stemming from neuropsychology is that investigating category specific semantic deficits. We have already seen that the majority of the evidence in favor of the existence of category specific impairments comes from HSVE patients (Warrington and Shallice, 1984). Another source of data are patients presenting focal lesions along the ventral visual path, due to ischemic or hemorrhagic strokes (Warrington and McCarthy, 1983; Warrington and McCarthy, 1987). Interesting dissociations in the behavioral performance of patients with semantic deficits have been observed since the dawn of the studies on semantic knowledge:

- patients with a selective impairment for stimuli referring to living items (e.g., animals) and spared performance for stimuli referring to non-living items (e.g., artefacts) (Warrington and Shallice, 1984; Pietrini et al., 1988; Sartori et al., 1993;

Caramazza and Shelton, 1998; Laiacona et al., 2003; Blundo et al., 2006);

- the opposite pattern, a spared performance for stimuli referring to living items and a deficit for non-living ones (Sacchett and Humphreys, 1992; Laiacona and Capitani, 2001);
- patients showing worse performance for stimuli referring to living inanimate things (e.g., vegetables) compared to living animate things (i.e., animals) (Hart et al., 1985; Hillis and Caramazza, 1991; Farah and Wallace, 1992; Crutch and Warrington, 2003; Samson and Pillon, 2003);
- the reverse situation, that is, worse performance for stimuli related with living animate things compared to living inanimate things (Hart and Gordon, 1992; Caramazza and Shelton, 1998);
- a selective deficit for stimuli referring to fruits and vegetables (as opposed to both other living items and non-living ones)(Hart et al., 1985; Samson and Pillon, 2003);
- a selective deficit for conspecifics (Ellis et al., 1989; Miceli et al., 2000).

These findings have triggered many different hypotheses on the structure and neural substrate of semantic memory, challenging the concept of a unitary semantic system and opening the investigation on possible internal sub-systems (Riddoch et al., 1988; Shallice, 1988; Caramazza et al., 1990; Hillis et al., 1995). These early theories can be assigned to one of two opposite perspectives (Capitani et al., 2003). One line of inquiry, the neural-structure principle, developed from the assumption that the representational constraints, determining which concepts cluster together, are internal to the brain itself. On one hand, it has been suggested that distinct semantic subsystems are specialized according to the type of information they process, giving rise to modality specific clusters of concepts (Warrington and Shallice, 1984). On the other hand, it has been advocated that there could be domain specific systems deputed to the processing of the information

linked with a given evolutionary relevant domain (e.g., animals) (Caramazza and Shelton, 1998). The opposite line of research, the correlated-structure principle, postulates that the representational constraints come from the statistical co-occurrence of object properties in the world and are not driven by neuro-anatomical constraints (Caramazza et al., 1990; Tyler and Moss, 1997; Garrard et al., 2001).

One key theory in the family of the neural-structure principle was proposed by Warrington and colleagues: the *sensory functional theory* (Warrington and Shallice, 1984; Warrington and McCarthy, 1987). The semantic system is broken down into modality-specific subsystems devoted to the analyses of visual/perceptual information (fundamental to process concepts related to living things) or functional/associative information (essential for concepts related to non-living things). A computational implementation of such a system has been developed in the early nineties (Farah and McClelland, 1991). This theory would not predict the existence of a dissociation within the category of living things, i.e., among items that share the same amount of contribution from the visual sub-system. However, as I reviewed above, dissociations of this kind have been observed: for instance, differences emerge inside the category of living things between animate (such as animals) and inanimate (such as vegetables) items (Hart et al., 1985; Hillis and Caramazza, 1991; Farah and Wallace, 1992; Hart and Gordon, 1992; Caramazza and Shelton, 1998; Crutch and Warrington, 2003; Samson and Pillon, 2003). Moreover, this theory would expect difficulties in the visual/perceptual domain to always correspond to deficits with living items, while, for instance, patients with deficits for colors, but not for vegetables have been reported (Miceli et al., 2001).

We have seen that, while keeping the perspective of a neural-structural principle as the origin of the organization of the semantic system, another interpretation is possible: the domain specific theory (Caramazza and Shelton, 1998; Caramazza and Mahon, 2003). This theory predicts, in line with the above-mentioned patients'

observations, that there is no association between a deficit for a given type or modality of knowledge (e.g., visual/perceptual) and a conceptual deficit for a specific category of objects (e.g., living things). Instead, the semantic system is thought to be subdivided in functionally dissociable neural circuits dedicated to evolutionary relevant domains: conspecifics, animals, fruit/vegetables, and tools.

As for the correlated-structure principle family of theories, I here synthesize the *Conceptual-structure account* (Tyler and Moss, 2001), which focuses on two observations. First, as compared to non-living items, living ones share more features, especially perceptual ones (e.g., having limbs, having eyes, having a mouth), that are strongly correlated as they frequently co-occur (i.e., typically, if something has eyes and limbs, it also has a mouth). Second, in living items, shared features are correlated with specific biological functions (e.g., it has wings = it can fly), while individual variations in form are usually not functionally significant (e.g., different kinds of wings). On the contrary, in non-living items the functional information is conveyed precisely by distinctive perceptual features (e.g., it has a blade = it is used to cut). The hypothesis made is that the more intercorrelated shared features concepts have, the more resistant to damage they are (Moss et al., 1998; Moss and Tyler, 2000; Tyler et al., 2000). An interaction between domain and distinctiveness is thus predicted. For living things, distinctive properties (e.g. the shape of the beak) should be more vulnerable than shared ones (e.g., having a beak), as they weakly correlated with other properties of the concepts. In the case of artifacts, shared properties (e.g., being made of plastic) are fewer and less inter-correlated (thus more vulnerable), while distinctive properties (e.g., having a handle) are protected by strong form–function correlations. However, the opposite prediction can be made: i.e., sharing many features weakens concepts (Gonnerman et al., 1997; Devlin et al., 1998). In fact, as we have seen, patients have been described with disproportionate deficits for non-living things (such as artefacts) with relatively intact performance for living things (such as animals), as well as the opposite pattern (Warrington and

Shallice, 1984; Pietrini et al., 1988; Sacchett and Humphreys, 1992; Sartori et al., 1993; Caramazza and Shelton,1998; Laiacona and Capitani, 2001; Laiacona et al., 2003; Blundo et al., 2006). However, each case should be examined separately as the theory predicts variation across categories within the same domains as a function of the inner structure of the category (e.g., vehicles have more highly correlated properties than tools, thus being closer to living items). It should be noticed that even this line of research converged on the ATL as a crucial site for semantic processing: in particular, the perirhinal cortex appears to be the area supporting fine-grained semantic processes across different tasks (Wright et al., 2015).

All the clinical evidence here reviewed leads to three general conclusions:

- generalized, multimodal and pervasive semantic deficits are observed in presence of lesions affecting the anterior temporal lobe;
- semantic dissociations can be elicited by appropriate testing when damage is confined to specific components of the semantic network;
- (motor)perceptual and conceptual variables differentially correlate with semantic categories and domains.

Any comprehensive theory that wishes to describe the cognitive structure and the neural substrate of semantic knowledge needs to be able to explain the full set of clinical findings. We will now see how neuroimaging data can be used to test the predictions made by the different theories proposed.

3.3 Neuroimaging Evidence

With the advent of neuroimaging techniques, the relation between cognitive functions and brain areas has been widely studied with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). We have seen that in neuropsychology the

first step is to define which performance (in which test) constitutes evidence for a semantic deficit. Similarly, when approaching the neuroimaging literature, three key features of this line of research should be kept in mind (see also Chap. 2):

- the observation that a given cognitive state (e.g., processing of a given stimulus) correlates with brain activity in a certain area does not imply a causal link between that area and the cognitive process being tested;
- the choice of which technique to use will depend on the tradeoff between the cognitive question investigated and the constraint imposed by the different methods (i.e., temporal and spatial resolution);
- different tasks (and stimuli) will allow for different conclusions according to the depth of semantic processing they require and possible confounding factors.

Applying the traditional subtraction method (Donders, 1968/1969), classical PET and fMRI paradigms to study semantic knowledge included comparison of the processing of different stimuli (e.g., pictures, words, sentences) during different tasks (e.g., silent naming/reading, categorization tasks). This method has proven successful in identifying cortical areas responding preferentially to different categories of visual stimuli such as: words in the left fusiform gyrus (Dehaene and Cohen, 2011) and numbers in the right fusiform gyrus (Abboud et al., 2015); objects (Lerner, 2001), bodies (Downing et al., 2007); faces in both the fusiform (Kanwisher et al., 1997) and the occipital face area (Gauthier et al., 2000); places, buildings, and large objects in the so called parahippocampal place area (Epstein and Kanwisher, 1998; Epstein et al., 1999).

Following the clinical evidence reviewed above concerning a possible organization of the semantic systems by categories, many neuroimaging investigations have revolved around the quest for specificity for semantic categories in the ventral visual path. The presentation of both pictures and words (Perani et al., 1995; Martin et al., 1996; Chao et al., 1999; Ishai et al., 1999) seems to elicit a double

dissociation: animal stimuli appear to recruit a lateral portion of the ventral visual path, while stimuli of tools appear to be processed in its medial portion. This finding has been extensively reviewed and mostly interpreted as evidence of separated semantic systems processing specific categories (Martin and Chao, 2001). However, while for pictorial presentations of objects the ventral partition of sub-areas preferring different categories of stimuli appears a solid finding, not all studies have been able to replicate the categorical findings when presenting words, and some authors have argued in favor of a unitary semantic system, undifferentiated by categories at the neural level (Devlin et al., 2002). The picture is further complicated by findings suggesting that the key factor determining whether stimuli are going to be processed laterally – as animate/living items – or medially –as inanimate/non-living items- is not their semantic category per se, but rather the interpretation done by the subjects as biological entities or not (Castelli et al., 2000; Martin and Weisberg, 2003).

As previously discussed, the alternative explanation is that of a feature-based organization of the semantic system. Early PET studies investigated attributes such as color and motion (Martin et al., 1995; Chao and Martin, 1999), and fMRI ones have tried to shed light on the interplay between the neural substrate of categorical and modality specific information (Thompson-Schill et al., 1999). Authors following this perspective have also shown how feature statistics can explain the clusters observed in the fusiform gyri, where objects with many shared features are associated with activity in the lateral portion of the gyri, whereas objects with fewer shared features activate predominantly the medial portion (Tyler et al., 2013). This kind of evidence has shifted the attention from the domain-specific latero-medial gradient to the postero-anterior one describing the shift from a coarse (i.e., categories) to a fine (i.e., individual concepts) processing of semantic information (Clarke et al., 2013).

Overall, since the first PET result (Petersen et al., 1988) on the neural substrate of semantic processing (comparing passive word

listening and reading with words repetition and words generation), numerous areas have been associated with an active role during semantic tasks:

1. Inferior frontal cortex (iFC), the so called Broca's area and its surroundings (BA 44,45, and 47), including the anterior inferior frontal gyrus (Demb et al., 1995; Wagner et al., 1997; Devlin, 2003; Goldberg et al., 2007)
2. Superior temporal cortex (sTC)
3. Inferior parietal cortex (iPC) and angular gyrus (Bonner et al., 2013; Price et al., 2015; Bonnici et al., 2016)
4. Inferior and middle temporal cortex (m/iTC) (Fairhall and Caramazza, 2013)
5. Anterior temporal cortex (aTC) or anterior temporal lobe (ATL) (Mion et al., 2010; Tsapkini et al., 2011)

A cautionary observation when reviewing neuroimaging literature. I have mentioned that different tasks will allow for different conclusions: what is the appropriate control task for a semantic experiment? At minima, it should entail the same cognitive effort without requiring access to semantic knowledge. Even when the choice of the task(s) is clear, problems arise with the stimuli selection. Let's say one decides to opt for a semantic categorization task (i.e., "is it an animal or a tool?"), which are the most appropriate stimuli: pictures or written names of the items? To decide that a cheetah is an animal, when presented with its picture, it's relatively easy: one quick look and all key features will be obvious (e.g., it has 4 legs, a tail, fur). One does not even need to know the actual name of the animal, categorization is possible simply based on the visual features. On the other hand, after reading the word "*cheetah*" I can correctly classify it only if I access the related concept. Two consequences follow: (1) the activity observed in parieto-frontal areas could be related with a differential load of attention, working memory and executive function between the semantic and the control task used (Van Doren et al., 2010; Whitney et al., 2011); (2) the activity in ventral occipito-temporal areas (known to be involved in high level visual processing)

could be driven by the nature of the stimuli used in most studies, i.e. pictures.

Overall, the most robust and consistent findings from multiple imaging techniques seem to converge with clinical evidence on a crucial role of the anterior temporal cortex. First, early PET studies (Gorno-Tempini and Price, 2001) suggested that the analyses of semantic attributes (for both famous faces and buildings) took place in the ATL. Second, fMRI data corroborated the idea that amodal semantic processing involved this portion of the temporal lobe (Tyler et al., 2004; Rogers et al., 2006; Spitsyna et al., 2006; Visser et al., 2010b; Peelen and Caramazza, 2012). Third, psychophysiological studies on semantic priming (Geukes et al., 2013; Lau et al., 2013) have complemented the traditional imaging results. Fourth, transcranial magnetic stimulation (TMS) can be used as a way to (temporarily) mimic the (chronic) effects of SD (Pobric et al., 2007; Binney et al., 2010; Pobric et al., 2010a; Pobric et al., 2010b). Healthy subjects whose normal ATL activity is shortly disturbed by magnetic impulses manifest cross-modal semantic deficits similar to those detected in SD patients. Hence, TMS provides invaluable causal evidence of the key role played by ATL in semantic memory.

Connectivity data from both comparative anatomical studies (Moran et al., 1987) and investigations in humans (Binney et al., 2012; Jung et al., 2016) have highlighted the powerful set of connections between ATL and unimodal regions. These connections seem to be at the origin of the distinct subregions of ATL that can be detected via functional (Pascual et al., 2015) as well as structural connectivity (Papinutto et al., 2016). The differential connectivity of ATL has been recently confirmed during both rest and an explicit semantic task (Jackson et al., 2016). The idea that ATL is best understood in terms of a heterogeneous portion of cortex is supported also by recent cytoarchitectonic differentiations (Ding et al., 2009). These observations have led to the hypothesis of a graded specialization within the ATL as a consequence of its differential connectivity with modality specific cortical regions (Rice et al., 2015).

Recently, it has been shown that different portions of ATL code respectively for input modality (e.g., written vs spoken words, aSTG) and input meaning (e.g., "loud" vs "shiny", vATL) (Murphy et al., 2016).

The results of decades of PET and fMRI studies tapping semantic processing have been extensively reviewed (Binder and Desai, 2011) and subjected to meta-analyses (Binder et al., 2009; Visser et al., 2010a). The first striking conclusion is that a wide portion of neocortex is involved: semantic processing appears to be an emerging property of a wide network. The second, corollary, observation is the involvement of modality-specific areas (devoted to processing sensory, motor, and emotion inputs) as well as multimodal associative areas (where multiple motor-perceptual processing streams converge). Much of the scientific effort should thus focus on characterizing the properties of this broad network of areas (see Fig. 20).

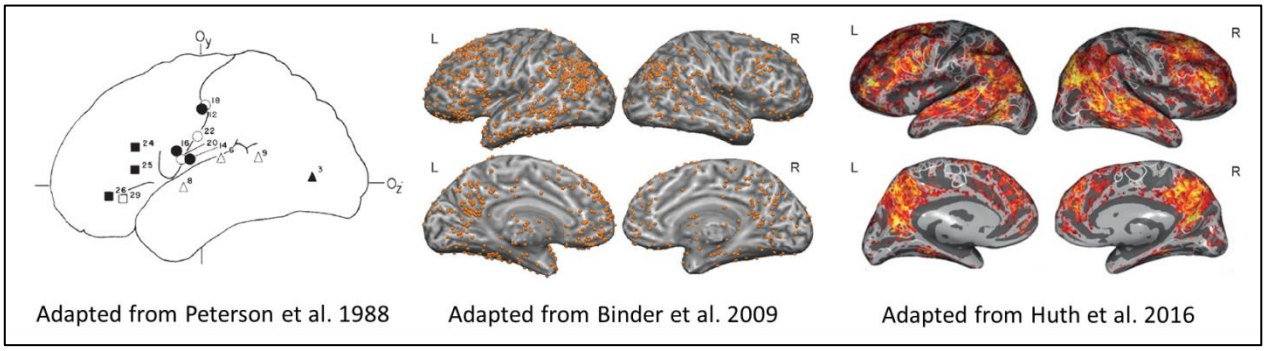


Figure 20 Topography of the semantic system. Since the first studies conducted with PET (an example on the left), numerous cortical areas have been associated with semantic processing of pictures, words and sentences. A comprehensive review of all the activation foci leads to the observation that a vast portion of the cerebral cortex is involved, mostly including inferior parietal, temporal and frontal areas (center). A recent study, deploying some of the most advance statistical techniques available, confirms the observation that the activity of a vast portion of the neocortex is involved in semantic processing (right).

It has been suggested that the integration of the semantic information distributed over the cortex takes place in dedicated multimodal hubs, so called *convergence zones* (Damasio et al., 1996; Tranel et al., 1997; Damasio et al., 2004) distributed within *convergence regions* (e.g. temporal pole, anterior IT, frontal operculum). These areas are thought to be innately dedicated to

performing conjunctive operations (i.e., available prior to any individual experience), but then being shaped by learning. Anatomical constraints are imposed on the location of convergence regions (e.g., due to white matter tracts connecting different areas), however, at the micro scale, the precise site of convergences zones is expected to change, even within the same individual, according to the type of stimuli and task demands.

A similar model, termed *hub-and-spokes*, posits that the integration needed in order to give rise to coherent, generalizable concepts takes place in a transmodal semantic hub that interacts with modality-specific sources of information (Rogers et al., 2004; Rogers et al., 2006; Lambon Ralph et al., 2007; Patterson et al., 2007; Lambon Ralph et al., 2010; Lambon Ralph, 2014). Different semantic features need to be combined in a nonlinear, modality invariant manner allowing:

(1) appreciation of both superficial (e.g., tomatoes and tennis balls are both round) and deep (e.g., tomatoes and bananas are both fruit) similarities,

(2) consistency through time and contexts (e.g., a tomato is such when entire as well as after having being cut), and

(3) adaptation and generalization whenever new information becomes available.

In light of all the functional and anatomical literature reviewed above, the region believe to correspond to such a semantic transmodal hub is the ATL.

To sum up, the classical neuroimaging findings here reviewed:

- leave open the debate on whether the organization of semantic knowledge in the brain is based on categorical (i.e., domain-specific clusters) or on featural constraints (i.e., clusters emerge due to correlations of sensory-motor and functional features);
- opens new questions on how information is integrated suggesting the need for semantic hub(s).

3.4 Grounded Cognition

In the last twenty years, the classical approach to the study of the neural substrate of semantic knowledge has been affected by the paradigm shift that swept across cognitive science: grounded cognition. As we have seen in the introduction of this chapter, the debate on the interaction and interdependency of body and mind, sensory-motor experiences and conceptual processing, is a very old one. In cognitive neuroscience, it has taken different declinations. The strong embodied perspective (i.e., perception, action and cognition are fused seamlessly) is at the extreme of a rich continuum of hypotheses on the relation between perception, action and language. Right before the turning of the century, Barsalou introduced the concept of perceptual symbol systems: there is no need for an additional amodal system, perceptual symbols are established in the same areas as the perceptual state they refer to (Barsalou, 2010). Other, more moderate, approaches have focused more specifically on the link between language and motor systems and how cognition can be grounded in perception-action systems (i.e., representations are shaped by the senses and the body).

The revolutionary discovery that paved the way for this line of research was that of so called mirror neurons: sensorimotor neurons in the ventral premotor cortex and inferior parietal lobe of monkeys' brain fire when the animal is acting as well as when it is simply seeing some act (Di Pellegrino et al., 1992; Rizzolatti et al., 1996a) or hearing the sound of the action (Kohler et al., 2002). Even if some authors have reported a similar mirror system in humans with PET (Grafton et al., 1996; Rizzolatti et al., 1996b; Decety et al., 1997), fMRI (Iacoboni, 1999; Buccino et al., 2001), M/EEG (Cochin et al., 1998; Hari et al., 1998), and TMS (Fadiga et al., 1995), other have failed to do so (Lingnau et al., 2009).

There are two main theoretical standpoints: Pulvermüller's distributed neuronal assemblies (Pulvermüller, 1999; Pulvermüller, 2013) and Gallese's neural parameters simulation (Gallese and Lakoff,

2005; Glenberg and Gallese, 2012). The first approach is based on the Hebbian learning rule "*fire together, wire together*". It is postulated that during language learning, word forms are mostly encountered when the objects they refer to are physically present or the action they refer to is being performed, thus the language perisylvian assembly of neurons connects with the sensory and motor ones by virtue of simultaneous firing. Once this higher order assembly is established, processing of linguistic inputs will activate sensory-motor cortices as well, grounding the meaning of symbols through their sensory-motor properties. The second, more radical, approach, affirms that language is an emerging multimodal faculty that exploits pre-existing properties of the sensory-motor system and can be completely resolved in their computations. The sensorimotor system is thought to provide all elements needed to implement the hierarchical structure that builds concepts, eliminating the need for an additional language module.

Evidence in support of these theories comes from different neuroimaging methods (see Fig. 21) and some clinical observations (as reviewed above). Many of the studies used fMRI, thus offering a precise localization of the effects, but an insufficient temporal resolution (see Chap. 2.2). Timing information (coming from M/EEG studies) and causal inference (possible only with lesion studies, but see Chap. 2.3) are critical to distinguish between a necessary and automatic activation of specific action-related networks, and an epiphenomenal consequence of a late, postlexical strategy to imagine or plan an action. In particular, this line of research has shown correspondence between brain areas activated by the conceptual processing, the observation, and the execution of actions and movements. The first studies involved reading words related to body parts (i.e., leg/mouth/arm words) (Hauk et al., 2004), listening to action-related sentences (i.e., "*I bite an apple*") (Tettamanti et al., 2005) and reading verb–object phrases related with body part actions (i.e., "*pressing the car brake*") (Aziz-Zadeh et al., 2006). The key finding is that motor and premotor areas appear to be involved in the conceptual processing in a body-part congruent way: leg, mouth and

arm words seem to have a precise and overlapping somatotopic organization, overlapping with that of action observation and execution. Follow-up studies have attempted to further dissociate the category of verbs by looking at how they can be described by the presence (or absence) of 5 distinct semantic components: *action, motion, contact, change of state and tool use*. They investigated the specific neural substrate of different classes of verbs, i.e. running verbs (e.g., *jog*), speaking verbs (e.g., *whisper*), hitting verbs (e.g., *poke*), cutting verbs (e.g., *slice*), and change of state verbs (e.g., *shatter*). It appears that the weight of the semantic components determined which brain areas would be involved (Kemmerer et al., 2008): M1 and M2, in a somatotopic fashion, for verbs of action; posterolateral temporal cortex for verbs of motion; intraparietal sulcus and inferior parietal lobule for verbs of contact; ventral temporal cortex for verbs of change of state; a temporal, parietal, and frontal network of regions for verbs implicating tool use. M/EEG studies have contributed to the debate by showing how fast the activation in the motor system is (see also paragraph 4). An initial report of a difference, around 240 ms after stimulus onset, between verbs related with leg (e.g., "*kick*") and mouth (e.g., "*speak*") actions (Pulvermüller et al., 2000), was corroborated by later studies. First, it appears that while subjects are reading words related to different body parts (e.g., leg/mouth/arm), a somatotopically coherent activation of the motor system can be observed as early as 220 ms after stimulus onset (Hauk and Pulvermuller, 2004). Secondly, if subjects are presented with auditory words, specific cortical topographies are observed even earlier: in frontocentral areas face-related stimuli elicited stronger activation than leg-related ones at 172–176 ms, while in superior central areas the opposite pattern was observed at 200 ms (Pulvermüller et al., 2005a).

The neural overlap between conceptual and modality specific processing is not observed exclusively in relation with the motor system: reading odor-related terms appears to activate primary olfactory cortex (González et al., 2006), while sound-related ones

activate the auditory cortex (Kiefer et al., 2008), and taste-related ones the gustatory cortex (Barros-Loscertales et al., 2011). Moreover, not only literal, but also idiomatic sentences have been shown to elicit a somatotopic involvement of the motor system (Boulenger et al., 2008a). Finally, abstract words have often been cited as the litmus test of embodied theories: how can words devoid of any concrete referent in the outside world be grounded in sensory-motor systems? They are often considered grounded in emotional (Kousta et al., 2011; Moseley et al., 2011; Vigliocco et al., 2013), introspective, or social information, perhaps via simulations of their metaphorical extension (Gallese and Lakoff, 2005; Gibbs, 2006; Jamrozik et al., 2016). In a given context, they acquire a specific sensory-motor instantiation either cataphorically (i.e., the abstract disembodied symbol is introduced and later explained) or anaphorically (i.e., a previous sensory-motor explanation is linked with an abstract symbol) (Zwaan, 2016).

While neuroimaging experiments can only show a correlation between the activity of a given area and some characteristic of the task or stimuli at hand (see also Chap. 2), lesions, whether real or virtual ones, can establish a causal link. For instance, Keifer's team has described a patient with a focal lesion in left posterior superior and middle temporal gyrus, who appears to be impaired in processing sound-related everyday objects (e.g., "bell"), while performance for non-sound-related everyday objects (e.g., "armchair") is spared. Interestingly, his performance with animals (irrespective of whether they are typically associated with a sounds (e.g., "cock") or not (e.g., "camel"), and musical instruments (e.g., "violin") was intact (Trumpp et al., 2013). Moreover, TMS experiments have revealed how stimulating hand and leg areas influences the processing of arm-related and leg-related words speeding up responses only for limb-specific words (Pulvermüller et al., 2005b). TMS can also be used to study motor evoked potentials (MEPs): MEPs recorded from hand muscles appear to be modulated by listening to hand-action-related

sentences, while MEPs from foot muscles by listening to foot-action-related sentences (Buccino et al., 2005).

Finally, Pulvermuller and colleagues have implemented a computational model that is able to simulate the clustering of object-related words vs action-related words due to the statistic of their learning (Garagnani and Pulvermuller, 2016). Subsequently, the same group of authors incorporated cortico-cortical connections (as they are documented by neuroanatomical studies) in the model and to provided information on the time-course of the understanding of concrete word meaning understanding (Tomasello et al., 2016).

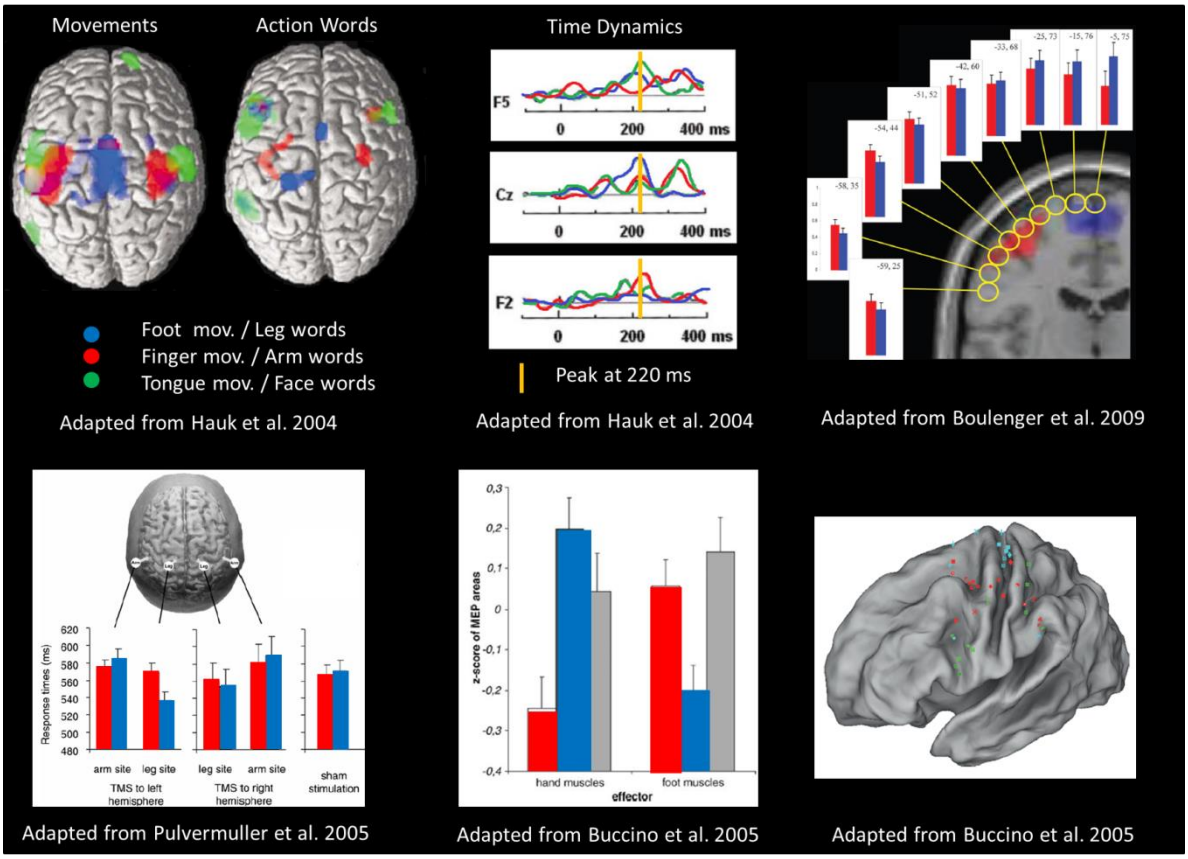


Figure 21 Review of some of the major results of the embodied perspective. Data from fMRI, M/EEG, and TMS converge in indicating that the same areas activated by motor tasks are recruited during semantic processing, in a somatotopic fashion. Throughout the figure red indicates movements of fingers/arm and semantic processing of words/verbs referring to those body parts, while blue is used for areas related with foot/leg, and green for tongue/face.

To sum up, partial support for the embodied theory of semantics comes from computational, neuroimaging and clinical data as it appears that sensory-motor areas are involved in conceptual processing. Potentially, embodied semantics solves the grounding issue: concept meaning is tightly linked with the sensory-motor

experiences that define our interaction with their referents. However, the extent to which this activation is necessary for concepts learning and storage is yet to be proven (see criticism in the following paragraph). Moreover, the omnicomprehensive deficit of SD patients is hardly accommodated in a completely distributed and embodied theory of semantics.

3.5 Latest Developments

The last few years saw the development of three (interconnected) axes: (1) the shifting of the attention towards how knowledge is acquired and what the elements necessary for a fruitful encoding of semantic information are; (2) the introduction of multivariate techniques for the analyses of neuroimaging data which allow new hypotheses to be put to test (see Chap. 2.4); (3) the spreading of computationally inspired models studying how semantic knowledge is distributed in the brain.

To study knowledge acquisition and manipulation is a key stepping-stone, which allows shedding light onto the weights assigned to different brain areas during the different stages of semantic processing. For instance, critics of the embodied theories have questioned the necessity of sensory-motor experience for the development of semantic knowledge. It has been shown that blind subjects, who never acquired any visual experience with animals or tools, present the same medial-to-lateral bias in the ventral visual path: nonliving stimuli elicit more activity in the medial fusiform gyrus, while living ones in lateral occipital cortex (Mahon et al., 2009). The observation of innate domain-specific constraints clashes not only with the embodied view, but also with any distributional theory of semantics based on frequency of co-occurrence of features and attributes. More specifically problematic for the embodied perspective is the observation of preserved conceptual processing in cases of motoric impairments. Deficits with motor-related semantics are not

observed in patients suffering from upper limb dysplasia (Vannuscorps and Caramazza, 2016), and corticobasal degeneration – even when followed longitudinally for three years (Vannuscorps et al., 2016). Conversely, simple training of pseudo words appears to be sufficient for the emergence of the domain specific dissociations between animals and tools in different semantic clusters, including the ATL (Malone et al., 2016). Finally, the anterior temporal lobe has been shown to play a crucial role in the acquisition of new conceptual knowledge through the integration of sensory features (Hoffman et al., 2014), while learning about new concepts (i.e., new animals) appears to tap into specific brain regions according to the feature learned (i.e., habitat in parahippocampal area and precuneus; eating habits in inferior frontal and post-central regions) (Bauer and Just, 2015). This set of results highlights the need (and feasibility) of studies aiming at discovering the neural organization of semantic knowledge in a dynamic way, paying attention to those elements that will turn out to be essential for the standard organization to be achieved.

Multivariate analyses permit to investigate whether the information represented in a given brain area is sufficient to discriminate a specific feature of the stimuli, for instance their semantic category. Critically, it allows the investigation of distributed patterns of information, as opposed to the massively univariate approach of classical methods (see Chap. 2.4). The first seminal paper that applied machine learning techniques to the study of concept organization in the brain used pictures as stimuli. It focused on the ventral visual path, known for being tessellated by a mosaic of areas selectively engaged for different kinds of stimuli (e.g., faces, letters, objects, etc...). Thanks to the new resolution provided by the method used, it was possible to show that the representations of faces and objects are distributed and overlapping (Haxby et al., 2001). Subsequently, similar approaches have been used to deepen our understanding of the neural correlates associated with the visual perception of different semantic categories (Carlson et al., 2003; Cox

and Savoy, 2003; Hanson et al., 2004; O'toole et al., 2005; Polyn et al., 2005; Hanson and Halchenko, 2008; Shinkareva et al., 2008; Connolly et al., 2012; Mur et al., 2012; Peelen and Caramazza, 2012; Carlson et al., 2014; Clarke and Tyler, 2014; Correia et al., 2014; Coutanche and Thompson-Schill, 2014; Connolly et al., 2016). One groundbreaking study used a computational model to predict the neural activation associated with written words presented with their relative picture (Mitchell et al., 2008). Another presented pictures and the relative written or spoken name (Akama et al., 2012). In all the above-mentioned studies, it is impossible to dissociate the contribution of low level properties of the physical input (i.e., the pictures used) from the pure semantic activation driven by the different concepts. Only recently have authors exploited multivariate methods to investigate neural processing of purely symbolic stimuli such as words. Some have compared the performance of classification methods when using pictorial stimuli as opposed to symbolic ones (Shinkareva et al., 2011; Devereux et al., 2013; Fairhall and Caramazza, 2013; Simanova et al., 2014), while very few have directly focused on words as stimuli (Just et al., 2010; Buchweitz et al., 2012; Bruffaerts et al., 2013; Correia et al., 2014; Liuzzi et al., 2015).

Some of the most interesting results of this line of research include:

- looking at global patterns (i.e., whole brain activity), it is possible to dissociate intra-categorical differences in the non-living domain (i.e., tools vs dwellings) irrespective of whether pictures or words are used as stimuli – even though the effect with words is less strong (Shinkareva et al., 2011). Moreover, factor analysis revealed how physical (i.e., word length) and semantic (i.e., can it be used for shelter? can it be manipulated? Is it food-related?) factors have differential loadings across the cortex (Just et al., 2010).
- local patterns, which can be investigated with a technique called searchlight (i.e., multiple ROIs covering the whole brain), have

highlighted the role of occipito-temporal cortex in semantic classification. Semantic category (i.e., animals vs tools) can be distinguished within and across 4 different modalities [visual verbal (i.e., written words), visual non-verbal (i.e., pictures), auditory verbal (i.e., spoken words), and auditory non-verbal (i.e., sounds)], with written words being the hardest task (Simanova et al., 2014).

- along the ventral visual path, it is possible to observe a posterior–to–anterior gradient of abstraction: stimuli are first represented according to their physical features (e.g., pixel similarity), then according to their perceptual features (e.g., visual similarity), finally according to conceptual information (e.g., location of use) (Peelen and Caramazza, 2012; Devereux et al., 2013; Carlson et al., 2014; Clarke and Tyler, 2014)
- semantic similarity between words correlates with the patterns of activity in left perirhinal cortex (Broadman areas 35 and 36) (Bruffaerts et al., 2013), even if this might be true only for written words, as the effect was not observed for spoken ones (Liuzzi et al., 2015). Moreover, the anterior portion of the superior temporal sulcus (STS) appears to be involved in the processing of language invariant semantic meaning (Correia et al., 2014). Finally, the anterior temporal lobe is confirmed as crucial region where visual properties converge and are integrated (Coutanche and Thompson-Schill, 2014).

Exploiting the latest methodological advances, both in terms of spatial resolution and statistical analyses, some authors are attempting to recover the neural substrate of the distributed organization of concepts postulated by featural, connectionists, and distributional models reviewed above. These distributed neuroimaging studies differ not only in the technical choices concerning data collection and analyses, but also with respect to the underlying hypothesis on the nature of the distributed representations (see Fig. 22). As we have seen, one can postulate that the different dimensions along which

concepts are organized are interpretable and can be addressed explicitly (Fernandino et al., 2015b; Fernandino et al., 2015a). The results of this distributed yet functionally localized perspective indicate that different portions of the semantic network encode distinct categories/features during semantic processing. On the other hand, one can hypothesize that knowledge is represented by a continuous semantic space mapped across a large extent of cortex (Huth et al., 2012; Huth et al., 2016). The results of this data-driven approach indicated that most areas within the semantic system represent clusters of related concepts, yet which features determine the emergence of each observed domain is not clear.

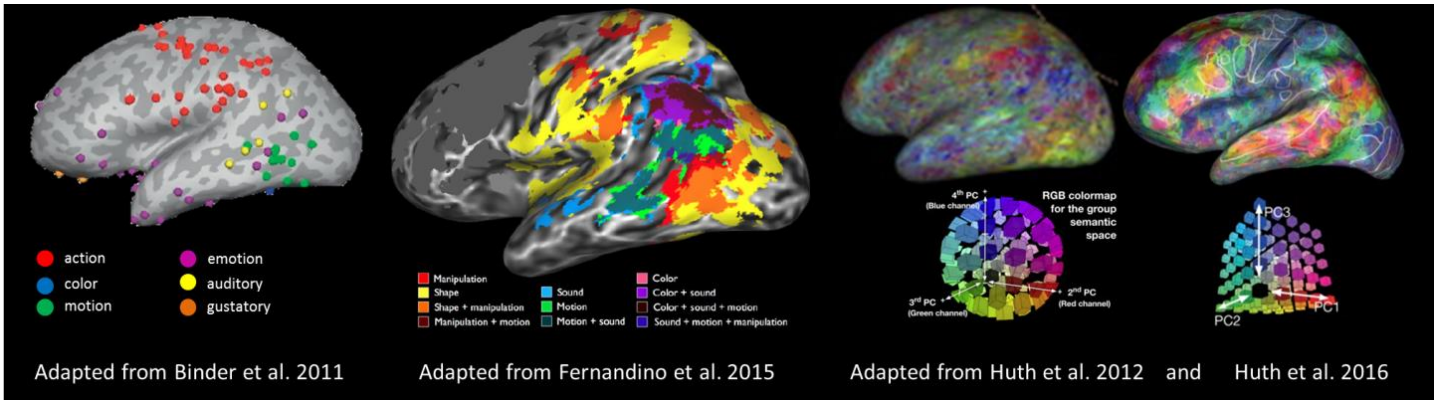


Figure 22 Topographical organization of different semantic dimensions. A review of the literature suggests that modality-specific activation peaks are distributed across the cortex in close proximity with the primary sensory-motor areas processing that kind of information (left). A recent investigation confirmed the results by investigating the distribution of 5 sensory-motor attributes (i.e., color, shape, visual motion, sound, and manipulation). It revealed that these aspects of conceptual knowledge are encoded in higher level unimodal and multimodal areas, the same areas involved in processing the corresponding types of information during perception and action (central). More data-driven studies have been able to show that the vast majority of the cortex responds to the semantic information presented visually or acoustically in naturalistic circumstances; however, in this case the dimensions are not directly interpretable even when dimensionality reduction techniques such as PCA are applied (right).

3.6 Open Questions and Future Directions

Given the current state of the art, it appears that some key questions have been answered and some are left open for future investigations.

Is semantic knowledge distributed across the cerebral cortex? Yes, it seems irrefutable that many areas contribute to semantic processing, but there seems to be a (yet to be properly described) functional specialization.

What is(are) the principle(s) organizing the neural representation of semantic knowledge? Whether the underlying

organizing principle is by domains, by features or by a combination of the two is an open question.

Is there a need for a convergence hub? Convergence hubs are not the only way the brain possess to integrate information, long range connections could potentially explain the observed activations as well as the detected deficits (Pulvermüller, 2013). However, it seems that much of the clinical evidence would not be accounted for by a theory excluding the existence of semantic hubs.

How many convergence hubs are there? What is their specific contribution to semantic processing? It is possible to presume that different hubs have different roles (e.g., integrating information from different sources). This is one of the most interesting open questions that neuroimaging studies can help elucidate.

Why are hubs located in those specific areas? Which kind of computations do they allow? It is unlikely that the hubs are located in random spots across the cortex. If (see previous point) they subserve different kinds of integrations, they likely are located where (1) they can easily access the information they are supposed to integrate; (2) they can perform the appropriate computations. A combination of computational and cognitive neuroscience is thought to answer this kind of question.

Which of the involved areas are actually necessary (and not just accessory) components of the semantic network? To date, the only viable way to gather data able to support causal inference is to expand the effort on clinical studies and virtual lesions ones (e.g., with TMS, see also Chap 2).

4. Temporal Dynamics of Semantic Representations

To study the timing of mental processes and representations, behavioral chronometric measures can be used. Traditionally, reaction times in different experimental conditions are considered a proxy of the duration and sequencing of cognitive operations. Regarding semantic processing, for instance, priming experiments have suggested the existence of processing in two phases: first, linguistic co-occurrences determine the content of the representation, then around 200 ms grounded perceptual simulation intervenes (Ostarek and Vigliocco, 2016). Later, while reviewing the behavioral methods available, I will explore in more detail the semantic priming paradigm (Chap 2.1.3). Moreover, I will present the results of our own priming experiment investigating how automatically different components of semantics are activated (Chap 3.4).

So far, in my overview of the neural substrate of semantic knowledge, I have focused on the topographical organization of such a system. However, the content of a representation in a given region might be changing dramatically over a short period, with different dimensions/features being activated at different time points: for instance, one could hypothesize that visual areas are involved in processing perceptual characteristics of the stimuli (e.g., word lengths) at T1 while at T2 they are replaying visual conceptual properties (e.g., the words refer to something red). For instance, Broca's area appears to code, in rapid succession, for lexical, grammatical and phonological features (Sahin et al., 2009) Moreover, different areas might be involved in this dynamic representation at different points in time: for instance, one could argue that during a given task, information coming from visual areas (T1) is read out by higher order cognitive areas in the temporal lobe (T2), which later provide inputs for complex computations happening in the frontal lobe (T3).

4.1 Temporal Representation

The neuroimaging techniques of choice when interested in fine-grained temporal dynamics are electroencephalography –EEG- and magnetoencephalography –MEG– (see Chap. 2.3). Overall, during reading, brain activation unfolds from occipital areas towards the anterior temporal pole (Marinkovic et al., 2003; Pammer, 2009). Similarly, listening elicits first activity in primary auditory areas and subsequently in supramodal temporal areas including the anterior temporal pole (Marinkovic et al., 2003; Salmelin, 2007). In both cases, the physical features of the stimuli are resolved within the first few milliseconds in modality specific areas (i.e., primary visual areas for written words, primary auditory areas for spoken words) and then converge in anterior temporal and inferior frontal cortices around 400ms (Marinkovic et al., 2003).

During the first 200 ms, analyses of the visual-orthographic feature, starting in primary visual cortex, spreads in a feed-forward wave along the inferior occipital gyrus and fusiform gyrus (Tarkiainen, 1999; Pammer et al., 2004). Likewise, the acoustic–phonetic analysis of spoken words takes places within the first 100 ms (N100) in non-primary auditory cortex (Kuriki and Murase, 1989; Parviainen et al., 2005). The language-specific phonetic and phonological analysis takes place in inferior frontal cortex and angular/supramarginal gyrus within the first 100-350 ms, when the mismatch negativity denotes access to phonological categories (Näätäneiv et al., 1997). Finally, between 200 and 500 ms, activity in superior and inferior temporal cortex, along with the inferior frontal one, denotes lexical-semantic processing (Kutas and Hillyard, 1980; Helenius et al., 2002).

Fine-grained features of semantic processing have been explored by studies investigating event-related potentials (or fields –ERP/ERF) following semantically charged stimuli such as sentences and single words. It is traditionally accepted that post-lexical semantic processes (i.e., those processes taking place after the meaning of the

word has been retrieved) are reflected by late components of ERP and ERF (Holcomb and Neville, 1990). Nevertheless, lexical effects (i.e., lexicality, word frequency, and word regularity) can be detected as early as 200 ms after stimuli onset (Sereno et al., 1998).

One of the most studied ERPs linked with semantic processing is the N400: a negative (N) deflection of the signal that starts around 300 ms and peaks around 400 ms (Kutas and Hillyard, 1980). It has been associated with the presentation (either auditory or visually) of words generating semantic violations such as *socks* in the following context: "*I like my coffee with cream and socks*" (Lau et al., 2008). Numerous factors have been shown to influence the shape of the N400, including:

- the degree of anomaly (e.g., in the example above, *socks* instead of *sugar*)
- the predictability (e.g., in the example above, *honey* instead of *sugar*, they are both semantically valid but one is very unlikely)
- the number of semantic features shared (e.g., in the example above, *salt* instead of *sugar* would produce a smaller N400 than *socks*)

The typical N400 effect is generally widespread across the scalp with a central-parietal tendency. Intracranial recordings suggest that the underlying sources of the N400 are located in the anterior-medial temporal lobe (McCarthy et al., 1995; Nobre and Mccarthy, 1995). Different interpretations of the N400 have been put forward. Some authors, following the so called integration view, posit that it reflects the incorporation of the words with its context (Brown and Hagoort, 1993). Other authors support a lexical view, thus suggesting that the N400 represents the activation in long term memory of the features associated with the critical word (Kutas and Federmeier, 2000). A seminal review concluded that there is strong evidence supporting the N400 as reflecting facilitated access, without discarding the role of integration mechanisms in building the predictions that facilitate access (Lau et al., 2008). However, as we will see next, there are recent indications that some aspects of word meaning might arise

much earlier than the N400 wave. These findings question the timing of semantic access and open the possibility that semantic content might be recovered not in a unitary fashion, but rather differentially according to the dimensions considered (in our lexicon: motor-perceptual vs conceptual ones) and the concurrent context (e.g., the task at end). The N400 is followed by a later negativity (N700) which seems to be most prominent when mental imagery is in place (West and Holcomb, 2000). This observation led to the investigation of possible differences in the profiles of the N400 and N700 generated by abstract and concrete words: once all other factors are controlled for, concrete words are associated with larger negativity waves (Barber et al., 2013).

Advocates of the embodied theory of semantics have tried to identify the first point in time when sensory-motor areas are recruited during conceptual processing. With a visual presentation of the stimuli, somatotopically coherent differences between verbs related to different body parts have been observed at 240 ms (Pulvermüller et al., 2000) and 220 ms (Hauk and Pulvermuller, 2004) after stimulus onset. When words are presented orally, specific cortical topographies appear earlier: 172–176 ms, 200 ms (Pulvermüller et al., 2005a). Moreover, authors have been able to show that both verbs and nouns can elicit characteristic somatotopic activations in motor cortex as early as ~80 ms after the acoustic disambiguation (i.e., the point when the words can be identified from the available acoustic information) (Shtyrov et al., 2014). However, this study is of difficult interpretation as they presented the same 6 words throughout the experiment (each seen 180 times), and the somatotopic distinctions across them correlated with the difference in their initial phonemes. It is thus possible that the early somatotopy observed as the product of the specific experimental conditions reflects the ultra-rapid semantic activity due to the particular experimental set. A double dissociation of word-categories has been reported at 150 ms: it appears that at that point action-related words most strongly activate fronto-central motor

areas while visual object-words activate the occipito-temporal cortex (Moseley et al., 2013). Furthermore, it appears that the motor cortex exhibits a higher mismatch negativity-like response and a higher predictive response (so called readiness potential) when single words are presented in body-part-incongruent sound contexts (e.g., "kiss" in the sound context of footstep) than in body-part-congruent contexts (e.g., "kiss" in whistle context) (Grisoni et al., 2016). Finally, the computational model mentioned above (3.4) has been able to replicate not only where, but also when semantic activation should take place (Tomasello et al., 2016): the central semantic hubs of the network activate slightly before modality-preferential areas carrying semantic information.

The multivariate techniques we have seen applied to fMRI data in the previous section have been rapidly extended to the analyses of M/EEG data as well. One of the first multivariate investigations of MEG data revealed that position of the stimuli could be decoded ~70ms after stimulus onset, classification based on low level visual features (i.e., objects vs textures) was possible at 110 ms, and finally semantic categories (i.e., faces vs cars) could be correctly classified at 135 ms (Carlson et al., 2011). Another group of authors has shown that the semantic category of pictures denoting animals or tools could be successfully decoded with both EEG and MEG signals using a preselected time-frequency bin (optimized thanks to the Common Spatial Patterns technique, CSP), ranging from 95 to 360 ms after stimulus onset, and from 4 to 18 Hz) (Murphy and Poesio, 2010). Subsequently, the same authors have deepened their exploration of EEG single trials decoding both at the individual subject level and at the group level (Murphy et al., 2011). Even in this case the optimal window was chosen thanks to CSP (100-370ms; 3-17Hz) whose spatial components indicate that a wide range of occipital, parietal and frontal areas played a role. Further attempts have focused on the possibility of dissociating different physical, perceptual and conceptual semantic features elicited by the conjoint presentation of

pictures of concrete items and their relative name (Sudre et al., 2012). Differences in the time course and locations of decodable semantic information were found: physical and perceptual features can be recovered earlier than conceptual ones, the former can be related with activity in posterior cortical areas, while the latter involves anterior-lateral ones. A separate group selected as stimuli images of 4 different categories (i.e., faces, scenes, bodies and tools), and applied multivariate analyses to the reconstructed sources of MEG signal (van de Nieuwenhuijzen et al., 2013). They were able to detect differences in visual category perception 85 ms after stimulus onset and to observe the evolution of the spatiotemporal dynamics: first, inferior occipital, inferior temporal and superior occipital gyrus sources are involved; then, additional sources in the anterior inferior temporal gyrus and superior parietal gyrus intervene. Both univariate (Clarke et al., 2011; Clarke et al., 2013) and multivariate (Clarke et al., 2015) findings suggest a coarse-to-fine model of category information processing: perceptual analyses in visual cortex is followed by early semantic effects (i.e., categorical distinction) within the first 120 ms; only after 200 ms conceptual differentiation and object identification take place in ventral temporal cortex.

Finally, only a handful of studies to date have attempted to recover semantic information from the electrophysiological signal evoked by symbolic stimuli. The first study demonstrated the possibility of achieving good single MEG trial classification of words, but without distinguishing between the contributions of their physical (e.g., the visual properties) and semantic (i.e., the meaning) properties (Guimaraes et al., 2007). Then, researchers focused on the possibility of recovering semantic category information from EEG and MEG data (Chan et al., 2011a) as well as intracranial macro- and micro-electrodes (Chan et al., 2011b). The decoding performance suggests that the representations of semantic categories is highly spatially distributed, involving in particular the anterior temporal, and inferior frontal cortices (Chan et al., 2011a). Furthermore, category-selective responses can occur at short latency (130 ms) and are detected in

measures sensitive to unit firing and synaptic activity (e.g., local field potentials and high gamma power) (Chan et al., 2011b). Moreover, the integration of lexical-semantic knowledge at different cortical scales (e.g., two visual attributes vs one visual and one auditory attribute) is reflected in frequency-specific oscillatory neuronal activity in unisensory and multisensory association networks (van Ackeren et al., 2014). Recently, it has been shown that even the category of internally generated words can be recovered from MEG single trials (Simanova et al., 2015), and that across-language generalization, denoting the activation of high-level semantic representations, appears to be possible around 550-600 ms (Correia et al., 2015).

Overall, the traditional semantic effects linked with the N400 (which, in our framework, appear to be mostly of conceptual nature) paired with the (few) early motor-perceptual activations observed, suggest that high-order conceptual integration follows re-activation of motor-perceptual features. However, so far no study has directly compared the representations of motor-perceptual and conceptual dimensions within the same subject, with the same stimuli and task. This kind of comparison, which controls for difference at the identification stage, is needed if one wishes to draw inferences on the relative temporal dynamics of different components of semantic representations.

4.2 Spectral Representation

The nature of M/EEG signals offers another precious tool: time-frequency analysis, which studies signals in both the time and frequency domains simultaneously (see Chap. 2.3). Frequency bands that have been associated with (different) key roles during language processing include:

- delta band (0.5 – 3.5 Hz) synchronization (i.e., power increase) is associated with inhibition of sensory afferences potentially interfering with the accomplishment of the task or attention

allocation during cognitive operations, including semantic tasks (Brunetti et al., 2013; Harmony, 2013; Kielar et al., 2015; Guntekin and Basar, 2016).

- theta (2-8 Hz) synchronization has been linked to lexical memory retrieval (Bastiaansen et al., 2005; Bastiaansen et al., 2008; Shahin et al., 2009; Maguire et al., 2010; Bakker et al., 2015; Kielar et al., 2015). Theta has also been associated with the integration of unimodal semantic features (van Ackeren et al., 2014) and the detection of semantic violations (Davidson and Indefrey, 2007).
- alpha (10–14 Hz) desynchronization (i.e., power decrease) has been associated with retrieval of lexical and semantic information (Shahin et al., 2009; Kielar et al., 2015), as well as the detection of grammatical violations (Davidson and Indefrey, 2007).
- lower beta band (17-20 Hz) desynchronization has been shown to be linearly related with the N400 ERP component (Wang et al., 2012) and has been generally linked with lexical-semantic processing (Davidson and Indefrey, 2007; Shahin et al., 2009; Bakker et al., 2015; Kielar et al., 2015)
- upper beta (25–30 Hz) synchronization has been observed during semantic tasks (Shahin et al., 2009).
- gamma (>30 Hz) has been associated with a series of combinatorial processes such as semantic unification (Braeutigam et al., 2001; Hagoort et al., 2004; Hald et al., 2006) and the combination of multimodal semantic information (van Ackeren et al., 2014). Gamma synchronization has also been shown to be sensible to repetition suppression effects (Matsumoto and Iidaka, 2008) and associated with semantic tasks (Shahin et al., 2009).

Generally speaking, it appears that:

a) desynchronization of alpha and lower beta is linked with attention processes and allocation of resources during cognitive task, including linguistic ones;
b) synchronization of slow frequencies, theta and delta, is associated with memory retrieval, including semantic memory;

c) synchronization of high frequencies, gamma and upper beta, is linked with unification processes, including linguistic ones.

With this respect, some authors have recently proposed a frequency-based segregation of syntactic and semantic unification processes (Bastiaansen and Hagoort, 2015): gamma band power appears to be linked with semantic unification (i.e., larger for semantically coherent than for semantically anomalous sentences), while lower beta band may signal syntactic unification (larger for syntactically correct sentences than for incorrect ones). Overall, it has been suggested that object representations may lie in the synchronized activity of cell assemblies representing different stimulus features (Tallon-Baudry and Bertrand, 1999). These cell assemblies (and thus the features they encode) appear to be distributed across different brain regions, and further studies are needed in order to successfully disentangle their contribution.

4.3 Long Range: Context and Experiences

Far from being resolved, the question "when?" can take an even larger range declination.

First of all, the depth and thoroughness of semantic processing will depend on the task at hand. It is possible to think that different circumstances will lead to a different load on embodied representations or, in the words of Zwaan (2014), to "*different levels of environmental embeddedness*". It is likely that different motor-perceptual and conceptual features are evoked only when needed, in an automatic yet task-conditioned fashion: for instance, reading "p a s h m i n a" during a Farsi class or on tag stripped from a sweater will activate different representations (i.e., the translation "woolen goods" in the first case, the soft feeling of wool in the second one).

Second, words are learned over a lifetime and many different experiences accompany the learning process. Therefore, it is plausible to expect individual differences to be molded on the different interactions one has with the items the words refer to: for example, the

concept of *wool* will be radically different for a sheep shepherd and for a shop assistant. For a thorough review on the contextual effects on conceptual processing at different timescales (from subject specific experience to current task goals), see (Yee and Thompson-Schill, 2016).

It appears that concepts are more dynamic than most existing theories can account for, thus we are in great need to develop explicit and testable predictions on when/why in certain situations different aspects of word meaning (more motor-perceptual or higher level conceptual or declarative) are expected to be activated. In light of the dissociation I introduced at the beginning of the chapter, between semantic representations and semantic processing, an extreme position would claim that stable, default, semantic representations do not exist, and that concepts are constructed online given one's previous experiences, the task to be solved and the goals to be satisfied.

To sum up, semantic information is readily available, already in the first 200-300 ms after stimulus onset. On one hand, differences in latency seem to be due to the level of processing more than to category/features. On the other hand, oscillations and frequency changes appear to play a role in feature integration. Overall, the role of timing appears crucial for the field: as recently stressed by Hauk, precise timing information will be key to the debate on the neural substrate of semantic information enabling us to differentiate the role of different distributed networks while distinguishing top-down from bottom-up processes, feed-forward from feedback ones (Hauk, 2016).

5. Format and Implementation of Semantic Representations

After having cleared the current views on *when* and *where what* we know is stored and retrieved in our brain, one challenge is

left for us. What is the nature of semantic representations? The matter of the discussion here is the format of the representation as well as its underlying neural code. To understand the concept of *format*, let's compare a bitmap and a vector graphic. They might have the same content (*what*), but they diverge in their formats (respectively a collection of pixels or of Bézier curves). Similarly, in the case of semantic representations, there are two competing views. On one hand there are those claiming that conceptual knowledge is stored in an abstract, amodal, propositional format (Mahon and Caramazza, 2008). On the other hand, there are those conjecturing that a modality specific analogical format is necessary and sufficient to represent semantic information (Barsalou, 2010). These are only the extremes of a continuum, which see in moderate positions those that suspend their judgment (a real ἐποχή, epokhē): as we will see, the question turns out to be an extremely ill-posed problem (Martin, 2015).

5.1 Relation between Geometry, Format and Neural Code

I have previously stated that the content of semantic representations are concepts, and, in a reductionist view, the meaning of words. The term representational geometry can be used to describe the organization of such content: it refers to the relationships (i.e., distances) between items (e.g., words) conceptualized as points in a multidimensional space. For instance, the bi-dimensional space described by the visual features of color and shape sees elongated orange-ish items (e.g., carrot), closer to elongated yellow-ish items (e.g., banana), and further apart from round yellow-ish items (e.g., lemon). The geometry – and thus the distances – would be different in a space dominated by conceptual taxonomic dimensions (e.g., lemon and banana would cluster together – being fruit – and would be far apart from carrot – a vegetable). Thanks to this higher-order layer, representations stemming from different sources can be compared (Kriegeskorte and Kievit, 2013): cognitive geometries derived from

behavioral data, predicted similarities as estimated by computational models, and neural representational spaces as resulting from neuroimaging observations.

A different problem is that of the representational format. As illustrated by the example of bitmap and vector images, given identical content, what tells apart one format from the other is which kinds of operations we can perform over the content. For instance, only vector drawings can be scaled without loss in quality. The quest for the appropriate descriptors of the format of neural representations (not only semantic ones) is still open.

Finally, the term neural code refers directly to the meaningful scheme of the activity of single neurons or of a populations of neurons (e.g., in the simplest models it is the firing rates, in more complex ones the precise temporal patterns of spikes) that allows encoding of (some feature of) the stimuli (Pouget et al., 2000). Crucially, the same area, through different neural codes, could encode multiple geometries of the same content (e.g., perceptual and conceptual similarities across the same items), or same geometry for different content (e.g., relative distances across different magnitudes).

To sum up, the format of a given representation and its neural implementation (i.e., the underlying neuronal code) should not be confused with its content/geometry (i.e., the aspect(s) of the semantic space that is encoded) or its localization (i.e., the brain region where the neural activity is observed) (Mahon and Hickok, 2016). The following debate stems from the more or less explicit assumptions made by different authors on the relation between content, geometry and format. Is the format determined by the content and/or by the location? Is there an isomorphism between what is represented and how it is represented?

5.2 Debate

In the early eighties, Paivio (1986) introduced the dual coding theory postulating the existence of multiple coding formats (i.e., a

verbal and a visual code) relying on different types of representational units (i.e. respectively logogens and imagens). The debate surrounding mental representations has been particularly heated in the field of mental imagery. The question at stake is whether images recreated via imagery are depictive (Kosslyn et al., 2001) or propositional (Pylyshyn, 2003) in nature. Recent methodological advances (see Chap. 2.4) have paved the way to new paradigms that might bring us closer to directly testing the hypothesis at stake. For instance, in primary visual areas algorithms trained on depictive sensory representation data have been shown to work once applied to data collected during imagery, demonstrating that the representation of perceived and imagined stimuli shares at least some of the same low level encoding characteristics (Naselaris et al., 2015; Pearson and Kosslyn, 2015).

Concerning semantic representations, I have mentioned that the continuum of possible theories sees two extremes. Some authors implicitly assume an isomorphism between brain localization (*where*), representational content (*what*) and format (*how*), and thus draw inferences on the content and the format of a representation from the localization of the brain region it activates. For instance, the observation of activation of the motor cortex during verb processing is taken to indicate that the content and the format of the representations are motoric (Barsalou, 1999). This would entail a geometrical configuration of different verbs that follows their relative distance along motoric dimensions (e.g., complexity of the movement), and the dependency on a neural code that also subserves the encoding of actual motor information during movement execution. Other author assume that the format of the representations can be entirely dissociated from both content (geometry) and localization (Mahon and Caramazza, 2008). For example, the code used to store motor and visual semantic representations can be identical, (e.g., purely abstract), even if implemented in different brain regions (e.g., motor cortex, visual cortex, or ATL).

Notwithstanding this fundamental difference across perspectives, it is not clear how the two views could be tested empirically, other than possibly through in vivo cellular recordings aiming at understanding the neural code associated with different representations. For instance comparing the pattern of neuronal spiking rates during (a) the reading of action verbs, and (b) the actual action execution, across the different areas potentially involved: the motor cortex, the visual cortex, and the ATL. However, down to its core, the problem is that all the conclusions we draw from our observations will be interpreted in light of our assumptions on the underlying coding scheme. To date, there is no ground truth corroborating those assumptions, not even for simple models such as visual representations in V1 (i.e., information could be carried by the pattern of activity across a large number of cells, the timing of the first wave of spikes, the timing or phase of continuous activity, synchrony across a population, etc… (de-Wit et al., 2016)).

5.3 Skeptical Epoché

In the interim, the conclusion one can reach is that suggested by Alex Martin (Martin, 2015): the question should be put aside as long as the field lacks appropriate cognitive descriptors to fit the neural substrate and agreed-upon procedures to determine the best proxy of the format of a representation. As of now, all the currently available neuroimaging methods lack adequate spatio-temporal resolution to tap directly on the format question. One possible compromise is to focus on the investigation of representational geometries through multivariate pattern analysis (or adaptation) across different brain areas (see Chap. 2.4). Indeed, representational geometries derived from imaging techniques can be seen as a proxy of the locally distributed population code, an intermediate level of description which supports the investigation of differential representational properties across the cortex (Kriegeskorte and Kievit, 2013). In the future, different theories on the organization of semantic

memory should make precise predictions on the kind of dissociations that one should expect when contrasting the representational geometry of critical cortical areas, at different time points.

6. Conclusions

After this broad overview, two key aspects should be clear. First, semantic knowledge lies at the very core of human nature. It is often used as a proxy of human-like intelligence with great implications for computer science, heating a debate that calls for philosophical reflections, as well as behavioral and imaging experiments. With varying degrees, it is in action in all our everyday activities: reading (understanding what we read!), using a mobile phone, cooking a traditional family dish. Its breakdown is thus highly disabling, causing great suffering to the patients and their families.

Second, semantic knowledge is a complex cognitive and neurological reality. Its neuropsychological aspects, having been the focus of extensive investigation, are well defined. On the contrary, its neural substrates require further exploration. Semantic knowledge appears to be distributed across the cortex in a specialized manner, involving modal, multimodal and heteromodal cortices. It is quickly recovered, differentially depending on the current context and previous experiences. Finally, the question on the neural code (or the neural format) of representations is currently considered an ill posed problem not limited to studies on semantics, but affecting all neuroimaging investigations of neural representation.

Recently, authors from virtually all the different perspectives here reviewed have somewhat agreed to disagree: it appears obvious that semantic representations need to be grounded (somehow), at the same time it is accepted that purely symbolic operations are indeed central to human cognition. In the words of an exponent of the embodied perspective (Zwaan, 2014): "We *need mental*

representations. At least some of them need to be grounded in perception and action. Not all processing requires representations that are directly grounded in perception and action […] *Which system dominates the comprehension process depends on the level of embeddedness."* As highlighted by one of the proponent of a *hybrid* theory (Martin, 2015), our daily life requires a: "[…] *dynamic interaction between higher-order conceptual, perceptual, and lower-order sensory regions in the service of specific task and bodily demands."* As summarized by (Mahon, 2015), an advocate for the necessity of abstract representations: "*We know that the conceptual system can "turn" the sensorimotor system and the sensorimotor system can "turn" the conceptual system. But we also know that conceptual processing can proceed unencumbered by the representation of the world and the body. […]It all depends on one's theory of activation dynamics—or information exchange— among representationally distinct processes."* The burning question, thus, is how do these different kinds of representations interact.

In the present thesis, representations of different kinds (i.e., perceptual and conceptual ones) are compared in terms of behavioral relevance (Chap 3), topographical organization (Chap 4) and temporal dynamics (Chap 5). I capitalized from recent methodological improvements (Chapt 2), that have the potential of widening the set of hypotheses that can be tested with neuroimaging techniques (Davis and Poldrack, 2013). I did so aware of the fact that any methodological progress should be accompanied by developments in the theoretical frameworks used to interpret the new findings (Coveney et al., 2016). For instance, results coming from the ever-improving neuroimaging techniques cannot be dissociated from the challenging clinical data. The work presented in this thesis appears thus timely and relevant for the clinical, theoretical and methodological dimensions of the quest for the neural correlates of semantic representations.

Bibliography

Abboud S, Maidenbaum S, Dehaene S, Amedi A (2015) A number-form area in the blind. Nature communications 6:6026.

Agosta F, Viskontas IV, Gorno-Tempini ML, Filippi M (2016) fMRI of Memory. 119:419-450.

Agosta F, Henry RG, Migliaccio R, Neuhaus J, Miller BL, Dronkers NF, Brambati SM, Filippi M, Ogar JM, Wilson SM, Gorno-Tempini ML (2009) Language networks in semantic dementia. Brain : a journal of neurology.

Agosta F, Galantucci S, Magnani G, Marcone A, Martinelli D, Antonietta Volonte M, Riva N, Iannaccone S, Ferraro PM, Caso F, Chio A, Comi G, Falini A, Filippi M (2015) MRI signatures of the frontotemporal lobar degeneration continuum. Human brain mapping 36:2602-2614.

Akama H, Murphy B, Na L, Shimizu Y, Poesio M (2012) Decoding semantics across fMRI sessions with different stimulus modalities: a practical MVPA study. Frontiers in neuroinformatics 6:24.

Anderson JR (1978) Arguments concerning representations for mental imagery. Psychological review 85:249–277.

Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M (2006) Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions. Current Biology 16:1818-1823.

Bak TH, Hodges JR (2004) The effects of motor neurone disease on language: Further evidence. Brain and Language 89:354-361.

Bak TH, Chandran S (2012) What wires together dies together: verbs, actions and neurodegeneration in motor neuron disease. Cortex 48:936-944.

Bak TH, O'Donovan DG, Xuereb JH, Boniface S, Hodges JR (2001) Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease–dementia–aphasia syndrome. Brain : a journal of neurology 124:103-120.

Bak TH, Yancopoulou D, Nestor PJ, Xuereb JH, Spillantini MG, Pulvermuller F, Hodges JR (2006) Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing. Brain : a journal of neurology 129:321-332.

Bakker I, Takashima A, van Hell JG, Janzen G, McQueen JM (2015) Changes in theta and beta oscillations as signatures of novel word consolidation. J Cogn Neurosci 27:1286-1297.

Barber HA, Otten LJ, Kousta ST, Vigliocco G (2013) Concreteness in word processing: ERP and behavioral effects in a lexical decision task. Brain Lang 125:47-53.

Barros-Loscertales A, Gonzalez J, Pulvermuller F, Ventura-Campos N, Bustamante JC, Costumero V, Parcet MA, Avila C (2011) Reading Salt Activates Gustatory Brain Regions: fMRI Evidence for Semantic Grounding in a Novel Sensory Modality. Cerebral Cortex 22:2554-2563.

Barsalou LW (1999) Perceptual symbol systems. Behavioral and brain sciences 22:577–660.

Barsalou LW (2010) Grounded cognition: past, present, and future. Topics in cognitive science 2:716-724.

Bastiaansen M, Hagoort P (2015) Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. J Cogn Neurosci 27:2095-2107.

Bastiaansen MC, Oostenveld R, Jensen O, Hagoort P (2008) I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. Brain Lang 106:15-28.

Bastiaansen MC, Van Der Linden M, Ter Keurs M, Dijkstra T, Hagoort P (2005) Theta responses are involved in lexical—Semantic retrieval during language processing. Journal of cognitive neuroscience 17:530-541.

Bauer AJ, Just MA (2015) Monitoring the growth of the neural representations of new animal concepts. Human brain mapping 36:3213-3226.

Binder JR, Desai RH (2011) The neurobiology of semantic memory. Trends in Cognitive Sciences 15:527-536.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cerebral Cortex 19:2767-2796.

Binder JR, Conant LL, Humphries CJ, Fernandino L, Simons SB, Aguilar M, Desai RH (2016) Toward a brain-based componential semantic representation. Cognitive neuropsychology:1-45.

Binney RJ, Parker GJ, Ralph MAL (2012) Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. Journal of cognitive neuroscience 24:1998-2014.

Binney RJ, Embleton KV, Jefferies E, Parker GJ, Ralph MA (2010) The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. Cereb Cortex 20:2728-2738.

Blundo C, Ricci M, Miller L (2006) Category-specific knowledge deficit for animals in a patient with herpes simplex encephalitis. Cognitive neuropsychology 23:1248-1268.

Bonner MF, Peelle JE, Cook PA, Grossman M (2013) Heteromodal conceptual processing in the angular gyrus. NeuroImage 71:175-186.

Bonnici HM, Richter FR, Yazar Y, Simons JS (2016) Multimodal Feature Integration in the Angular Gyrus during Episodic and Semantic Retrieval. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:5462-5471.

Borghesani V, Piazza M (under review) The neuro-cognitive representations of symbols: the case of concrete words. . Neuropsychologia.

Borghesani V, de Hevia L, Viarouge A, Pinheiro Chagas P, Eger E, Piazza M (2016) Comparing magnitudes across dimensions: a univariate and multivariate approach. . International Workshop on Pattern Recognition in Neuroimaging.

Boulenger V, Hauk O, Pulvermuller F (2008a) Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. Cerebral Cortex 19:1905-1914.

Boulenger V, Mechtouff L, Thobois S, Broussolle E, Jeannerod M, Nazir TA (2008b) Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. . Neuropsychologia 46:743-756.

Bozeat S, Ralph MAL, Patterson K, Garrard P, Hodges JR (2000) Non-verbal semantic impairment in semantic dementia. Neuropsychologia 38:1207-1215.

Braeutigam S, Bailey AJ, Swithenby SJ (2001) Phase-locked gamma band responses to semantic violation stimuli. Cognitive Brain Research 10:365-377.

Brambati SM, Rankin KP, Narvid J, Seeley WW, Dean D, Rosen HJ, Miller BL, Ashburner J, Gorno-Tempini ML (2009) Atrophy progression in semantic dementia with asymmetric temporal involvement: a tensor-based morphometry study. Neurobiology of aging 30:103-111.

Bréal M (1904) Essai de sémantique: science des significations. .

Broca P (1861) Remarques sur le siège de la faculté du langage articulé, suivis d'une observation d'aphémie (perte de la parole). Bulletin de La Société Anatomique 6:330–357.

Broca P (1865) Sur le siège de la faculté du langage articulé. Bulletins de la Société d'anthropologie de Paris 6:377-393.

Brown C, Hagoort P (1993) The processing nature of the N400: Evidence from masked priming. Journal of Cognitive Neuroscience 5:34-44.

Bruffaerts R, Dupont P, Peeters R, De Deyne S, Storms G, Vandenberghe R (2013) Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:18597-18607.

Brunetti E, Maldonado PE, Aboitiz F (2013) Phase synchronization of delta and theta oscillations increase during the detection of relevant lexical information. Frontiers in psychology 4:308.

Buccino G, Colage I, Gobbi N, Bonaccorso G (2016) Grounding Meaning in Experience: A Broad Perspective on Embodied Language. Neuroscience and biobehavioral reviews.

Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, Rizzolatti G (2005) Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. Cognitive Brain Research 24:355-363.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. European journal of neuroscience 13:400-404.

Buchweitz A, Shinkareva SV, Mason RA, Mitchell TM, Just MA (2012) Identifying bilingual semantic neural representations across languages. Brain Lang 120:282-289.

Buxbaum LJ, Saffran EM (2002) Knowledge of object manipulation and object function: dissociations in apraxic and nonapraxic subjects. Brain and Language 82:179-199.

Capitani E, Laiacona M, Mahon B, Caramazza A (2003) What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. Cognitive neuropsychology 20:213-261.

Cappelletti M, Butterworth B, Kopelman M (2001) Spared numerical abilities in a case of semantic dementia. Neuropsychologia 39:1224-1239.

Caramazza A (1986) On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. Brain and cognition 5:41-66.

Caramazza A, Shelton JR (1998) Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Journal of Cognitive Neuroscience 10:1–34.

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

Caramazza A, Hills AE, Rapp BC, Romani C (1990) The multiple semantics hypothesis: Multiple confusions? Cognitive neuropsychology 7:161-189.

Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. J Cogn Neurosci 15:704-717.

Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR (2014) The emergence of semantic meaning in the ventral temporal pathway. J Cogn Neurosci 26:120-131.

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. Journal of vision 11.

Castelli F, Happe F, Frith U, Frith C (2000) Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12:314-325.

Chan AM, Halgren E, Marinkovic K, Cash SS (2011a) Decoding word and category-specific spatiotemporal representations from MEG and EEG. Neuroimage 54:3028-3039.

Chan AM, Baker JM, Eskandar E, Schomer D, Ulbert I, Marinkovic K, Cash SS, Halgren E (2011b) First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. Journal of Neuroscience 31:18119-18129.

Chao LL, Martin A (1999) Cortical regions associated with perceiving, naming, and knowing about colors. Journal of Cognitive Neuroscience 11:25-35.

Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nature neuroscience 2:913-919.

Clarke A, Tyler LK (2014) Object-specific semantic coding in human perirhinal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:4766-4775.

Clarke A, Taylor KI, Tyler LK (2011) The evolution of meaning: spatio-temporal dynamics of visual object recognition. Journal of Cognitive Neuroscience 23:1887-1899.

Clarke A, Devereux BJ, Randall B, Tyler LK (2015) Predicting the Time Course of Individual Objects with MEG. Cereb Cortex 25:3602-3612.

Clarke A, Taylor KI, Devereux B, Randall B, Tyler LK (2013) From perception to conception: how meaningful objects are processed over time. Cereb Cortex 23:187-197.

Cochin S, Barthelemy C, Lejeune B, Roux S, Martineau J (1998) Perception of motion and qEEG activity in human adults. Electroencephalography and clinical neurophysiology 107:287-295.

Collins AM, Quillian MR (1969) Retrieval time from semantic memory. Journal of verbal learning and verbal behavior 8:240-247.

Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. Psychological review 82:407.

Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV (2012) The representation of biological classes in the human brain. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:2608-2618.

Connolly AC, Sha L, Guntupalli JS, Oosterhof N, Halchenko YO, Nastase SA, di Oleggio Castello MV, Abdi H, Jobst BC, Gobbini MI, Haxby JV (2016) How the Human Brain Represents Perceived Dangerousness or "Predacity" of Animals. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:5373-5384.

Corbett F, Jefferies E, Ehsan S, Lambon Ralph MA (2009) Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. Brain : a journal of neurology 132:2593-2608.

Correia J, Formisano E, Valente G, Hausfeld L, Jansma B, Bonte M (2014) Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:332-338.

Correia JM, Jansma B, Hausfeld L, Kikkert S, Bonte M (2015) EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. Frontiers in psychology 6:71.

Cotelli M, Borroni B, Manenti R, Zanetti M, Arévalo A, Cappa SF, Padovani A (2007) Action and object naming in Parkinson's disease without dementia. European Journal of Neurology 14:632-637.

Cotelli M, Borroni B, Manenti R, Alberici A, Calabria M, Agosti C, Arévalo A, Ginex V, Ortelli P, Binetti G, Zanetti O, Padovani A, Cappa SF (2006) Action and object naming in frontotemporal dementia, progressive supranuclear palsy, and corticobasal degeneration. Neuropsychology 20:558-565.

Coutanche MN, Thompson-Schill SL (2014) Creating Concepts from Converging Features in Human Cortex. Cereb Cortex.

Coveney PV, Dougherty ER, Highfield RR (2016) Big data need big theory too. Philosophical transactions Series A, Mathematical, physical, and engineering sciences 374.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19:261-270.

Crutch SJ, Warrington EK (2003) The selective impairment of fruit and vegetable knowledge:amultiple processing channels account of fine-grain category specificity. Cognitive neuropsychology 20:355-372.

Cummins R (1989) Meaning and mental representation. Cambridge, MA: MIT Press.

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR (1996) A neural basis for lexical retrieval. Nature.

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

Daniele A, Giustolisi L, Silveri MC, Colosimo C, Gainotti G (1994) Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. Neuropsychologia 32:1325-1341.

Daniele A, Barbier A, Di Giuda D, Vita MG, Piccininni C, Spinelli P, Tondo G, Fasano A, Colosimo C, Giordano A, Gainotti G (2013) Selective impairment of action-verb naming and comprehension in progressive supranuclear palsy. Cortex 49:948-960.

Davidson DJ, Indefrey P (2007) An inverse relation between event-related and time-frequency violation responses in sentence processing. Brain research 1158:81-92.

Davie CA (2008) A review of Parkinson's disease. British medical bulletin 86:109-127.

Davies RR, Hodges JR, Kril JJ, Patterson K, Halliday GM, Xuereb JH (2005) The pathological basis of semantic dementia. Brain : a journal of neurology 128:1984-1995.

Davies RR, Kipps CM, Mitchell J, Kril JJ, Halliday GM, Hodges JR (2006) Progression in frontotemporal dementia: identifying a benign behavioral variant by magnetic resonance imaging. Archives of neurology 63:1627-1631.

Davis T, Poldrack RA (2013) Measuring neural representations with fMRI: practices and pitfalls. Annals of the New York Academy of Sciences 1296:108-134.

de-Wit L, Alexander D, Ekroll V, Wagemans J (2016) Is neuroimaging measuring information in the brain? Psychonomic bulletin & review.

Decety J, Grezes J, Costes N, Perani D, Jeannerod M, Procyk E, Grassi F, Fazio F (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. Brain : a journal of neurology 120:1763-1777.

Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. Trends Cogn Sci 15:254-262.

Demb JB, Desmond JE, Wagner AD, Vaidya CJ, Glover GH, Gabrieli JD (1995) Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. The Journal of Neuroscience 15:5870-5878.

Desgranges B, Matuszewski V, Piolino P, Chételat G, Mézenge F, Landeau B, De La Sayette V, Belliard S, Eustache F (2007) Anatomical and functional alterations in semantic dementia: a voxel-based MRI and PET study. Neurobiology of aging 28:1904-1913.

Devereux BJ, Clarke A, Marouchos A, Tyler LK (2013) Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:18906-18916.

Devlin JT, Gonnerman LM, Andersen ES, Seidenberg MS (1998) Category-specific semantic deficits in focal and widespread brain damage: A computational account. Journal of cognitive Neuroscience 10:77-94.

Devlin JT, Russell RP, Davis MH, Price CJ, Moss HE, Fadili MJ, Tyler LK (2002) Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. Neuropsychologia 40:54-75.

Devlin JT, 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. . Journal of Cognitive Neuroscience 15:71-84.

Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. Experimental brain research 91:176-180.

Diehl J, Grimmer T, Drzezga A, Riemenschneider M, Forstl H, Kurz A (2004) Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. Neurobiology of aging 25:1051-1056.

Ding SL, Van Hoesen GW, Cassell MD, Poremba A (2009) Parcellation of human temporal polar cortex: a combined analysis of multiple cytoarchitectonic, chemoarchitectonic, and pathological markers. The Journal of comparative neurology 514:595-623.

Donders FC (1968/1969) On the speed of mental processes. Acta psychologica 30: 412-431.

Downing PE, Wiggett AJ, Peelen MV (2007) Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:226-233.

Ellis AW, Young AW, Critchley EMR (1989) Loss of memory for people following temporal lobe damage. Brain and Language 112:1469–1483.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature Reviews Neuroscience 392:598-601.

Epstein R, Harris A, Stanley D, Kanwisher N (1999) The parahippocampal place area: Recognition, navigation, or encoding? Neuron 23:115-125.

Fadiga L, Fogassi L, Pavesi G, Rizzolatti G (1995) Motor facilitation during action observation: a magnetic stimulation study. Journal of neurophysiology 73:2608-2611.

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:10552-10558.

Farah MJ (2004) Visual agnosia. Cambridge: MIT press.

Farah MJ, McClelland JL (1991) A computational model of semantic memory impairment: modality specificity and emergent category specificity. Journal of Experimental Psychology 120:339.

Farah MJ, Wallace MA (1992) Semantically-bounded anomia: Implications for the neural implementation of naming. . Neuropsychologia 30:609-621.

Fernandino L, Humphries CJ, Seidenberg MS, Gross WL, Conant LL, Binder JR (2015a) Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia 76:17–26.

Fernandino L, Binder JR, Desai RH, Pendl SL, Humphries CJ, Gross WL, Conant LL, Seidenberg MS (2015b) Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. Cereb Cortex 26:2018–2034.

Frege G (1892) On Sense and Reference [Über Sinn und Bedeutung]. Zeitschrift für Philosophie und philosophische Kritik 100:25-50.

Galantucci S, Tartaglia MC, Wilson SM, Henry ML, Filippi M, Agosta F, Dronkers NF, Henry RG, Ogar JM, Miller BL, Gorno-Tempini ML (2011) White matter damage in primary progressive aphasias: a diffusion tensor tractography study. Brain : a journal of neurology 134:3011-3029.

Gallese V, Lakoff G (2005) The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. Cognitive neuropsychology 22:455-479.

Gallistel CR (2001) Mental representations, psychology of. In: International Encyclopedia of the social and behavioural sciences (P.B.Baltes NJSa, ed), pp 9691-9695. New York: Elsevier.

Galton CJ, Patterson K, Graham K, Lambon-Ralph MA, Williams G, Antoun N, Sahakian BJ, Hodges JR (2001) Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. Neurology 57:216-225.

Garagnani M, Pulvermuller F (2016) Conceptual grounding of language in action and perception: a neurocomputational model of the emergence of category specificity and semantic hubs. The European journal of neuroscience 43:721-737.

Garrard P, Lambon Ralph MA, Hodges JR, Patterson K (2001) Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. Cognitive neuropsychology 18:125-174.

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform "face area" is part of a network that processes faces at the individual level. Journal of cognitive neuroscience 12:495-504.

Geukes S, Huster RJ, Wollbrink A, Junghofer M, Zwitserlood P, Dobel C (2013) A large N400 but no BOLD effect--comparing source activations of semantic priming in simultaneous EEG-fMRI. PloS one 8:e84029.

Gibbs RW (2006) Metaphor interpretation as embodied simulation. Mind Lang 21: 434–458.

Glenberg AM, Gallese V (2012) Action-based language: a theory of language acquisition, comprehension, and production. Cortex 48:905-922.

Goldberg RF, Perfetti CA, Fiez JA, Schneider W (2007) Selective retrieval of abstract semantic knowledge in left prefrontal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:3790-3798.

Goldstein MN (1974) Auditory agnosia for speech ("pure word-deafness"): A historical review with current implications. Brain and Language 1:195-204.

Gonnerman LM, Andersen ES, Devlin JT, Kempler D, Seidenberg MS (1997) Double dissociation of semantic categories in Alzheimer's disease. Brain and language 56:254-279.

González J, Barros-Loscertales A, Pulvermüller F, Meseguer V, Sanjuán A, Belloch V, Ávila C (2006) Reading cinnamon activates olfactory brain regions. NeuroImage 32:906-912.

Gorno-Tempini ML, Price CJ (2001) Identification of famous faces and buildings. Brain : a journal of neurology 124:2087-2097.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, Grossman M (2011) Classification of primary progressive aphasia and its variants. Neurology 76:1006-1014.

Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK, Weiner MW, Miller BL (2004) Cognition and anatomy in three variants of primary progressive aphasia. Annals of neurology 55:335-346.

Grafton ST, Arbib MA, Fadiga L, Rizzolatti G (1996) Localization of grasp representations in humans by PET: 2.Observation compared with imagination. Experimental Brain Research 112:103–111.

Grisoni L, Dreyer FR, Pulvermuller F (2016) Somatotopic Semantic Priming and Prediction in the Motor System. Cereb Cortex 26:2353-2366.

Grossman M (2010) Primary progressive aphasia: clinicopathological correlations. Nature reviews Neurology 6:88-97.

Grossman M, Anderson, C., Khan, A., Avants, B., Elman, L. and McCluskey, L., (2008) Impaired action knowledge in amyotrophic lateral sclerosis. Neurology 71:1396-1401.

Guimaraes MP, Wong DK, Uy ET, Grosenick L, Suppes P (2007) Single-trial classification of MEG recordings. IEEE Transactions on Biomedical Engineering 54:436-443.

Guntekin B, Basar E (2016) Review of evoked and event-related delta responses in the human brain. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 103:43-52.

Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004) Integration of word meaning and world knowledge in language comprehension. Science 304:438-441.

Hald LA, Bastiaansen MC, Hagoort P (2006) EEG theta and gamma responses to semantic violations in online sentence processing. Brain Lang 96:90-105.

Hanson SJ, Halchenko YO (2008) Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. Neural Computation 20:486-503.

Hanson SJ, Matsuka T, Haxby JV (2004) Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage 23:156-166.

Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, Rizzolatti G (1998) Activation of human primary motor cortex during action observation: a neuromagnetic study. Proceedings of the National Academy of Sciences 95:15061-15065.

Harmony T (2013) The functional significance of delta oscillations in cognitive processing. Frontiers in integrative neuroscience 7:83.

Harnad S (1990) The symbol grounding problem. . Physica D: Nonlinear Phenomena 42:335-346.

Harnad S (2003) The Symbol Grounding Problem. In: Encyclopedia of Cognitive Science: Nature Publishing Group/Macmillan.

Hart J, Gordon B (1992) Neural subsystems for object knowledge. Nature 359:60-64.

Hart J, Berndt RS, Caramazza A (1985) Category-specific naming deficit following cerebral infarction. Nature 316:439-440.

Hauk O (2016) Only time will tell - why temporal information is essential for our neuroscientific understanding of semantics. Psychonomic bulletin & review.

Hauk O, Pulvermuller F (2004) Neurophysiological distinction of action words in the fronto-central cortex. Human brain mapping 21:191-201.

Hauk O, Johnsrude I, Pulvermüller F (2004) Somatotopic representation of action words in human motor and premotor cortex. Neuron 41: 301-307.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Helenius P, Salmelin R, Connolly JF, Leinonen S, Lyytinen H (2002) Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. The Journal of neuroscience 22:2936-2944.

Hillis AE, Caramazza A (1991) Category-specific naming and comprehension impairment: A double dissociation. Brain : a journal of neurology 114:2081-2094.

Hillis AE, Rapp B, Caramazza A (1995) Constraining claims about theories of semantic memory: More on unitary versus multiple semantics. Cognitive neuropsychology 12:175-186.

Hodges JR, Patterson K, Oxbury S, Funnell E (1992) Semantic dementia. Brain : a journal of neurology 115:1783-1806.

Hodges JR, Bozeat S, Ralph MAL, Patterson K, Spatt J (2000) The role of conceptual knowledge in object use evidence from semantic dementia. Brain : a journal of neurology 123:1913-1925.

Hoffman P, Evans GA, Lambon Ralph MA (2014) The anterior temporal lobes are critically involved in acquiring new conceptual knowledge: evidence for impaired feature integration in semantic dementia. Cortex 50:19-31.

Holcomb PJ, Neville HJ (1990) Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. Language and cognitive processes 5:281-312.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology 148:574-591.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210-1224.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453-458.

Iacoboni M (1999) Cortical Mechanisms of Human Imitation. Science 286:2526-2528.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. Proceedings of the National Academy of Sciences 96:9379-9384.

Jackson RL, Hoffman P, Pobric G, Lambon Ralph MA (2016) The Semantic Network at Work and Rest: Differential Connectivity of Anterior Temporal Lobe Subregions. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:1490-1501.

Jamrozik A, McQuire M, Cardillo ER, Chatterjee A (2016) Metaphor: Bridging embodiment to abstraction. Psychonomic bulletin & review 23:1080-1089.

Johnson M (2006) Mind incarnate: from Dewey to Damasio. Daedalus 135:46-54.

Jung J, Cloutman LL, Binney RJ, Lambon Ralph MA (2016) The structural connectivity of higher order association cortices reflects human functional brain networks. Cortex.

Just MA, Cherkassky VL, Aryal S, Mitchell TM (2010) A neurosemantic theory of concrete noun representation based on the underlying brain codes. PloS one 5:e8622.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of neuroscience 17:4302-4311.

Kapur N, Barker, S., Burrows, E.H., Ellison, D., Brice, J., Illis, L.E.E.A., Scholey, K., Colbourn, C., Wilson, B. and Loates, M., (1994) Herpes simplex encephalitis: long term magnetic resonance imaging and neuropsychological profile. . Journal of Neurology,Neurosurgery & Psychiatry 57:1334-1342.

Kemmerer D, Castillo JG, Talavage T, Patterson S, Wiley C (2008) Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. Brain and Language 107:16-43.

Kiefer M, Sim EJ, Herrnberger B, Grothe J, Hoenig K (2008) The Sound of Concepts: Four Markers for a Link between Auditory and Conceptual Brain Systems. Journal of Neuroscience 28:12224-12230.

Kielar A, Panamsky L, Links KA, Meltzer JA (2015) Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. Neuroimage 105:507-524.

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846-848.

Kosslyn SM, Ganis G, Thompson WL (2001) Neural foundations of imagery. Nature Reviews Neuroscience 2:635-642.

Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E (2011) The representation of abstract words: Why emotion matters. Journal of Experimental Psychology: General 140:14-34.

Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci 17:401-412.

Kuriki S, Murase M (1989) Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds. Experimental Brain Research 77:127-134.

Kutas M, Hillyard SA (1980) Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological psychology 11:99-116.

Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. Trends in cognitive sciences 4:463-470.

Laiacona M, Capitani E (2001) A case of prevailing deficit of nonliving categories or a case of prevailing sparing of living categories? Cognitive neuropsychology 18:39-70.

Laiacona M, Capitani E, Caramazza A (2003) Category-specific semantic deficits do not reflect the sensory/functional organization of the brain: a test of the "sensory quality" hypothesis. Neurocase 9:221-231.

Lambon Ralph MA (2014) Neurocognitive insights on conceptual knowledge and its breakdown. Philosophical transactions of the Royal Society of London Series B, Biological sciences 369:20120392.

Lambon Ralph MA, Lowe C, Rogers TT (2007) Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. Brain : a journal of neurology 130:1127-1137.

Lambon Ralph MA, Cipolotti L, Manes F, Patterson K (2010) Taking both sides: do unilateral anterior temporal lobe lesions disrupt semantic memory? Brain : a journal of neurology 133:3243-3255.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2016) The neural and computational bases of semantic cognition. Nature reviews Neuroscience.

Landauer TK, Dumais ST (1997) A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. , 104(2), p.211. Psychological review 104:211-240.

Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics:(de) constructing the N400. Nature Reviews Neuroscience 9:920-933.

Lau EF, Gramfort A, Hamalainen MS, Kuperberg GR (2013) Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:17174-17181.

Laurent G (1996) Dynamical representation of odors by oscillating and evolving neural assemblies. Trends in neurosciences 19:489-496.

Lee SE, Rabinovici GD, Mayo MC, Wilson SM, Seeley WW, DeArmond SJ, Huang EJ, Trojanowski JQ, Growdon ME, Jang JY, Sidhu M, See TM, Karydas AM, Gorno-Tempini ML, Boxer AL, Weiner MW, Geschwind MD, Rankin KP, Miller BL (2011) Clinicopathological correlations in corticobasal degeneration. Annals of neurology 70:327-340.

Leigh PN, Ray-Chaudhuri K (1994) Motor neuron disease. Journal of Neurology, Neurosurgery & Psychiatry 57:886-896.

Lerner Y, Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001) A hierarchical axis of object processing stages in the human visual cortex. Cerebral Cortex 11:287-297.

Lichtheim L (1885) Über Aphasie. . Deutsches Archiv für klinische Medicin 36:204–268.

Lingnau A, Gesierich B, Caramazza A (2009) Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. Proceedings of the National Academy of Sciences of the United States of America 106:9925-9930.

Liuzzi AG, Bruffaerts R, Dupont P, Adamczuk K, Peeters R, De Deyne S, Storms G, Vandenberghe R (2015) Left perirhinal cortex codes for similarity in meaning between written words: Comparison with auditory word input. Neuropsychologia 76:4-16.

Lund K, Burgess C (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, & Computers 28:203-208.

Luzzi S, Snowden JS, Neary D, Coccia M, Provinciali L, Lambon Ralph MA (2007) Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia, and corticobasal degeneration. Neuropsychologia 45:1823-1831.

Maguire MJ, Brier MR, Ferree TC (2010) EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships. Brain Lang 114:16-25.

Mahon BZ (2015) The burden of embodied cognition. Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale 69:172-178.

Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of physiology, Paris 102:59-70.

Mahon BZ, Hickok G (2016) Arguments about the nature of concepts: Symbols, embodiment, and beyond. Psychonomic bulletin & review 23:941-958.

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

Malone PS, Glezer LS, Kim J, Jiang X, Riesenhuber M (2016) Multivariate Pattern Analysis Reveals Category-Related Organization of Semantic Representations in Anterior Temporal Cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 36:10089-10096.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing. Neuron 38:487–497.

Marr D (1982) Vision. New York:: Freeman & Co.

Martin A (2015) GRAPES-Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. Psychonomic bulletin & review.

Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. Current opinion in neurobiology 11:194-201.

Martin A, Weisberg J (2003) Neural foundations for understanding social and mechanical concepts. Cognitive neuropsychology 20:575-587.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. Nature 379:649-652.

Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG (1995) Discrete cortical regions associated with knowledge of color and knowledge of action. Science 270.

Masson ME (1995) A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 21.

Matsumoto A, Iidaka T (2008) Gamma band synchronization and the formation of representations in visual word processing: evidence from repetition and homophone priming. Journal of cognitive neuroscience 20:2088-2096.

McCarthy G, Nobre AC, Bentin S, Spencer DD (1995) Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. The Journal of Neuroscience 15:1080-1089.

McRae K, Cree GS (2002) Factors underlying category-specific semantic deficits. In: Category-specificity in brain and mind (Forde EME, Humphreys GW, eds), pp 211–249. East Sussex, England: Psychology Press.

Mesulam M (1982) Slowly progressive aphasia without generalized dementia. Annals of neurology 11:592-598.

Mesulam M (1987) Primary progressive aphasia—differentiation from Alzheimer's disease. Annals of neurology 22:533-534.

Miceli G, Capasso R, Daniele A, Esposito T, Magarelli M, Tomaiuolo F (2000) Selective deficit for people's names following left temporal damage: An impairment of domain-specific conceptual knowledge. Cognitive neuropsychology 17:489-516.

Miceli G, Fouch E, Capasso R, Shelton JR, Tomaiuolo F, Caramazza A (2001) The dissociation of color from form and function knowledge. Nature neuroscience 4:662-667.

Mion M, Patterson K, Acosta-Cabronero J, Pengas G, Izquierdo-Garcia D, Hong YT, Fryer TD, Williams GB, Hodges JR, Nestor PJ (2010) What the left and right anterior fusiform gyri tell us about semantic memory. Brain : a journal of neurology 133:3256-3268.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191-1195.

Moran MA, Mufson EJ, Mesulam MM (1987) Neural inputs into the temporopolar cortex of the rhesus monkey. Journal of Comparative Neurology 256:88-103.

Moseley R, Carota F, Hauk O, Mohr B, Pulvermuller F (2011) A Role for the Motor System in Binding Abstract Emotional Meaning. Cerebral Cortex 22:1634-1647.

Moseley RL, Pulvermuller F, Shtyrov Y (2013) Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. Scientific reports 3:1928.

Moss HE, Tyler LK (2000) A progressive category-specific semantic deficit for non-living things. . Neuropsychologia 38:60-82.

Moss HE, Tyler LK, Durrant-peatfield M, Bunn EM (1998) 'Two Eyes of a See-through': Impaired and Intact Semantic Knowledge in a Case of Selective Deficit for Living Things. Neurocase 4:291-310.

Mur M, Ruff DA, Bodurka J, De Weerd P, Bandettini PA, Kriegeskorte N (2012) Categorical, yet graded--single-image activation profiles of human category-selective cortical regions. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:8649-8662.

Murphy B, Poesio M (2010) Detecting semantic category in simultaneous EEG/MEG recordings. In: Proceedings of the naacl hlt 2010 first workshop on computational neurolinguistics, pp 36-44: Association for Computational Linguistics.

Murphy B, Poesio M, Bovolo F, Bruzzone L, Dalponte M, Lakany H (2011) EEG decoding of semantic category reveals distributed representations for single concepts. Brain and Language 117:12-22.

Murphy C, Rueschemeyer SA, Watson D, Karapanagiotidis T, Smallwood J, Jefferies E (2016) Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. Neuroimage.

Näätäneiv R, Lehtokoski A, Lennest M, Luuki A, Alliki J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432-434.

Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL (2015) A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. Neuroimage 105:215-228.

Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Freedman M, Kertesz A, Robert PH, Albert M, Boone K, Miller BL, Cummings J, Benson DF (1998) Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. Neurology 51:1546-1554.

Negri GA, Rumiati RI, Zadini A, Ukmar M, Mahon BZ, Caramazza A (2007) What is the role of motor simulation in action and object recognition? Evidence from apraxia. Cognitive neuropsychology 24:795-816.

Neininger B, Pulvermüller F (2003) Word-category specific deficits after lesions in the right hemisphere. . Neuropsychologia 41:53-70.

Nestor PJ, Fryer TD, Hodges JR (2006) Declarative memory impairments in Alzheimer's disease and semantic dementia. Neuroimage 30:1010-1020.

Nobre AC, Mccarthy G (1995) Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. The Journal of Neuroscience 15:1090-1098.

Noppeney U, Patterson K, Tyler LK, Moss H, Stamatakis EA, Bright P, Mummery C, Price CJ (2007) Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. Brain : a journal of neurology 130:1138-1147.

O'toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. Journal of Cognitive Neuroscience 17:580-590.

Ostarek M, Vigliocco G (2016) Reading Sky and Seeing a Cloud: On the Relevance of Events for Perceptual Simulation. Journal of experimental psychology Learning, memory, and cognition.

Paivio A (1986) Mental representations: a dual coding approach. Oxford. England: Oxford University Press.

Pammer K (2009) What can MEG neuroimaging tell us about reading? Journal of Neurolinguistics 22:266-280.

Pammer K, Hansen PC, Kringelbach ML, Holliday I, Barnes G, Hillebrand A, Singh KD, Cornelissen PL (2004) Visual word recognition: the first half second. Neuroimage 22:1819-1825.

Papeo L, Negri GA, Zadini A, Rumiati RI (2010) Action performance and action-word understanding: evidence of double dissociations in left-damaged patients. Cognitive neuropsychology 27:428-461.

Papinutto N, Galantucci S, Mandelli ML, Gesierich B, Jovicich J, Caverzasi E, Henry RG, Seeley WW, Miller BL, Shapiro KA, Gorno-Tempini ML (2016) Structural connectivity of the human anterior temporal lobe: A diffusion magnetic resonance imaging study. Human brain mapping 37:2210-2222.

Parviainen T, Helenius P, Salmelin R (2005) Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. Cereb Cortex 15:1054-1063.

Pascual B, Masdeu JC, Hollenbeck M, Makris N, Insausti R, Ding SL, Dickerson BC (2015) Large-scale brain networks of the human left temporal pole: a functional connectivity MRI study. Cereb Cortex 25:680-702.

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

Pazzaglia M, Pizzamiglio L, Pes E, Aglioti SM (2008a) The sound of actions in apraxia. Current biology : CB 18:1766-1772.

Pazzaglia M, Smania N, Corato E, Aglioti SM (2008b) Neural underpinnings of gesture discrimination in patients with limb apraxia. The Journal of neuroscience : the official journal of the Society for Neuroscience 28:3030-3041.

Pearson J, Kosslyn SM (2015) The heterogeneity of mental representation: Ending the imagery debate. Proceedings of the National Academy of Sciences of the United States of America 112:10089-10092.

Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:15728-15736.

Penfield W, Boldrey E (1938) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. . Brain : a journal of neurology 15:389-443.

Perani D, Cappa SF, Bettinardi V, Bressi S, Gorno-Tempini M, Matarrese M, Fazio F (1995) Different neural systems for the recognition of animals and man-made tools. Neuroreport 6:1637-1641.

Pereira F, Gershman S, Ritter S, Botvinick M, A (2016) A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. Cognitive neuropsychology 33:175–190.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331:585-589.

Phan TG, Donnan GA, Wright PM, Reutens DC (2005) A digital map of middle cerebral artery infarcts associated with middle cerebral artery trunk and branch occlusion. Stroke 36:986-991.

Pietrini V, Nertempi P, Vaglia A, Revello MG, Pinna V, Ferro-Milone F (1988) Recovery from herpes simplex encephalitis: selective impairment of specific semantic categories with neuroradiological correlation. Journal of Neurology, Neurosurgery & Psychiatry 51:1284-1293.

Pignatti R, Ceriani F, Bertella L, Mori I, Semenza C (2006) Naming abilities in spontaneous speech in Parkinson and Alzheimer's disease. Brain and Language 99:124-125.

Piwnica-Worms KE, Omar R, Hailstone JC, Warren JD (2010) Flavour processing in semantic dementia. Cortex 46:761-768.

Pobric G, Jefferies E, Ralph MA (2007) Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. Proceedings of the National Academy of Sciences of the United States of America 104:20137-20141.

Pobric G, Jefferies E, Lambon Ralph MA (2010a) Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. Current biology : CB 20:964-968.

Pobric G, Jefferies E, Ralph MA (2010b) Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. Neuropsychologia 48:1336-1342.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963-1966.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nature Reviews Neuroscience 1:125-132.

Price AR, Bonner MF, Peelle JE, Grossman M (2015) Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:3276-3284.

Price CJ, Hope TT, Seghier ML (2016) Ten problems and solutions when predicting individual outcome from lesion site after stroke. Neuroimage.

Pulvermüller F (1999) Words in the brain's language. Behavioral and brain sciences 22:253-279.

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

Pulvermüller F, Härle M, Hummel F (2000) Neurophysiological distinction of verb categories. . Neuroreport 11:2789-2793.

Pulvermüller F, Shtyrov Y, Ilmoniemi R (2005a) Brain Signatures of Meaning Access in Action Word Recognition. Journal of Cognitive Neuroscience 17:884-892.

Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ (2005b) Functional links between motor and language systems. European Journal of Neuroscience 21:793-797.

Putnam H (1975) The meaning of "meaning". Minnesota Studies in the Philosophy of Science 7:131-193.

Pylyshyn Z (2003) Return of the mental image: are there really pictures in the brain? Trends in Cognitive Sciences 7:113-118.

Quillian MR (1967) Word concepts: A theory and simulation of some basic semantic capabilities. Behavioral science 12:410-430.

Quillian R (1966) Semantic Memory. Unpublished doctoral dissertation. Carnegie Institute of Technology

Rapp B, Caramazza A (1993) On the distinction between deficits of access and deficits of storage: A question of theory. Cognitive neuropsychology 10:113-141.

Rauschecker JP (1998) Cortical processing of complex sounds. Current opinion in neurobiology 8:516-521.

Reed CL, Caselli, R.J. and Farah, M.J., (1996) Tactile agnosia. Brain : a journal of neurology 119:875-888.

Rice GE, Hoffman P, Lambon Ralph MA (2015) Graded specialization within and between the anterior temporal lobes. Annals of the New York Academy of Sciences 1359:84-97.

Riddoch MJ, Humphreys GW, Coltheart M, Funnell E (1988) Semantic systems or system? Neuropsychological evidence re-examined. Cognitive neuropsychology 5:3-25.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996a) Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3:131-141.

Rizzolatti G, Fadiga L, Matelli M, Bettinardi V, Paulesu E, Perani D, Fazio F (1996b) Localization of grasp representations in humans by PET: 1. Observation versus execution. Experimental Brain Research 111.

Rodriguez-Ferreiro J, Menendez M, Ribacoba R, Cuetos F (2009) Action naming is impaired in Parkinson disease patients. Neuropsychologia 47:3271-3274.

Rogers TT, McClelland JL (2004) Semantic cognition: A parallel distributed processing approach. . Cambridge: MIT press.

Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K (2004) Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychological review 111:205-235.

Rogers TT, Hocking J, Noppeney U, Mechelli A, Gorno-Tempini ML, Patterson K, Price CJ (2006) Anterior temporal cortex and semantic memory: reconciling findings from neuropsychology and functional imaging. Cognitive, Affective, & Behavioral Neuroscience 6:201-213.

Rosen HJ, Gorno–Tempini ML, Goldman WP, Perry RJ, Schuff N, Weiner M, Feiwell R, Kramer JH, Miller BL (2002) Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology 58:198-208.

Rumelhart DE, Todd PM (1993) Learning and connectionist representations. . In: Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience. (Meyer DE, Kornblum S, eds), pp 3-30. Cambridge: MIT press.

Russell RP (1910) Knowledge by Acquaintance and Knowledge by Description. Proceedings of the Aristotelian Society 11: 108-128.

Sacchett C, Humphreys GW (1992) Calling a squirrel a squirrel but a canoe a wigwam: a category-specific deficit for artefactual objects and body parts. Cognitive neuropsychology 9:73-86.

Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science 326:445-449.

Salmelin R (2007) Clinical neurophysiology of language: The MEG approach. Clinical Neurophysiology 118:237-254.

Samson D, Pillon A (2003) A case of impaired knowledge for fruit and vegetables. Cognitive neuropsychology 20:373-400.

Sartori G, Job R, Miozzo M, Zago S, Marchiori G (1993) Category-specific form-knowledge deficit in a patient with herpes simplex virus encephalitis. Journal of clinical and experimental neuropsychology 15:280-299.

Searle JR (1980) Minds, brains, and programs. Behavioral and brain sciences 3:417-424.

Sereno MI, Dale AM, Reppas JB, Kwong KK (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268.

Sereno SC, Rayner K, Posner MI (1998) Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. Neuroreport 9:2195-2200.

Shahin AJ, Picton TW, Miller LM (2009) Brain oscillations during semantic evaluation of speech. Brain and cognition 70:259-266.

Shallice T (1988) Specialisation within the semantic system. Cognitive neuropsychology 5:133-142.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. Neuroimage 54:2418-2425.

Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA (2008) Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PloS one 3:e1394.

Shtyrov Y, Butorina A, Nikolaeva A, Stroganova T (2014) Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. Proceedings of the National Academy of Sciences of the United States of America 111:E1918-1923.

Silveri MC, Ciccarelli N (2007) The deficit for the word-class "verb" in corticobasal degeneration: linguistic expression of the movement disorder? Neuropsychologia 45:2570-2579.

Simanova I, Hagoort P, Oostenveld R, van Gerven MA (2014) Modality-independent decoding of semantic information from the human brain. Cereb Cortex 24:426-434.

Simanova I, van Gerven MA, Oostenveld R, Hagoort P (2015) Predicting the semantic category of internally generated words from neuromagnetic recordings. J Cogn Neurosci 27:35-45.

Smith EE, Shoben EJ, Rips LJ (1974) Structure and process in semantic memory: A featural model for semantic decisions. Psychological review 81:214.

Snowden JS, Goulding PJ, Neary D (1989) Semantic dementia: a form of circumscribed cerebral atrophy. . Behavioural Neurology.

Spatt J, Bak T, Bozeat S, Patterson K, Hodges JR (2002) Apraxia, mechanical problem solving and semantic knowledge. Journal of Neurology, Neurosurgery & Psychiatry 249:601-608.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ (2006) Converging language streams in the human temporal lobe. The Journal of neuroscience : the official journal of the Society for Neuroscience 26:7328-7336.

Steele JC, Richardson JC, Olszewski J (1964) Progressive supranuclear palsy: a heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia. Archives of neurology 10:333-359.

Sudre G, Pomerleau D, Palatucci M, Wehbe L, Fyshe A, Salmelin R, Mitchell T (2012) Tracking neural coding of perceptual and semantic features of concrete nouns. Neuroimage 62:451-463.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends in cognitive sciences 3:151-162.

Tarkiainen A, Helenius, P., Hansen, P.C., Cornelissen, P.L. and Salmelin, R., (1999) Dynamics of letter string perception in the human occipitotemporal cortex. Brain : a journal of neurology 122:2119-2132.

Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, Fazio F, Rizzolatti G, Cappa SF, Perani D (2005) Listening to action-related sentences activates fronto-parietal motor circuits. Journal of cognitive neuroscience 17:273-281.

Thompson-Schill SL, Aguirre GK, Desposito M, Farah MJ (1999) A neural basis for category and modality specificity of semantic knowledge. Neuropsychologia 37:671-676.

Tomasello R, Garagnani M, Wennekers T, Pulvermuller F (2016) Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. Neuropsychologia.

Tranel D, Damasio H, Damasio AR (1997) A neural basis for the retrieval of conceptual knowledge. Neuropsychologia 35:1319-1327.

Trumpp NM, 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.

Tsapkini K, Frangakis CE, Hillis AE (2011) The function of the left anterior temporal pole: evidence from acute stroke and infarct volume. Brain : a journal of neurology 134:3094-3105.

Tulving E (1972) Episodic and semantic memory Organization of Memory London: Academic 381.

Tyler LK, Moss HE (1997) Functional properties of concepts: Studies of normal and brain-damaged patients. Cognitive neuropsychology 14:511-545.

Tyler LK, Moss HE (2001) Towards a distributed account of conceptual knowledge. Trends in cognitive sciences 5:244-252.

Tyler LK, Moss HE, Durrant-Peatfield MR, Levy JP (2000) Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. Brain Lang 75:195-231.

Tyler LK, Stamatakis EA, Bright P, Acres K, Abdallah S, Rodd JM, Moss HE (2004) Processing objects at different levels of specificity. Journal of cognitive neuroscience, 16:351-362.

Tyler LK, Chiu S, Zhuang J, Randall B, Devereux BJ, Wright P, Clarke A, Taylor KI (2013) Objects and categories: feature statistics and object processing in the ventral stream. J Cogn Neurosci 25:1723-1735.

van Ackeren MJ, Schneider TR, Musch K, Rueschemeyer SA (2014) Oscillatory neuronal activity reflects lexical-semantic feature integration within and across sensory modalities in distributed cortical networks. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:14318-14323.

van de Nieuwenhuijzen ME, Backus AR, Bahramisharif A, Doeller CF, Jensen O, van Gerven MA (2013) MEG-based decoding of the spatiotemporal dynamics of visual category perception. Neuroimage 83:1063-1073.

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

Vandenberghe R (2016) Classification of the primary progressive aphasias: principles and review of progress since 2011. Alzheimer's research & therapy 8:16.

Vannuscorps G, Caramazza A (2016) Typical action perception and interpretation without motor simulation. Proceedings of the National Academy of Sciences 113:86-91.

Vannuscorps G, Dricot L, Pillon A (2016) Persistent sparing of action conceptual processing in spite of increasing disorders of action production: A case against motor embodiment of action concepts. Cognitive neuropsychology:1-29.

Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M (1997) Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. Science 277:376-380.

Vigliocco G, Kousta ST, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF (2013) The Neural Representation of Abstract Words: The Role of Emotion. Cerebral Cortex 24:1767-1777.

Visser M, Jefferies E, Ralph ML (2010a) Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. Journal of cognitive neuroscience 22:1083-1094.

Visser M, Embleton KV, Jefferies E, Parker GJ, Ralph MA (2010b) The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. Neuropsychologia 48:1689-1696.

Wagner AD, Desmond JE, Demb JB, Glover GH, Gabrieli JD (1997) Semantic repetition priming for verbal and pictorial knowledge: A functional MRI study of left inferior prefrontal cortex. Journal of Cognitive Neuroscience 9:714-726.

Walker LC, LeVine H (2000) The cerebral proteopathies. Molecular neurobiology 21:83-95.

Wang L, Jensen O, van den Brink D, Weder N, Schoffelen JM, Magyari L, Hagoort P, Bastiaansen M (2012) Beta oscillations relate to the N400m during language comprehension. Human brain mapping 33:2898-2912.

Warrington EK (1975) The selective impairment of semantic memory. The Quarterly journal of experimental psychology 27:635-657.

Warrington EK, McCarthy R (1983) Category specific access dysphasia. Brain : a journal of neurology 106:859-878.

Warrington EK, Shallice T (1984) Category specific semantic impairments. Brain : a journal of neurology 107:.829-853.

Warrington EK, McCarthy RA (1987) Categories of knowledge. Further fractionations and an attempted integration. Brain : a journal of neurology 110:1273-1296.

Wernicke C (1874/1977) Der aphasische symptomencomplex: eine psychologische studie auf anatomischer basis. In: Wernicke's works on aphasia: a sourcebook and review (Eggert GH, ed), pp 91–145. The Hague: Mouton.

West WC, Holcomb PJ (2000) Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. Journal of Cognitive Neuroscience 12:1024-1037.

Whitley RJ, Gnann JW (2002) Viral encephalitis: familiar infections and emerging pathogens. The Lancet 359:507-513.

Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E (2011) The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. Cereb Cortex 21:1066-1075.

Wright P, Randall B, Clarke A, Tyler LK (2015) The perirhinal cortex and conceptual processing: Effects of feature-based statistics following damage to the anterior temporal lobes. Neuropsychologia 76:192-207.

Yee E, Thompson-Schill SL (2016) Putting concepts into context. Psychonomic bulletin & review 23:1015-1027.

Zwaan RA (2014) Embodiment and language comprehension: reframing the discussion. Trends Cogn Sci 18:229-234.

Zwaan RA (2016) Situation models, mental simulations, and abstract concepts in discourse comprehension. Psychonomic bulletin & review 23:1028-1034.

CHAPTER 2:
INVESTIGATING COGNITIVE AND NEURAL REPRESENTATIONS

All models are wrong, some are useful.
All models are right, most are useless.
[George Box vs Thad Tarpey]

In this chapter I present the methods used in the literature, and in particular in this thesis, to investigate the cognitive and neural substrate of semantic representations. Three different behavioral methods (distance measure, feature listing, and semantic priming), and two neuroimaging techniques (functional magnetic resonance imaging - fMRI, and magnetoencephalography - MEG) are introduced. Key methodological and statistical aspects are identified, in order to simplify their exposition in the following chapters. In the last part, attention is drawn to multivariate analyses of neuroimaging data as they constitue a great advance in the field and a major component of this thesis.

Highlights:

- Behavioral evidence provides insights, and potentially causal links, about the mind-brain relation.
- The topographical organization of the neural substrates of cognitive phenomena can be study with fMRI.
- The temporal and spectral features of those neural substrates can be studied with M/EEG.
- Multivariate analyses of neuroimaging data broaden the set of testable cognitive hypotheses.
- Pattern decoding/encoding demonstrates mutual information between brain responses and stimulus properties.
- Patterns geometries can be compared across areas, modalities and theoretical models.

1. Behavior to Look into Cognition

Let's imagine your television is not working properly. The very first thing you can do (and often, the only thing you *can* do) is to define how so. Is it a problem of sound emission? Are the colors on the screen not correct? Or is the signal simply missing from time to time? By simply looking at what is going on, what is not working (i.e., deficit) and what is working fine (i.e., spared properties), one can (1)

understand the problem at hand; (2) generate hypotheses on possible causes. In other words, when faced with a complex system, by looking at the output of an impaired process, one can try to understand what the potential underlying software/hardware problems are and how its computations are performed under normal circumstances. Behavioral responses of human beings, what they do and say (or don't do and don't say), can reveal important details on the cognitive functions involved (i.e., software), and on the underlying neural implementation (i.e., hardware). This is true not only in the case of deficits following brain trauma, but also when considering performance under normal circumstances.

The goal of this thesis was to shed light on the cognitive and neural underpinnings of semantic representations. As a first step, I aimed at understanding cognitive representations, which meant investigating their content and how they are internally organized. The inner structure of a representation (i.e., its geometry, see Chap. 1), can be conceptualized as the relation across its constituents. In the case of semantic representations, one is interested in describing the relations across concepts. Thus I set out to probe the internal structure of semantic representations by studying subjects' judgments of semantic distance.

1.1 Semantic Distance Judgment

Perhaps the most intuitive way to access subjects' representations of the semantic distance between different concepts is to directly ask them to rate it in a Semantic Distance Judgment, SDJ). Presented with pairs of words, subjects are asked to define how semantically far (i.e., different) or close (i.e., similar) they think the two words to be (see Fig. 23 upper part). The judgment can be provided via a rating scale, which can take several values, typically from 1 to 7 or 9. Given that the interest lies in relative judgment across pairs of words, individual ratings are then normalized (e.g., between 0 and 1) to correct for possible inter-individual differences in the

ranking scaled adopted. For instance, one subject might never consider any pair of words to fall at a distance "9", thus using *de facto* a scale from 1 to 8. The normalized data of each subject are then re-arranged in an *n x n* matrix describing the pairwise semantic distance between the *n* words tested. Single subject matrices can then be averaged producing a final semantic distance matrix depicting the average distance at which each pair of words fell in the cognitive semantic space of that population. In the final stage, dimensionality reduction techniques such as multidimensional scaling (MDS) and hierarchical clustering can be used to obtain a graphical representation of such space (Shepard, 1980). This geometric model of mental representations has a long history (Coombs, 1954; Torgerson, 1965) and wide applications, notwithstanding known issues of asymmetry and contextual effects. First, the order in which the pairs are presented can potentially alter the answer received: for instance, "*tiger*" appears more associated with ''*leopard*'' than ''*leopard*'' is to "*tiger*", thus potentially different judgments would be collected when presenting the pair "*tiger - leopard*" or the pair "*leopard - tiger*" (Tversky, 1977). Second, the distance across items depends upon the list of items presented to the subjects: by adding to the list a very distant element (e.g., "*elephant*") the distance between "*tiger*" and "*leopard*" can be shrunk, while adding a closer one (e.g., "*jaguar*") would amplify it (Goldstone et al., 1997). Using this method the number of pairwise comparisons may quickly become extremely high, thus whenever they become pragmatically prohibitive to test, alternative procedures, always based on judgment of perceived distance, can be used. For instance, so-called inverse-MDS requires subjects to directly arrange subsets of items, freely grouping them according to the perceived semantic distance (Kriegeskorte and Mur, 2012). The process is repeated multiple times until the full semantic space has been mapped.

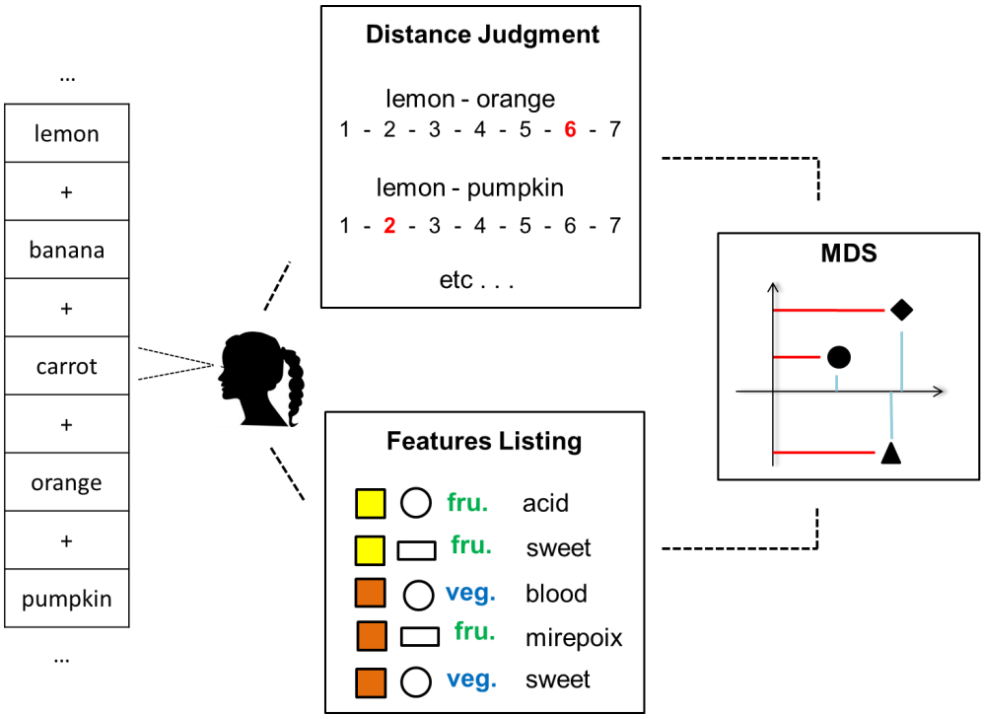


Figure 23 Schematic representation of the work flow of Distance Judgment and Feature Listing tasks. Subjects are asked to judge the distance perceived among pairs of items (upper), or to list perceptual and functional features of each single item (lower). Either method can be used to estimate how close items are in a multidimensional representation of the cognitive semantic space (right). In this example, items are described by the subjects in terms of their color, shape, semantic category (i.e., fruits or vegetables), and flavor.

1.2 Semantic Feature Listing

We have seen that distance measures focus the attention of the subjects directly on the semantic distance across concepts, without explicitly investigating the different features of the items on the basis of which subjects perform the similarity judgements. Another approach that could more directly tap into the question of the features governing the semantic space is the so-called Semantic Feature Listing (SFL): subjects are asked to list the features that define a given concept without being asked to compare it to another one. In this paradigm, single words are presented and subjects are asked to list a number of features that would describe, define, those words (see Fig. 23 lower part). They are usually encouraged to think about any distinctive feature of the item, in terms of not only its perceptual properties (e.g., feelings when touching it, seeing it, hearing it), but also more elaborate ones (e.g. where it is usually found, how and for what it is usually used). If *n* words are presented to the subjects, a *n x n* similarity matrix can be created by counting how many features are shared by any pair of words. The similarity matrix can be easily

reversed, as to describe differences, i.e. distances. Then, similarly to what can be done with the matrix derived from the semantic distance task, it can be normalized, averaged across subjects and analyzed by means of MDS or hierarchical clustering. SFL has been extensively used (Garrard et al., 2001; Cree and McRae, 2003; McRae et al., 2005) and shown to have fair correspondence with the complementary Feature Rating (Tranel et al., 1997; Gainotti et al., 2009), where subjects are presented with different features (e.g., it is colored, it has a rounded shape, etc…) and are asked to rate how much that feature applies to each item. A direct comparison of the two approaches has revealed that ratings, not relying exclusively on participants' verbal ability to describe the items, might provide more information on features such as motion, likely capturing better the overall sensory-motor knowledge available for each concept (Hoffman and Lambon Ralph, 2013).

1.3 Priming Paradigm

Another paradigm adopted in numerous language studies to probe semantic representations is that of priming. A positive priming effect is observed when the presentation of a prior stimulus (prime) facilitates the processing of a second one (target) (see Fig. 24). Such facilitation (usually very short lived, but see Becker et al., 1997) can be measured in terms of lower error rate (i.e. more accurate processing) and/or lower reaction time (i.e. higher speed of processing). Priming experiments vary in terms of:

1. The relation between prime and target. For instance, since early investigation, authors have distinguished between semantically associated (e.g., *cat – dog*, the association is based on statistical co-occurrences) and semantically related (e.g., *nurse – wife*, there is a semantic relation) words (Fischler, 1977) .
2. The nature of the task to be performed on the target. Classical studies have asked subjects to perform a semantic categorization

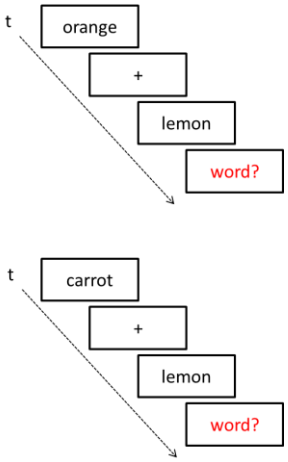


Figure 24 Priming Paradigm. The performance in the main task (e.g., deciding whether the target stimulus is a word or non-word) is analyzed with respect to the relation between the target and a previous stimulus, called prime. An improvement in time and/or accuracy of the response is called priming effect. In the example given, one can expect "orange" to be a better prime for "lemon" than "carrot", in virtue of being two citrus fruit.

or lexical decision on the target or, more rarely, on both prime and target (Meyer and Schvaneveldt, 1971).

3. The nature of the stimuli. Prime and target could be pictures, spoken words, or written words. The nature of prime and target could also differ, giving rise to what has been called multimodal priming (Swinney et al., 1979).
4. The visibility of the prime. Prime words can be visible and thus consciously perceived by the subjects, or subliminal, thus not overtly reported by the subjects (Marcel, 1983). This effect can be reached, for example, by forward or backward masking of the prime.

Priming (often in its subliminal variant) has been widely used to test the automaticity of access of different aspects of word processing such as orthography (Kouider et al., 2007), phonology (Wilson et al., 2011), and morphosyntax (e.g,, gender congruency between article and name (Ansorge et al., 2013)). Semantic aspects, for instance the existence of direct (e.g. milk-cow) and indirect (milk-bull) associations (De Groot, 1983), have been investigated as well. As reviewed in the Chap 1, priming effects have been considered indices of the internal structure of the semantic system (as example consider Masson, 1995), and are often used in concert with neuroimaging techniques (e.g., Holcomb and Neville, 1990; Geukes et al., 2013; Grisoni et al., 2016). In this regard, a plausible neural underpinning of priming is the neurophysiological phenomenon known as repetition suppression (Henson and Rugg, 2003): when two identical (or very similar) stimuli are presented in rapid succession, there is a reduction in neural activity (see also paragraph 4 of this chapter).

In a behavioral semantic priming paradigm, as in any typical priming paradigm, data analysis generally follows these steps:

1. Data are often cleaned from outliers eliminating trials whose RTs fall 2 or 3 standard deviations away from the subject specific average. The rationale of this choice is that if RT is too short or too long, it likely denotes processes other that the one the study focuses on: if on average subjects' reaction is around 200 ms, a

response before 50 ms is likely a false positive (e.g. the subject pressed before reading the target word), similarly a response after 800 ms would likely be contaminated by unrelated distracting factors.

2. For each subject and each condition of interest, one computes the number of errors and the mean (or median) RTs on correct trials.
3. The difference between the condition of interest in terms of number of errors and RT is statistically assessed with the appropriate test (e.g., for normally distributed data, paired *t*-test if only 2 conditions of interest, repeated measure ANOVA if more than 2 conditions of interest). A priming effect can thus be observed in a significantly reduced RT (or in a significantly reduced number of errors) in the congruent condition (e.g. in a semantic priming experiment, prime and target belong to the same semantic category: *cat – dog*), as compared with the incongruent condition (e.g. prime and target belong to the different semantic categories: *cup – dog*).

Two different analyses are possible: one that considers the subjects as the random factor (also known as by-subject-analyses); one that considers the items as the random factor (by-item-analyses) (Hutchison et al., 2008). Moreover, it is possible to investigate the temporal development of the effect by running a time-bin analysis of the RTs. In this latter case, trials are sorted in multiple temporal bins according to the subjects' specific distribution of RTs (Balota et al., 2008). Finally, while selecting the stimuli and designing the experiment, attempts should be made to avoid a simple stimulus-response mapping and provide evidence of a true semantic priming (Damian, 2001).

Positive priming (i.e., a facilitation of processing in congruent prime-target pairs compared to incongruent ones) is not the only possible outcome, the complementary result has also been observed. A negative priming effect is observed when the presentation of a prior stimulus (prime) impairs the processing of a subsequent one (target). Such an effect can be measured in terms of a higher error rate (i.e. less

accurate processing) and/or longer reaction time (i.e. slower speed of processing). Negative priming effects can be used to demonstrate how deep and automatic the processing of unattended stimuli is across tasks differentially tapping semantic processing (Tipper and Driver, 1988; Damian, 2000). For a review on possible explanations see (Mayr and Buchner, 2007).

The effect of negative priming can be related to tasks studying interference effects. In this case, the execution of two concurrent tasks can result in an interference effect with detrimental effects on how stimuli are processed (again, in terms of errors and/or RTs). Interpreting the interference effect in terms of competition for the same underlying cognitive (and neural) resources (or representations), it is possible to use this paradigm to investigate the functional role of sensory-motor systems in conceptual processing. For instance, it has been shown that engaging in motor activity impacts tool naming (Witt et al., 2010), that previous manual experience affects the degree of motor interference during a conceptual task (Yee et al., 2013), and that, during lexico-semantic processing, it is possible to elicit verb-effector compatibility effects (it is easier to respond to a verb –e.g., "*kick*" - with the congruent effector – i.e., feet – than with an incongruent one – e.g., hand) (Andres et al., 2015). However, when comparing a congruent and an incongruent condition (as in the examples above), the valuation of the interference effects is difficult as a proper baseline condition is lacking: the observed performance is congruent both with a facilitation for the congruent case, and with an interference for the incongruent one.

To sum up, the behavioral methods here reviewed can help investigate semantic representations in many complementary ways. With Semantic Distance Judgment and Semantic Feature Listing one can study the organization of subjects' cognitive semantic space. However, given the differences between the two methods, it is legitimate to wonder what their relation is: does the same

representational space emerge when implicit (SFL) and explicit (SDJ) measures are used?

Moreover, priming paradigms (both in terms of facilitation and interference) can be used to assess how automatic the retrieval of specific components of language is, in general, and semantics, in particular. Finally, having observed which dimensions describe the cognitive semantic space, one can wonder if they are coded differentially in the brain, a hypothesis that will require neuroimaging testing.

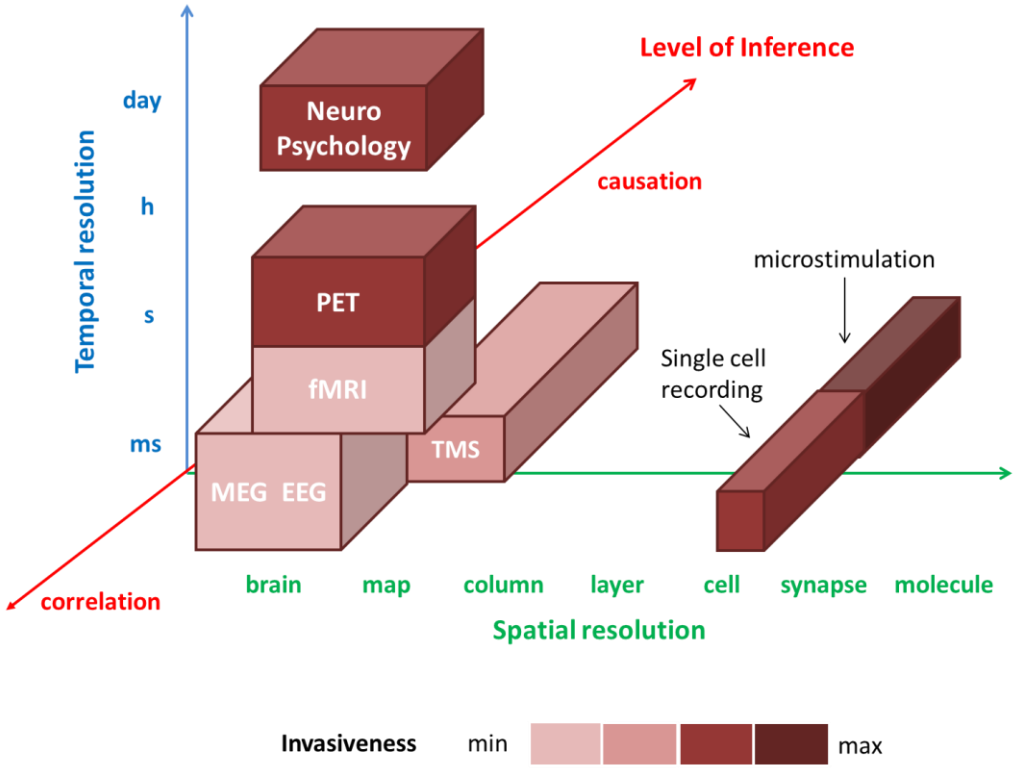


Figure 25 Multidimensional space of brain activity measurement modalities.

Many different tools are available nowadays to investigate the neural underpinning of cognitive processes and representations. All available techniques sample from a subset of the complex spatio-temporal dynamic of brain activity, facing a tradeoff between temporal resolution, spatial resolution, level of inference, and invasiveness. Each of them has advantages and drawbacks, making it suitable to the study of specific cognitive questions, while being unsuitable for others.

2. functional Magnetic Resonance Imaging (fMRI)

Remember the broken television I introduced at the beginning of the chapter? If you were to bring it to an expert technician, alongside your precious information on what's going on (i.e., your diagnosis) and your ideas on what it means (i.e., your functional hypothesis), he/she would be able to test (some of) them. Looking inside the apparatus, it would be possible to better understand the

relation between the software installed by your cable company (i.e., cognitive functions) and the hardware that operates underneath it (i.e., neural substrate). Different methods have been developed to investigate the brain (see Fig. 25), all offering different advantages and disadvantages. Ideally one would like to have (1) whole brain spatial coverage, as likely most cognitive processes recruit more than one cortical area; (2) high temporal resolution, as the dynamics of mental processes are indisputably rapid; (3) high spatial resolution, as it has been shown that functional information in many brain areas is coded at a very fine grained neurobiological level; (4) minimal invasiveness, as to respect the human beings volunteering for the experiment, guarantying maximal comfort and protection from any possible (even indirect) harm; (5) maximal causal inference, as most of the techniques will only show a correlation between a given cognitive process and activity in one brain area, not allowing any kind of causal inference. Thus, cognitive neuroscientist working with neuroimaging techniques are always faced with a trade off in a four dimensional space: time x space (in terms of both resolution and coverage) x causality x invasiveness (see Fig. 25).

In the case of the malfunctioning television, if you'd like to know exactly which small component (e.g., portion of a cable, tiny chip) is broken, your focus would be on the spatial resolution. In neuroimaging, this corresponds to choosing functional magnetic resonance imaging (fMRI). Compared to other neuroimaging techniques, fMRI has a good-to-great spatial resolution and a rather poor temporal resolution. The spatial resolution is affected by the strength of the magnetic field (i.e., whether the magnet is 7, 3 or 1.5 Tesla) as well as the sequence used to acquire brain volumes (e.g., whether it is possible to acquire multiple slices at the same time, so called multiband sequences). The temporal resolution is limited by the nature of the signal measured (i.e. it depends on blood flow, see below) and by the choice of parameters for the acquisition sequence (e.g., repetition time or TR).

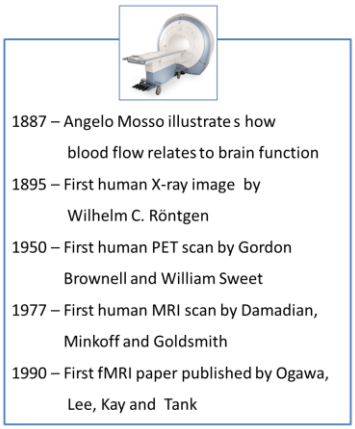


Figure 26 Brief history of fMRI. More than 100 years and about 6416 km separate the discovery of the link between cognitive functions and blood circulation by the Italian physiologist Mosso, and the first fMRI experiment conducted at the American Telephone & Telegraph (AT&T) Laboratories in New Jersey.

2.1 Acquisition

fMRI does not measure brain activity. It measures, indirectly, the blood flow in the brain, under the assumption that highly active neurons have higher metabolic demands than less active ones (so called metabolic imaging). The invention of modern fMRI traces back to the early nineteen-nineties when researchers discovered that hemoglobin can be used as natural contrast agent for magnetic resonance imaging (see Fig. 26), provided its effects are accentuated by multiple excitation pulses in high fields (Ogawa et al., 1990). The contrast of interest[1], called Blood Oxygenation Level-Dependent (BOLD), is that between de-oxygenated (paramagnetic) and oxygenated (diamagnetic) hemoglobin. Sampling several million of neurons per voxel, the BOLD effect has been shown to correlate with neural activity, and especially with population-level electrical activity (Logothetis et al., 2001). For an in-depth review on the methodology, highlighting the boundaries of possible interpretations of fMRI results, the reader is referred to (Logothetis, 2008).

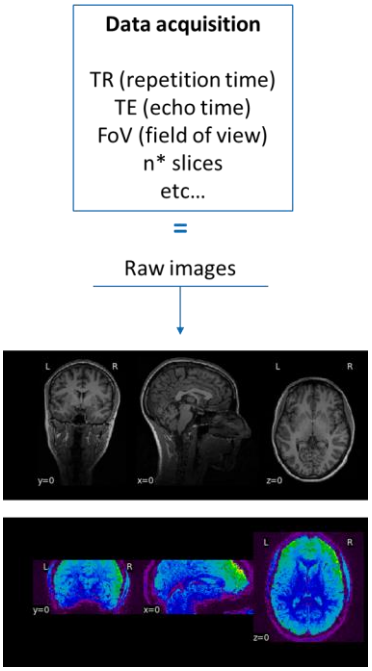


Figure 27 fMRI data acquisition. Once all sequence parameters are established, it is possible to acquire raw anatomical images (upper, black and white scale) and raw functional images (lower, green-blue scale). This example from our data illustrates how the tradeoff between space and time resolution can lead the researcher to opt for functional images that cover the brain only partially: about 3 cm of the parietal lobe are sacrificed in order to guarantee good coverage of the temporal lobe, with high resolution (1.5 cm isotropic voxels) and fast TR (2.3s).

A typical session of an fMRI experiment starts with the acquisition of anatomical images, followed by one or more runs of functional acquisitions. Data are recorded while subjects are lying down with their head inside the scanner. Thanks to a mirror system mounted on the head coil, subjects can see stimuli that are projected on the screen. Similarly, they can be prompted with audio stimuli thanks to MRI-compatible headphones. The overall quality of the images is determined by the interplay of keys parameters that the researcher can tune in accordance with the goal of the investigation (for examples of raw images see Fig. 27), determining the temporal and spatial resolution of the images acquired. Among them:

- Repetition time (TR), the time between two successive applications of the radiofrequency pulse (i.e., interval between

[1] BOLD fMRI is not the only viable option for cognitive imaging. Another functional MRI method is arterial spin labeling (ASL), relying on arterial water as tracer to measure cerebral blood flow. It offers lower signal to noise ratio, but higher spatial and temporal resolution [Detre JA, Wang J (2002) Technical aspects and utility of fMRI using BOLD and ASL. Clinical Neurophysiology 113:621-634, Borogovac A, Asllani I (2012) Arterial Spin Labeling (ASL) fMRI: advantages, theoretical constrains, and experimental challenges in neurosciences. International journal of biomedical imaging 2012:818456.]. Through this manuscript we use the name fMRI to denote functional MRI based on the BOLD signal.

successive data read out from the same location), measured in milliseconds;

- Echo time (TE), the brief time between the radiofrequency pulse and the acquisition of data (i.e., interval between the pulse and the peak of the echo), measured in milliseconds;
- Field of view (FoV), the physical size of the image (i.e., square image area that contains the object of interest), measured in cubic millimeters;
- Slice parameters: the number of slices to be acquired, the direction of the acquisition (e.g., sequential or interleaved), and whether multiple slices are obtained at the same time (i.e., so called multiband sequences).

2.2 Pre-Processing

Raw images need to undergo a series of preprocessing steps before they can be statistically analyzed. While many options are possible, I here focus on pipelines similar to the one I adopted, which include steps such as (see Fig. 28):

1. Slice timing. Different brain volumes are not acquired at the same time. Thus signal detection can be enhanced by adjusting (via time interpolation) the images to the true time of acquisition.
2. Realignment. Functional images are all aligned to a reference image (e.g., the first scan of the first run), thanks to rigid body transformations (i.e., 3 translations and 3 rotations). This step adjusts for head movement between slices.
3. Co-registration. Images of different modalities (i.e., anatomical and functional images) need to be aligned in the same space (within subjects).
4. Segmentation. Anatomical images can be separated according to the tissue types: white matter, gray matter and cerebrospinal fluid.

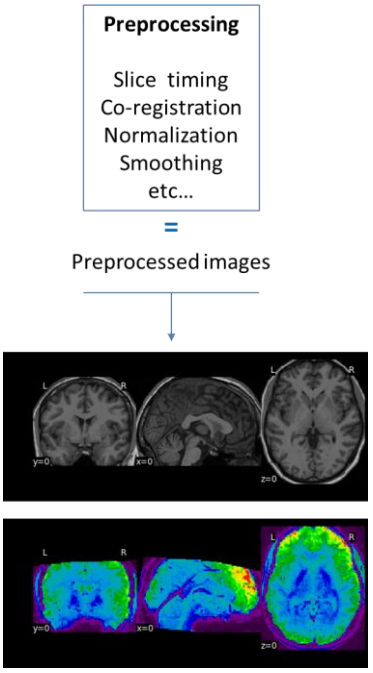


Figure 28 fMRI data preprocessing. Raw images undergo a series of preprocessing steps, some of which are optional and need to be tailored to the scientific question investigated. Here we illustrate the effect of co-registration and normalization on the same anatomical and functional images as in Fig. 27. Data are taken from our experiments.

5. Normalization. Whenever group level analyses are foreseen and/or one wishes to analyze the data with respect to regions or coordinates of interest derived from the literature, the brain space of single subjects needs to be normalized in a common reference space. The images are warped (stretched and squeezed) as to match a standardized anatomical template (e.g., Talairach atlas or MNI template).
6. Smoothing. Signal-to-noise ratio and functional overlap across subjects can be improved by spatial smoothing of the images, at the expenses of spatial resolution. It is implemented via a convolution with a 3D Gaussian kernel of specified width.

2.3 Standard Univariate Analyses

After pre-processing, fMRI time-series data are statistically analyzed in order to detect reliable activations. Since 1995, the General Linear Model (GLM), an adaptation of multiple regression analysis, has been the most widely adopted practice (Friston et al., 1995; Worsley and Friston, 1995). It allows the decomposition of the overlapping BOLD signals based on the experimental conditions (so called design matrix). The underlying assumption is that the observed BOLD signal (y) results from the multiplication of the design matrix (X) times unknown parameters (β), plus an error term (ε): $y = X\beta + \varepsilon$. Each voxel is considered as an independent observation and X can contain regressors quantifying (i.e., continuous predictors) and/or classifying (i.e., binary predictors) the experimental conditions.

Given the known properties of the hemodynamic response, the BOLD signal is expected to peak approximately 5 seconds after stimulation, and being followed by an undershoot that lasts as long as 30 seconds. The representation of an ideal, noiseless response to an infinitesimally brief stimulus is called Hemodynamic Response Function (HRF). The signal of each voxel can be treated as a linear

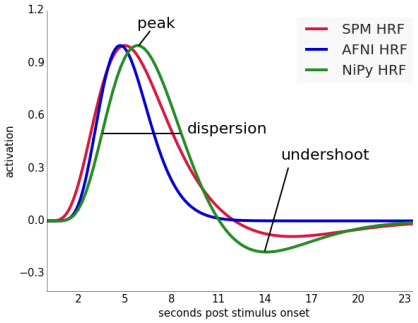


Figure 29 Example of canonical HRF shapes. Different software currently used for fMRI data analyses rely on slightly different HRF shapes.

superposition of multiple HRFs once assumptions on their shape are made, namely describing the time to peak, the dispersion (i.e., width of the curve), and the final undershoot. The canonical HRF is characterized by two gamma functions, one modelling the peak and one modelling the undershoot (see Fig. 29). To allow for variations from the canonical form, partial derivatives of the canonical HRF with respect to its peak delay (i.e., temporal derivative) and dispersion (i.e., dispersion derivative) can be added as further basis functions. These alterations can help capture differences in the latency and in the duration of the peak response. Hence, regressors of the GLM are convolved with the chosen HRF (with or without derivatives). The widespread use of the canonical HRF is challenged by studies showing how responses modeled after a data-driven estimation can outperform the classical approach while relaxing some of the assumptions (for instance forcing it to be equal across experimental condition, yet letting it differ across voxels (Pedregosa et al., 2015)). However, in most cases the statistical advantage is yet not sufficient to justify the extra computational cost.

Among the different regressors that enter into the GLM design matrix, it is usually possible to distinguish between regressors of interest (those that will be the focus of the following analyses, which are convolved with the HRF) and of no-interest (which are not convolved with the HRF). Among the latter, we find for instance so called movements regressors, estimated during the co-registration pre-processing, accounting for slight movements of the subject in the scanner. If data have been acquired in multiple sessions or runs, regressors accounting for these different time points of data acquisition can be added. It is important to model all known variables, even if not experimentally interesting, as this minimizes the variance of the residual error (ε) and adjusts the means of the effects-of-interest.

The outputs of the GLM are beta maps (one β for each column of X) that describe, for each voxel individually, the strength of the effect of that particular experimental condition. Following the

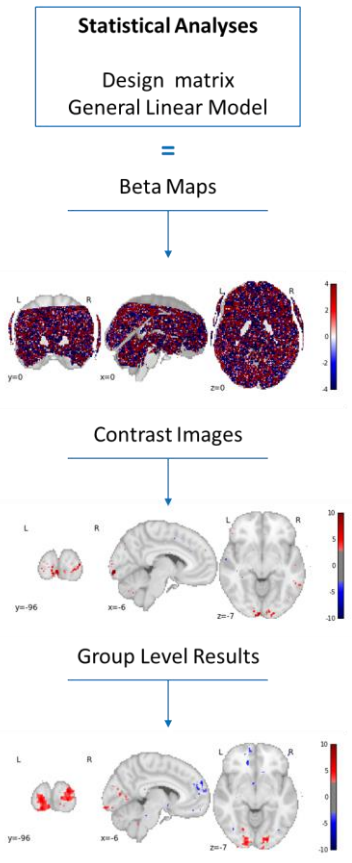


Figure 30 fMRI data analyses. Example of a beta map (upper), a contrast map at the single subject level (middle) and at the group level (lower). Data are taken from our experiments.

standard pipeline of analyses, beta maps are then contrasted according to the experimental design: for example, the map corresponding to left hand movement can be compared with the one corresponding to right hand movement. Such statistical testing is performed at the voxel level (hence the name mass-univariate test), usually by means of t or F tests. Under the null hypothesis (e.g., no difference across the two conditions), the values in the resulting map are distributed according to the respective probability density function, Student t or the F distribution. This t-map (or F-map) is the usually plotted threshold as to reveal which voxels (or cluster of voxels) exhibit significance differences. The contrasts generated for the individual subjects can be analyzed at the group level either with a Fixed Effects Analysis (FFX, the time series of different subjects are concatenated), or with a Random Effects Analysis (RFX, comparing the group effect to the between-subject variability) (see Fig. 30). As multiple statistical tests are performed simultaneously (one for each voxel), the possibility of false positive is inflated and the results need to be corrected for multiple comparisons, with methods such as the familywise error rate (FWER) or the false discovery rate (FDR) (Bennett et al., 2009).

To enhance statistical power and allow more specific inferences, the analyses are often restricted to given areas of interest (ROIs), in which the BOLD signal (averaged or voxelwise), is compared between different conditions (see Fig. 31). Particular care should be paid to avoid so called "double dipping", i.e., a circular analysis where statistical inference is drawn from the same dataset that was used to select the region of interest to begin with (Kriegeskorte et al., 2009; Vul et al., 2009; Kriegeskorte et al., 2010). For instance, if wishing to test the hypothesis that a given region X responds to condition A more than B, selection of voxels as belonging to X should not be based on the A-B contrast. Ideally, ROIs tailored to the cognitive function investigated should be identified either on the basis of anatomical constraints or with an independent localizer (see for instance Fedorenko et al., 2010), depending on the precise hypotheses at stake.

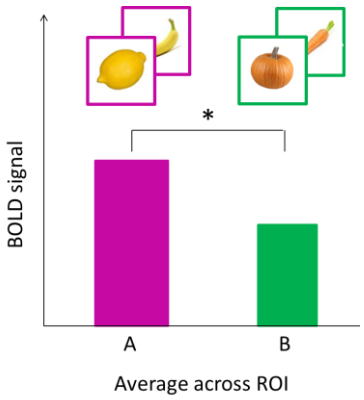


Figure 31 fMRI univariate analyses. Schematic illustration of the classical univariate analyses comparing, in a given ROI, the average BOLD signal for condition A (e.g., yellow items) vs condition B (e.g., orange items). Synthetic data.

2.4 Discussion

The classic, univariate, approach to fMRI data analyses here described led to major discoveries in cognitive neuroscience. For instance, in vision, highly selective regions (i.e., exhibiting a preferential response for a given category of images) have been discovered for objects (Malach et al., 1995), places (Epstein and Kanwisher, 1998), faces (Kanwisher et al., 1997), and words (Dehaene and Cohen, 2011). Moreover, it has been instrumental in studying networks involved in high level cognitive functions such as attention (Fan et al., 2005) or emotions (Ochsner et al., 2002).

The analysis steps here reviewed can be implemented with different software, of which the most widely used are SPM (http://www.fil.ion.ucl.ac.uk/spm/), FSL (http://fsl.fmrib.ox.ac.uk/), AFNI (https://afni.nimh.nih.gov/afni/), and BrainVoyager (http://www.brainvoyager.com/). Notwithstanding an overall standardization of the practices, substantial wiggle room is left to the researcher in terms of parameter tuning, calling for careful reporting of all methodological choices at all stages as these greatly impact the final results (Carp, 2012b, a; Pauli et al., 2016).

Current improvements are following two directions: concerning data acquisition, improving spatial and temporal resolution; regarding statistical analyses, making inference at the sub- or multi- voxel level. For instance, parallel imaging can push the spatial resolution without sacrificing the temporal one (e.g., (Feinberg et al., 2010; Moeller et al., 2010). Further attempts to improve the time resolution of fMRI include so called slice-based fMRI (Janssen et al., 2016). Overall, it seems that the time and space resolutions will be pushed significantly in the upcoming years. Recently, authors have reported fMRI responses to stimuli oscillating at up to 0.75 Hz (Lewis et al., 2016), while in terms of spatial resolutions, innovative findings are stemming from investigation with ultra-high field MRI (7 tesla) (Harvey et al., 2015).

Finally, we have seen that the standard univariate approach allows inferences only at the voxel level. As we will see in the last section of this chapter, since the nineties, researchers have tried to go beyond the voxel resolution and one of the most promising ways of analyzing brain data is thanks to multivariate techniques.

3. MagnetoEncephaloGraphy (MEG)

Going back to our television metaphor, let's say we are interested in understanding not only which parts are broken, but also how this breaks the normal flow of information within the appliance. We are thus willing to sacrifice the spatial resolution in order to get the best temporal resolution possible. In neuroimaging, this corresponds to choosing techniques such as ElectroEncephaloGraphy (EEG) and MagnetoEncephaloGraphy (MEG). While fMRI can be seen as a specific form of image processing, these time-resolved methods belong to the domain of signal processing.

Neurons are current generators: when an assembly of neurons fires, it generates microscopic electric currents. If neurons in the assembly are oriented in parallel and each of them forms an electric dipole, their post-synaptic potentials sum up to a current dipole, which in turn generates electric and magnetic fields strong enough to be measurable on the surface of the head (the scalp). The technique developed to measure the scalp's electric field, EEG, was one of the first used to measure brain activity (Berger, 1929) for both clinical and research oriented goals. The magnetic field can also be measured thanks to a younger technique, MEG (see Fig. 32). Thus, contrary to fMRI, both MEG and EEG directly measure brain activity, specifically post-synaptic potentials (which are slower but stronger than action potentials), believed to be mostly generated by pyramidal neurons (which are parallel to each other, and oriented perpendicular to the surface)(Hansen et al., 2010). Theoretically, the topographies

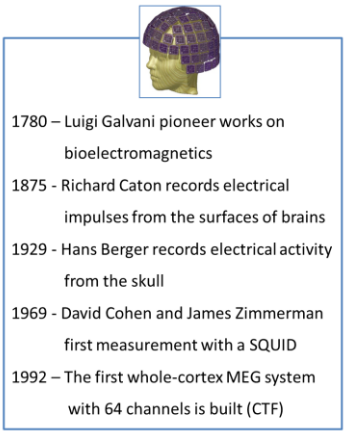


Figure 32 Brief history of M/EEG. We here report only a few of the many methodological and theoretical developments of both physics and physiology that have brought about M/EEG.

recorded with the two methods are orthogonal, it suffices to think of how the direction of the electric current determines the direction of the corresponding magnetic field (i.e., right-hand rule). In reality the complementarity is more indirect, as the two techniques are potentially sensitive to different magnetic sources. Moreover, there are some key differences that might lead one to prefer one method over the other. EEG suffers more from spatial smearing effects due to the poor conductivity of the skull: electrical currents are distorted by the bones more than the corresponding magnetic fields. Additionally, usually MEG systems offer more data points than EEG caps (e.g., 306 sensors instead of 64 electrodes), even if EEG caps with up to 256 electrodes are now available. The combination of these two weak points of EEG is a problem especially when attempting to locate the source of the electrical activity inside the brain: source reconstruction is generally less accurate with EEG compared to MEG. On the other hand, MEG cannot see radial components (i.e., as they do not give rise to an external magnetic field), which are instead detectable with EEG. Moreover, it has a weak sensitivity for signals originating in deep sources. For my experiment, not foreseeing the need to explore deep sources, I chose MEG (see Chap. 5); therefore, I will now focus only on this technique.

3.1 Acquisition

The MEG signal is recorded while subjects are sitting or lying down in an isolated room. The main goal of the isolation is that of magnetically shielding the MEG gantry (i.e. the structure hosting the recording equipment) from external sources of magnetic noise, which would washout the brain signals' weak magnetic field (~10^{-12} T, as a reference, earth's magnetic field is 10^{-4} T). Isolation is also instrumental to attenuate the sounds from the outside (which could potentially interfere with the experimental task) and to ensure that the luminosity is kept constant and controlled.

Currently produced MEG systems can have a different number and kind of sensors or SQUIDs (i.e., superconducting quantum interference devices, very sensitive instruments able to detect extremely subtle magnetic fields). For instance, Elekta systems have 102 magnetometers and 204 axial gradiometers. Magnetometers are the simplest pickup coil, measuring the component of the magnetic vector which is normal to its plane called B_z (the unit of measure is Tesla). Gradiometers measure the difference of B_z in the axial direction z (i.e., axial gradiometers), or in the tangential direction y (i.e., planar gradiometers), thus their unit of measure is Tesla over meter. Gradiometers suppress distant noise (i.e., background signal), thus providing a better measure of the local magnetic field (focal sensitivity) at the expense of a reduced sensitivity to distant sources (i.e., capturing signal from only quite superficial cortical sources).

While acquiring data, the following parameters need to be defined: the sampling frequency (e.g., 1 kHz, it needs to ensure adequate acquisition of the signals of interest); the online filters (e.g., band-pass between 0.03 Hz and 330 Hz, low-pass filter at one half or less of the sampling frequency to avoid aliasing and a high-pass filter to minimize effects of large low-frequency signals); whether to apply an active compensation of the external noise (i.e., ambient magnetic distortions are measured by a magnetometer placed outside of the magnetically shielded room and ad hoc compensation signals are sent to three coil pairs mounted on the outer surface of room); whether to continuously record the head position of the subjects (i.e., head position can be tracked by continuously exciting references coils place on the skull of the subject at frequencies higher than the typical brain signal).

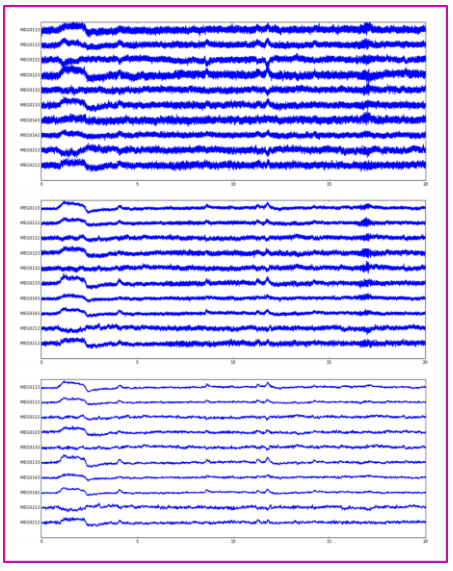


Figure 33 Example of MEG data at different stages of preprocessing. The same 20 seconds of data (from the experiment presented in Chap. 5) are shown: as they are acquired with Maxshield active noise compensation on (upper panel); after the application of Maxfilter (middle panel); after low-pass filtering at 40 Hz.

3.2 Pre-Processing

As for fMRI, MEG data need to undergo a series of preprocessing steps before statistical analyses can be correctly carried out (see Fig. 33). Again, several different pipelines can be

followed; I here focus on the steps I took. First, visual inspection of the data can be used to select "bad channels": some sensors can be damaged (always showing recurrent artifacts), other can be temporarily malfunctioning (for instance due to a local temperature change). Second, if active compensation of external noise (e.g., as implemented in the magnetic shielding system MaxShield of the Neuromag by Elekta) was on during the acquisition, one needs to remove such magnetic interference (running the NeuroMag MaxFilter program). There are three classical steps performed by MaxFilter:

1. Signal space separation (SSS) to suppress external magnetic interference. It consist in the application of spatial filtering based on the different sources of signals: from inside the subjects' head (of interest) and outside the subjects' head (noise). It is possible to take into account temporal information as well (so called tSSS).
2. Interpolation of noisy MEG sensors. Despite maintenance and tuning, some coils can exhibit a permanently noisy behavior (for instance because of magnetic flux trapped during the previous MEG cool down). Moreover, during each single session (or even each single experimental run), some channels can manifest recurrent jumps or other artifacts. Bad/noisy channels can be visually identified and manually declared and/or automatically detected.
3. Realignment of MEG data into a subject-specific head position. Subjects might slightly move their head beneath the gantry causing a misalignment of the different sensors across the experimental runs.

Next, some typical artifacts can be detected (with more or less automated procedures) and removed. These include eye blinks and hearth beats, which show a very stereotyped time profile and topography (see Fig. 34). Data can then be filtered. The goal of this step is, first and foremost, to remove line noise and its harmonics (in the EU, this artifact is seen at 50 Hz, in the US at 60 Hz). Depending on the kind of analyses planned, additional filtering can be performed:

for analysis aiming at detecting event-related fields (see ERF below), data are usually low pass filtered at 30/40 Hz (to remove line noise and high-frequency artifacts), while for time-frequency analysis (see TFA below) one usually wishes to keep higher frequencies as well (as they may contain relevant effects). Optionally, data can also be down-sampled, mostly to ease the computational costs. In most cases, data are epoched according to the experimental conditions, for instance breaking the continuous signal from ~200 ms before to ~600 ms after stimuli onset.

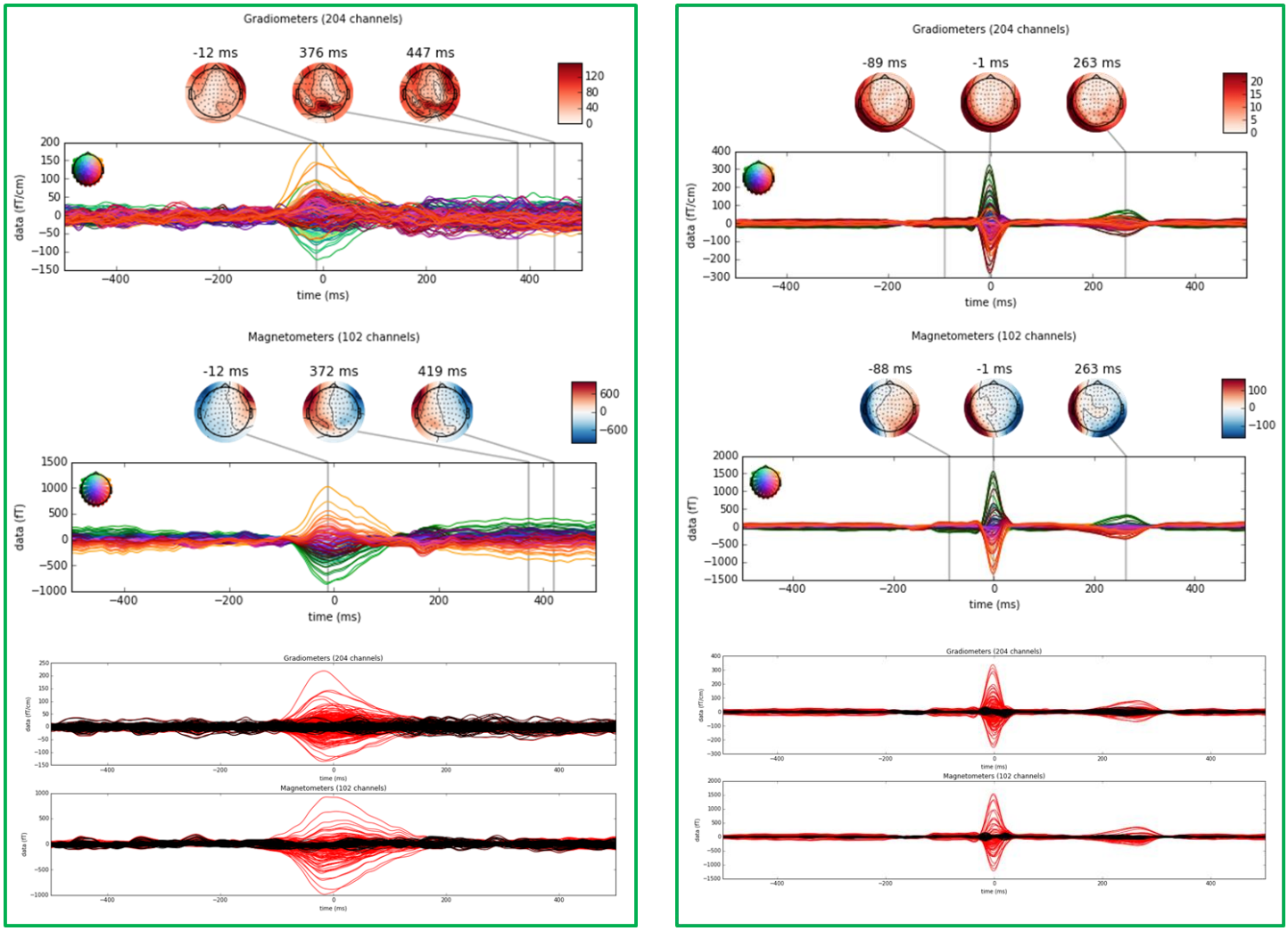


Figure 34 Artifacts identification and removal. Blinks (on the left) and heart beat (on the right) can be identified and subsequently removed. Note the different (and characteristic) time profile and topography: blinks are slower and present a typical frontal distribution; heart beats are faster and more lateralized. In both cases, it appears clear that the electrical source of the magnetic field recorded is not cortical. Data from one of the participants of the experiment in Chap. 5.

Baseline correction can be applied to isolate the changes in signal due to the stimulation from those associated with random low-frequency

noise from sensors and slow field fluctuations. To this end, it is necessary to select a range of time where it is reasonable to assume that no-stimulus related activity was being produced, typically the 100-200 ms pre-stimulus interval. Then, one commonly computes, for each recording channel, the mean signal over this interval, and subtracts it from the signal at all following time points. Some authors opt for statistical analyses on non-baseline corrected data, which they have high pass filtered at 1 or 2 Hz, aiming at discarding irrelevant low fluctuations without relying on the definition of a baseline period (which should not contain event-related fields, while being close in time to the events) (Gross et al., 2013).

3.3 Standard Univariate Analyses

The multidimensional nature of the signal acquired through MEG enables rich and diverse analyses. The first observation is that electrophysiological signals contain information on both "evoked" and "induced" neural activity. The so called evoked responses are aligned in phase (i.e., phase-locked) while the so-called induced responses are not (i.e., they are modulations of ongoing oscillatory processes that are not phase-locked). When epochs of different conditions are averaged and compared to one another, the event-related fields (see below ERF) one obtains represent only of phase-locked neural activity, as the averaging procedure cancels out any modulation that across trials is out of phase. Non-phase-locked activity can be appreciated as event-related changes in the power of neuronal oscillations (see paragraph on time-frequency analysis, TFA). The spectral representation of the signal offers the possibility to study evoked (thus time-locked and phase-locked) changes too (see paragraph on inter-trial phase coherence, ITC). Finally, the underlying sources of the observed effects can be tentatively reconstructed to improve spatial localization of the cognitive processes/representations of interest (see paragraph on source reconstruction).

ERF. In a standard ERF analysis, the epochs obtained at the end of the pre-processing, are averaged, subject by subject, condition by condition, and then analyzed to look for spatio-temporal clusters (i.e., multiple nodes extending in time and space) in which they statistically differ (see Fig. 35). As for fMRI (and even more so given the multidimensional nature of the signal), the statistical analysis of MEG data faces a serious problem of multiple comparisons: the effect of interest (i.e., a difference between experimental conditions, e.g., yellow vs orange stimuli) is evaluated for an extremely large number of channel/time-pairs. One widespread solution is to calculate a cluster-based test statistic (i.e., the effect of interest is quantified on the basis of temporal, spatial and spectral adjacency) and then compute its significance probability with non-parametric procedures (i.e., random partitions of the data are used to determine the distribution of the results under the null hypothesis) (Maris and Oostenveld, 2007). ERFs are simple and fast to compute, they offer a high temporal precision and accuracy, and they can be easily interpreted (at least when providing positive results). However, as we have seen time-locked but not phase-locked activity is lost due to the averaging procedure. Moreover, there are well known nonlinear neural activity patterns that elude investigation by means of ERF, such as synchronization and cross-frequency coupling (Cohen, 2014).

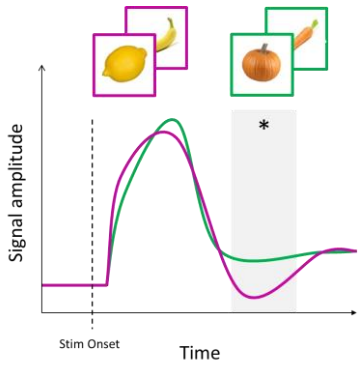


Figure 35 MEG univariate analyses: ERF. Schematic illustration of the classic univariate analyses comparing, in a spatio-temporal cluster of sensors, the average signal for conditions A vs condition B. Synthetic data.

TFA. MEG signals can be transformed from the time domain to the frequency domain thanks to, for instance, the Fourier transforms. This step, a so called spectral analysis, allows a better investigation of oscillatory signal components. Following this transformation, measures of power changes (TFA, this paragraph) as well as measures focusing on phase properties (ITC, next paragraph) can be computed and compared across conditions.

Time-frequency analysis (or TFA) aims at quantifying frequency specific neural activity that is not necessarily phase-locked to an event (see Fig. 36). Since the first EEG studies, different power bands have been identified and differentially linked to neural functions

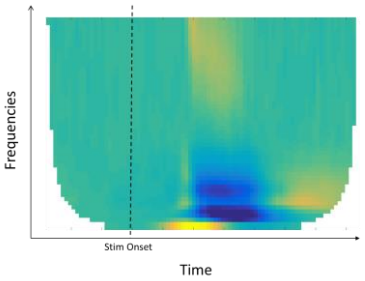


Figure 36 MEG univariate analyses: TFA. Example of the TFA reconstruction of the MEG signal from left occipital sensors after the presentation of a written word. Data derived from the experiment presented in Chap. 5.

and cognitive processes: theta band, approximately from 2 to 8 Hz; alpha band, approximately from 8 to 13 Hz; beta band, approximately from 13 to 30 Hz; low gamma band, approximately from 30 to 70 Hz; high gamma band, approximately from 70 to 120 Hz. When a change in a given frequency band is observed, for instance following the presentation of a stimulus, it can be characterized as a decreases relative to the pre-stimulus baseline (event-related desynchronization, ERD) or increase (event-related synchronization, ERS) (Pfurtscheller and Da Silva, 1999). According to the frequency band affected, these power changes, reflecting coupling or uncoupling of populations of neurons, may indicate either activation or deactivation of a given brain region: in the gamma band ERD reflects a reduction in processing, while in the alpha band it reflects increased processing (Pfurtscheller and Da Silva, 1999). To date, oscillations are the most promising window on neural processes, linking results from different neuroscientific approaches and allowing exploration of an added dimension, frequency. However, with respect to ERF analyses, the wiggle room (i.e., the different choices that the experimenter has to take while tuning parameters) increases, and the temporal precision is somehow undermined by the process of time-frequency decomposition (Cohen, 2014). As for ERF, cluster-based permutation tests are used to assess whether there is a significant difference between experimental conditions (i.e., whether the data come from different, non-exchangeable, distributions).

ITC We have seen that TFA concerns the power of the spectral features irrespective of their phase (time-locked but not phase-locked). However, the spectral representations of different experimental conditions (e.g., yellow vs orange items) might also significantly differ over sensors, time and frequencies in a time-locked and phase-locked manner. ITC (inter-trial phase coherence or phase locking factor (Tallon-Baudry et al., 1996)) analysis identifies clusters (in time, space and frequency) where there are differences in phase consistency over trials. The results denote evoked effects similar to

those captured by the ERF, but with additional information derived from the decomposition into its constituent phase-locked frequency bands (e.g., Shah et al., 2004). Again, statistical analyses are conducted using minimal distributional assumptions thanks to parametric testing.

Source Reconstruction. Given a dipole in a conductive medium (i.e., some neurons firing in a given brain location), the electrical current spreads through this medium (i.e., the brain, the skull, the scalp) and it is possible to deterministically know the distribution of the induced magnetic field (linear forward model). The opposite process, i.e., to identify the location of a dipole given its superficial electric potential (EEG) or magnetic field (MEG) is a difficult mathematical problem (i.e., linear problem with more unknowns than observations). Different methods have been developed to solve the so-called inverse problem and provide a good estimate of the source of M/EEG signals. First, an accurate head model is needed, accounting for the properties of the different tissues (i.e., skin, skull, brain) in order to explain the conductivity patterns. Simple sphere models provide fast analytical solutions, while more realistic models such as the boundary element method (BEM) are harder to solve but offer finer reconstructions. Then, attempts can be made to solve the inverse problem, while taking into account priors such as the pre-computed anatomical constraints. Generally speaking, two perspectives can be taken: (1) make specific modelling assumptions on the number of focal sources and their approximate location, as done in so called dipole models (or discrete source approaches); (2) do not restrict the effects to a set of focal sources but rather distribute a large number of dipoles throughout the brain volume (ideal when dipoles number and locations cannot be predicted), as done in distributed source models. These models aim at estimating the dipole originating from the observed signal, an underdetermined problem as there are more unknown (dipoles) than data points (MEG channels). The different widespread distributed source models (e.g., Minimum-

norm, LORETA) vary in the way they choose to minimize the sum of the dipoles across all voxels (i.e., minimal estimate that can explain the measurements).

3.4 Discussion

The analyses of M/EEG data here reviewed can be considered "classic", as they exploit univariate statistical methods. They have led to major discoveries in cognitive neuroscience, allowing the description of key ERP/Fs (as an example consider the N400 (Lau et al., 2008)) and of the role of oscillations in specific frequency bands (e.g., induced gamma activity as construction of object representation (Tallon-Baudry and Bertrand, 1999)). For a review on the contribution of neurophysiological techniques to the study of language see (Salmelin, 2007). Moreover, synchronization has been proposed as a mechanism for establishing communication between brain areas and has been linked with cortical interactions underlying, for instance, multimodal associative learning (Miltner et al., 1999; Palva and Palva, 2012). Finally, one approach not detailed here but contributing to the popularity of time-resolved neuroimaging methods is that of frequency tagging. The aim is to achieving an easier dissociation of the signal of interest from the endogenous activity, i.e., higher signal to noise ratio than classic ERP. To do so, it exploits the fact that if stimuli are presented at a constant rate, the associated neural population will oscillate with the same period (Buiatti et al., 2009; Kosem et al., 2014).

Generally speaking, current improvements include better source estimates (Bekhti et al., 2016), automatization of artifacts detection and repair (Jas et al., 2016), and the development of cheaper and more economical MEG systems (for instance thanks to low-maintenance sensors (Knappe et al., 2014).

The analysis steps here reviewed can be implemented with different software, the most widely used are SPM (http://www.fil.ion.ucl.ac.uk/spm/), Brainstorm

(http://neuroimage.usc.edu/brainstorm), FieldTrip (http://www.fieldtriptoolbox.org/), MNE (http://martinos.org/mne), and NUTMEG (https://www.nitrc.org/projects/nutmeg/). As for fMRI - and perhaps even more so given the relative recent spread of the method - standardization of pipelines and thorough reporting of all steps (from data acquisition to data analyses) are warranted (Gross et al., 2013; Keil et al., 2014).

Finally, notwithstanding the inherently multidimensional nature of the signal, traditional M/EEG investigations have relied on standard statistical inferences (with the noteworthy exception of brain-computer interface studies, e.g., Blankertz et al., 2007). We will now see how multivariate techniques of data analyses are revolutionizing not only fMRI, but also MEG investigations of the neural correlates of cognitive representations (Stokes et al., 2015).

4. Multivariate Analyses of Neuroimaging Data

In the previous sections of this chapter, we have seen how univariate analyses can be used to detect, during the execution of a particular task, which brain regions are engaged (fMRI) and when (MEG). The reasons why this approach is widely exploited are two-fold. First, it answers core questions in cognitive neuroscience (e.g., is area A engaged more during condition X than during condition Y? is activity linked with condition X observed earlier than activity linked with condition Y?). Second, it is easily implemented thanks to the standardized pipelines offered by many softwares. With respect to fMRI studies, the main drawback of univariate analysis is that it fails to reveal two kinds of representational codes (see below): those that are distributed across multiple voxels and those that are encoded at below-voxel resolution. The same issue affects M/EEG data, where similar observations apply not only to the spatial resolution (i.e., in this case the minimal units are the sensors, not the voxels), but also to

the time and frequency dimension (i.e., statistical analyses rely on differences between conditions of interest at a given time point, in a given frequency range). Different methods have been proposed in the last 20 years to overcome the shortcomings of univariate analyses.

4. 1 Resolutions of Representational Codes

In the previous chapter (see Chap. 1. 5.1), we have explored the relationship between representational content (which information is encoded in a given brain area), representational geometry (the representational space in which the information is organized), and representational format (the corresponding neural code). Our understanding of possible neural codes, i.e., meaningful schemes of the activity of single neurons or of a population of neurons, is still limited. However, we do know that univariate analysis gives access only to the information that is expressed in terms of changes of the BOLD signal at the level of single voxels, missing information that is distributed across multiple voxels and/or that is encoded at below-voxel resolution. First, let's say that within a given region there are voxels consistently responding with high activation to dangerous animals (and no activation to harmless ones). If these voxels are intermingled with others consistently showing the opposite behavior, the univariate average activity would show no differences when a tiger and a rabbit are presented. Yet in principle the consistency of the underlying pattern could be detected. Second, let's imagine that within a given voxels there are neurons tuned to (i.e., preferentially responding to) different levels of height (short, medium and tall). The univariate activity recorded after the presentation of a cat, a wolf and a zebra would be the same, as the overall activity would show no preference.

This last case, representational code below-voxel resolution, was the first to be tackled thanks to an approach called adaptation or repetition suppression. It consists in measuring the difference in

activation within a voxel as a function of the relation between subsequent stimuli. The underlying principle is that the repeated presentation of the same stimulus produces a decrease in the response. By varying the features of the stimuli that are repeated, across trials, it is possible to investigate which feature repetition produces a reduction in the amount of activation, thus determining the (set of) features of the stimulus that are encoded in any given voxel. This method allows for the investigation of representations coded across neurons encompassed within a single voxel. In other words, it permits the detection of differences across experimental conditions even when these can be appreciated only at sub-voxel resolution. Thinking of the previous example with height-coding neurons, even if at the population level no univariate difference can be appreciated when stimuli are presented in isolation, an adaptation effect could be detected: the presentation of two stimuli of the same height (e.g., a wolf and a dog) would yield a smaller response than the presentation of two stimuli of different height (a cat and a zebra). Thus, adaptation allows to (indirectly) study the "tuning curve" of neurons with fMRI (Piazza et al., 2004): what are the stimulus features the neurons within a given voxel care for? Different models have been proposed to explain the mechanisms behind fMRI-adaptation: less overall neuronal firing (fatigue model), activity of fewer neurons (sharpening model), or shorter processing time (facilitation model) – for a review see (Grill-Spector et al., 2006). fMRI-adaptation paradigms have been pivotal in describing cortical areas selectively tuned to low level perceptual representations (Vuilleumier et al., 2002), as well as high level conceptual ones - including semantic associations (Wheatley et al., 2005). Adaptation paradigms have also been applied to M/EEG, looking at the effects of repetitions on ERP/Fs (Schweinberger et al., 2004) and power in different frequency bands (Gruber and Muller, 2005). The adaptation paradigm can be generalized into so called carry-over designs: an unbroken sequence of stimuli is analyzed in terms of how the response to a given stimulus is modulated by the previously presented ones (Aguirre, 2007). This setting permits the

simultaneous detection of potential differences between the mean neural responses to different stimuli (e.g., univariate effect of yellow vs red items), as well as the carry-over effect, in other words whether the responses to a stimulus is affected by its relation to the preceding ones (e.g., comparing red items when preceded by yellow ones vs orange ones).

Recall the hypothetical area which contains voxels consistently responding to dangerous (e.g., tiger) or harmless (e.g., rabbit) animals: it is possible that information is carried by the pattern of activity across voxels and not by the overall average level of activation. In the attempt to investigate representations distributed across multiple voxels, researchers have developed so called multivariate methods. These new approaches have been highly influential in the field of cognitive neuroscience first with their application to fMRI data and more recently with their exploitation in M/EEG paradigms as well (Grootswagers et al., 2016). Not surprisingly, the nomenclature of this collection of techniques (hereafter MVPA) changed from "*multi-voxel pattern analyses*" (Norman et al., 2006) to the more general "*multi-variate pattern analyses*" (Haxby et al., 2014). While fMRI and M/EEG data differ in many important ways, with respect to multivariate analyses the central idea is the same: to take into account the activity in multiple units at the same time. Spatially, the smallest unit of fMRI data are voxels (of varying size according to the resolution of the scanner and the sequence used), while with MEG data they can either be sensors (if working in sensor space) or, again, voxels (if working in the reconstructed source space). Additionally, in the case of MEG, temporal (i.e., which point(s) in time?) and spectral (which frequency band(s)?) dimensions of the unit need to be defined as well. Once the characteristics of the basic unit have been established, MVPA methods can be applied to fMRI or MEG data set with little variations.

The core idea behind this multivariate approach may be traced back to computational neuroscience's concept of population coding:

content is represented by the distributed activation of different representational units (Pouget et al., 2000). In principle, it should thus be possible to investigate the link between cognitive representations and the corresponding multi-units patterns of activity. Two approaches have been particularly successful in the neuroimaging literature: pattern decoding and the analyses of pattern geometries (so called representational similarity analyses - RSA). Moreover, voxel-wise modeling (so called encoding), an approach that can be seen as complementary with respect to decoding (Naselaris et al., 2011), is often included among MVPA methods, even if (as we will see later in the chapter) virtually all its steps (excluding possible validation practices) do not take into consideration the activity of multiple units (i.e., it is a univariate modeling of brain activity). In-depth reviews have been published on the different algorithms that can be used (Pereira et al., 2009), the many possible cognitive applications (Norman et al., 2006; Tong and Pratte, 2012), the underlying hypothesis on neural coding (Serences and Saproo, 2012; Kriegeskorte and Kievit, 2013; Haxby et al., 2014), and some of the challenges and pitfalls (Davis and Poldrack, 2013; Haynes, 2015). I will here introduce the main features of the three approaches: pattern classification, pattern correlation and voxel-wise modeling.

4. 2 Pattern Decoding

This approach relies heavily on supervised machine learning models to test the hypotheses at stake. It emphasizes diagnostic information that can help discriminate one kind of stimulus/condition from another.

Core concepts and key steps

The core concept is the search for a function $f()$, which takes as input X and returns y, where:

- X, i.e. the brain imaging data (being MEG recordings or fMRI scans). X has shape $n * m$ where n is the number of samples

available (i.e., observations, e.g., beta maps, single trials or average thereof), and m is the number of features (i.e., individual measurable properties, e.g., voxels, sensors);

- y is a vector of length n that assigns to each sample the label corresponding to its experimental condition.

Often, we want to test whether two different conditions (e.g., orange stimuli vs yellow stimuli) elicit reliably different patterns of activation (see Fig. 37). Thus we would like to assess whether the information in the distributed pattern of activation is sufficient to classify a given brain scan as belonging to category A (e.g., yellow) or B (e.g., orange).

Two commonly used algorithms to estimate a linear decision boundary between the two categories are linear discriminant analysis (LDA, (Fisher, 1936)) and linear support vector machines (SVMs, (Boser et al., 1992; Cortes and Vapnik, 1995)). LDA projects data onto a lower-dimensional space that maximizes class separation, in other words identifies projection weights that maximize the between-class to within-class variance. SVMs focus on the points that are most difficult to discriminate (i.e., support vectors) and attempts to draw a hyperplane that maximizes the margin, i.e., the distance between the hyperplane and the nearest data point from either class. Different models are classified according to the loss function they minimize (i.e., the function representing the price paid for inaccurate predictions). In certain cases, response distributions cannot be partitioned sufficiently well using single linear decision boundaries and thus nonlinear approaches (e.g., non-linear classifiers and multilayer neural networks) can be used (see Fig. 38).

Models can be extended to accommodate multiple classes, for instance if one wishes to classify stimuli as belonging to 3 classes (e.g., orange vs yellow vs red stimuli). Any binary classification method (e.g. SVMs) can be extended to multiclass classification via decomposition into binary classification problems thanks to a schema known as "one-vs-rest" (i.e., one classifier is built for each class and fit against all the others) and "one-vs-all" (i.e., one classifier is built

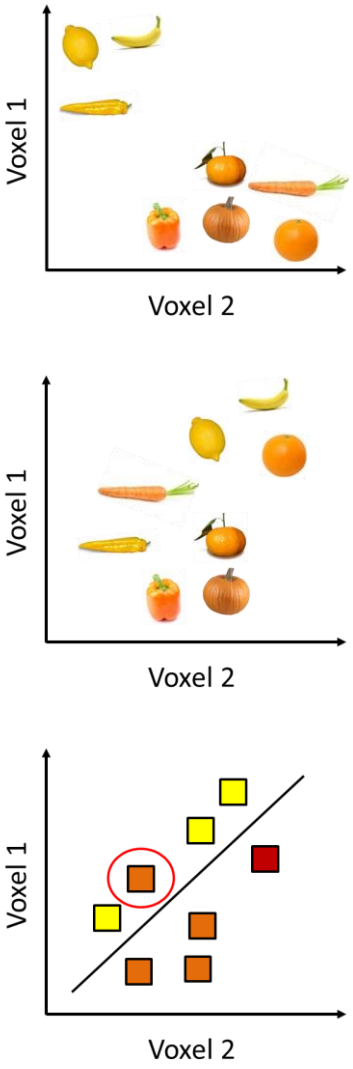


Figure 37 Exemplification of the difference between univariate and multivariate analyses. The upper panel shows two voxels whose univariate activation profiles distinguish between condition A (yellow items) and B (orange items). Specifically, voxel 1 shows high BOLD signal for category A and low activity for category B; voxel 2 presents the opposite pattern. The middle panel introduces a more complex situation in which only the multivariate pattern (i.e., taking into account both voxels at the same time) permits discrimination across the two categories. The lower panel shows which hyperplane can split the data according to the color label. Note how the classification is not perfect – the error is circled in red (corresponds to the carrot in the explicit representation above.)

for each pair of classes, e.g. red vs yellow, red vs orange, orange vs yellow). Sometimes, the different stimuli or experimental conditions we would like to compare do not simply belong to different classes, but to classes that are ordered, ranked. For instance, if stimuli are words or melodies, they can be ranked according to, respectively, how many letters and how many tones constitute each of them. Wishing to associate different brain recordings to one of the different classes, we can exploit multivariate regression approaches. Notwithstanding the algebraic nature of the problems (i.e., classification or regression), the key steps and the core issues of these methods borrowed from machine learning are the same. I will now highlight them.

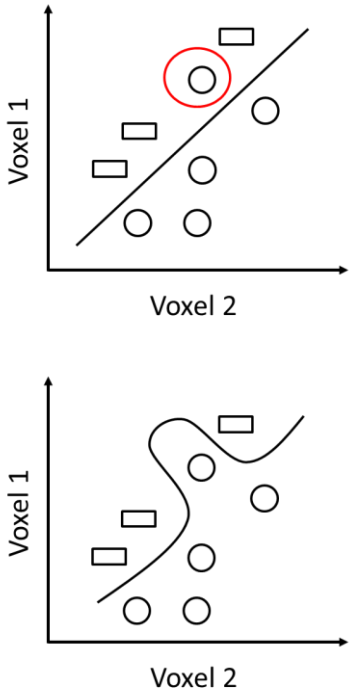


Figure 38 Example of the difference between a linear and a non-linear classifier. The shape classification problem can be solved via a linear model (upper panel) or a non-linear one (lower panel). In this latter case, perfect accuracy is reached, but this likely constitutes a case of overfitting: having over-learned from the train set, the model will perform badly when unseen data are introduced.

Feature Selection

The first decision that has to be made is which data will be fed to the classifier. Usually, when analyzing neuroimaging data, high-dimensional spaces (i.e., having many features), correspond to few available data points (i.e., scarse samples), which is problematic when wishing to assess statistical significance (the so called "curse of dimensionality"). As an example, in a typical fMRI setting the features (which determine the dimensionality of the problem) are the different voxels selected, and even a small ROI will likely include ~500 voxels. The data points available are the beta maps one wishes to learn from (and test on), which will usually be less than 300 (even if each beta map corresponds to a few/one trial). Features thus need to be selected and, according to the imaging technique of choice, this will involve the spatial dimension (i.e., voxels in fMRI, sensors in MEG) as well as the temporal and spectral dimensions (i.e., in MEG frequencies and time points should be selected as well). Different neuro-cognitive questions should drive the choices, tailoring feature selection to the goal of the paradigm. One possibility is the implementation of ROI analysis, thus selecting in space/time/frequency of clusters of data of interest. In order to avoid circularity, such selection should be based on independent observations (e.g., thanks to a functional localizer or anatomically

defined boundaries). Another option for dimensionality reduction is that of retaining the best features based on a univariate statistical test (e.g., F-test), as long as it is computed on a different contrast with respect to the one that will be investigated with decoding (e.g., one wishing to discriminate between orange and yellow items, could retain all voxels responding to the presentation of any visual stimulus vs baseline). Finally, an alternative approach which is now rapidly spreading is the application of a searchlight procedure: a sphere of arbitrary radius is centered in each and every voxel, essentially defining multiple ROIs moving in space (and/or time/frequency). The result of the classification is assigned to the voxel at the center of the sphere [for review of the results obtained with a searchlight in fMRI over the last 10 years see (Woolgar et al., 2016)].

Cross-validation

A standard way to estimate the predictive power of a decoding model is via cross-validation (or CV). For each round of CV, data are split in two sets and the model is trained (i.e., fit) on one, before being tested (i.e., attempted to predict) on the other. After multiple rounds of CV, the validation scores (i.e., the performance of the model) are averaged, leading to an estimation of the ability of the decoder to generalize to unseen data. Different CV schemes can be selected and nested-CV can be used to tune specific parameters of the models such as the regularization parameter C in SVMs (for a thorough review of CV approaches in neuroimaging see Varoquaux et al., 2016). A fundamental aspect of CV is that complete independence between train and test dataset has to be assured: there should be no leakage of information from the train to the test set, a particularly delicate topic in neuroimaging (consider for instance the temporal-autocorrelation in fMRI). In neuroimaging, a frequently selected CV scheme is Leave-One-Out, where for each fold data coming from one run are left aside as test set. Far from being flawless (Rao et al., 2008; Varoquaux et al., 2016), CV is an essential component of decoding analyses pipelines.

Overfitting

A potential risk in decoding is to over-learn characteristics of the train test (i.e., the model describes the noise instead of the real trend), resulting in a poor performance once tried on the test set (i.e. overreacts to minor fluctuations). The more complex the model (i.e., the more parameters are estimated), the more likely it will overfit the train set, thus resulting in worse generalization accuracy (see Fig. 38). In order to minimize overfitting, the first and best strategy is to test the model on unseen data (as done in cross-validation). Other actions consist in contrasting the "curse of dimensionality" through feature selection (see previous section) and explicitly penalizing excessively complex models (i.e. adding a complexity penalty to the loss function, e.g. the C parameter in SVMs).

Statistical significance of the results

The performance of a classifier is assessed by comparing it against what could have been achieved by pure chance. In a binary classification test (i.e., class A vs class B), assuming accuracy as the score function, theoretical chance level lies at 50%, provided the two classes are equally likely. Similarly, in a 4 class classification test it lies at 25%, while in a regression test it is close to 0 (with r^2 as score function). In principle, the performance of the classifier could then be compared (for instance with a *t*-test) against this theoretical chance level. However, given the numerous assumptions underlying parametric tests (e.g., normality of the distribution, homogeneity of variance, and independence of the samples), non-parametric tests should be preferred whenever possible. For instance, with permutation testing, the distribution of scores that would be obtained under the null hypothesis can be estimated from the data by randomly permuting the labels. Then the original, true score can be compared against the dataset specific, estimated null distribution, determining the probability that it was obtained by chance, and thus potentially allowing rejection of the null hypothesis. Permutation tests have

proven to be more valid than binomial or t- tests (Nichols and Holmes, 2002; Schreiber and Krekelberg, 2013; Stelzer et al., 2013). Especially when in presence of unbalanced classes (i.e., more samples are available for a given class, for instance orange stimuli are more frequent than yellow ones), receiver-operating curve (ROC) analysis might be preferred as a test statistic, having the advantage of being less computationally expensive than permutations while protecting from biased results due to unbalanced labels in the test set. In this case, results are summarized as area under the curve (AUC), where a value of 50% implies that true positive predictions (e.g. orange stimuli predicted as orange) and false positive predictions (e.g. yellow stimuli predicted as orange) are equally probable.

Interpretations and implications

What can we conclude on brain functions if we can decode a property of a stimulus (e.g., whether it represents a fruit or a vegetable) from a given brain area, at a given time point? Certainly, successful decoding implies that, pooling together all the selected units (being voxels, sensors, time points, etc…), there is enough information in the resulting pattern of brain activity to be able to classify the stimuli according to the labels we provided. However, as in the standard univariate activation detection setting, this does not necessarily imply that the information (present at *that* time point in *that* area) is either necessary or sufficient to support the cognitive task performed by the subject (or the cognitive representation involved). Moreover, as always, negative results of decoding are hard to interpret: not being able to decode information from a certain brain area, at a certain time point, could be due to lack of sensitivity, for instance because of low signal-to-noise ratio. Finally, decoders are by construction powerful tools exploiting any bias that can enable the required discrimination: they are thus extremely sensitive to any factor that co-varies with the stimulus features that we are trying to decode (and experimental conditions at large) (see Fig. 39). As will be discussed later, other multivariate approaches do not suffer from this

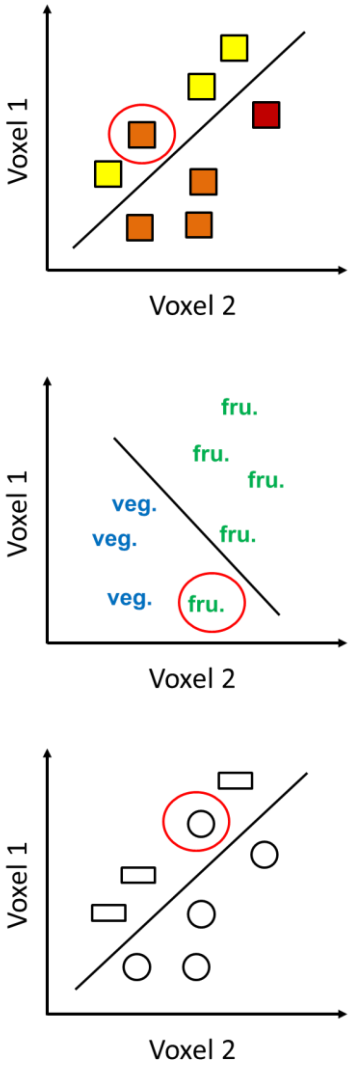


Figure 39 Example of confusions between covarying factors. The same neural representation (upper panel) can be used to decode other information about the stimuli such as their conceptual categorization as vegetables or fruits (middle panel), or their shape, round or elongated (lower panel). Categorization errors are circled in red. Note how the hyperplane changes to accomplish the different classification tasks.

ambiguity, as they rely on explicit models of the representation (Naselaris and Kay, 2015).

History and examples

As I have mentioned in the first chapter, pattern classification has been used since the early 2000s to decode information from brain activity on, for instance, visual stimuli orientation (Haynes and Rees, 2005; Kamitani and Tong, 2005), category (Haxby et al., 2001; Cox and Savoy, 2003) and exemplars (Eger et al., 2008). Moreover, it has been applied to higher order cognitive processes and functions such as numerical cognition (Eger et al., 2009; Eger et al., 2015), short term memory and enumeration (Knops et al., 2014), mental arithmetic (Knops et al., 2009), and word meaning across languages (Buchweitz et al., 2012). Similarly, in the MEG setting, first results concerned low level sensory-motor processes (Waldert et al., 2008; Carlson et al., 2011; Ramkumar et al., 2013), then, more recently, higher level representations of motion (Tucciarelli et al., 2015). Recent key findings include the detection of responses to novelty even in non-communicative patients (King et al., 2013) and the maintenance of seen and unseen information (King et al., 2016). Finally, it appears that decoding can push the spatial resolution of MEG: information available at the level of cortical columns (e.g., edges orientation) can be extracted with multivariate analyses of M/EEG signals (Cichy et al., 2015).

Criticism and future directions

As hinted at before, when attempting to investigate neural representations with classification techniques, one of the main issues is that of correlated variables. It is implicitly assumed that the decoded features are independent (e.g., being a fruit or a vegetable, being yellow or orange). However, this is clearly false, as often features are related. For instance, consider the case of size and grip type: very small items are also the ones you can handle with a precision grip, while larger ones are handled with a power grip. While using

decoding one might discriminate between “small” and “large” tools, it is possible that in fact the decoder is using information from a brain region that is not coding for size, but rather gripping type. Furthermore, this relationship takes the form of nested structures: for instance, while classifying wild animals, domesticated animals, house tools and gardening tools, two layers are nested one in the other: only animals can be wild or domesticated, only tools belong to the house or the garden. For a recent example of how this issue can be tackled via hierarchical logistic regression (i.e., a combination of multiple logistic regression models) see (Huth et al., 2016a).

When linear classifiers are used, the weight associated with each unit directly reflects its contribution to the classification result, thus it is possible to plot weight maps illustrating which units have been considered relevant by the classifier. However it should be noticed that only the performance of the classifier (as a whole) is statistically tested against chance level, no inference on the contribution of single units can be drawn. One possible way out is to re-evaluate the classifier, for instance after the exclusion of certain units to detect a possible decrease in performance (Pereira et al., 2009; Haynes, 2015). Recent efforts to increase neurophysiological interpretability of the recovered weights include attempts to recover more stable weights (Hoyos-Idrobo et al., 2015), and conversion of the backward model extraction filters (linear decoding weights) into activation patterns of the corresponding forward model (Haufe et al., 2014).

Finally, for time resolved techniques (M/EEG), brain dynamics can be explored with the temporal generalization method which offers the possibility to distinguish serial from parallel processes, continuous flow from discrete stages (King and Dehaene, 2014). Generalization across time is tested by training a classifier at each time point (e.g., t1) and testing it on all other time points (e.g., t1,t2,t3, etc…). Successful performance suggests the presence of representational codes which are stable in time. Similarly, generalization across conditions can be attempted: the classifier

trained to discriminate between two conditions (e.g., seeing orange vs yellow items) is tested on data from two other conditions (e.g., imagining orange vs yellow items). However, generalization is a weak logical control as SVMs, purely discriminant models, do not fit the variance in the data (i.e., there is no estimate of the amount of variance explained). In other words, the classification can be driven by purely incidental factors which might differ from one condition/time point to the other.

Finally, the main issue regards the under-specification of the underlying representational space: often the cognitive question one is exploring requires a finer description than the coarse binary discrimination provided by categorical classification. To better describe the representational space, two possible solutions can be used as a proxy for more explicit models (such as RSA and encoding, see next paragraphs):

- single pair-wise item classification can be performed, where accuracy scores can be used as a measure of the distance between items in the representational space (i.e., the harder it is to discriminate between two items, the closer they are) (Weber et al., 2009; Cichy et al., 2014);
- confusion matrices from one-versus-rest multiclass classifications can be analyzed, again under the assumption that the harder the classification problem (i.e., the lower the accuracy score) the closer in representational space the classes are (n.b., one-versus-all method where one classifier is built for each pair of classes would lead to the same result as above).

4. 3 Pattern Geometry

Significant decoding accuracy of particular classes in a given brain region only provides evidence that there is enough information to detect a difference among them. It does not entail that the neural representational space recovered from that region corresponds to a relevant cognitive one. The goal of Representational Similarity

Analyses (or RSA) is that of investigating the representational content of an area providing a principled way to test how well specific computational/behavioral models fit with the distributed activity pattern observed. RSA tests whether the similarity between the brain responses to different stimuli (e.g., similarities between beta maps) matches the similarity between the stimuli themselves as estimated with behavioral measures (e.g., subjects' perceptual judgements) and/or thanks to specific cognitive/computational models.

Core concepts and key steps

The first step is the computation of similarities (or dissimilarities) among distributed brain activations. In principle, any distance measure can be used (e.g., Pearson product moment correlation, Spearman rank-order correlation) as the goal is purely the description of the representational geometry of the area investigated. Neural similarity matrices can be derived from virtually any source of brain activity data: brain maps from fMRI, brain signals in M/EEG, intracranial recordings, etc…Once the neural features have been selected (e.g., set of beta maps in a given areas), their pairwise similarities are estimated (for instance, they are correlated across experimental conditions) resulting in a similarity matrix (see Fig. 40). The process can be repeated multiple times, i.e. across different spatio-temporal ROIs and/or in a searchlight fashion (Su et al., 2012).

Then, these neural representational spaces (expressed in terms of neural similarities or dissimilarities, 1-similarity) can be compared with the cognitive, psychological or perceptual space(s) that can be derived from behavioral data or modeled based on known, salient, characteristics of the stimuli. For instance, predicted similarity matrices can be built considering semantic features of the stimuli: perceptual attributes such as color and shape, as well as conceptual attributes such as the taxonomic category (see Fig 41).

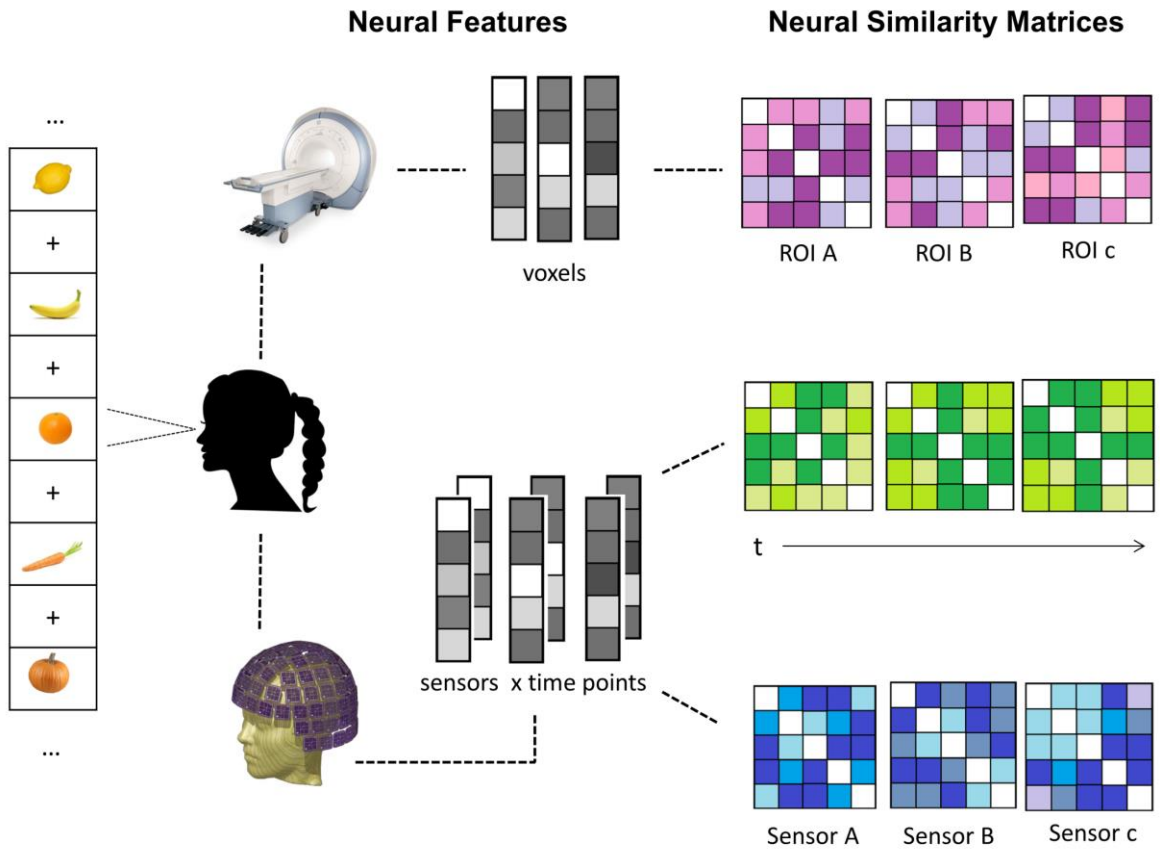


Figure 40 Schematic representation of how to build neural similarity matrices from fMRI/MEG. In the case of fMRI, neural similarity matrices are computed estimating the proximity (e.g., via correlation) among voxels, and one can obtain a similarity matrix for each region of interest (ROI). Thanks to a searchlight, one sphere can be centered on each voxel and one small ROI can be draw around each of them, leading to a whole brain map where each voxels contains the neural similarity matrix of the corresponding spherical ROI. With MEG data, we have at least two options. First, time-resolved matrices can be obtained computing similarities across all sensors at each time point, e.g., building one matrix for each time point by computing the correlation of the different experimental conditions across sensors. Second, space-resolved matrices can be obtained computing similarities across all time points for each sensor, i.e., building one matrix for each sensor by computing the correlation of the different experimental conditions across time points. As for fMRI, different ROIs can be selected and a searchlight can be used to explore a wider portion of the data. Contrary to fMRI though, the selection does not concern only the spatial information (which sensors?), but also the time (which time points?), and the spectral domain (which frequencies?).

The proximity/similarity between the two representational spaces -the neural and the predicted ones - can be assessed and compared across brain areas and across models. Note how different neural matrices, derived from very diverse techniques (imaging, behavioral measures and computational models) can be directly compared and integrated (Cichy et al., 2014), even across species (Kriegeskorte et al., 2008b), thus bridging the gap between different neuroscientific tools (Kriegeskorte et al., 2008a; Mur et al., 2013).

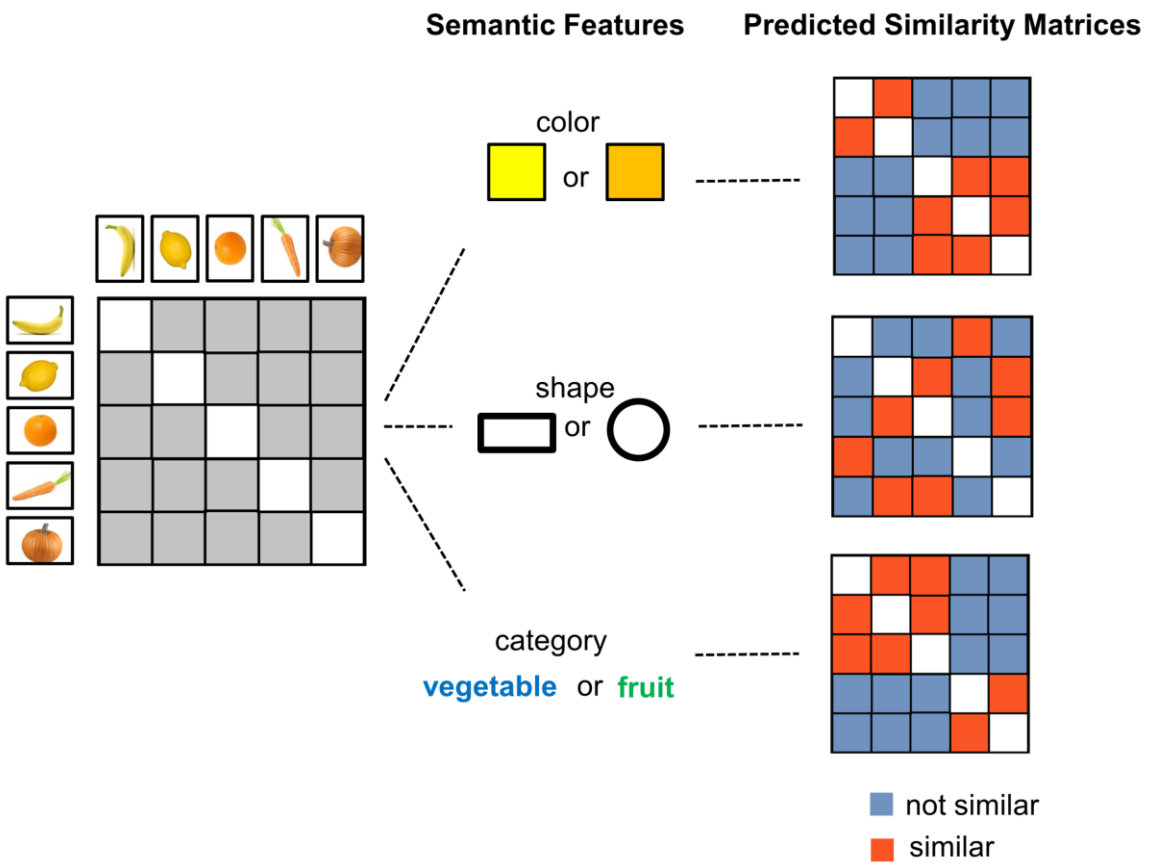


Figure 41 Schematic representation of how to build predicted similarity matrices. The chosen stimuli can be compared along different dimensions. Pairwise comparison for color will lead to a similarity matrix that expresses how similar the items are concerning that particular visual feature (upper row). Similarly, one can compare stimulus shapes (middle row) or their conceptual categorization (lower row). The three comparisons will lead to three different predicted matrices. To simplify the exposition, we are presenting fictional dichotomous situations (e.g., stimuli can be either orange OR yellow), but RSA is typically used in continuous circumstances (e.g., asking subjects to rate the "yellowness" of all the stimuli, thus building a graded representation of the color space, faithful to subject's perception.

Statistical significance of the results

Early studies relied on parametric testing, for instance the correlation scores obtained were z-scored and then tested against zero using one-sample t tests with participants as a random factor. However, as observed in the case of pattern classification, parametric testing of the null hypothesis (i.e., the correlation between the predicted and the neural matrices is zero) relies on rigid assumptions likely not met by brain imaging data.

Appropriate statistical inference can be performed, for instance, by means of randomization testing. Randomly permuting the stimulus labels (i.e., reordering rows and columns of the similarity matrices) for a sufficient number of times one can simulate the null distribution. With 1.000 permutations, if none of the permuted results

exceeds the true score, the smallest possible p-value is 0.001 (1/1000). In other words, If the correlation for correctly labeled matrices falls within the top α % of the simulated null distribution, the null hypothesis can be rejected with a p-value of α.

Ideally, one would also benefit from the estimation of how fully the model tested (i.e., the predicted matrix) explains the data, aiming at determining the amount of non-noise variance left unexplained. To this end, authors have suggested the computation of the noise ceiling whose upper bound expresses the highest accuracy any model can achieve, while the lower bound represents the minimal expected correlation of a good model (Nili et al., 2014; Khaligh-Razavi et al., 2016). The upper bound can be approximated by the correlation between the average neural matrix across all subjects and the subject specific matrices (thus it is overfitted to the single subject ones). The lower bound can be estimated thanks to a leave-one-subject-out approach, computing each single-subject matrix correlation with the average of the other subjects.

Interpretations and implications

Pattern decoding offers an insight into which information in a given brain area (at a given time point) is amenable to (linear) readout. RSA additionally sheds light into the representational geometry which, in principle, can vary across regions (and time points) even if the decoding accuracies are the same (Kriegeskorte and Kievit, 2013). For instance, it has been proven that while decoding the color of a set of stimuli from neural activity, the highest accuracy scores are reached in V1. However, in this region the representational space is dramatically different from the perceptual space: the similarity in neuronal patterns did not match the similarity in the perceptual space. Contrarily, in V4 perceptually similar colors evoked the most similar responses, if yielding a lower decoding score (Brouwer and Heeger, 2009).

History and examples

RSA has been formalized by (Kriegeskorte et al., 2008a) and it stems from antecedents such as correlation-based classification and the analyses of the geometries of representational spaces. Since (Haxby et al., 2001), the comparison of correlations has been used as a way to classify brain data as belonging to one category or the other. (Edelman et al., 1998) showed that the perceptual judgement of similarities among visually presented objects can be compared with the similarities of what they called "the voxel-space representation", i.e., the neural similarity matrix.

RSA has now been widely adopted in both fMRI and MEG settings to study cognitive representations spanning different domains (for a review see Kriegeskorte and Kievit, 2013). In fMRI, some noteworthy findings include subjects' specific idiosyncrasies in the perception of similarities among objects (Charest et al., 2014), memory consolidation (i.e., fear learning) expressed as changes in neural patterns (Visser et al., 2013), the representation of sound categories above and beyond low-level feature models (Giordano et al., 2013), and the detection of fine-grained emotional distinctions not reducible to affective primitives such as valence and arousal (Skerry and Saxe, 2015). Efforts with MEG data include the investigation of the temporal dynamic similarities during processing of visual stimuli (Wardle et al., 2016), syntactic ambiguities (Tyler et al., 2013), and communicative gestures (Redcay and Carlson, 2015).

As mentioned above, RSA offers the unique opportunity of being able to correlate neural representational distances to behavioral measures such as subjects' reaction times (e.g., Carlson et al., 2014), as well as to computational measures such as those retrieved from appropriately trained deep neural networks (e.g., Cichy et al., 2016b). Finally, fMRI and MEG data can be directly compared (Cichy et al., 2014) and used conjonitly after being combined in one metric (Cichy et al., 2016a).

Criticisms and future directions

The main issue with RSA is that the development and assessment of correct methodological choices is still undergoing. One open question concerns which class of dissimilarity measure should be preferred. Comparing classification accuracy, Euclidean/Mahalanobis distance, and Pearson correlation distance, (Walther et al., 2016) showed that (1) continuous dissimilarity measures (e.g., Euclidean distance) are substantially more reliable than classification accuracy, (2) performance can be improved by assuring that noise is isotropic or making it so by spatial pre-whitening, (3) crossvalidated Mahalanobis distance has the advantage of offering a meaningful zero point and interpretable ratios between distances. A second, related, dispute is which statistical testing should be adopted with respect to the chosen distance metric in order to avoid inflation of false positives. For instance, it has been shown that with correlations, the rate of false positives greatly increases when data are not independent and identically distributed (i.e., noise is heteroscedastic) even when permutation testing is performed (Thirion et al., 2015).

Compared to pattern classification, pattern correlation techniques offer the possibility of explicitly studying the representational space. However, they do not offer the possibility of investigating single voxel "tuning curves" (given the intrinsic multivariate nature), nor the chance of predicting how new stimuli and conditions would be represented (not being a generative model). One possible solution is to adopt a hybrid approach exploiting some of the key features of voxel-wise modeling (see next paragraphs) (Khaligh-Razavi et al., 2016; Kriegeskorte and Diedrichsen, 2016)

4.4 Pattern Encoding

We have seen that decoding (i.e., pattern classification or regression), aims at predicting the stimuli given the recorded brain activity, $f: X \rightarrow y$. The reverse direction of inference can be attempt: is it possible to predict brain activation given an accurate-enough description of the stimuli, $g: y \rightarrow X$ (see Fig. 42)? On one hand, decoding, i.e. the backward model, allows the mapping between stimuli (features) and brain activation (patterns), focusing on discriminative information, without further specification (latent features space, black box). On the other hand, encoding, i.e., the forward model, specify the principles (or intervening variables) that mediate the mapping, thus making it possible to predict the activation pattern for new stimuli (hence the name generative/predictive model).

Core concepts and key steps

Encoding models have been mostly applied to fMRI data (i.e. voxel-wise modeling), and in the following description of the method I will refer to that setting, considering voxels as the smallest units of brain recorded activity on which the analyses are carried on. However, applications to MEG data are possible, as we will see later.

First, stimuli need to be described (i.e., labeled) with the appropriate features according to the question at stake. For instance, words can be labeled according to semantic features, whether behaviorally collected or extracted from corpora. Second, in a portion of the dataset, a regularized linear regression (e.g., ridge regression) is used to estimate for each voxel the relative effect (weights) of each feature. Third, such weights are used to predict the responses to the held out data. Finally, one can compute the correlation between the predicted and the actual responses for each voxel.

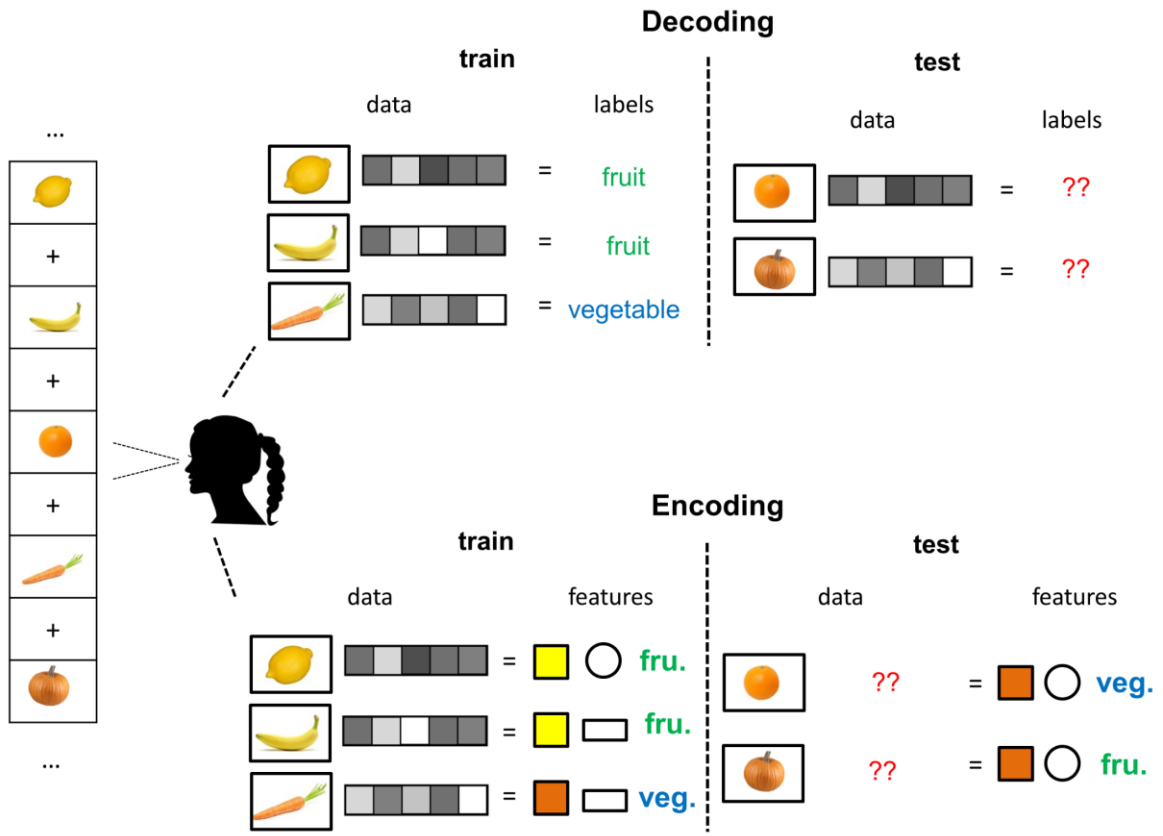


Figure 42 Difference between decoding and encoding. Both approaches start by splitting the data in a train and a test set. In the case of decoding (upper panel), what is learned during the train phase is the relation between the brain activation patterns (data) and the experimental conditions (labels). In the case of encoding (lower panel), what is learned is the relation between the provided descriptors of the stimuli (features) and the brain activation patterns (data). Note the key difference during the crucial test phase: in the decoding setting what is observed is the brain activation data and what is predicted is the experimental label, while in the encoding setting what is observed is the feature set and what is predicted are the brain activations.

Statistical significance of the results

To evaluate the performance of the encoding model (i.e., whether a voxel was predicted significantly above chance level) permutation based or bootstrap procedures can be used. Under the null hypothesis, predicted and actual responses should not be correlated.

Frequently, a decoding step (multivariate) follows the prediction step (univariate). Voxels whose responses are predicted accurately by the model can be fed to a classifier testing for discriminative information.

Interpretations and implications

It should be noticed that both the model fitting and the following prediction are done at the voxel level, which leads to two

consequences. On one hand, it allows the study of the "tuning curve" of single voxels. On the other hand, it constrains neuroscientific interpretations to voxel resolution. In a second step, the predicted responses can be fed to a classifier if one wishes to investigate the information contained in the multivoxel pattern of activity of a certain area.

The main advantage of encoding models is their generative nature. They allow prediction on brain responses to new stimuli (e.g., unseen images or words) as long as they can be described by the features the model has learned (e.g., semantic properties). This generalization power enables the study of complex stimulus sets (even under naturalistic circumstances), alleviating the cost of multiple iterations over highly selected and controlled stimuli (required when aiming at mapping broad stimulus spaces).

History and main results

The possibility of moving beyond stimulus classification and perform stimulus reconstruction was demonstrated by pivotal experiments with retinotopy (Thirion et al., 2006) and simple visual stimuli, such as geometric or alphabetic shapes (Miyawaki et al., 2008), handwritten digits (Van Gerven et al., 2010) or characters (Schoenmakers et al., 2013). Subsequently, prediction of the semantic content of word-picture pairs (Mitchell et al., 2008), as well as natural image identification (Kay et al., 2008) and reconstruction (Naselaris et al., 2009) were successfully performed. Finally, dynamic visual stimuli (i.e., short movies) were tackled in terms of their motion energy properties (Nishimoto et al., 2011) as well as semantic features (Huth et al., 2012). The latest development was the investigation of whole brain cortical responses to natural speech (Huth et al., 2016b).

Encoding models have been used to study, for instance, the categorical organization of visual areas in naturalistic circumstances, illustrating that the animate vs inanimate contrast might be more spatially smooth than previously thought (Naselaris et al., 2012). Moreover, they have shown that co-occurrence statistics in the real

world can explain the representation of complex visual scenes in visual cortex (Stansbury et al., 2013). The most compelling results is perhaps the observation of how attention can cause a tuning shift that alters the representation of the stimuli according to which category is being attended to (Cukur et al., 2013).

Finally, the voxel-wise encoding approach here presented can be adapted to an MEG setting as illustrated by (Clarke et al., 2015). Being interested in the differential contribution of low level and semantic features to the temporal dynamics of object processing, they followed the two classical steps: (1) fitting of the different models with non-regularized multiple linear regression, (2) prediction of the signal for unseen stimuli and attempt to classify them. They were able to show that performance after 200 ms significantly increases when semantic features are taken into account.

Criticisms and future directions

Two interrelated aspects of voxel-wise encoding models should be highlighted. First, by definition, the model operates (i.e., aims at fitting and predicting) at the voxel level. This means that stimuli/conditions should be modeled according to features whose combination has, presumably, an impact at the single voxel level. Such a constraint works well for perceptual representations, where low level features are known, or can be fairly easily estimated/derived from computational models. A similar modeling is way harder in case of higher order representations lacking such a detailed level of description.

Hence, the second issue concerns the interpretability of encoding results aiming at describing the cortical organization of complex naturalistic stimuli (for instance spoken or written words) on the basis of data-driven features. For an example of an attempt to recover distributed activation patterns associated with interpretable sensory-motor features see (Fernandino et al., 2015).

4.5 Discussion

Over the last twenty years, cognitive neuroscience has witnessed a progressive shift of interest from univariate activation-based approaches to multivariate information-based ones. Classically, engagement of a given brain area, at a given time, is taken as a sign of its involvement in a particular cognitive process or representation. However, current debates among cognitive theories require more in-depth descriptions of what it means for one area (or network thereof) to be recruited during a particular task. MVPA can shed light onto the information carried by distributed patterns of activity in terms of both what can be decoded from them, and which representational geometry they describe. At the beginning of the chapter I mentioned the tradeoff between temporal and spatial resolution any researcher in cognitive neuroscience needs to face. The triangle is closed by a third key issue: that of conceptual resolution[2], greatly improved by multivariate analyses. MVPA broadens the set of hypotheses that can be empirically tested, for instance, allowing the investigation of whether a certain cognitive disorder is due to impaired access or to degraded representations (e.g., Boets et al., 2013).

However, some open questions and pressing issues concern all multivariate analyses techniques here reviewed, and will now be discussed. First, when testing multiple ROIs, results should be corrected for multiple comparisons. This matter becomes particularly relevant when a searchlight is used, as the number of ROIs equals the numbers of voxels within the chosen brain mask (fMRI), and/or the number of time points and frequency bands selected (MEG). Problematic for a searchlight is also the choice of the sphere's radius, which will determine spatial, temporal, and/or frequency resolution according to the kind of data analyzed. Theoretically speaking, one should make a principled decision based on the prior on the sparsity of the representation. How broad is it reasonable to think the pattern will

[2] term adopted by Op de Beeck @ PRNI2016

be? While for some representations accurate predictions are possible (e.g., low level visual information in V1), for others there is little evidence guiding the guess of the corresponding sparsity. Moreover, when facing a comparison between different factors (e.g., testing for different semantic features), it is plausible that they will be represented at a different granularity scale. Virtually, researchers could compare the effects of many different spheres' radii, but overall a searchlight would be best considered an exploratory tool, which needs to be backed up by confirmatory tests (Etzel et al., 2013).

Second, while univariate group tests control for potential confounds (e.g., faulty randomization across subjects) as the sign of the effect would be randomly distributed across subjects, such confounds could drive the performance of MVPA group-level tests as these are usually based on single-subject summary statistics, which discards the sign/direction of the effects (Todd et al., 2013). For a discussion on how to deal with these potential confounds such as reaction time differences, see also (Woolgar et al., 2014). Overall, the heightened sensitivity of the methods calls for a careful design of the experimental setting and cautious interpretation of the results. Multivariate methods are highly opportunistic (i.e., will exploit any available bias in the data), thus extra attention should be paid to stimuli selection and randomization, tasks balance in terms of cognitive load, and any other potential source of cognitive confound.

Third, even if it is generally acknowledged that multivariate methods offer higher sensitivity than univariate ones, neurophysiological evidence suggests to lighten the conclusions on MVPA power to detect fine-scale informational content of brain regions (Kriegeskorte and Bandettini, 2007). As a matter of fact, it has been shown that effects recorded at the single cell level might be missed by MVPA: unsurprisingly, stimulus aspects that are poorly spatially clustered are intrinsically hard to decode from the BOLD response (Dubois et al., 2015). While the advantages of the multi-units analyses are clear, claims on sub-unit resolution of MVPA should be scaled down.

Fourth, comparisons of the results of univariate and multivariate analyses have been conducted. Authors suggested that univariate regional averages might denote the engagement of basic, core processing due to the task, while MVPA likely detects representational content, sub-processes differentiating across stimuli (Jimura and Poldrack, 2012). However, an interpretation of the dimensionality of the representation cannot be supported by MVPA analysis alone: a unidimensional representational space would still be appreciated only by MVPA if highly variable across subjects and highly consistent within subjects (Davis et al., 2014). The take-home message is that univariate and multivariate analyses provide complementary answers as they tackle complementary issues: the choice should depend on the cognitive hypothesis one wishes to test. Moreover, if aiming at understanding the relative contributions of multi-voxel and univariate sources of information, careful exploration of the plausible alternative explanations should be conducted (Coutanche, 2013).

Generally speaking, univariate and multivariate analyses share one core assumption: overall consistency of the functional specialization of cortical areas across subjects. This universality assumption (key to classic cognitive neuroscience as well (Caramazza and Coltheart, 2006)) is at the core of both forward (Henson, 2006) and reverse (Poldrack, 2006) inference. In an information-based setting, as feature correspondence across brains is virtually impossible, usually new multivariate models are fitted and tested for each subject's brain. Aiming at building a general model of the representational space, relying on a common set of response-tuning functions, a new method (called hyperalignment), has been introduced. It aligns patterns of neural responses across subjects into a common, high-dimensional space in a given ROI (Haxby et al., 2011) or whole brain thanks to a searchlight procedure (Guntupalli et al., 2016). Instead of working on a common cortical topography, one is thus dealing with a common representational space.

It should be stressed that none of the methods here presented permits causal inferences: they can, at best, suggest a representational function (for a Bayesian account of the causal power of multivariate analyses, according to the direction of the inference and the general setting, see Weichwald et al., 2015). Only lesion studies (whether real ones in patients or virtual ones temporarily simulated with neuromodulation techniques[3]) can demonstrate whether a given brain region is causally involved in a cognitive process/representation (and when so). Moreover, even if attention is shifted from univariate data points to multivariate patterns, basic shortcomings of fMRI and MEG apply to MVPA setting as well. It would be impossible to appreciate any representation coded in ways that do not relate to BOLD responses or to detectable magnetic effects. As an example consider temporally demanding coding schemes such as burstiness coding or synchronous firing, and representations coded as within (e.g. differential weights in membrane potentials) or across (e.g., functional connectivity) neuron changes (for instance compare synaptic and connectivity accounts of working memory (Mongillo et al., 2008; Stokes, 2015)). The mirror observation holds as well, given the observation that stable representations are possible despite activity variations: not every change in neuronal activity corresponds to a change in the stimulus representation (Druckmann and Chklovskii, 2012).

I have briefly mentioned that from a neuro-cognitive point of view, the ideal method would combine the main advantages of encoding and RSA. On one hand, a generative model would allow generalization to new stimuli and study of tuning curves. On the other hand, the higher order summary statistics provided by RSA enables

[3] Non-invasive brain stimulation techniques include: transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), transcranial alternating current stimulation (tACS), and transcranial random noise stimulation (tRNS).

the comparison of geometries across different models. Current developments of RSA such as probabilistic RSA (Kriegeskorte and Diedrichsen, 2016) and mixed RSA (Khaligh-Razavi et al., 2016) are going precisely in this direction. The need for explicit models of the representational spaces, such as those provided by RSA and encoding, has been recently highlighted by (Naselaris and Kay, 2015) while reviewing the three kinds of ambiguities faced by MVPA research. First, geometrical ambiguity is due to the fact that the activity patterns (multivariate vectors) can be discriminated thanks to a difference in length (overall activation, detected by univariate analyses too) or orientation (actual geometry of the representation, captured only by MVPA). As previously stressed, this observation calls for in-depth comparison of univariate and multivariate results before theoretical conclusions are drawn. Second, spatial ambiguity is linked with the dangerous interpretation of model weights and to the shortcoming of searchlight analyses detailed earlier. This opacity will be minimal in those settings where opposite theories, making clear topographical predictions, are directly compared, especially if ROIs can be defined via functional localizers. Third, representational ambiguity originates from the difficulty to establish which features of the stimuli are driving the performance of the multi-variate model. Highly controlled experiments can help ensure alternative features do not correlate with the one under investigation; however performance of multivariate methods relying on latent feature representations (i.e., decoding) could still be heavily biased. A possible solution is to test and compare explicit models of the representations via encoding or RSA. While first critical observations focused on the importance for multivariate models to be biologically plausible (for instance advocating for the use of linear models (DiCarlo and Cox, 2007)), current theoretical remarks highlight the need for psychologically plausible models: to evaluate multivariate results and fruitfully exploit them to understand the mind-brain link, one needs to observe a link with behavioral performance (Williams et al., 2007; Ritchie and Carlson, 2016) and study how multiple channels (not just BOLD-fMRI) interact.

Finally, two topics deserve to be highlighted: the concept of information and the metaphor of the brain as its own decoder. From an information theory point of view, we as researchers are not on the receiving end of the flux of information within the brain, other brain areas are (de-Wit et al., 2016). It should not be taken for granted that the pattern of activity that enables successful classification (i.e., we are the receivers) is actually used by the rest of the brain to perform a task or represent a stimulus (i.e., the cortex is the receiver).

Moreover, the frequent assimilation of the brain to a decoder, leads to stimulating questions (many of which already spelled out in (King and Dehaene, 2014): as the brain has all the information available at the same time, should we give up localization attempts? If not, how spread out should the population of interest be? Does the brain suffer from overfitting issues, and if so, does it use regularization or penalization schemes? Does the brain suffer from a "curse of dimensionality", and if so, how are feature selected / how many data points are needed? Does the information need to be explicitly read out, such that linearity becomes a biological constraint? Nonlinear methods are spreading, for instance one rapidly rising approach is the integration of neuroscience and representation-learning methods such as deep learning (i.e., fed with raw data, the machine automatically discover the representations needed for classification, thanks to multiple non-linear modules that transform the initial input into progressively more abstract levels (LeCun et al., 2015)). This perspective might be useful in deepening our understanding of information processing in the brain (Kriegeskorte, 2015; Marblestone et al., 2016), especially in the case of sensory cortices (Eickenberg et al., 2016; Yamins and DiCarlo, 2016).

To conclude, we have seen that there are different types of pattern-information techniques (supported by different mathematical frameworks), but their neuroscientific implications are overlapping. The direction of the model (i.e., from the brain response to the

stimulus set or vice-versa) is irrelevant: what the success of these methods is indicating is a statistical dependency (i.e. the presence of mutual information) between stimuli and brain response pattern. The differences across methods and studies, to be kept in mind while evaluating their conclusions and the implications, are:

1. the degree of generalization (i.e., are there implications for novel stimuli or a different types of mental states?);
2. the stimulus-space complexity (i.e., does the method scale to higher dimensional feature spaces?);
3. how explicit is the description of the relation linking brain activity and stimulus features.

5. Conclusions

The behavioral and neuroimaging methods here presented are a non-exhaustive excursus of the methodologies available to contemporary cognitive neuroscientists, all of which have specific advantages and disadvantages. Notwithstanding the constant technical and mathematical progress that is pushing them forward, it is likely that no method will ever be intrinsically either better or sufficient on its own to explore the neural substrate of semantic knowledge (or any other high order cognitive function). The election of one method over the others should always be based on the theoretical questions one wishes to answer, tailoring any operational and statistical choice on the variables at stake. Overall, behavioral and neuropsychological data will keep providing fundamental insights on the mind-brain relation, suggesting new theories. Neuroanatomical data will contribute with additional, critical biological constraints. Finally, computational models along with all imaging techniques will serve to tests the hypothesis derived from the formal theories.

Concerning neuroimaging research, recently general methodological concerns have been highlighted and some

solutions/good practiced suggested: low statistical power, uncontrolled analytic flexibility, multiple comparison issues, potential software errors, insufficient study reporting and lack of independent replications (Poldrack et al., 2016). It is worth noticing that these matters apply to all kinds of neuroscientific research and ultimately to science in general. Similarly, the reproducibility crisis in psychological sciences emphasized last year (Open Science, 2015) concerns many (all?) fields, not only experimental psychology. Overall, authors are manifesting the need to change the reward system and the scientific culture at large (Wiener et al., 2016), encouraging practices of data and code sharing, open review and open access publishing. As a minimal example of how this would greatly benefit the field, consider that sharing data and code would not only ease reproducibility, but also improve overall quality, as software is not flawless and debugging would be greatly strengthened by independent iterations over the code (Eklund, 2016).

Progress will come from the interplay of data-driven studies with naturalistic stimuli and theory-driven studies with controlled stimuli. The first ones, by spanning vast representational spaces, will provide us with observations crucial to develop new theories; the second ones, will afford the opportunity to test explicit hypotheses on the geometries of those spaces. It is important to notice that purely mapping approaches, aiming at describing connections and activations devoid of the inferential power that come from comparing different predictions, would be useless in the study of the mind-brain relation. As recently illustrated (Jonas and Kording, 2016), even knowing all the details of the hardware (neural level), the software (cognitive level) would represent a mystery if neuroscientific methods are applied blindly and the correct questions (testing appropriate theories) are not asked. Throughout this chapter, I have dealt with the question of how we can investigate neural and cognitive representations. Attempting to answer such a query one faces a deeper related issue: what does it mean to understand a system such as the brain? More or

less explicitly, I adopted the interpretation of understanding a system as "*being able to fix it*". A completely different perspective is that of understanding as "*being able to reproduce*". Do you recall our broken television? As a cognitive neuropsychologist, I mentioned my interest in knowing what, where and when it was broken, aiming at fixing it. A computer scientist interested in artificial intelligence would perhaps aim at replicating the properties of the systems (more than knowing how to improve them in case of flaws). Consequentially, he/she would be happy with any model succeeding in imitating the performance of the original system, irrespective of whether it provides an explicit answer to the key questions we explored in the Chap. 1: what, where, when and how.

Given the complexity of the hypotheses tested by this thesis, work has been conducted along the three different axes here detailed:

- behavioral testing in order to deepen our understanding of the perceptual and conceptual dimensions organizing the cognitive semantic space, and to test the automaticity of their access (Chap. 3);
- an fMRI experiment to shed light onto the neural topographical organization of the different semantic dimensions investigated (Chap. 4);
- an MEG experiment to explore the temporal dynamics of their activation (Chap. 5).

Bibliography

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480-1494.

Andres M, Finocchiaro C, Buiatti M, Piazza M (2015) Contribution of motor representations to action verb processing. Cognition 134:174-184.

Ansorge U, Reynvoet B, Hendler J, Oettl L, Evert S (2013) Conditional automaticity in subliminal morphosyntactic priming. Psychol Res 77:399-421.

Balota DA, Yap MJ, Cortese MJ, Watson JM (2008) Beyond mean response latency: Response time distributional analyses of semantic priming. Journal of Memory and Language 59:495-523.

Becker S, Moscovitch M, Behrmann M, Joordens S (1997) Long-term semantic priming: a computational account and empirical evidence. Journal of Experimental Psychology: Learning, Memory, and Cognition 23.

Bekhti Y, Strohmeier D, Jas M, Badeau R, Gramfort A (2016) M/EEG source localization with multi-scale time-frequency dictionaries. In: 6th International Workshop on Pattern Recognition in Neuroimaging.

Bennett CM, Wolford GL, Miller MB (2009) The principled control of false positives in neuroimaging. Social cognitive and affective neuroscience 4:417-422.

Berger H (1929) Über das elektrenkephalogramm des menschen. European Archives of Psychiatry and Clinical Neuroscience 87:527.

Blankertz B, Dornhege G, Krauledat M, Muller KR, Curio G (2007) The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects. Neuroimage 37:539-550.

Boets B, Op de Beeck HP, Vandermosten M, Scott SK, Gillebert CR, Mantini D, Bulthe J, Sunaert S, Wouters J, Ghesquiere P (2013) Intact but less accessible phonetic representations in adults with dyslexia. Science 342:1251-1254.

Borogovac A, Asllani I (2012) Arterial Spin Labeling (ASL) fMRI: advantages, theoretical constrains, and experimental challenges in neurosciences. International journal of biomedical imaging 2012:818456.

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory 144-152.

Buchweitz A, Shinkareva SV, Mason RA, Mitchell TM, Just MA (2012) Identifying bilingual semantic neural representations across languages. Brain Lang 120:282-289.

Buiatti M, Pena M, Dehaene-Lambertz G (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage 44:509-519.

Caramazza A, Coltheart M (2006) Cognitive Neuropsychology twenty years on. Cognitive neuropsychology 23:3-12.

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. Journal of vision 11.

Carlson TA, Ritchie JB, Kriegeskorte N, Durvasula S, Ma J (2014) Reaction time for object categorization is predicted by representational distance. J Cogn Neurosci 26:132-142.

Carp J (2012a) On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. Frontiers in neuroscience 6:149.

Carp J (2012b) The secret lives of experiments: methods reporting in the fMRI literature. Neuroimage 63:289-300.

Charest I, Kievit RA, Schmitz TW, Deca D, Kriegeskorte N (2014) Unique semantic space in the brain of each beholder predicts perceived similarity. Proceedings of the National Academy of Sciences of the United States of America 111:14565-14570.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nature neuroscience 17:455-462.

Cichy RM, Ramirez FM, Pantazis D (2015) Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? Neuroimage 121:193-204.

Cichy RM, Pantazis D, Oliva A (2016a) Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. Cereb Cortex 26:3563-3579.

Cichy RM, Khosla A, Pantazis D, Oliva A (2016b) Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. Neuroimage.

Clarke A, Devereux BJ, Randall B, Tyler LK (2015) Predicting the Time Course of Individual Objects with MEG. Cereb Cortex 25:3602-3612.

Cohen MX (2014) Analyzing neural time series data: theory and practice: MIT Press.

Coombs CH (1954) A method for the study of interstimulus similarity. Psychometrika 19:183-194.

Cortes C, Vapnik V (1995) Support-vector networks. Machine learning 20:273-297.

Coutanche MN (2013) Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? Cognitive, affective & behavioral neuroscience 13:667-673.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19:261-270.

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

Cukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nature neuroscience 16:763-770.

Damian MF (2000) Semantic negative priming in picture categorization and naming. Cognition 76:B45-B55.

Damian MF (2001) Congruity effects evoked by subliminally presented primes: automaticity rather than semantic processing. Journal of Experimental Psychology: Human Perception and Performance 27.

Davis T, Poldrack RA (2013) Measuring neural representations with fMRI: practices and pitfalls. Annals of the New York Academy of Sciences 1296:108-134.

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA (2014) What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage 97:271-283.

de-Wit L, Alexander D, Ekroll V, Wagemans J (2016) Is neuroimaging measuring information in the brain? Psychonomic bulletin & review.

De Groot AM (1983) The range of automatic spreading activation in word priming. Journal of verbal learning and verbal behavior 22:417-436.

Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. Trends Cogn Sci 15:254-262.

Detre JA, Wang J (2002) Technical aspects and utility of fMRI using BOLD and ASL. Clinical Neurophysiology 113:621-634.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333-341.

Druckmann S, Chklovskii DB (2012) Neuronal circuits underlying persistent representations despite time varying activity. Current biology : CB 22:2095-2103.

Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:2791-2802.

Edelman S, Grill-Spector K, Kushnir T, Malach R (1998) Toward direct visualization of the internal shape representation space by fMRI. Psychobiology 26:309-321.

Eger E, Pinel P, Dehaene S, Kleinschmidt A (2015) Spatially invariant coding of numerical information in functionally defined subregions of human parietal cortex. Cereb Cortex 25:1319-1329.

Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G (2008) fMRI activity patterns in human LOC carry information about object exemplars within category. Journal of cognitive neuroscience 20:356-370.

Eger E, Michel V, Thirion B, Amadon A, Dehaene S, Kleinschmidt A (2009) Deciphering cortical number coding from human brain activity patterns. Current biology : CB 19:1608-1615.

Eickenberg M, Gramfort A, Varoquaux G, Thirion B (2016) Seeing it all: Convolutional network layers map the function of the human visual system. Neuroimage.

Eklund A, Nichols, T.E. and Knutsson, H., (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences:201602413.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature Reviews Neuroscience 392:598-601.

Etzel JA, Zacks JM, Braver TS (2013) Searchlight analysis: promise, pitfalls, and potential. Neuroimage 78:261-269.

Fan J, McCandliss BD, Fossella J, Flombaum JI, Posner MI (2005) The activation of attentional networks. Neuroimage 26:471-479.

Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: defining ROIs functionally in individual subjects. Journal of neurophysiology 104:1177-1194.

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PloS one 5:e15710.

Fernandino L, Humphries CJ, Seidenberg MS, Gross WL, Conant LL, Binder JR (2015) Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia 76:17–26.

Fischler I (1977) Semantic facilitation without association in a lexical decision task. Memory & cognition 5:335-339.

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of eugenics 7:179-188.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995) Analysis of fMRI time-series revisited. Neuroimage 2:45-53.

Gainotti G, Ciaraffa F, Silveri MC, Marra C (2009) Mental representation of normal subjects about the sources of knowledge in different semantic categories and unique entities. Neuropsychology 23:803-812.

Garrard P, Lambon Ralph MA, Hodges JR, Patterson K (2001) Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. Cognitive neuropsychology 18:125-174.

Geukes S, Huster RJ, Wollbrink A, Junghofer M, Zwitserlood P, Dobel C (2013) A large N400 but no BOLD effect--comparing source activations of semantic priming in simultaneous EEG-fMRI. PloS one 8:e84029.

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P (2013) Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex 23:2025-2037.

Goldstone RL, Medin DL, Halberstadt J (1997) Similarity in context. Memory & Cognition 25:237-255.

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14-23.

Grisoni L, Dreyer FR, Pulvermuller F (2016) Somatotopic Semantic Priming and Prediction in the Motor System. Cereb Cortex 26:2353-2366.

Grootswagers T, Wardle SG, Carlson TA (2016) Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. J Cogn Neurosci:1-21.

Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R, Parkkonen L, Taylor JR, van Wassenhove V, Wibral M, Schoffelen JM (2013) Good practice for conducting and reporting MEG research. Neuroimage 65:349-363.

Gruber T, Muller MM (2005) Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. Cereb Cortex 15:109-116.

Guntupalli JS, Hanke M, Halchenko YO, Connolly AC, Ramadge PJ, Haxby JV (2016) A Model of Representational Spaces in Human Cortex. Cereb Cortex 26:2919-2934.

Hansen P, Kringelbach M, Salmelin R (2010) MEG: an introduction to methods. : Oxford university press.

Harvey BM, Fracasso A, Petridou N, Dumoulin SO (2015) Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. Proceedings of the National Academy of Sciences of the United States of America 112:13525-13530.

Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, Biessmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87:96-110.

Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. Annual Review of Neuroscience 37:435-456.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72:404-416.

Haynes JD (2015) A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron 87:257-270.

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature neuroscience 8:686-691.

Henson R (2006) Forward inference using functional neuroimaging: dissociations versus associations. Trends Cogn Sci 10:64-69.

Henson RNA, Rugg MD (2003) Neural response suppression, haemodynamic repetition effects, and behavioural priming. Neuropsychologia 41:263-270.

Hoffman P, Lambon Ralph MA (2013) Shapes, scents and sounds: quantifying the full multi-sensory basis of conceptual knowledge. Neuropsychologia 51:14-25.

Holcomb PJ, Neville HJ (1990) Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. Language and cognitive processes 5:281-312.

Hoyos-Idrobo A, Schwartz Y, Varoquaux G, Thirion B (2015) Improving sparse recovery on structured images with bagged clustering. In: 5th International Workshop on Pattern Recognition in Neuroimaging.

Hutchison KA, Balota DA, Cortese MJ, Watson JM (2008) Predicting semantic priming at the item level. Q J Exp Psychol (Hove) 61:1036-1066.

Huth A, Lee T, Nishimoto S, Bilenko N, Vu A, Gallant J (2016a) Decoding the semantic content of natural movies from human brain activity. Frontiers in systems neuroscience 10.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210-1224.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016b) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453-458.

Janssen N, Hernández-Cabrera JA, Ezama Foronda L (2016).

Jas M, Engemann D, Raimondo F, Bekhti Y, Gramfort A (2016) Automated rejection and repair of bad trials in MEG/EEG. . In: 6th International Workshop on Pattern Recognition in Neuroimaging.

Jimura K, Poldrack RA (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50:544-552.

Jonas E, Kording K (2016) Could a neuroscientist understand a microprocessor? bioRxiv, p059188.

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nature neuroscience 8:679-685.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of neuroscience 17:4302-4311.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.

Keil A, Debener S, Gratton G, Junghofer M, Kappenman ES, Luck SJ, Luu P, Miller GA, Yee CM (2014) Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology 51:1-21.

Khaligh-Razavi S-M, Henriksson L, Kay K, Kriegeskorte N (2016) Fixed versus mixed RSA: Explaining visual representations by fixed and mixed feature sets from shallow and deep computational models. bioRxiv, p059188 009936.

King J-R, Pescetelli N, Dehaene S (2016) Selective maintenance mechanisms of seen and unseen sensory features in the human brain.

King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203-210.

King JR, Faugeras F, Gramfort A, Schurger A, El Karoui I, Sitt JD, Rohaut B, Wacongne C, Labyt E, Bekinschtein T, Cohen L, Naccache L, Dehaene S (2013) Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. Neuroimage 83:726-738.

Knappe S, Sander T, Trahms L (2014) Optically-Pumped Magnetometers for MEG. Magnetoencephalography:993-999.

Knops A, Thirion B, Hubbard EM, Michel V, Dehaene S (2009) Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. Science 324:1583-1585.

Knops A, Piazza M, Sengupta R, Eger E, Melcher D (2014) A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:9857-9866.

Kosem A, Gramfort A, van Wassenhove V (2014) Encoding of event timing in the phase of neural oscillations. Neuroimage 92:274-284.

Kouider S, Dehaene S, Jobert A, Le Bihan D (2007) Cerebral bases of subliminal and supraliminal priming during reading. Cereb Cortex 17:2019-2029.

Kriegeskorte N (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annual Review of Vision Science 1:417-446.

Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage 38:649-662.

Kriegeskorte N, Mur M (2012) Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. Frontiers in psychology 3:245.

Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci 17:401-412.

Kriegeskorte N, Diedrichsen J (2016) Inferring brain-computational mechanisms with models of activity measurements. Philosophical transactions of the Royal Society of London Series B, Biological sciences 371.

Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in systems neuroscience 2:4.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nature neuroscience 12:535-540.

Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E (2010) Everything you never wanted to know about circular analysis, but were afraid to ask. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 30:1551-1557.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126-1141.

Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics:(de) constructing the N400. Nature Reviews Neuroscience 9:920-933.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436-444.

Lewis LD, Setsompop K, Rosen BR, Polimeni JR (2016) Fast fMRI can detect oscillatory neural activity in humans. Proceedings of the National Academy of Sciences of the United States of America.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature Reviews Neuroscience 412:150-157.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences 92:8135-8139.

Marblestone AH, Wayne G, Kording KP (2016) Toward an Integration of Deep Learning and Neuroscience. Frontiers in computational neuroscience 10:94.

Marcel AJ (1983) Conscious and unconscious perception: Experiments on visual masking and word recognition. Cognitive psychology 15:197-237.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. Journal of neuroscience methods 164:177-190.

Masson ME (1995) A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 21.

Mayr S, Buchner A (2007) Negative Priming as a Memory Phenomenon. Zeitschrift für Psychologie / Journal of Psychology 215:35-51.

McRae K, Cree GS, Seidenberg MS, McNorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. Behavior research methods 37:547-559.

Meyer DE, Schvaneveldt RW (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology 90:227-234.

Miltner WH, Braun C, Arnold M, Witte H, Taub E (1999) Coherence of gamma-band EEG activity as a basis for associative learning. Nature 397:434-436.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191-1195.

Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, Sadato N, Kamitani Y (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60:915-929.

Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Ugurbil K (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 63:1144-1153.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543-1546.

Mur M, Meys M, Bodurka J, Goebel R, Bandettini PA, Kriegeskorte N (2013) Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. Frontiers in psychology 4:128.

Naselaris T, Kay KN (2015) Resolving Ambiguities of MVPA Using Explicit Models of Representation. Trends in Cognitive Sciences 19:551-554.

Naselaris T, Stansbury DE, Gallant JL (2012) Cortical representation of animate and inanimate objects in complex natural scenes. Journal of physiology, Paris 106:239-249.

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400-410.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902-915.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping 15:1-25.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. PLoS Comput Biol 10.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Current biology : CB 21:1641-1646.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424-430.

Ochsner KN, Bunge SA, Gross JJ, Gabrieli JD (2002) Rethinking feelings: an FMRI study of the cognitive regulation of emotion. Journal of cognitive neuroscience 14:1215-1229.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences 87:9868-9872.

Open Science C (2015) PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 349:aac4716.

Palva S, Palva JM (2012) Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. Trends Cogn Sci 16:219-230.

Pauli R, Bowring A, Reynolds R, Chen G, Nichols TE, Maumet C (2016) Exploring fMRI Results Space: 31 Variants of an fMRI Analysis in AFNI, FSL, and SPM. Frontiers in neuroinformatics 10:24.

Pedregosa F, Eickenberg M, Ciuciu P, Thirion B, Gramfort A (2015) Data-driven HRF estimation for encoding and decoding models. NeuroImage 104:209-220.

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199-209.

Pfurtscheller G, Da Silva FL (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology 110:1842-1857.

Piazza M, Izard V, Pinel P, Le Bihan D, Dehaene S (2004) Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron 44:547-555.

Poldrack R, Baker CI, Durnez J, Gorgolewski K, Matthews PM, Munafo M, Nichols T, Poline J-B, Vul E, Yarkoni T (2016) bioRxiv, p059188.

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10:59-63.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nature Reviews Neuroscience 1:125-132.

Ramkumar P, Jas M, Pannasch S, Hari R, Parkkonen L (2013) Feature-specific information processing precedes concerted activation in human visual cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:7691-7699.

Rao RB, Fung G, Rosales R (2008) On the Dangers of Cross-Validation. An Experimental Evaluation. SIAM International Conference on Data Mining: 588-596.

Redcay E, Carlson TA (2015) Rapid neural discrimination of communicative gestures. Social cognitive and affective neuroscience 10:545-551.

Ritchie JB, Carlson TA (2016) Neural Decoding and "Inner" Psychophysics: A Distance-to-Bound Approach for Linking Mind, Brain, and Behavior. Frontiers in neuroscience 10:190.

Salmelin R (2007) Clinical neurophysiology of language: The MEG approach. Clinical Neurophysiology 118:237-254.

Schoenmakers S, Barth M, Heskes T, van Gerven M (2013) Linear reconstruction of perceived images from human brain activity. NeuroImage 83:951-961.

Schreiber K, Krekelberg B (2013) The statistical analysis of multi-voxel patterns in functional imaging. PloS one 8:e69328.

Schweinberger SR, Huddy V, Burton AM (2004) N250r: a face-selective brain response to stimulus repetitions. NeuroReport 15:1501-1505.

Serences JT, Saproo S (2012) Computational advances towards linking BOLD and behavior. Neuropsychologia 50:435-446.

Shah AS, Bressler SL, Knuth KH, Ding M, Mehta AD, Ulbert I, Schroeder CE (2004) Neural dynamics and the fundamental mechanisms of event-related brain potentials. Cereb Cortex 14:476-483.

Shepard RN (1980) Multidimensional scaling, tree-fitting, and clustering. . Science 210:390-398.

Skerry AE, Saxe R (2015) Neural representations of emotion are organized around abstract event features. Current biology : CB 25:1945-1954.

Stansbury DE, Naselaris T, Gallant JL (2013) Natural scene statistics account for the representation of scene categories in human visual cortex. Neuron 79:1025-1034.

Stelzer J, Chen Y, Turner R (2013) Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage 65:69-82.

Stokes MG (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn Sci 19:394-405.

Stokes MG, Wolff MJ, Spaak E (2015) Decoding Rich Spatial Information with High Temporal Resolution. Trends Cogn Sci 19:636-638.

Su L, Fonteneau E, Marslen-Wilson W, Kriegeskorte N (2012) Spatiotemporal Searchlight Representational Similarity Analysis in EMEG Source Space.97-100.

Swinney DA, Onifer W, Prather P, Hirshkowitz M (1979) Semantic facilitation across sensory modalities in the processing of individual words and sentences. Memory & Cognition 7:159-165.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends in cognitive sciences 3:151-162.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. The Journal of Neuroscience 16:4240-4249.

Thirion B, Pedregosa F, Eickenberg M, Varoquaux G (2015) Correlations of correlations are not reliable statistics: implications for multivariate pattern analysis. In: ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins, ed).

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S (2006) Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33:1104-1116.

Tipper SP, Driver J (1988) Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. Memory & Cognition 16:64-70.

Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: Theory and rule representation case study. Neuroimage 77:157-165.

Tong F, Pratte MS (2012) Decoding patterns of human brain activity. Annual review of psychology 63:483-509.

Torgerson WS (1965) Multidimensional scaling of similarity. Psychometrika 30:379-393.

Tranel D, Logan CG, Frank RJ, Damasio AR (1997) Explaining category-related effects in the retrieval of conceptual and lexical knowledge for concrete entities: Operationalization and analysis of factors. Neuropsychologia 35:1329-1339.

Tucciarelli R, Turella L, Oosterhof NN, Weisz N, Lingnau A (2015) MEG Multivariate Analysis Reveals Early Abstract Action Representations in the Lateral Occipitotemporal Cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:16034-16045.

Tversky A (1977) Features of similarity. Psychological review 84.

Tyler LK, Cheung TP, Devereux BJ, Clarke A (2013) Syntactic computations in the language network: characterizing dynamic network properties using representational similarity analysis. Frontiers in psychology 4:271.

Van Gerven MA, De Lange FP, Heskes T (2010) Neural decoding with hierarchical generative models. Neural computation 22:3127-3142.

Varoquaux G, Raamana P, Engemann D, Hoyos-Idrobo A, Schwartz Y, Thirion B (2016) Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. . arXiv 1606.05201.

Visser RM, Scholte HS, Beemsterboer T, Kindt M (2013) Neural pattern similarity predicts long-term fear memory. Nature neuroscience 16:388-390.

Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nature neuroscience 5:491-499.

Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. Perspectives on psychological science : a journal of the Association for Psychological Science 4:274-290.

Waldert S, Preissl H, Demandt E, Braun C, Birbaumer N, Aertsen A, Mehring C (2008) Hand movement direction decoded from MEG and EEG. The Journal of neuroscience : the official journal of the Society for Neuroscience 28:1000-1008.

Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137:188-200.

Wardle SG, Kriegeskorte N, Grootswagers T, Khaligh-Razavi SM, Carlson TA (2016) Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. Neuroimage 132:59-70.

Weber M, Thompson-Schill SL, Osherson D, Haxby J, Parsons L (2009) Predicting judged similarity of natural categories from their neural representations. Neuropsychologia 47:859-868.

Weichwald S, Meyer T, Ozdenizci O, Scholkopf B, Ball T, Grosse-Wentrup M (2015) Causal interpretation rules for encoding and decoding models in neuroimaging. Neuroimage 110:48-59.

Wheatley T, Weisberg J, Beauchamp MS, Martin A (2005) Automatic priming of semantically related words reduces activity in the fusiform gyrus. . Journal of Cognitive Neuroscience 17:1871-1885.

Wiener M, Sommer FT, Ives ZG, Poldrack RA, Litt B (2016) Enabling an Open Data Ecosystem for the Neurosciences. Neuron 92:617-621.

Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. Nature neuroscience 10:685-686.

Wilson LB, Tregellas JR, Slason E, Pasko BE, Rojas DC (2011) Implicit phonological priming during visual word recognition. Neuroimage 55:724-731.

Witt JK, Kemmerer D, Linkenauger SA, Culham J (2010) A functional role for motor simulation in identifying tools. Psychol Sci 21:1215-1219.

Woolgar A, Golland P, Bode S (2014) Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. Neuroimage 98:506-512.

Woolgar A, Jackson J, Duncan J (2016) Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. J Cogn Neurosci 28:1433-1454.

Worsley KJ, Friston KJ (1995) Analysis of fMRI time-series revisited—again. Neuroimage 2:173-181.

Yamins DL, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience 19:356-365.

Yee E, Chrysikou EG, Hoffman E, Thompson-Schill SL (2013) Manual experience shapes object representations., p.0956797612464658. Psychological science.

CHAPTER 3:
BEHAVIORAL EVIDENCES OF MULTIVARIATE SEMANTIC REPRESENTATIONS

All animals are equal, but some animals are more equal than others.
[George Orwell, 1945]

In this chapter I illustrate the outcome of the behavioral experiments I ran. First, I present the results of the Semantic Distance Judgment (SDJ) and the Semantic Features Listing (SFL) experiments. Notably, they were instrumental to the definition of the cognitive semantic space of our volunteers (i.e., French and Italian native speakers), validating the selection of the stimuli for our following neuroimaging experiments. Second, I report the results of a series of priming experiments aiming at elucidating the degree of automaticity of the retrieval of different semantic dimensions.

Highlights:

- Word meanings lie in a multidimensional semantic space describing how close concepts are to each other.
- Semantic distance judgments are stable in time and consistent across subjects.
- Different behavioral measures lead to the description of similar multidimensional semantic spaces.
- Linguistic databases metrics provide converging accounts of the relation (distance) between concepts.
- Perceptual semantic priming appears to be possible, yet greatly interacts with tasks' characteristics.

The behavioral experiments here detailed had a two-fold objective. First, we aimed at exploring Semantic Distance Judgment and Semantic Features Listing as tools to investigate the geometry of the cognitive semantic space. Would these two different ways of assessing relations between concepts lead to the reconstruction of the same representational space? The second aim was to validate the stimuli for our following neuroimaging experiments. For pragmatic reasons (not theoretical ones), the first imaging experiment (see Chap. 4), was going to be conducted with Italian mother tongue participants, while our second one (see Chap. 5) with French participants. Hence, we selected, validated, and deeply analyzed stimuli in both languages.

Moreover, we tailored the pre-selection of the stimuli to the main goals of the two different imaging experiments envisaged: compare nested semantic classification in the first case, control for perceptual semantic dimensions in the second one. Unless otherwise specified, all analyses were run with Matlab (https://www.mathworks.com/products/matlab).

1. Study 1

The first neuroimaging study, intended as a follow up of a preceding EEG study run in Italy (Buiatti et al., 2012), was going to be an fMRI study with Italian mother tongue speakers. The main goal was to shed light onto the neural correlates of semantic distance (i.e., how close/far words are in the semantic space) across and within categories.

1.1 Stimuli Selection

We selected 24 Italian words belonging to two different semantic categories: animal (12 words) and tool (12 words). Moreover, inside each category, we carefully chose words which introspectively fell in different sub-clusters: some stimuli referred to domesticated animals (e.g., *cow*), while some to wild ones (e.g. *giraffe*); some to weapons (e.g., *spear*), and some to tools (e.g., *hammer*). Stimuli also differ on many perceptual semantic dimensions, for instance words referred to rather big (e.g., *whale* and *sword*) or rather small (e.g., *shrimp* and *nail*) items. However, as this study initially focused on higher order taxonomical classification, we did not explicitly control for these perceptual dimensions. All stimuli are listed in Table 1.

Words (EN)	Words (IT)	Length (n# letters)	Implied Real World Size	Category	Cluster
whale	balena	6	24	animal	1
dolphin	delfino	7	22	animal	1
seal	foca	4	18	animal	1
octupus	polipo	6	12	animal	2
squid	calamaro	8	11	animal	2
shrimp	gambero	7	4	animal	2
giraffe	giraffa	7	23	animal	3
camel	cammello	8	21	animal	3
zebra	zebra	5	19	animal	3
cow	mucca	5	20	animal	4
sheep	pecora	6	16	animal	4
goat	capra	5	17	animal	4
spear	lancia	6	15	tool	5
saber	sciabola	8	14	tool	5
sword	spada	5	13	tool	5
nail	chiodo	6	1	tool	6
hammer	martello	8	10	tool	6
pincer	tenaglia	8	9	tool	6
brush	spazzola	8	8	tool	7
comb	pettine	7	7	tool	7
hairpin	forcina	7	2	tool	7
pastel	pastello	8	5	tool	8
pencil	matita	6	6	tool	8
pencil sharpener	temperino	9	3	tool	8

Table 1 Stimuli used for feature listing and distance rating in Study 1. Stimuli were pre-selected by the authors as to span two semantic categories, each of which could be subdivided in four semantic clusters.

1.2 Stimuli Psycholinguistic Validation

First, we ensured that differences across semantic categories and clusters were not correlated with differences in psycholinguistic variables known to influence word processing, such as number of letters, number of syllables, gender, accent, and frequency of use (retrieved from *Corpus e Lessico di Frequenza dell'Italiano Scritto* – COLFIS, http://linguistica.sns.it/CoLFIS/Home.htm).

All these psycholinguistic variables did not significantly differ across the two semantic categories (two–sample t–test of frequency: $t = -0.35$, $p = 0.73$; number of letters: $t = -1.99$, $p = 0.06$; number of syllables: $t = -0.34$, $p = 0.74$; chi–square of gender $\chi = 0.34$, $p = 0.56$; accent $\chi = 3.0$, $p = 0.08$) or across the four semantic clusters (Kruskal–

Wallis test for small sample size of frequency: h = 10.44, p = 0.17; number of letters: h = 8.38, p = 0.30; number of syllables: h = 9.34, p = 0.23; chi–square test of gender: χ = 6.0, p =0.54; accent: χ = 2.44, p = 0.93). The analyses were run with the statistical functions provided by Python's library SciPy (https://docs.scipy.org/doc/scipy/reference/stats.html).

1.3 Stimuli Psychological Validation

In order to recover the internal representation of our stimuli in the general population, and to confirm that the general clustering of words was universally shared, we proceeded with two experiments exploiting different methods to investigate cognitive semantic representations: Semantic Distance Judgment and Semantic Feature Listing.

Semantic Distance Judgment

As seen in the Chapt. 2.1.1, one possible way of investigating subjects' semantic space is that of explicitly asking them to rate the distance (dissimilarity) between word pairs. We recruited fifty subjects, naïve to the goal of the experiment, and we tested them with an internet-based questionnaire. Stimuli were arranged in 132 pairs, and consisted all possible combinations of the within-category words. We then asked subjects to rate how similar the concepts referred to by the words were on a Likert scale from 1 (not similar at all, very far in meaning) to 7 (very similar, very close in meaning). We decided to present only within category combinations in order to prevent the large difference across categories from overshadowing the smaller, but relevant, differences within them (Goldstone et al., 1997). For the same reason, we presented tool word pairs and animal word pairs in separate blocks. The order of presentation of the different pairs inside each category was randomized for each subject, whereas the order of presentation of the two categories was pseudo–randomized across

subjects: half of the subjects rated animals before tools and the other half did the opposite.

All subjects' scores were normalized (i.e., scaled between 0 and 1), in order to correct for possible inter–individual differences in the ranking scale adopted. Normalized data were then re–arranged to create two 12x12 matrices describing the pairwise semantic distance between words for animals and tools separately. Next, for both categories we computed the two mean distance matrices averaging across all subjects. We then applied multidimensional scaling analysis (MDS, 2 dimensions, criterion: metric stress) to obtain a graphical representation of the cognitive semantic space of our subjects. The analyses leading to the choice of stress and number of dimensions are included as supplementary materials (Appendix, 1.1).

This visual representation show 4 sub-categorical clusters in each of the two categories (see Fig. 43). In the animals set the clusters were domesticated land animals (*cow, sheep, and goat*), wild land animals (*zebra, camel and giraffe*), sea mammals (*whale, dolphin and seal*), and not–mammal sea animals (*squid, shrimp and octopus*). In the tools set the clusters were weapons (*spear, saber and sword*), office/schools tools (*pencil, pastel, pencil sharpener*), work appliances (*hammer, nail, and pincer*), and hair instruments (*comb, brush, and hairpin*). K-Means clustering (with k = 4) confirms the assignment of the single words to the four clusters. Figures illustrating the centroids positions can be found in the supplementary materials (Appendix, 1.1).

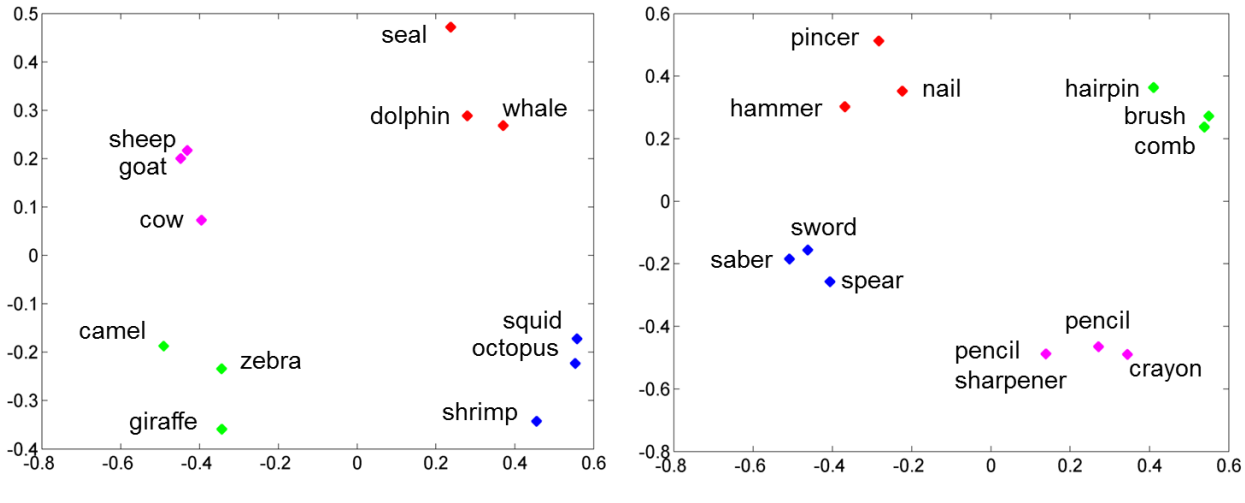


Figure 43 Semantic Distance Judgment results of Study 1. Multidimensional scaling visualization of the results of our first SDJ experiment with Italian words. Subjects' judgments lead to the validation of 4 semantic clusters in each of the two categories (left =animals, right = tools).

Semantic Distance Judgment Retest

We wished to assess whether the SDJ measure would be reliable and consistent enough to show the same results at a following re-test. In order to asses this, we asked 20 out of the 50 subjects who participated in the previous experiment to complete the similarity judgment task a second time after about 6 months. Again, they completed an online questionnaire. They received the same instruction as the first time with the added remark that it was not a memory task and they should not have tried to remember the answer given 6 months before.

On the data collected, the same pipeline of analyses described above was applied. A simple visualization of the results shows a striking similarity with the previous measurement (see Fig. 44), as confirmed by the following statistical analyses. The correlation between subjects' similarity matrices was used as a measure of inter–subject variability, while the correlation within subjects was used to estimate the intra–subject consistency. All pairwise across subjects correlations were statistically significant: the average correlation coefficient was 0.73±0.07 for animals and 0.56±0.08 for tools at the first evaluation, and 0.68±0.09 for animals and 0.52±0.05 for tools at the second evaluation. Subjects were also consistent across sessions: all showed a significant and positive correlation between their two judgments for both sets, with an average of 0.78±0.14 for animals and 0.60±0.15 for tools.

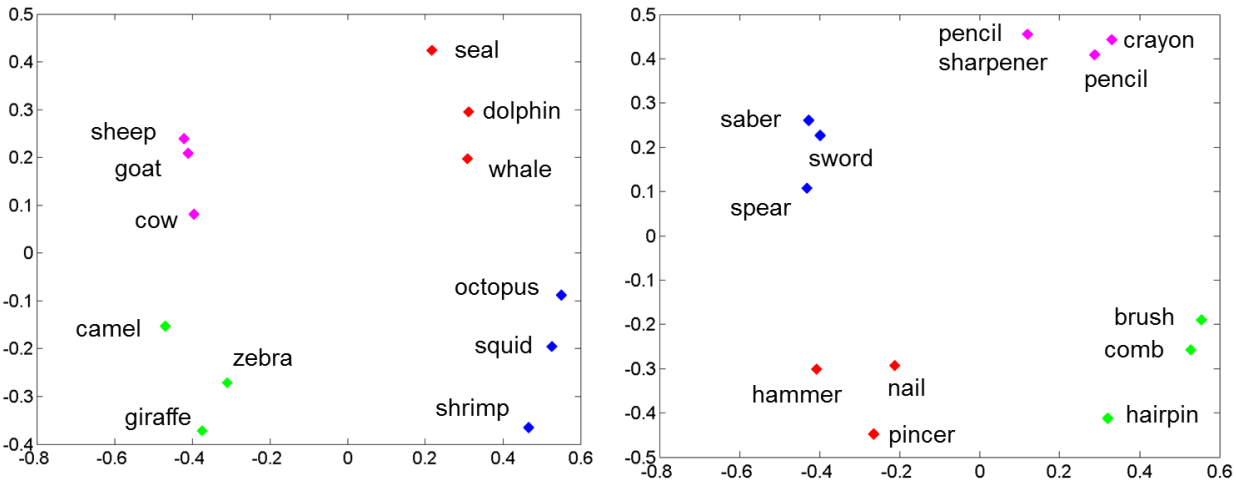


Figure 44 Re-test of the SDJ. Multidimensional scaling visualization of the results of our second SDJ experiment with Italian words. Comparing these results with the first experiment, it appears that the representational space of both categories appears to be organized in the same clusters of semantically related words. This indicates consistency across time of the cognitive semantic representations (left =animals, right = tools).

Semantic Feature Listing

The proximity (similarity) in semantic space can also be computed as the number of shared features (see Chap. 2.1.2). In this case, subjects are not asked to explicitly rate the semantic distance across pairs of words, but rather to list the features they spontaneously associate with each of single words. The number of features that are common across words is taken as indirect measure of semantic proximity.

Eighty subjects, naïve to the goal of the experiment and that had not taken part to any previous related experiment, were recruited. Again, testing was performed via an internet-based questionnaire. Subjects were asked to list between 5 and 10 characteristics or properties of each of the 24 target stimuli. They were explicit instructed to think about both the perceptual properties (in terms of view, touch, hearing, etc…), the functional properties (e.g. where it is usually found, how and for what is usually used), as well as any other property that could be considered important to describe the concepts the word referred to.

A similarity matrix between the words was created computing the number of shared features across all pairs of words belonging to the same category. The subsequent steps (i.e. normalization, conversion in distance matrices and MDS application) were the same as for the SDJ task. Results strongly confirmed the presence of 4 clusters of semantically related words in each category (see Fig. 45).

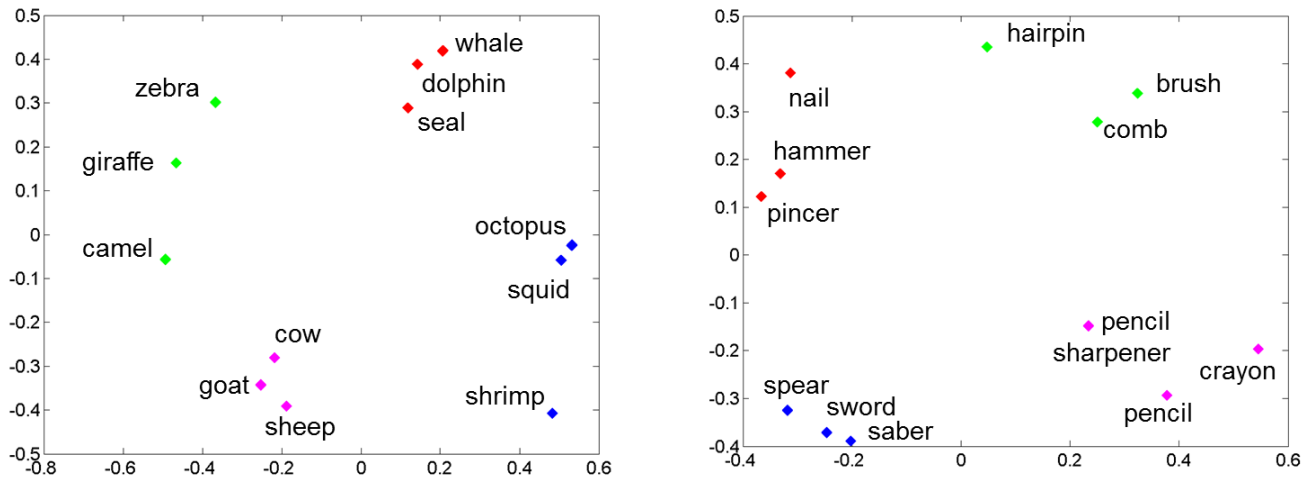


Figure 45 Semantic Features Listing results of Study 1. Multidimensional scaling visualization of the results of our SFL experiment with Italian words. The comparison of shared features lead to the same 4 semantic clusters in each of the two categories (left =animals, right = tools).

2. Study 2

The second neuroimaging study was going to be run with French participants and aimed at elucidating the temporal features of the neural correlates of semantic dimensions as detected with MEG. The main goal was to investigate how perceptual and conceptual dimensions of the stimuli space interact to determine the neural representational geometries.

2.1 Stimuli Selection

We constructed a stimulus set where the taxonomical (i.e. conceptual) and perceptual dimensions orthogonally varied. We selected 32 French words that refer to two broad semantic categories (i.e., living and non-living items), each of which could be potentially subdivided in semantic sub-clusters (e.g., wild animals vs domesticated animals). Moreover, we selected words varying along two perceptual semantic features: their real world size (i.e., words could refer to rather small or rather big items) and their real world auditory properties (i.e., words could refer to items that are strongly associated with a prototypical sound or not). For instance, the word *giraffe* corresponds to a big animal not associated with any particular prototypical sound (at least for western college students), while the word *cricket* corresponds to a small animal strongly associated with a prototypical sound. All stimuli are listed in Table 2.

Words (EN)	Words (FR)	Length (# of letters)	Implied Real World Size	Implied Real World Sound	Category	Cluster
gorilla	gorille	7	big	typical sound	living	wild
elephant	éléphant	8	big	typical sound	living	wild
giraffe	girafe	6	big	silent	living	wild
lama	lama	4	big	silent	living	wild
marmoset	ouistiti	8	small	typical sound	living	wild
parrot	perroquet	9	small	typical sound	living	wild
scorpion	scorpion	8	small	silent	living	wild
chameleon	caméléon	8	small	silent	living	wild
cow	vache	5	big	typical sound	living	domesticated
sheep	mouton	6	big	typical sound	living	domesticated
bull	taureau	7	big	silent	living	domesticated
chamois	chamois	7	big	silent	living	domesticated
cricket	cricket	7	small	typical sound	living	domesticated
cock	coq	3	small	typical sound	living	domesticated
ant	fourmi	6	small	silent	living	domesticated
rabbit	lapin	5	small	silent	living	domesticated
vacuum cleaner	aspirateur	10	big	typical sound	not-living	indoor
washing machine	lave-linge	10	big	typical sound	not-living	indoor
wardrobe	armoire	7	big	silent	not-living	indoor
sofa	sofa	4	big	silent	not-living	indoor
blender	mixeur	6	small	typical sound	not-living	indoor
alarm clock	réveil	6	small	typical sound	not-living	indoor
pillow	oreiller	8	small	silent	not-living	indoor
fork	fourchette	10	small	silent	not-living	indoor
helicopter	hélicoptère	11	big	typical sound	not-living	outdoor
motorbike	moto	4	big	typical sound	not-living	outdoor
bike	vélo	4	big	silent	not-living	outdoor
canoe	canoë	5	big	silent	not-living	outdoor
car stereo	autoradio	9	small	typical sound	not-living	outdoor
rotating beacon	gyrophare	9	small	typical sound	not-living	outdoor
roller	roller	6	small	silent	not-living	outdoor
boots	bottes	6	small	silent	not-living	outdoor

Table 2 Stimuli used for feature listing and distance rating in Study 2. Stimuli were pre-selected by the authors as to span two semantic categories and four semantic clusters. Moreover two perceptual semantic dimensions were manipulated: implied real world size and prototypical sound.

2.2 Stimuli Psycholinguistic Validation

As for Study 1, we ensured differences across semantic categories and dimensions were not correlating with differences in the low-level psycholinguistic variables. Words belonging to the different semantic categories, semantic clusters, and perceptual clusters (e.g., big vs small) were well matched for number of letters, number of syllables, number of phonemes, gender, frequency of use in books and in movies (retrieved from Lexique, http://lexique.org).

These psycholinguistic variables did not differ across the two semantic categories (Mann-Whitney rank test for number of letters: u = 109.5, p = 0.25; number of syllables: u = 105, p = 0.17; number of

phonemes: u = 98.5, p = 0.13; frequency of use in books: u = 120, p = 0.39; frequency of use in movies: u = 126, p = 0.48; chi–square test of gender: χ = 0.14, p = 0.70) nor across semantic clusters (Kruskal–Wallis test for small sample size of number of letters: h = 3.93, p = 0.27; number of syllables: h = 6.67, p = 0.08; number of phonemes: h = 6.39, p = 0.09; frequency of use in books: h = 3.87, p = 0.28; frequency of use in movies: h = 2.08, p = 0.56; chi–square test of gender: χ = 0.43, p = 0.93). Similarly, they did not differ across the visual-perceptual semantic property (Mann-Whitney rank test for number of letters: u = 103, p = 0.17; number of syllables: u = 121, p = 0.39; number of phonemes: u = 91, p = 0.08; frequency of use in books: u = 111.5, p = 0.27; frequency of use in movies: u = 103, p = 0.18; chi–square test of gender: χ = 0.14, p = 0.71), nor across the audio-perceptual semantic property (Mann-Whitney rank test for number of letters: u = 89.5, p = 0.07; number of syllables: u = 103.5, p = 0.15; number of phonemes: u = 91, p = 0.08; frequency of use in books: u = 104.5, p = 0.19; frequency of use in movies: u = 126, p = 0.48; chi–square test of gender: χ = 1.29, p =0.26). These analyses were run with the statistical functions provided by Python's library SciPy (https://docs.scipy.org/doc/scipy/reference/stats.html).

2.3 Stimuli Psychological Validation

As for Study 1, we proceeded with the validation of our stimuli set thanks to both Semantic Distance Judgment and Semantic Feature Listing.

Semantic Distance Judgment

We recruited sixty-five subjects, naïve to the goal of the experiment. We collected the data with an internet-based questionnaire, then the same pipeline of analyses described above was performed. We applied multidimensional scaling analysis (MDS, 2 dimensions, criterion: stress) to obtain a graphical representation of the cognitive semantic space of our subjects. Supplementary materials

in Appendix, 1.1 include in-depth description of the choice of stress and number of dimensions, as well as the K-means algorithm centroids.

Unsurprisingly, in contrast with Study 1 (which was design to highlight sub-clusters of words within the same semantic category) the organization of the semantic space is in this case less fragmented. Indeed, the focus of this study was to contrast high order categorical (animals vs. tools) with perceptual features (large vs. small, and prototypical sound vs. silent), therefore the pre-selection of the stimuli focused more on highlighting those dimensions than to the definition of nested classifications. Both visual inspection and k-means attempts to assign items to 2, 3 or 4 clusters lead to unstable solutions. One, introspectively sound, possible partition is between two sub-categorical clusters in each of the two categories (see Fig. 46). In the animals set, domesticated animals (*bull, sheep, cow, chamois, rabbit, rooster, ant, and cricket*) can be opposed to exotic animals (*elephant, giraffe, gorilla, lama, marmoset, parrot, chameleon, and scorpion*). In the non-living set, house appliances (*fork, wardrobe, sofa, pillow, washing machine, vacuum cleaner, blender, and alarm clock*), can be contrasted with objects linked with means of transportation (*canoe, boots, roller, bike, motorcycle, helicopter, car stereo, and rotating beacon*).

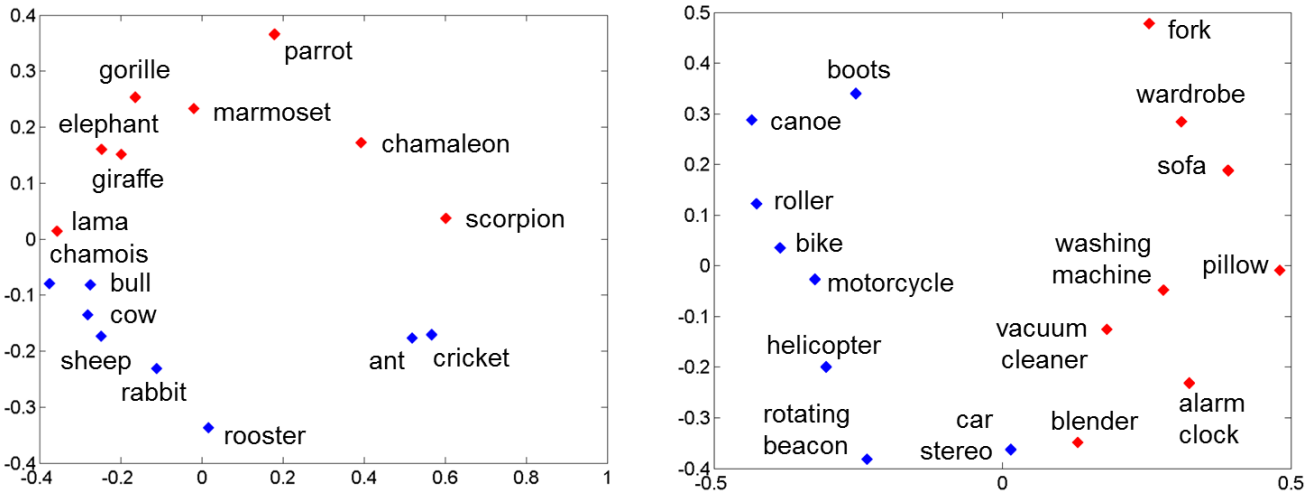


Figure 46 Semantic Distance Judgment results in Study 2. Multidimensional scaling visualization of the results of our SDJ experiment with French words. Subjects' judgments lead to the emergence of a rather distributed organization of the semantic space (left =animals, right = tools). One of the possible clustering solution (i.e., domesticated animals, wild animals, house appliances, means of transportation) is highlighted with red and blue colors.

Semantic Feature Listing

Sixty-six French native speakers, naïve to the goal of the experiment, were recruited. Following the same protocol, via an internet-based questionnaire, they were asked to list between 5 and 10 characteristics or properties of each of the 24 target stimuli, in terms of both perceptual properties and functional properties. Again, a similarity matrix was created computing the number of shared features. The subsequent steps (i.e. normalization, conversion in distance matrices and MDS application) were the same as for the SDJ task. As observed with the Italian stimuli, the semantic space described by the SFL experiment closely resembles the one derived from the previously described SDJ experiment (see Fig. 47).

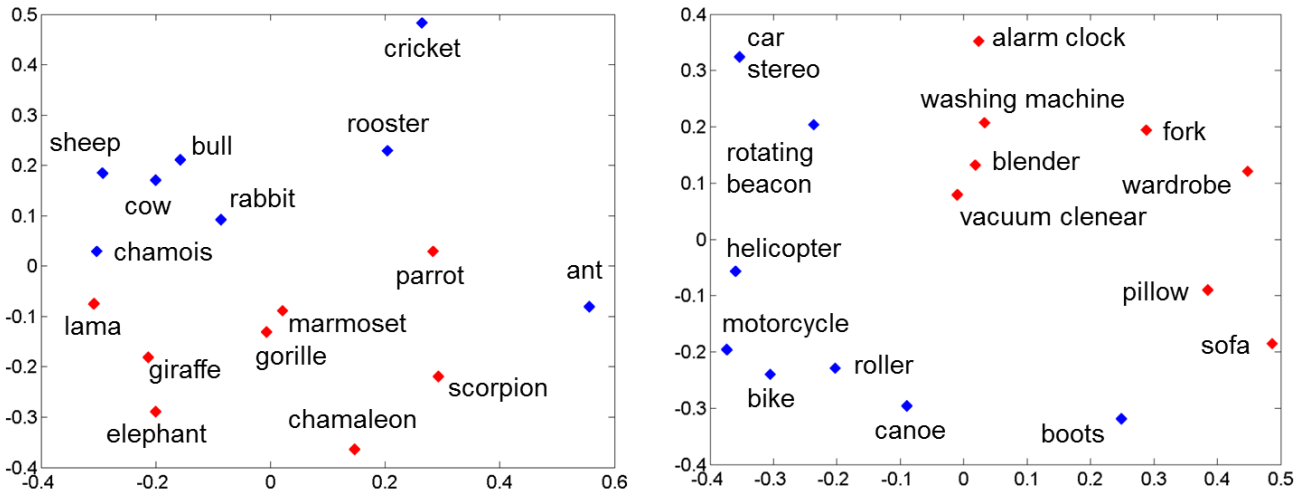


Figure 47 Semantic Features Listing results in study 2. Multidimensional scaling visualization of the results of our SFL experiment with French words. As for SDJ, the result depicts a rather distributed semantic space (left =animals, right = tools). One of the possible clustering solution (i.e., domesticated animals, wild animals, house appliances, means of transportation) is highlighted with red and blue colors.

3. One Space, Many Metrics?

Thus far we have used the distance matrices recovered from the two behavioral experiments only to display the corresponding semantic geometry in a two-dimensional space. Nevertheless, there are other ways in which the richness of these datasets can be exploited.

First, distance matrices enable the comparison of the representational spaces described by the two metrics (i.e. Semantic Distance Judgment and Semantic Feature Listing).Visual inspection of the MDS plot reported above can be used to perform a first, qualitative, comparison. For instance, in Study 1 both techniques lead to the emergence of the same four clusters of related words in the two semantic categories. However, partitions of the semantic space are not always univocally defined (see for instance the results of Study 2) and visual exploration of the MDS plot might mislead judgments, under- or over- estimating differences. Aiming at understanding whether the two different methods lead to the description of the same representational space, one needs a more precise quantification of their similarities.

Second, behavioral distance matrices can be compared with those stemming from other sources. In the past decades, many neuroimaging studies have resort to linguistic corpora and databases to describe the semantic space they investigated (for a prominent example, see Huth et al., 2012). Hence, it would be useful to compare the representational space(s) obtained via behavioral testing with the ones derived from these linguistic databases.

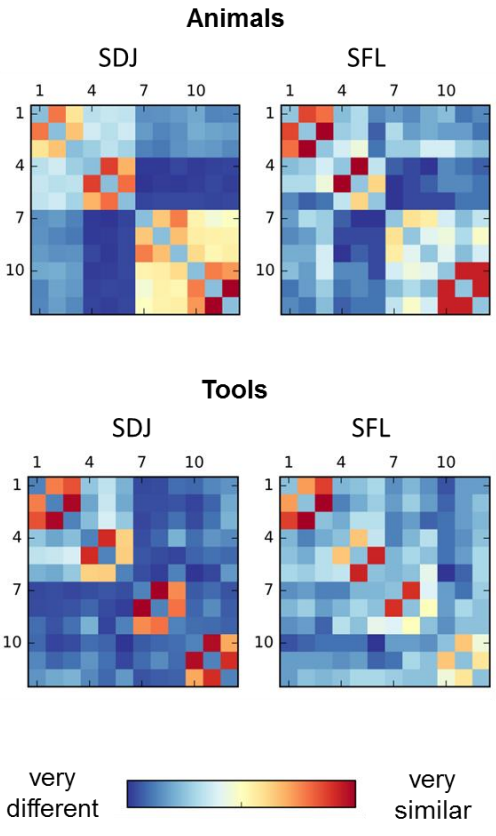


Figure 48 Similarity Matrices for Study 1. Matrices describing the representational spaces as reconstructed from the two behavioral tasks concerning 24 Italian words (12 animals and 12 tools names). [SDJ = semantic distance judgment. SFL = semantic feature listing].

3.1 Distance Judgment vs Features Listing

We sought to quantify the difference between SDJ and SFL by comparing the correlation between their representational spaces. As the matrices are symmetrical around a meaningless diagonal (i.e., representing the null distance of one concept with itself), only value of the upper triangular part of the matrices were used to compute the correlations. Finally, the similarity matrices scores (bound from 0 to 1) were z-transformed before computing the correlations.

All pairwise correlations between the representational spaces are highly significant, for both semantic categories (i.e., animals and tools), and for both studies. Table 3 reports the r scores and the p values for the four comparisons. One way of visualizing the results is by plotting the similarity matrices produced by the different methods. Similarity matrices for the first study are reported in Fig. 48, while for second one in Fig. 49. For visualization purposes, the meaningless diagonal is arbitrarily set to the median value.

corr(SDJ, SFL)	r	p
Animals Study 1	0.86	$<10^{-19}$
Tools Study 1	0.86	$<10^{-19}$
Animals Study 2	0.69	$<10^{-17}$
Tools Study 2	0.70	$<10^{-18}$

Table 3 Correlation between SDJ and SFL representational spaces. R score and p value of the 4 pairwise comparisons. All correlations are highly significant, with the two concerning the set of stimuli of Study 1 being slightly higher. [SDJ = semantic distance judgment. SFL = semantic feature listing]

Thus, it appears that both methods can used interchangeably when aiming at describing the representational geometry of the cognitive semantic spaces. Semantic clusters emerge spontaneously from subjects' judgments not only when they are to judge explicitly semantic similarity across word pairs (SDJ) but also when they have to evaluate words individually (SFL). Clearly, the richness of the feature based metric lies in the possibility to go beyond distances estimations. The reported features can be used to detect distinctions across cluster of words, for instance in data from Study 1, we observed that reference to the implied real world size was present for all categories, while reference to color were disproportionately more frequent for animals than for tools. For an example of how exhaustive this kind of analyses can be see (Hoffman and Lambon Ralph, 2013).

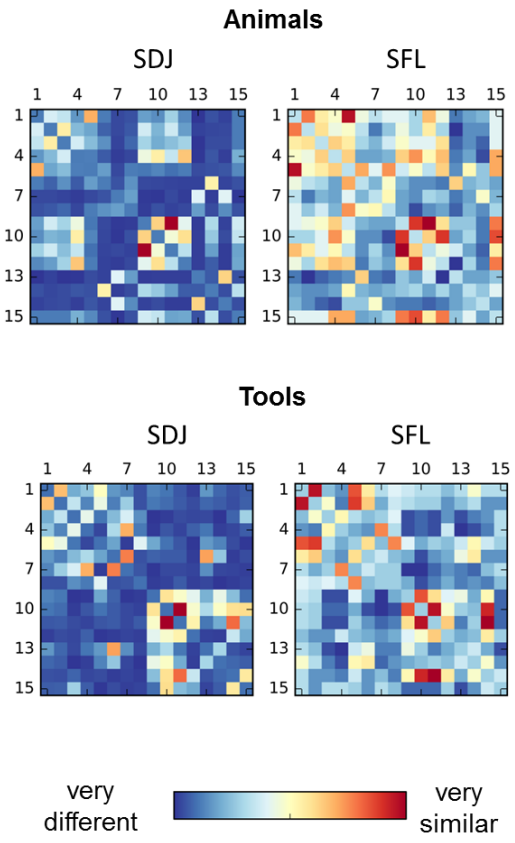


Figure 49 Similarity Matrices for Study 2. Matrices describing the representational spaces as reconstructed from the two behavioral tasks concerning 32 French words (16 animals and 16 tools names). [SDJ = semantic distance judgment. SFL = semantic feature listing]

3.2 Comparison with a Linguistic Database

Among the different linguistic databases available, WordNet, an English machine-readable lexical database developed at Princeton University, is perhaps the most widely used. It is organized by

meanings: it groups words into sets of synonyms called *synsets*, which are connected one another by means of semantic relations. It includes different lexical categories (e.g., nouns, verbs, adjectives and adverbs). Verbs and nouns are organized into hierarchies defined by hypernym or IS-A relationships.

We looked for the English translation of the words used in our experiment in WordNet as included in Natural Language Toolkit (Bird et al., 2009), which can be found at http://www.nltk.org/.

Different distance measures can be automatically derived from WordNet. We examined three of them:

- Wu-Palmer Similarity (WPS), it estimates of how close two words are based on the depth of their tree in the taxonomy and most specific ancestor node [2*depth(lcs) / (depth(s1) + depth(s2)) where s1 and s2 are the two words nodes and lcs the Least Common Subsumer, i.e., most specific ancestor node];
- Path Similarity (PS), it denotes how close two words are based on the shortest path that connects them in the IS-A taxonomy;
- Leacock-Chodorow Similarity (LCS), it combines the previous estimation of the

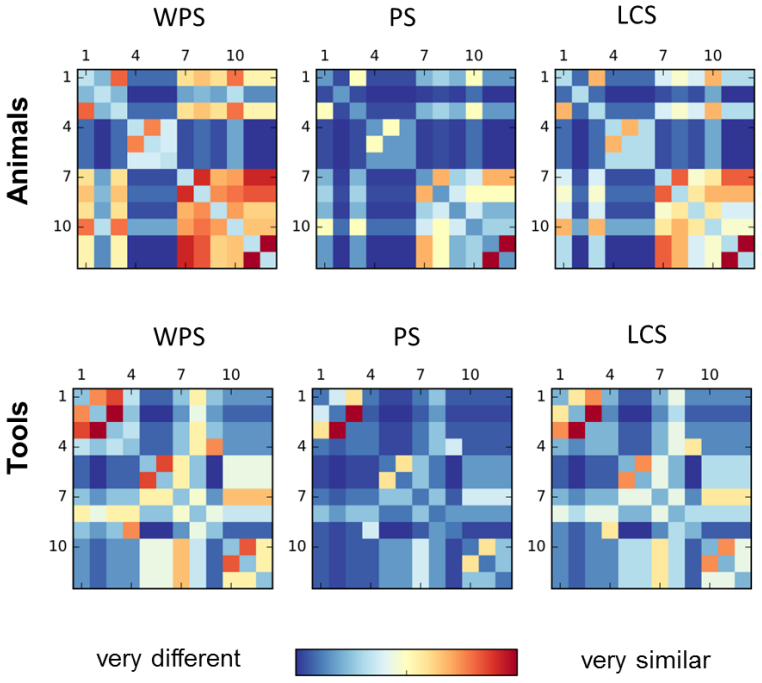


Figure 50 WordNet similarities matrices for Study 1. Matrices describing the representational spaces as reconstructed from the three distance metrics available in WordNet for the 24 Italian words (12 animals and 12 tools names). [WPS = Wu-Palmer similarity, PS = path similarity, LCS = Leacock-Chodorow similarity]

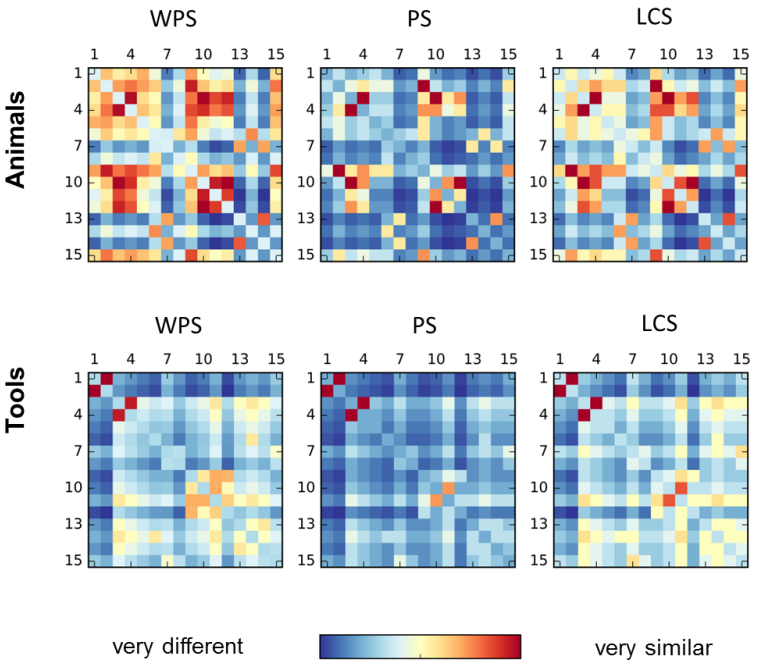


Figure 51 WordNet similarities matrices for Study 2. Matrices describing the representational spaces as reconstructed from the three distance metrics available in WordNet for the 32 French words (16 animals and 16 tools names). [WPS = Wu-Palmer similarity, PS = path similarity, LCS = Leacock-Chodorow similarity]

shortest path with the maximum depth of the taxonomy in which the words are found [-log(p/2d) where p is the shortest path length and d the taxonomy depth].

For both set of stimuli, we computed the similarity matrices according to the three distance metrics. Results for the 24 stimuli used with the Italian subjects (Study 1) are reported in Fig. 50, while the matrices for the 32 stimuli used with French subjects (Study 2) in Fig. 51.

The correlations between the three distance measures derived from WordNet and the two distance measures obtained from subjects' judgments are reported in Table 4. Similarity scores (bound from 0 to 1) were z-transformed before computing the correlations. Overall, it appears that all measures are significantly correlated. The only exception is the Semantic Feature Listing matrix for the words from Study 2, which does not appear to be correlated with the corresponding Leacock-Chodorow Similarity. In the cases where there was a significant correlation, however, the correlation coefficients were not very high, suggesting perhaps that the semantic space derived from WordNet is not entirely overlapping with the subject psychological space, thus not fully reflecting it. Unfortunately, this conclusion is somehow weakened by the fact that while subjects in our experiments evaluated the words in French or Italian (their mother tongue), the corpus-based data we analyzed come from English corpora. It is thus possible that the translation to English contributed to the less refined structure recovered from WordNet. However, even if corpora-based tools similar to WordNet for Italian and French have been proposed, none of them is as developed and as widely used as WordNet. Moreover, by exploiting the same database and thus the same build-in distance metrics, results can be directly compared across set of stimuli. It should also be appreciated that WordNet classification

corr(SDJ, WUS)	r	p
Animals Study 1	0.59	$<10^{-6}$
Tools Study 1	0.37	<0.01
Animals Study 2	0.68	$<10^{-17}$
Tools Study 2	0.35	$<10^{-4}$
corr(SDJ, PS)	**r**	**p**
Animals Study 1	0.60	$<10^{-7}$
Tools Study 1	0.47	$<10^{-4}$
Animals Study 2	0.62	$<10^{-13}$
Tools Study 2	0.35	$<10^{-4}$
corr(SDJ, LCS)	**r**	**p**
Animals Study 1	0.60	$<10^{-07}$
Tools Study 1	0.41	<0.001
Animals Study 2	0.64	$<10^{-14}$
Tools Study 2	0.32	<0.001

corr(SFL, WUS)	r	p
Animals Study 1	0.53	$<10^{-05}$
Tools Study 1	0.51	$<10^{-4}$
Animals Study 2	0.62	$<10^{-13}$
Tools Study 2	0.19	=0.03
corr(SFL, PS)	**r**	**p**
Animals Study 1	0.52	$<10^{-05}$
Tools Study 1	0.64	$<10^{-8}$
Animals Study 2	0.49	$<10^{-7}$
Tools Study 2	0.21	=0.01
corr(SFL, LCS)	**r**	**p**
Animals Study 1	0.53	$<10^{-05}$
Tools Study 1	0.57	$<10^{-6}$
Animals Study 2	0.52	$<10^{-9}$
Tools Study 2	0.14	=0.12

Table 4 Correlations between WordNet and behavioral measures. [SDJ = Semantic Distance Judgement , SFL = Semantic Feature Listing, WPS = Wu-Palmer similarity, PS = path similarity, LCS = Leacock-Chodorow similarity]

is based on senses (i.e., concepts meaning) and not lexical entries (i.e., words), thus allowing disambiguation of polysemic entries.

These analyses were run with the statistical functions provided by Python's library SciPy (https://docs.scipy.org/doc/scipy/reference/stats.html).

3.3 Conclusions

Overall, it appears that the correlations of the three WordNet-based similarities with the subject-based ones is smaller (and in one case non-significant), as compared to the different subject-based judgements (SDJ and SFL).

Two caveats undermining the fairness of the comparison should be acknowledged. First of all, shades of meaning might have been lost in the translation from Italian/French to English, thus future comparison should be based on metrics derived from Italian and French databases. Second, we did not include a comparison with measures from computational linguistic corpora, where distances are computed based on statistical co-occurrences of the words in text. As mentioned in Chap. 1, it has been suggested that these measure can recover semantic spaces that closely approximate behaviorally retrieved ones (for an in-depth analyses of currently available tools see (Pereira et al., 2016)).

4. A Space to Prime

Thanks to the experiments here reported, we have been able to appreciate the multidimensional nature of the cognitive semantic space of our volunteers. While in the past, in the psychological and psycholinguistic literature there has been an important emphasis on taxonomy as the most important dimension organizing the semantic space (as also reflected in the WordNet), we have suggested, in line

with the most recent research that the organization of the semantic space also reflect perceptual components of word meaning (see Chap. 1). A crucial question is whether these perceptual component of word meaning are automatically retrieved when subjects are processing single words. One way to assess the degree of automaticity of such representations is to test possible priming effects (see Chap. 2.1.3).

4.1 Semantic Priming

Priming studies have contributed to all major discussions revolving around the organization of semantic representations. Crucially, they have been instrumental in questioning the classical model of concepts as interconnected nodes, in favor of a distributed semantic network perspective (see for instance (Masson, 1995)). Central to our interests, priming has also been exploited to investigate the organization of such semantic network. Is the semantic space more likely organized around semantic features links or associative ones? Different classes of models can be contrasted on the kind of pairs of words that are expected to generate priming effects. Company-based models provided a measure of associative relatedness (e.g., (Postman and Keppel, 1970)), a normative description of how words are used (e.g., the word *dog* is frequently associated with the word *leash*). Instead, attribute-based models (e.g., the featural model proposed by (Smith et al., 1974)) attempt to provide a measure reflecting primarily word meaning, i.e. semantic relatedness (e.g., the word *dog* is semantically close to the word *wolf)*. Interestingly, association can be asymmetrical: e.g., *leash* is strongly associated with *dog* (for instance, it would likely be the first associated word that one spontaneously retrieve), while this doesn't hold in the other direction – from *dog* to *leash*. On the contrary, semantic relations purely based on features sharing cannot be but symmetrical, once agreement on which features matter is reached (e.g., the number of features share by *dog* and *wolf* is constant). Thus, since early investigations, authors have tried to distinguish between frequently associated and semantically related

words, with mixed findings. Some researchers found evidence of pure semantic priming effects, i.e., with pairs of words semantically similar but not associated (Fischler, 1977). Others were able to detect automatic priming for pairs of words that were only semantically associated (i.e., pure associative priming), and not for word pairs that were semantically related yet not associated (Shelton and Martin, 1992). After 30 years of research, a meta-analysis concluded in favor of the evidence of a pure semantic priming effect, claiming evidence of purely associative priming is non substantial (Lucas, 2000). However, a subsequent meta-analysis highlighted that both association strength and feature overlap appear to contribute to automatic priming, thus stressing the need for further investigations in order to understand their interplay (Hutchison, 2003). Therefore, if aiming at isolating the contribution of shared features to the priming effect, one should attempt to control for purely associative links between the stimuli.

Traditionally, studies have focused on conceptual relations between prime and target, for instance contrasting conditions where prime and target either belong or do not belong to the same semantic category (e.g. *bread-cake vs. bell-cake* (Meyer and Schvaneveldt, 1971; Fischler, 1977)). Subsequently, some authors have attempted to explore the effects of semantic features overlap following two premises. First, concepts are conceptualized as point in a multidimensional space, where each dimension corresponds to biologically and psychologically relevant semantic features. This leads to the observation that two concepts will be closer in representational space the more features they share. Second, the full representation of a concept, including all its relevant features, is automatically activated whenever the corresponding word is read. If this is the case, words sharing motor-perceptual and conceptual features should prime one another. This kind of priming effects have been reported for words referring to items that have similar visual shape (e.g. *apple-ball* (Schreuder et al., 1984)), associated movement (e.g., *piano-typewriter* (Myung et al., 2006)), or color (e.g., *emerald-cucumber* (Yee et al., 2012)). Nonetheless, doubts persist on the automaticity of the retrieval

of motor-perceputal features, as such priming effects have been shown only in specific circumstances (e.g., when subjects' attention has been directed to the targeted feature just beforehand (Pecher et al., 1998; Yee et al., 2012), see below for more detailed discussions). Thus, collection of more evidence is needed before any definitive decision can be reached. This is especially relevant in light of the current debate on the nature of neural semantic representations (see Chap. 1): is the retrieval of motor-perceptual feature an automatic and necessary component of word meaning understanding?

4.2 Perceptual Priming

The first studies reporting priming effects for words sharing perceptual features (i.e., shape similarity) were those by (Schreuder et al., 1984; d'Arcais et al., 1985). They compared unrelated pair of words, words having a conceptual relation (e.g., banana-cherry), words having a perceptual relation (e.g., ball-cherry), or both (e.g., apple-cherry). During a lexical decision task, strong priming for conceptual congruency was observed, while only a weak one emerged for perceptual congruency. The situation was reversed in the setting of a reading task, where a strong effect of conceptual priming, but no effect of perceptual priming was found. Observing how reaction times (RTs) of the lexical decision task were considerably longer than those for the reading task, the authors suggested that perceptual dimensions are accessed faster and in a transitory way, while conceptual ones are accessed only at later stages (Schreuder et al., 1984). To test this timing hypothesis, a follow up study compared a speeded lexical decision task (thus speeding up stimuli processing), and a reading task with degraded target (thus slowing down stimuli processing). In the first case, priming for both conceptual and perceptual dimensions was found, with the conceptual one being smaller than formerly reported. In the second case, conceptual priming was observed, while perceptual one only approached significance. The manipulations were thus successful in overturning previous findings, indicating that weather

perceptual or conceptual priming is found does not depend on the task itself, but rather on the latency at which stimuli are processed. At short latencies, perceptual effects are more prominent than conceptual ones; at long latencies, the pattern is reversed (i.e., conceptual effects are more prominent than perceptual ones).

From this hypothesis, it follows that manipulations of the interval between prime and target should affect the perceptual priming: it should occur only if prime and target are presented close enough in time. A recent study explored this aspect, while investigating whether words would prime the identification of a target picture as a function of two factors (Ostarek and Vigliocco, 2016). First, they manipulated the relation between the prime word and the target image depicted in the picture (e.g., the word "*star*" followed by the picture of the moon). Second, they presented the target image in a position that was congruent (or not) with the location implied by the prime word (e.g., "*rainbow*" implies a position with is high, thus might prime attention to the upper part of the screen, while "*carpet*" the lower one). The effect of the conceptual dimension (i.e., event congruency) and the one of the perceptual dimensions (i.e., spatial congruency) where compared at different stimulus onset asynchrony (SOA): 100 ms, 250 ms,, 800 ms. While the conceptual priming effect was evident in all three conditions, perceptual priming effect emerged only with SOA of 250 ms. This finding suggests a specific window for perceptual effects to be observed: after 100 ms but before 800 ms (Ostarek and Vigliocco, 2016). Overall, timing appears to be a crucial factor when investigating perceptual priming effects.

Notwithstanding their relevance in raising the issue of a dissociation between perceptual and conceptual aspects of semantic priming, (Schreuder et al., 1984; d'Arcais et al., 1985) studies have been heavily criticized with respect to shortcomings in the experimental setting adopted. The first group of remarks concerns stimuli presentation: (1) prime and target were presented simultaneously on the screen, which could have promoted active comparison; (2) in case of an incorrect response, the word pair was

repeated later on, which, as repetition affects RT and error rates may differ for different conditions, might have influenced priming effects. The second set of remarks involves material selection: stimuli acting as prime of perceptually compatible, conceptually compatible, and unrelated pairs were different, and some words were used both as prime and as target, with some primes being repeated with different targets. This lead to conditions that differ not only in the type of relation between prime and target, but also in the identity of the primes and in their frequency of occurrence. (Pecher et al., 1998) overcome these limitations while comparing perceptually and/or conceptually related words pairs in six different experiments (requiring subjects to make a lexical decision or to read aloud). Crucially, the priming experiment could be preceded by a task directing subjects' attention towards perceptual features (i.e., asking them to judge shape of the items the words refer to) or not. Moreover, the presence of associatively related prime-targets pairs was controlled. They found perceptual priming only when (1) subjects' focus had been directed toward visual properties of the items and (2) the whole stimulus set was devoid of associatively related pairs. Retrieval of perceptual dimensions appears thus less automatic and more strategic than previously thought: it occurs only when those features are made salient and only if no stronger direct link between the words can be perceived. Contextual effects have been highlighted even with a stroop-like paradigm, and in cases where color is the perceptual feature that words share (or don't share). In the first case, (Rubinsten and Henik, 2002) compared the effect of semantic size congruity (e.g., lion and bull are both big animals) and physical size congruity (e.g., words could be written with smaller or bigger fonts: lion vs ant). They found a physical size congruity effect for both semantic (i.e., *is this animal bigger than the other?*) and physical (i.e., *is this word written with a bigger font?*) judgments, while the semantic size congruity effect was observed only during the conceptual ones. In the second case, (Yee et al., 2012) described effects of priming for pair of words sharing the same color during

categorization task. Nevertheless, the effect was observed only for those subjects that performed a stroop-task before the categorization one. Finally, one study found subliminal priming effect during a size judgment task on words denoting concrete objects, however only when prime words were also members of the response set (i.e., they also act as targets). This suggest that the effect could arise thanks to an acquired mapping between targets and response keys, which is applied to subliminal stimuli too (Damian, 2001). Nonetheless, few examples of automatic retrieval of perceptual dimensions even during orthogonal tasks have been reported. (Setti et al., 2009), for example, have been able to detect priming for words sharing one perceptual feature (i.e., the implied real world size) even in absence of an explicit focus onto the perceptual properties of the stimuli. Even in this case, though, the effect appeared stronger when subjects were actively instructed to use mental imagery.

Given this panorama, we set out to test the effects of perceptual features sharing and task focus with a series of priming experiments. We investigated two perceptual features, one visual (i.e., the average size) and one auditory (i.e., the sound emitted), thus selecting words referring to items orthogonally spanning from very small to very big, from very loud to very silent. We selected implied real world size as visual perceptual dimension as studies from object recognition demonstrate that real-world size is an automatic property of object representation (Konkle and Oliva, 2012), and that processing of both physical and conceptual magnitude in object perception is automatic (Gliksman et al., 2016). We opt for audio as additional perceptual dimension as neuroimaging studies suggest that audio properties of symbolic stimuli are retrieved rapidly even when not explicitly required by the task (Kiefer et al., 2008). We hypothesized that words referring to items of similar real world size (e.g., *sofa* and *wardrobe*) would prime each other, while no priming would be detected for words of different relative size (e.g., *sofa* and *alarm clock*). Similar prediction was made for words referring to items

emitting a prototypical sound (e.g., *alarm clock* and *hoover*) or not (e.g., *pillow* and *sofa*). Moreover, we controlled the explicit focus of the subjects over the different features by contrasting four tasks, tapping either into the perceptual (e.g., *s it big? does it make a prototypical sound?*), or into the conceptual (e.g., *is it an animal? is it a color?*) features of the items.

4.2 Stimuli

The selection and validation of the stimuli followed three steps. First, we pre-selected the stimuli according to semantic criteria: we included words referring to non-living items and belonging either to a domestic environment (i.e., typically found and used within the house) or to an outdoor environment (i.e., typically found and used outside the house). Moreover, the objects referred to by the words varied orthogonally along two perceptual dimensions: size, they could be rather big or rather small (i.e., could or could not fit in a regular size drawer), and sound, they could either emit a prototypical sound or not. The preselection led to 32 words of which: 16 referred to indoor items and 16 to outdoor items. Orthogonally, 16 words referred to objects associated with a prototypical sound and 16 did not. Moreover, always orthogonally, 16 words referred to rather big objects (i.e., bigger than an average-sized sheep) and 16 to rather small ones (see Table 5).

Second, we verified that psycholinguistic variables such as length, and frequency of use did not significantly differ between compatible (e.g., sharing a given dimension) and incompatible (e.g., being different along that dimension) prime-target pairs. We implemented six t–tests and no statistically significant difference was found for:

- length of words across the visual dimension ($T(1,254)= -0.623$, $p=0.5343$), the auditory dimension ($T(1,254)= -1.8784$,

p=0.06147), and the conceptual dimension (T(1,254)= -0.2073, p=0.83594);

- frequency of use across the visual dimension (T(1,254)= -1.1660, p=0.24472), the auditory dimension (T(1,254)= -1.7996, p=0.24472), and the conceptual dimension (T(1,254)= -0.5224, p=0.60184);

Third, we validate the hypothesized perceptual and conceptual semantic dimensions via an internet-based questionnaire. The same questionnaire was used to control other psycholinguistic factors that could potentially affect our results: associative links between our stimuli and differences in familiarity. Thirty subjects underwent 4 short tasks in order to assess:

a) Visuo-perceptual semantic dimension. For each given word, subjects had to answer to the following question: "*Is the object this word refers to smaller than this drawer? Could it fit in the drawer?*". For instance, the object the word "*blender*" refers to can fit in a drawer, while the word "*dishwasher*" cannot. Only yes or no answers were allowed.
b) Audio-perceptual semantic dimension. As above, but this time the question was: "*Is the object this word refers to associated with any prototypical sound?*". As example of object associated to a characteristic sound, consider "*whistle*", as silent object, consider "*compass*".
c) Conceptual semantic dimension. This question concerned the natural location in which the item is encountered: "*Is the object this word refers to typically used, found in the house?".* For instance, "*binoculars*" are usually used outdoor, while a "*vacuum cleaner*" indoor.
d) Association. Subjects wrote the first word that came to their mind in association with the word presented. For instance, many subjects wrote "*remote control*" in response to "*television*".
e) Familiarity. Subjects were asked to indicate on a Likert scale how familiar they were with the item referred to by the word: 1

meant not familiar at all, 7 very familiar. For instance, the word "*bed*" received on average a score of 7, "*motorcycle*" a 3.

Words (EN)	Words (FR)	Set	Implied Real World Size	Implied Real World Sound	Category
dishwasher	lavastoviglie	1	big	sound	indoor
television	televisore	1	big	sound	indoor
table	tavolo	1	big	no sound	indoor
armchair	poltrona	1	big	no sound	indoor
phone	telefono	1	small	sound	indoor
blender	frullatore	1	small	sound	indoor
brush	spazzola	1	small	no sound	indoor
fork	forchetta	1	small	no sound	indoor
helicopter	elicottero	1	big	sound	outdoor
tractor	trattore	1	big	sound	outdoor
bench	panchina	1	big	no sound	outdoor
parachute	paracadute	1	big	no sound	outdoor
car stereo	autoradio	1	small	sound	outdoor
megaphone	megafono	1	small	sound	outdoor
boots	scarponi	1	small	no sound	outdoor
skates	pattini	1	small	no sound	outdoor
vacuum cleaner	aspirapolvere	2	big	sound	indoor
washing machine	lavatrice	2	big	sound	indoor
bed	letto	2	big	no sound	indoor
wardrobe	armadio	2	big	no sound	indoor
phon	phon	2	small	sound	indoor
alarm clock	sveglia	2	small	sound	indoor
glass	bicchiere	2	small	no sound	indoor
sponge	spugna	2	small	no sound	indoor
van	furgone	2	big	sound	outdoor
motorcycle	motocicletta	2	big	sound	outdoor
street lamp	lampione	2	big	no sound	outdoor
canoe	canoa	2	big	no sound	outdoor
horn	clacson	2	small	sound	outdoor
whistle	fischietto	2	small	sound	outdoor
compass	bussola	2	small	no sound	outdoor
binoculars	binocolo	2	small	no sound	outdoor

Table 5 Stimuli used for the four priming experiments. Three dimensions were orthogonally manipulated: a conceptual semantic dimension (i.e.,location of typical use), a visuo-perceptual semantic dimension (i.e., implied real world size), and an audio-perceptual semantic dimension (i.e., whether the item is associated with a prototypical sound or not).

Overall, there was a very high agreement across subjects on the perceptual and conceptual dimension of our stimuli (tasks a-c). All items were classified as expected by the significant majority of the subjects along all dimensions (as verified with a binomial test), except for the word "*skates*", which was considered associated to a prototypical sound by 50% of the subjects. These results led us to the introduction of a screening test for all participants recruited for the priming experiments. In case of disagreement between our suggested

classification and that proposed by the subject (e.g., a subject considering *television* silent instead of associated with sound), coding of the conditions was adapted to the subject specific classification (e.g., the pair *television* – *pillow*, would be re-coded as sharing the auditory property).

No associations between our stimuli emerged from the association task (d). As additional control over possible associative link between our stimuli, we used the Italian web-based corpus Web Infomap (http://clic.cimec.unitn.it/infomap-query/info.html) to check that none of our stimuli would appear within the first 20 semantic neighbors of all the other stimuli.

The only statistically significant difference was in the familiarity score between pairs compatible (e.g., "*street lamp*" – "*bench*") vs incompatible (e.g., "*bench*" – "*sofa*") along the conceptual dimension (T(1,254)=-13.0679, p=3.4093e-30). This would have been a problem as potential conceptual priming effect (i.e., when prime and target share the typical location of use) would have been confounded by the familiarity effect (i.e., whether prime and target are equally familiar). Therefore, we decided to never pair stimuli across different conceptual domains: thus, for all pairs, prime and target stimuli were either both indoor (so both highly familiar) of both outdoor (thus both less familiar) terms.

4.3 Method

We tested speed of single word processing during four different tasks with a between subjects design. Number of subjects, randomization and timing presentation of the stimuli were the same across tasks.

Experimental tasks. The first two tasks required an explicit access to the perceptual features investigated (hereafter, Explicit Tasks):

a) In the sound-related task (hereafter Audio task), subjects were asked to judge whether the target stimulus is associated with a prototypical sound (i.e. attention explicitly directed to the auditory property)
b) In the size-related task (hereafter-Video task), subjects were asked to judge whether the target stimulus could fit in a small basket (i.e. attention explicitly directed to the visual property: the implied real world size).

With the other two tasks, we diverted subjects' attention away from the visual and auditory properties, in order to investigate if we could find traces of an implicit access to the non-attended dimensions (hereafter, Implicit Tasks):

c) In the Animals task, subjects were asked to determine if the target stimulus belong to the category of animals or not. We therefore added to 32 animal names, making sure that only half of them were associated with a prototypical sound and, orthogonally, only half of them were bigger than the reference used for the Video task.
d) In the Colors task, subjects were asked to determine if the target stimulus belong to the category of names of colors. We therefore added 32 names of color (e.g., turquoise, vermilion, ocher).

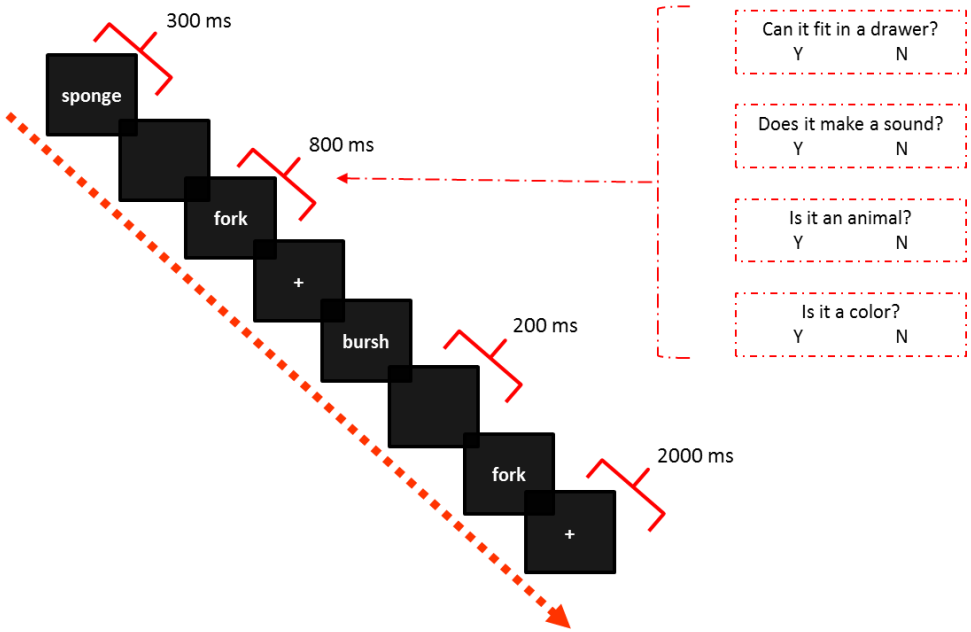


Figure 52 : Experimental setting. Example of a sequence of stimuli during the priming experiments. Irrespective of the task to be performed on the target stimulus, each trial followed the same structure and timing here reported. Subjects were instructed to pay attention to the target stimuli and, according to the different experimental conditions, to answer to one of the four questions depicted on the right.

Stimuli randomization and presentation. We split the 32 words in two sets of 16 stimuli, taking care that they both included two exemplars of each combination of our three factors (i.e., location, visual and auditory properties). Words of set A were used to prime words of set B and the reverese. Each of the 32 words appeared 8 times as prime and 8 times as target. Stimuli were paired only within the same conceptual category (i.e., indoor or outdoor items). In total, 256 pairs were presented, half of which concerned indoor stimuli, half outdoor ones. Between prime and target, four possible relations were possible: (1) not sharing the value of any perceptual dimension; (2) sharing only the value of the visual dimension (i.e., same size); (3) sharing only the value of the auditory dimension (i.e., both associated (or not) to a prototypical sound); (4) sharing the value of both perceptual dimensions (i.e., same size and same auditory association). Each target appeared twice in each of the four conditions, thus we collected 64 (2*64) observations in total for each condition (within each participant). The total sets of trials were divided 4 blocks, pseudo-randomizing the trials as to assure that trials belonging to the different conditions were presented with the same proportion. In the case of the Animals task and Colors task, additional trials were added:

each of the 32 words appeared once per block as prime for a randomly chosen name of animal (or color) , thus reaching a total of 96 trials per block.

For the four tasks, the structure of a trial was as follow (see Fig. 52). After 2000 ms during which only the fixation cross was presented, the prime was flashed for 300 ms, followed by a 200 ms blank screen (inter-stimuli-interval). Then, the target was presented for 800 ms, and subjects' response was recorded (up to 1000 ms post target onset). The fixation cross was left on screen until the begin of the following trial 2 s after the target offset. Stimulus Onset Asynchrony (SOA) between prima and target was thus of 500 ms, while 3300 ms elapsed between one prime and the following one.

Subjects provided their answer thanks to a Qwerty keyboard whose Z and M keys had been replaced by two "YES" and "NO" labels. The assignment of the labels was randomized across subjects, thus half of the subjects answered positively with their right hand, half with their left hand. Stimuli were presented with Matlab Psychophysics toolbox (http://psychtoolbox.org).

Subjects. Ninety-six students of the University of Trento participated in the experiments in exchange for a monetary reward (5 euros) or university credits. Randomly, 24 students were assigned to each of the four experimental conditions.

4.4 Results

Subjects by subjects, data were cleaned from RTs at more than 3 std from the subject specific mean. The different conditions were then compared with respect to the average number of errors (accuracy of processing) and the average RT in correct trials (speed of processing). As there were very few errors (see Table 6 and 7), we concentrated our analyses on response times.

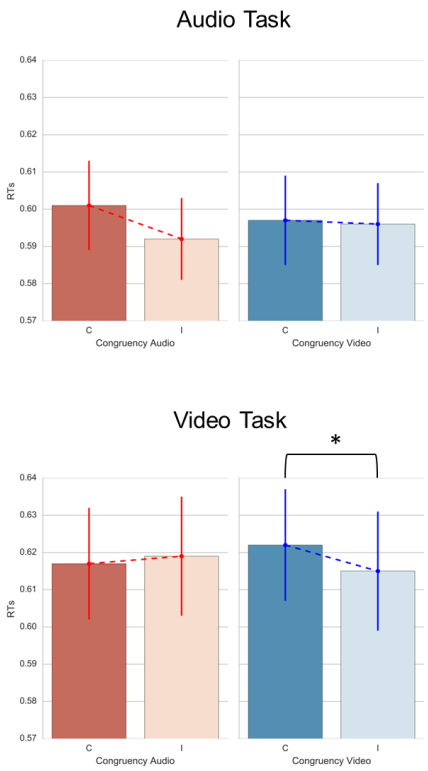


Figure 53 Explicit Tasks. Results of the two tasks directly tapping the perceptual dimensions. A task specific interference effect can be appreciated: subjects are significantly slower when prime and target share the same value of the perceptual feature. [C=congruent, I=incongruent]

First we analyzed the two explicit tasks together via a mixed ANOVA with three variables, two within subjects (audio and video congruency) and one across subjects (task), resulting in a 2 (video congruency) x 2 (audio congruency) x 2 (task) design. A significant interaction congruency audio * task was detected [F(1,46)= 9,712, p=0,003], indicating that the Audio congruency level differentially modulated the two tasks (see Fig. 53). We therefore analyzed the two tasks separately through a 2 (audio congruency) X 2 (video congruency) repeated measures design.

Audio task. Subjects were slightly slower in trials where prime and target shared the same audio-perceptual feature compared with when they did not, [F(1,23)=4,239, p= 0,051], However, there was no main effect of Video congruency and crucially no interaction.

Video Task. There was a significant main effect of video congruency [F(1,23)=20,671; p=0,000], indicating that subjects were significantly slower in trials where prime and target shared the same video-perceptual feature. However, there was no main effect video, and crucially no interaction.

Audio Task

Video	Mean RTs	Std RTS	Mean Error	Std Error
Congruent	0.597	0.012	3.98	0.5
Incongruent	0.596	0.011	3.19	0.47
Audio	**Mean RTs**	**Std RTS**	**Mean Error**	**Std Error**
Congruent	0.601	0.012	3.67	0.53
Incongruent	0.592	0.011	3.5	0.44

Video Task

Video	Mean RTs	Std RTS	Mean Error	Std Error
Congruent	0.622	0.015	3.98	0.5
Incongruent	0.615	0.016	3.19	0.47
Audio	**Mean RTs**	**Std RTS**	**Mean Error**	**Std Error**
Congruent	0.617	0.015	3.98	0.5
Incongruent	0.619	0.016	3.19	0.47

Table 6 Results Explicit Tasks. Audio (right) and Video (left). Mean RTs (and std), and mean number of errors (and std) across the 24 subjects.

As for the two Explicit Tasks, we first analyzed the two Implicit task together via a mixed ANOVA 2 (video congruency) x 2 (audio congruency) x 2 (task). A significant interaction congruency audio * task was detected [F(1,46)=13,719, p=0,001] (see Fig. 54). We therefore split the two tasks and analyzed them separately with a within subjects ANOVA 2 (video congruency) x 2 (audio congruency).

Animal Task. There was a significant main effect of audio congruency [F(1,23)=8,879, p=0,007]. Subjects were significantly faster (i.e., classical priming pattern) in trials where prime and target shared the same audio-perceptual feature (i.e., both associated with a prototypical sound or both silent). However, there was no main effect video and no interaction.

Color Task. There was a significant main effect of audio congruency [F(1,23)=6,690, p=0,017], in that subjects were significantly slower in trials where prime and target shared the same audio-perceptual feature (i.e., both associated with a prototypical sound or both silent). There was no main effect video and no interaction.

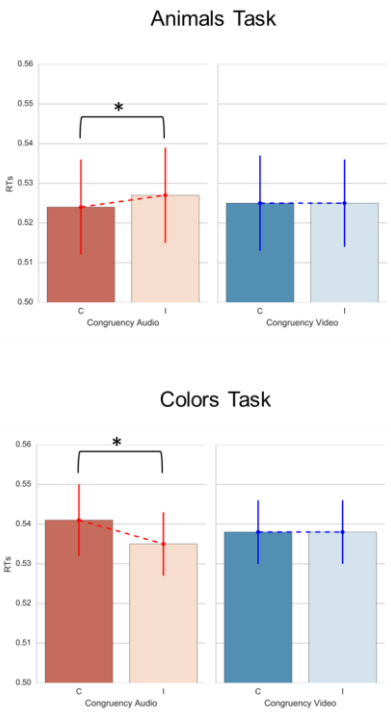


Figure 54 Explicit Tasks. When attention of the subjects if focused on conceptual semantic categories (i.e., animals vs tools), a significant priming for pairs sharing the same value of the auditory feature is observed. When attention is brought onto perceptual semantic categories (e.g., colors or not), a significant interference effect is found for pairs sharing the same value of the auditory feature. [C=congruent, I=incongruent]

Animals Task

Video	Mean RTs	Std RTS	Mean Error	Std Error
Congruent	0,529	0,009	1.35	0.5
Incongruent	0,53	0,009	1.5	0.52
Audio	**Mean RTs**	**Std RTS**	**Mean Error**	**Std Error**
Congruent	0,528	0,009	1.35	0.43
Incongruent	0,531	0,009	01.5	0.58

Colors Task

Video	Mean RTs	Std RTS	Mean Error	Std Error
Congruent	0,538	0,009	0.69	0.19
Incongruent	0,538	0,009	0.56	0.15
Audio	**Mean RTs**	**Std RTS**	**Mean Error**	**Std Error**
Congruent	0,541	0,009	0.60	0.18
Incongruent	0,535	0,008	0.64	0.13

Table 7 Results Implicit Tasks. Animals (right) and Color (left). Mean RTs (and std), and mean number of errors (and std) across the 24 subjects.

As for the two Explicit Tasks, we directly compared the Implicit Tasks via a mixed ANOVA 2 (video congruency) x 2 (audio congruency) x 2 (task). A significant interaction congruency audio * task was detected [F(1,46)=13,719, p=0,001].

Additional analyses. We analyzed the performance of the subjects in the two Implicit Tasks separately for indoor and outdoor stimuli, in order to investigate if the conceptual dimension interacted with the perceptual ones. A within subjects ANOVA 2 (video congruency) x 2 (audio congruency) was performed for each task (Animals vs Colors) and each sub-set of stimuli (Indoor and outdoor). In the Animals task, no significant difference emerged comparing priming in Indoor vs. Outdoor items: in both cases, a tendency towards a facilitation effect was observed (see Fig. 55). On the contrary, in the case of the Colors task, a significant main effect of audio congruency (in an interference direction) was detected [F(1,23)=4,820, p=0,038] for Indoor stimuli, while tendency for Outdoor stimuli was not significant (see Fig. 56). Crucially, again, no interaction was detected.

In summary, we found that in tasks directly tapping perceptual properties (Explicit tasks) subjects are slower when prime and target share the same value of a given feature. Conversely, when the task is orthogonal to the auditory properties, sharing this perceptual feature enhances subjects' performance (Animal task). Finally, if the task does not explicitly tap into the chosen visual property (size), but on a potentially correlated one (color), then sharing the same visual features result in a decrease in performance (Color task). However, the effects observed are very small (few milliseconds) and not qualified by significant interaction, thus strong conclusions should be hold until further testing is performed.

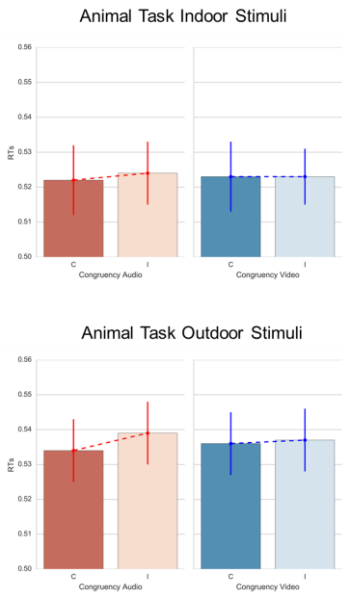


Figure 55 Impact of the conceptual dimension: Animal Task. After data are divided by conceptual category, no significant effects are found. For both indoor and outdoor stimuli, pairs sharing the value of the auditory feature show a trending priming effect. [C=congruent, I=incongruent]

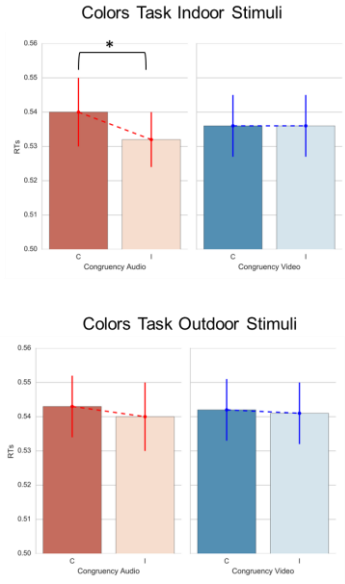


Figure 56 Impact of the conceptual dimension: Color Task. After data are divided by conceptual category, a significant interference effect is found for indoor stimuli sharing the same value of the auditory feature. The same effect is only trending for outdoor stimuli. [C=congruent, I=incongruent]

4.5 Discussion

The two Explicit Tasks suggested a dimension specific interference effect in that responses were slower when prime and target were congruent along the dimensions tapped by the task. However, the absence of a significant interaction between the interference effect and the two dimensions does not strongly support its specificity. Nevertheless, the results reported are interesting and suggest a spontaneous and automatic recovery of the perceptual dimension relevant for the task. The results are the opposite direction compared to what was predicted (interference vs. priming). This could reflect some form of strategic inhibition: upon elaboration of the prime, subjects might automatically prepare a response. However, since a response to the prime is not required by the task, it would need to be inhibited. The following response to the target would then be slowed down if coherent with the one on the prime. Thus, these findings are in line with evidences that, when relevant for the task at hand or when re-activated by the immediately preceding task, perceptual dimensions of word meaning are automatically recovered during reading (Pecher et al., 1998; Yee et al., 2012).

The first Implicit Task, requiring the detection of animals' names, showed an actual priming effect for the auditory (but not for the visual) dimension. Again, however, this effect was not qualified by an interaction, hence missing evidence of its specificity. Thus, even if not relevant for the task at hand, the perceptual dimension was reactivated. Even if this phenomenon is observed more rarely, there are previous examples of a neutral task eliciting the recovery of perceptual dimensions of word meaning (Schreuder et al., 1984; d'Arcais et al., 1985; Rubinsten and Henik, 2002; Setti et al., 2009). The absence of an effect for the visual dimension could be explained by an accentuated bipolar description of the auditory feature (i.e., stimuli either make a sound or not) as compared with the gradual differences along the visual dimension (i.e., stimuli's size vary greatly).

The second Implicit Task, requiring the detection of colors' names, showed an interference effect along the audio dimension. Post-hoc analyses revealed that such effect appears to be mainly driven by indoor stimuli. As previous findings suggest the interference effect to be related with the inhibition of a response to the prime stimulus, we examined the possibility that pairs sharing the same auditory property also shared color-related features. As a matter of fact, especially in the case of indoor stimuli, the auditory dimension correlates with the prototypical color: noisy items tend to be white or gray (i.e., made of plastic or metal, such as dishwasher, washing machine) while silent items tend to be brown (i.e., made of wood, such as table). One very speculative explanation would be that the perceptual feature indirectly tapped by the task (i.e., color) was automatically retrieved for both prime and target, causing the need to inhibit the response and hence slowing down subjects' performance.

Previous priming experiments have suggested that perceptual priming can be elicited (Schreuder et al., 1984; d'Arcais et al., 1985; Setti et al., 2009), even if it appears that specific conditions need to be met. First of all, subjects focus has to be directed towards the perceptual features of interest (Pecher et al., 1998; Yee et al., 2012). Second, timing of stimuli presentation and responses collection should be carefully chosen as it appears that perceptual dimensions are retrieved only transitorily in an early window of word processing (Schreuder et al., 1984; d'Arcais et al., 1985; Ostarek and Vigliocco, 2016). Priming is not the only paradigm used to look for automatic retrieval of perceptual information. For instance, a distance effect on comparison tasks (a paradigm extensively used in the literature on numerical cognition), has been not found for size (Hoedemaker and Gordon, 2014), yet was observed for shape (Zeng et al., 2016).

Overall, these evidences suggest a complex interplay between the task (in terms of both timing and attentional focus) and the perceptual/conceptual dimensions investigated. Future investigation should first attempt to replicate our current results, whose effects are extremely small, and obtained with a relatively small sample size (24

subjects per each experimental group). Then, aiming at dissociating possible alternative explanations, one should focus on the possibility to better control the set of stimuli and the demands of the task. For instance, new experiments could be devised with stimuli that further increase the differences along visual (i.e., size and color) and auditory (i.e., make a sound or not) dimensions.

Statistical analyses were performed with IBM SPSS (http://www.ibm.com/analytics/us/en/technology/spss/).

5. Conclusions

The cognitive semantic space of French and Italian native speakers seems to be organized around multiple perceptual and conceptual dimensions. The setting chosen, MDS on the distance metrics derived from SDJ and SFL, does not permit to fully interpret the dimension characterizing such space. However, two things should be highlighted. First, the representational spaces retrieved with SDJ and SFL appear to be highly correlated and overall consistent within and across subjects. Second, both methods significantly correlate with corpora-based measures, while providing a more fine-grained illustration of the cognitive semantic spaces of native speakers.

The series of priming experiments we conducted suggests that perceptual dimensions of word meaning (such as implied real world color and sound) are recovered during reading in an automatic way. Perceptual features are recovered for words that are not the target of the task at hand (i.e., the prime stimuli), and even when the task does not explicitly requires it (e.g., the Implicit Animal task). Further investigations are needed in order to establish (1) which perceptual features are consistently retrieved, and (2) which factors determine whether their retrieval will interact in a positive (priming) or negative (interference) way with the task.

It is perhaps necessary to conclude with a critical remark. The evidence coming from perceptual semantic priming experiments is

sometimes used to support a sensory-motor view of the cognitive (and neural) semantic system. However, priming effects can be interpreted as fast spreading of activation in a purely symbolic system capable of sensorimotor representations (Mahon and Caramazza, 2008): they do not necessarily entail the activation of sensory-motor representations/areas/formats. Recently, those who have the scope of supporting a sensory-motor view of the cognitive (and neural) semantic system have shifted towards interference paradigms which can have stronger implications for the causal role played by sensory-motor representations in semantics (Yee et al., 2013). The reasoning of these studies is as follow: if two simultaneous representations/tasks engage the same neural substrate, then performance should suffer (in terms of RTs and/or errors). Thus, if accessing meaning of words requires retrieval of perceptual features, concomitant tasks should interfere with subjects' performance proportionally to the involvement of related sensory-motor features. For instance, understanding words with a strong auditory component should be affected by concomitant auditory tasks, while performance with words with strong visual components by a visual task.

Acknowledgements:

Concerning the priming experiments, I would like to thank Serena Melison for her help with stimuli selection, data collection and data analyses. These experiments constitute part of her Master Thesis in Psychology (defended September 2016 at the University of Trento). I gratefully acknowledge M. Fabre and C. Pallier for their help with stimuli selection.

Bibliography

Bird S, Loper E, Klein E (2009) Natural Language Processing with Python. : O'Reilly Media Inc.

Buiatti M, Finocchiaro C, Caramazza A, Dehaene S, Piazza M (2012) Word meaning in the human brain: evidence for distinct category specific neural semantic spaces. In: Biomag, 18th International Conference on Biomagnetism. . Paris, France. .

d'Arcais GF, Schreuder R, Glazenborg G (1985) Semantic activation during recognition of referential words. Psychological research 47:39-49.

Damian MF (2001) Congruity effects evoked by subliminally presented primes: automaticity rather than semantic processing. Journal of Experimental Psychology: Human Perception and Performance 27.

Fischler I (1977) Semantic facilitation without association in a lexical decision task. Memory & cognition 5:335-339.

Gliksman Y, Itamar S, Leibovich T, Melman Y, Henik A (2016) Automaticity of Conceptual Magnitude. Scientific reports 6.

Goldstone RL, Medin DL, Halberstadt J (1997) Similarity in context. Memory & Cognition 25:237-255.

Hoedemaker RS, Gordon PC (2014) Embodied language comprehension: encoding-based and goal-driven processes. Journal of experimental psychology General 143:914-929.

Hoffman P, Lambon Ralph MA (2013) Shapes, scents and sounds: quantifying the full multi-sensory basis of conceptual knowledge. Neuropsychologia 51:14-25.

Hutchison KA (2003) Is semantic priming due to association strength or feature overlap? A microanalytic review. Psychonomic bulletin & review 10:785-813.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210-1224.

Kiefer M, Sim EJ, Herrnberger B, Grothe J, Hoenig K (2008) The sound of concepts: four markers for a link between auditory and conceptual brain systems. The Journal of neuroscience : the official journal of the Society for Neuroscience 28:12224-12230.

Konkle T, Oliva A (2012) A familiar-size Stroop effect: real-world size is an automatic property of object representation. Journal of experimental psychology Human perception and performance 38:561-569.

Lucas M (2000) Semantic priming without association: A meta-analytic review. Psychonomic bulletin & review 7:618-630.

Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of physiology, Paris 102:59-70.

Masson ME (1995) A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 21.

Meyer DE, Schvaneveldt RW (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology 90:227-234.

Myung JY, Blumstein SE, Sedivy JC (2006) Playing on the typewriter, typing on the piano: manipulation knowledge of objects. Cognition 98:223-243.

Ostarek M, Vigliocco G (2016) Reading Sky and Seeing a Cloud: On the Relevance of Events for Perceptual Simulation. Journal of experimental psychology Learning, memory, and cognition.

Pecher D, Zeelenberg R, Raaijmakers JG (1998) Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. Journal of Memory and Language 38:401-418.

Pereira F, Gershman S, Ritter S, Botvinick M, A (2016) A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. Cognitive neuropsychology 33:175–190.

Postman L, Keppel G (1970) Norms of word associations. . New York: Academic Press.

Rubinsten O, Henik A (2002) Is an ant larger than a lion? Acta Psychologica, 111:141-154.

Schreuder R, d'Arcais GBF, Glazenborg G (1984) Effects of perceptual and conceptual similarity in semantic priming. Psychological Research 45:339-354.

Setti A, Caramelli N, Borghi AM (2009) Conceptual information about size of objects in nouns. European Journal of Cognitive Psychology 21:1022-1044.

Shelton JR, Martin RC (1992) How semantic is automatic semantic priming? Journal of Experimental Psychology: Learning, memory, and cognition 18.

Smith EE, Shoben EJ, Rips LJ (1974) Structure and process in semantic memory: A featural model for semantic decisions. Psychological review 81:214.

Yee E, Ahmed SZ, Thompson-Schill SL (2012) Colorless green ideas (can) prime furiously. Psychological science 23:364-369.

Yee E, Chrysikou EG, Hoffman E, Thompson-Schill SL (2013) Manual experience shapes object representations., p.0956797612464658. Psychological science.

Zeng T, Zheng L, Mo L (2016) Shape Representation of Word Was Automatically Activated in the Encoding Phase. PloS one 11:e0165534.

CHAPTER 4:
TOPOGRAPHICAL FEATURES OF SEMANTIC DIMENSIONS

It will be possible […]
to project the image of any object one conceives in thought on a screen and make it visible.
If this could be done it would revolutionize all human relations.
I am convinced that it can and will be accomplished.
[Tesla, 1919]

In this chapter I review the work I conducted to investigate the topographycal organization of the neural representations of different semantic dimensions. Portions of the results here presented have been published in

Borghesani, V., Pedregosa, F., Eger, E., Buiatti, M., & Piazza, M. (2014). A perceptual-to-conceptual gradient of word coding along the ventral path. *International Workshop on Pattern Recognition in Neuroimaging* https://hal.inria.fr/hal-00986606/document

and

Borghesani, V., Pedregosa, F., Buiatti, M., A. Alexis, Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage* http://dx.doi.org/10.1016/j.neuroimage.2016.08.068.

Highlights:

- Perceptual semantic dimensions (e.g. implied size) are coded in early sensory areas.
- Conceptual semantic dimensions are coded in higher level anterior temporal regions.
- Different brain areas encode complementary dimensions of the semantic space.

1. Introduction

We have seen how word meaning is a key component of conceptual knowledge, i.e. the ability to acquire, store, update and retrieve semantic representations of the world we live in (see Chap.1). Many cognitive tasks that we face daily rely on semantic memory and

especially on our ability to manipulate and combine the abstract symbolic forms it can assume: words. It is perhaps the most uniquely human aspect of this peculiar kind of memory, and for decades cognitive neuroscientist have attempt to shed light onto its neural substrate. Recently, thanks to the development of multivariate methods (see Chap. 2.4), new cognitive hypotheses on its topographical organization can be tested.

1. 1 The Topography of Word Reading in the Brain

Word reading, i.e. the process of extracting meanings from symbols, requires the sophisticated interplay of different brain regions. As highlighted by neuroimaging studies, thanks to the fine temporal resolution of magnetoencephalography (MEG), brain activation unfolds from occipital areas towards the anterior temporal pole (Marinkovic et al., 2003). The classical view of the brain as a feed-forward information processor hypothesizes that this continuous stream of activation along the ventral stream may be dissected into multiple stages where information is represented with increasing levels of complexity and abstractness. From this perspective, the first steps permit the perceptual analysis of the stimuli words as purely visual shape, a process which culminates at the level of the visual word form area (VWFA) in a case, position, and size invariant representation of letter strings (Dehaene and Cohen, 2011). More anterior regions of the temporal lobe support more abstract word representations: semantic concepts. Lesions studies seem to confirm this view, with patients showing cortical blindness (Aldrich et al., 1987), pure alexia (Dejerine, 1892; Epelbaum et al., 2008), or semantic deficits (Gorno-Tempini et al., 2004) accordingly to the location of their lesions along the posterior-to-anterior axe in the occipito-temporal cortex. When it comes to understanding how the meaning of words instantiated in the brain many open questions are

left unanswered. Where and how strings of letters (i.e., percepts) become meaningful semantic entities (i.e., concepts)?

1. 2 Cognitive and Neural Semantic Geometries

Making sense of symbols involves retrieving from long term memory the corresponding semantic representations. One way to think about such representations is to consider them as points in a multidimensional space, where each dimension represents a specific property of the concept denoted by the word. In the case of words referring to concrete entities, the semantic space includes dimensions such as prototypical size, shape or sound, but also taxonomic class and functional information. Following what we introduced in Chap 1.2.5, we distinguish between perceptual dimensions (i.e., those along which physical properties of the objects are stored) and conceptual dimensions (i.e., those long which more complex, higher order features of the objects are stored). Storing both perceptual and conceptual features of object concepts is key for making sense of the word surrounding us: it is their combination that allows us to generalize across conceptually similar but perceptually different objects (e.g., a cat and a tiger), and differentiating between perceptually similar but conceptually different ones (e.g., a lemon and a tennis ball) (Rogers et al., 2004).

To understand how these dimensions mold representational geometry, consider the words “mouse”, “clownfish”, “giraffe”. Thanks to the multidimensional nature of the semantic space, we immediately know that the first two refer to animals that are closer in size (being rather small), compared to the third one. At the same time we can appreciate that the last two have a similar color (orange-ish) and that the first and the last one are close in taxonomy (both are terrestrial mammals, compared to the second one, a fish). A representational geometry that highlights visual attributes would weight distances on those dimensions more than any other: *mouse* and *giraffe* would be very distant as they do not share implied real world

size nor color. On the contrary, a representational geometry emphasizing conceptual dimensions would be mostly described by distances along taxonomic dimensions: *mouse* and *giraffe* would be very close as they are both terrestrial mammals. This toy example is clearly an over-simplification as many more dimensions are concurrently involved in the definition of a concept, but it stress how the representational space can be governed by complex geometries.

In Chapt. 3, we illustrated that both perceptual and conceptual semantic dimensions are relevant for the organization of the cognitive semantic space, highlighting possible dissociations. In the work here presented, we sought to investigate whether perceptual and conceptual semantic dimensions are neurally dissociable, i.e. preferentially encoded in different brain areas. We aim at doing so by mapping different representational geometries (e.g., dominated by perceptual or conceptual dimensions) onto brain activity in different cortical regions.

1.3 Neural Correlates of Semantic Representations

Even though the quest for the neural underpinning of semantics has a longstanding traditions (as we have seen in Chap 1), neither neuropsychology nor functional neuroimaging research have provided conclusive evidence on how different perceptual and conceptual semantic dimensions defining single concepts are encoded in the brain. Clinical data so far suggest that semantic knowledge is neurally coded in a distributed fashion, as it can be degraded by lesions to sensory–motor brain regions (Pulvermüller and Fadiga, 2010), and profoundly disrupted by lesions to higher–level associative regions (especially the anterior temporal lobe) (Gorno-Tempini et al., 2004; Hodges and Patterson, 2007; Lambon Ralph, 2014). Similarly, functional neuroimaging data indicate that during processing of object-related words there is an increased activation not only in high–level associative cortices (sometimes referred to as “semantic hubs”

(Patterson et al., 2007)) such as the inferior frontal cortex (Devlin et al., 2003), the anterior temporal cortex (Mion et al., 2010), or the inferior parietal cortex (Bonner et al., 2013), but also in primary and secondary sensory–motor cortices, in a way that appears proportional to the relevance of perceptuo-motor attributes (Pulvermuller, 2013).

Researchers capitalizing from both machine learning techniques and Representational Similarity Analysis (RSA) frameworks have shown that it is possible to discriminate between words belonging to different semantic categories (e.g., animals vs tools) as well as sub-categorical clusters (e.g., mammals vs insects) using distributed patterns of brain activation (Shinkareva et al., 2011; Bruffaerts et al., 2013; Devereux et al., 2013; Fairhall and Caramazza, 2013; Simanova et al., 2014). However, they did not determine if such discriminations were driven by conceptual or/and by correlated perceptual information, as we mentioned in Chap. 2.4, explicit (and complete) models are needed if one wishes to draw conclusion of the geometry of a given representations (Naselaris and Kay, 2015).

Finally, encoding approaches (modelling and predicting voxel-wise activation for different stimuli according to their defining set of features) has been successfully applied to predict brain activation during the elaboration of images and movies (Naselaris et al., 2009; Nishimoto et al., 2011), and only very recently to words (Fernandino et al., 2015a) and sentences (Anderson et al., 2016; Huth et al., 2016). Previous groundbreaking work used a computational model (trained on words data from text corpus) to predict the neural activation associated with written words, but always presented words together with their relative picture, thus being unable to dissociate the contribution of low level properties of the physical input from the pure semantic activation driven by the symbolic stimulus (Mitchell et al., 2008). More recent studies that used words and sentences as stimuli, do not distinguish between perceptual and conceptual features, being unsuitable to provide a clear picture of the brain topography involved in encoding each of the different features involved.

Overall, the brain regions which are thought to be crucially involved, and thus will be here explored, are the ventral visual path and the anterior temporal lobe. The ventral visual path, and in particular the left fusiform gyrus (Dehaene and Cohen, 2011), is not only involved in low level processing of the physical attributes of words, but it is also a good candidate for the encoding of visuo-perceptual semantic dimensions. First of all, the ventral occipito-temporal path (VOT) has been connected with the encoding of specific visual features, e.g. color (Beauchamp et al., 1999), and more generally it has been suggested it hosts the kind of computations that enable visuo-perceptual categorization (Grill-Spector and Weiner, 2014). Second, studies in the domain of object recognition have reported representational geometries tuned to perceptual semantic dimensions in VTO (Peelen and Caramazza, 2012; Devereux et al., 2013; Clarke and Tyler, 2014).

As we have seen in depth in Chap. 1, numerous converging evidences point to a key role of the ATL in high level semantic processing: from clinical data to neurophysiology, from univariate to multivariate analyses. Above all, the ATL appears to be conveniently connected, both structurally and functionally, to a distributed network of cortical regions (Binney et al., 2012; Pascual et al., 2015). Recently, it has been shown that ATL can be parcellated based on its structural connectivity with other key cortical areas (Papinutto et al., 2016), supporting the hypothesis of a graded specialization within the ATL as a consequence of its differential connectivity with modality specific cortical regions (Rice et al., 2015). It appears thus as ideal location for a semantic, supramodal hub.

1. 4 Present Study Hypotheses

We were interested in studying the representations evoked by purely symbolic stimuli, i.e. written words, spanning different semantic dimensions. We test the hypothesis that perceptual and conceptual dimensions of the word meaning, for which behavioral

studies suggest that they are automatically activated during word reading (Rubinsten and Henik, 2002; Zwaan et al., 2002; Setti et al., 2009), are coded partially independently in the brain. If that was the case, then we should observe brain regions of which the response profiles reflect dimension–specific metrics, resulting in a double dissociation: some areas should present activation patterns more consistent with the perceptual dimensions of the stimulus space and less with the more conceptual ones (e.g., size, but not taxonomic class), while other areas should present the complementary activation patterns (e.g., more related to taxonomic class and less to size).

We presented adult subjects with written words varying parametrically along three different dimensions (see Fig. 57): one low level, purely physical (the number of letters), one perceptual-semantic (the average real–word size of the objects referred to by the words), and one conceptual-semantic (at two levels of granularity, consisting in 2 semantic categories, each subdivided in 4 sub-categorical clusters). Our aim was to investigate to what extent the representational geometry of different regions along the ventral visual stream matched the dimension-specific cognitive representational geometry of the stimuli. We predicted that the visual–perceptual semantic dimension of the semantic space would be primarily encoded in early visual regions of the ventral stream (Pulvermuller, 2013), while the conceptual dimensions would be primarily encoded further anteriorly in the temporal lobe (Peelen and Caramazza, 2012).

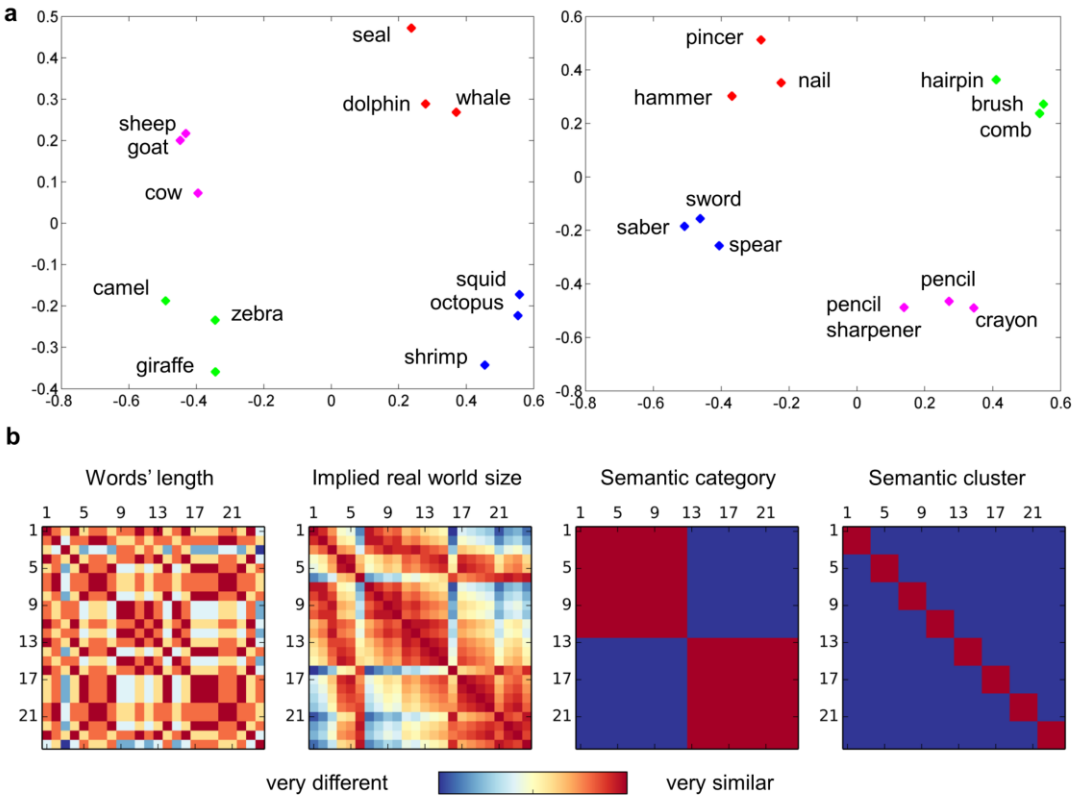


Figure 57 Word meaning describes a multidimensional semantic space. (a) The words used as stimuli in behavioral (a similarity judgment task and a feature generation task) and fMRI experiments. Multidimensional scaling technique was used to visualize the semantic distances perceived between the 12 words denoting animals (left) and the 12 words denoting tools (right). Four clusters of semantically close words are detectable in each of the two semantic categories: domesticated land animals, wild land animals, mammal sea animals, not–mammal sea animals, weapons, office/schools tools, work appliances, and hair instruments. Here shown: the MDS retrieved from the similarity judgment task. (b) Predicted similarity matrices modeling the similarities across stimuli along the four dimensions investigated. The words' length matrix depicts all pairwise differences in terms of number of letters between the stimuli. The implied real–world size matrix is built computing the distances in ranking position between all pairs of stimuli. The semantic category matrix indicates which pairs of stimuli belong to the same category (e.g. both animals) and which do not. The semantic cluster matrix designates which pairs of stimuli belong to the same semantic cluster (e.g. cluster of domesticated land animals: cow, sheep and goat) and which do not.

2. Materials and Methods

2.1 Subjects

Sixteen healthy adult volunteers (five males, mean age 30.87 ± 5.34) participated in the fMRI study. All participants were right–handed as measured with the Edinburgh handiness questionnaire, had normal or corrected–to–normal vision, and were Italian native speakers. All experimental procedures were approved by the local ethical committee and each participant provided signed informed consent to take part in the study. Participants received a monetary compensation for their participation. A seventeenth volunteer was

excluded from the analyses for not complying with the task (see Testing procedures).

2.2 Stimuli

In order to validate the target stimuli for the fMRI experiment (i.e. 24 words, 12 names of animals and 12 names of tools) we ran two behavioral experiments that involved 130 Italian native speakers, tested through internet–based questionnaires. These experiments are described in details in Chap. 3.1 and are here only briefly summarized.

In the first experiment, fifty subjects rated how similar the concepts indicated by the words were (Semantic Distance Judgment, n=50). In the second experiment, a group of new subjects listed between 5 and 10 characteristics or properties of each of the 24 target stimuli (Semantic Feature Listing, n=80). Data from both experiments were used as indicators of semantic proximity: two related concepts are closer in semantic space (semantic distance judgment) and share more features (features generation task). Separately for both experiments, mean distance matrices across subjects were computed and multidimensional scaling used to obtain a graphical representation of the cognitive semantic space of native Italian speakers. Results from the two experiments converge well in pointing to 4 sub-categorical clusters in each of the two categories. In the animals set the clusters were: domesticated land animals (cow, sheep, and goat), wild land animals (zebra, camel and giraffe), sea mammals (whale, dolphin and seal), and not–mammal sea animals (squid, shrimp and octopus). In the tools set the clusters were weapons (spear, saber and sword), office/schools tools (pencil, pastel, pencil sharpener), work appliances (hammer, nail, and pincer), and hair instruments (comb, brush, and hairpin).

Words belonging to the different semantic categories and clusters were well matched across several psycholinguistic variables such as number of letters, number of syllables, gender, accent and frequency of use (see Chap. 3.1).

2.3 Testing Procedures

In order to obtain a measure of the subject specific cognitive semantic space and verify the validity of the pre-defined clusters for the subjects participating in our fMRI experiment, we asked our participants to complete the same similarity judgment questionnaire as described above. The experimental session of the main experiment was divided into two parts: first, subjects underwent the fMRI experiment (being totally naïve with respect to the type of stimuli that were going to be presented), then they completed the similarity questionnaire. The analyses of the questionnaires followed the same steps as we used to validate the stimuli. To assess the consistency of each subject's judgement with the semantic space that had emerged from our prior behavioral experiments, we computed the correlation between the subject specific normalized distance matrix for animals and tools and the average ones obtained from the fifty subjects that had participated in the first behavioral study. Because one subject failed to comply with the instruction of the task (pressing the response keys according to a numerical progression (1, then 2, then 3, etc...) regardless of the pair of words presented), we excluded his data (both behavioral and fMRI) from further analysis. All sixteen remaining subjects showed highly positive and significant correlations with the behavioral group average: 0.84 ± 0.08 and 0.84 ± 0.10 for the animals and tools respectively. Because there was very little inter-subject variability in the ratings we decided that it was to worth applying a subject specific similarity space in the subsequent fMRI analyses.

During the fMRI experiment, subjects were instructed to silently read the target stimuli (i.e. 12 names of tools and 12 names of animals) and to perform semantic decisions only on extremely rare odd stimuli (Fig. 58). The odd stimuli appeared on average on 16% of the trials and consisted either in a picture or in a triplet of words referring to one of the targets, thus promoting both a depictive and a declarative comparison. Subjects pressed a button with the left or the

right hand to indicate whether the odd stimulus was related or not to the previously seen target word (1–back task). The hand–answer mapping was counterbalanced within subjects: half of the subjects answered *yes* with the left hand in the first half of the fMRI runs and then *yes* with the right hand in the last half; the other half of the subjects followed the reverse order. The triplets of words defining the target stimuli did not contain any verbs, in order not to stress the functional differences between animals and tools. Such a 1–back oddball task was orthogonal to the dimensions investigated (i.e., it did not consist in judging the items relative to their size, category, or cluster), and this allowed us to disentangle task–dependent processes from the spontaneous mental representations of the words (Cukur et al., 2013). Target stimuli were flashed in the center of the screen three times in a row (each time in a different font among Lucida Fax, Helvetica and Courier, to avoid adaptation): each presentation lasted 0.5 s and the interval between them was 0.2 s for a total of 1.9s for each target stimulus. The goal for this multiple flashed presentation was to ensure that subjects well read the word but at the same time did not have time to make eye movements. The inter target interval was randomly chosen between three values (1.7s, 1.8s and 1.9s, mean = 1.8 s). The odd events were presented differently according to their nature: images were shown for 2.0 s while definitions appeared as a series of three words, presented in a sequence, each for 0.5 s with an interval of 0.2 s between them. The interval after each odd event was randomly chosen between three values (1.7s, 2s and 2.3s, mean = 2s). The average accuracy in the oddball task was very high = 92.64% (missed = 2.06%, errors = 5.2 %). Within a given fMRI session, participants underwent 6 runs of 9 min and 40 sec each. Each run contained 4 repetitions of each of the 24 targets, 16 odd stimuli, and 24 rest periods (only fixation cross present on screen for 1.5s). Stimuli were completely randomized for each subject and each run, the only constraint being that odd stimuli would appear every 6-to-10 target stimuli. This ensures that, notwithstanding the (minimal) memory component of the task, we can exclude that the results reflect any

systematicity due to the stimulus sequence. They were presented with Matlab Psychophysics toolbox (http://psychtoolbox.org/).

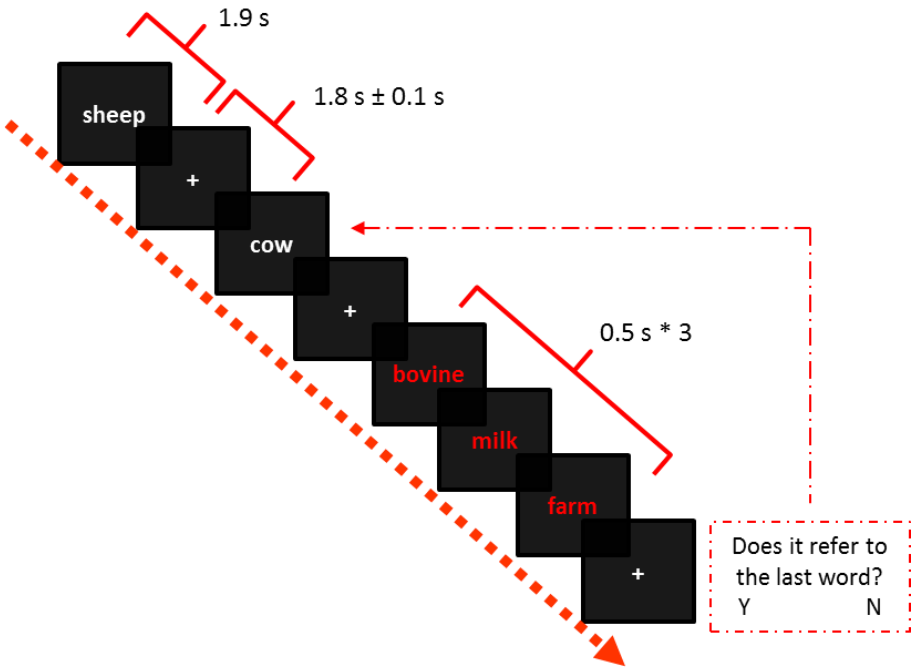


Figure 58 : Experimental setting. Example of a sequence of stimuli: during the fMRI experiment, subjects were instructed to silently read the target stimuli and to press a button at the presentation of rare odd stimuli. The odd stimuli consist either in a picture or in a triplet of words referring to one of the targets.

2.4 MRI and fRMI Protocols

Data were collected at Neurospin (CEA–Inserm/Saclay, France) with a 3 Tesla Siemens Magnetom TrioTim scanner using a 32–channel head coil. Each subject underwent one session that started with one anatomical acquisition followed by six functional runs. Anatomical images were acquired using a T_1–weighted MP–RAGE sagittal scan (voxels size 1x1x1.1mm, 160 slices, 7 minutes). Functional images were acquired using an echo–planar imaging (EPI) scan over the whole brain (repetition time = 2.3s; echo time = 23ms; field of view = 192mm; voxel size = 1.5x1.5x1.5mm; 235 repetitions; 82 slices, multi–band acceleration factor 2, GRAPPA 3)(Feinberg et al., 2010; Moeller et al., 2010). The acquisition used a phase encoding direction from posterior to anterior (PA) and an inclination of –20° with respect to the subject's specific AC/PC line.

2.5 Data Pre-Processing and First Level Model

Pre–processing of the raw functional images was conducted with Statistical Parameter Mapping toolbox (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). It included realignment of each scan to the first of each given run, co–registration of anatomical and functional images, segmentation, and normalization to MNI space. No smoothing was applied. For each subject individually, functional images were then analyzed within the framework of a general linear model (GLM). For each of the 6 runs, 35 regressors were included: 24 regressors of interest (corresponding to the onset of the 12 names of animals and 12 names of tools), 4 regressors of no–interest (corresponding to the onset of the odd events – definitions and images – subdivided into those receiving a left hand vs right hand response from the subject), 6 head–motion regressors (i.e. the six–parameter affine transformation estimated during motion correction in the pre–processing) and 1 constant. Fixation baseline was modeled implicitly and regressors were convolved with the standard hemodynamic response function without derivatives. Low–frequency drift terms were removed by a high–pass filter with a cutoff of 128s. Thus, one beta map was estimated for each target event (i.e. words stimuli) for each run. Both subsequent multivariate analyses – decoding and RSA – had as input data the 24 x 6 beta maps corresponding to the target stimuli normalized across conditions separately run by run (i.e. within each run the values for each given voxel were normalized across conditions to have zero mean and unit variance).

2.6 Regions of Interest

Given our hypothesis and the absence of principled functional localizers, to avoid circularity regions of interests (ROIs) were defined only based on anatomical criteria thanks to SPM toolbox PickAtlas (Fig. 59). Proceeding from the occipital lobe to the anterior temporal lobe (ATL), we selected six Brodmann areas along the ventral visual

pathway: BA 17 – primary visual area, BA 18 – secondary visual areas, BA 19 – lateral and superior occipital gyri, BA 37 – occipito–temporal cortex (includes the posterior fusiform gyrus and the posterior inferior temporal gyrus), BA20 – inferior temporal gyrus, and BA 38 – temporal pole. We included homologue areas from both hemisphere and the average number of voxels of each ROI were: BA17 (13940 voxels), BA18 (69617 voxels), BA19 (65248 voxels), BA37 (65248 voxels), BA20 (28026 voxels), BA38 (27254 voxels). Given the known signal drop out problems in ATL and following previous similar studies (Peelen and Caramazza, 2012), for each subject we calculated the signal–to–fluctuation–noise–ratio (SFNR) map by dividing the mean of the time series (of the first run) by the standard deviation of its residuals once detrended with a second order polynomial (Friedman et al., 2006). This analysis was carried out with the python library nipype (http://nipy.org/nipype). We then computed the average SFNR in each of our ROIs and verified that in all regions this value was above the value of 20 which is usually considered to be the limit for meaningful signal detection (Binder et al., 2011). The average SFNR across the 16 subjects for BA17 was 49.76 ± 5.63, BA18 = 49.34 ± 4.89, BA19 = 52.76 ± 4.7, BA37 = 42.78 ± 3.65, BA20 = 32.87 ± 2.69, and BA38 = 30.99 ± 2.43.

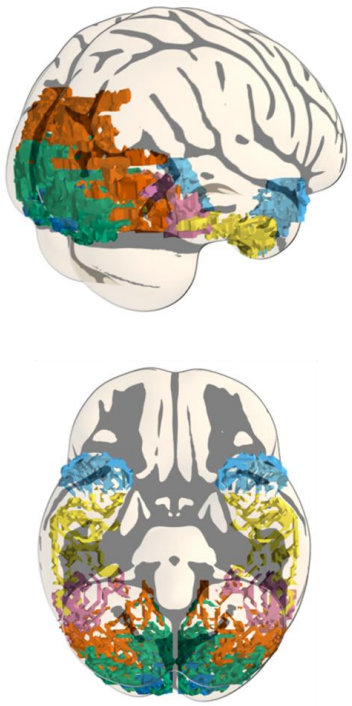


Figure 59 Regions of interest. ROIs were defined based on anatomical criteria. Proceeding from the occipital lobe to the temporal pole: Brodmann area 17 (primary visual area), Brodmann area 18 (secondary visual areas), Brodmann area 19 (lateral and superior occipital gyri), Brodmann area 37 (occipito–temporal cortex), Brodmann area 20 (inferior temporal gyrus), and Brodmann area 38 (temporal pole).

2.7 Univariate Analyses

For the univariate analyses only, beta maps were smoothed (kernel [4,4,4]). First, two random effects analyses were run searching for regions in which activity was linearly modulated by length of words and implied real world size. Second, random effects analysis was applied to the contrast animals vs tools. Unsurprisingly, the only significant result was a linear effect of length of words in 5 occipital clusters (extent threshold = 100 voxel, $p<0.001$ FEW corrected) comprising primary and secondary visual cortices. This is in line with the literature on categorical effects in the ventral stream that shows

less consistent results when words stimuli are used (as compared with pictures) [for a recent review on the topic: (Bi et al., 2016)].

2.8 Multivariate Pattern Analyses

None of the semantic variables of interest resulted in a dissociation at the univariate analysis level, thus we used multivariate pattern analysis (MVPA) which investigates differences in the distributed patterns of activity over a given cortical region (Davis and Poldrack, 2013). In this framework, the decoding approach aims at predicting one or more classes of stimuli (i.e. "classification problem") or a continuous target (i.e. "regression problem") based on the pattern of brain activation elicited by the stimuli. The models are fitted on part of the data (i.e. train set) and tested on left–out data (i.e. test set). Previous studies of semantic representations used this method to decode the semantic category of words from brain activations patterns, and generalize this categorical discrimination across different input formats (from pictures to words and vice versa) (Shinkareva et al., 2011; Simanova et al., 2014). These studies, however, are limited because: (1) they evaluate the decoding model on the full brain volume, which fails to provide evidence in favor or against the differential contribution of different regions in coding sensory and/or conceptual information (Shinkareva et al., 2011), or (2) they contrast two broad semantic categories (i.e. animals vs tools), without investigating which dimensions of the meaning of the words (i.e. conceptual vs perceptual) drove the observed discriminations (Simanova et al., 2014). A second approach, representational similarity analysis (RSA) (Kriegeskorte et al., 2008), compares the similarity between different stimuli and the one observed between the multivoxel activation patterns elicited by them (i.e. neural similarity). To our knowledge, this approach was deployed only a few times to investigate the processing of symbolic stimuli (words), and no one investigated at the same time the organization of concepts inside and across semantic categories (Bruffaerts et al., 2013; Devereux et al.,

2013). Contrary to previous studies, we estimated the similarity of our stimuli considering multiple dimensions at the same time: a low-level physical dimension (number of letters), and three semantic dimensions (a perceptual–semantic: the size of the objects referred to by the words, and two conceptual–semantic dimensions: the category and sub–categorical cluster). An advantage of RSA is that it permits the investigation of the neural coding of several different dimensions even when those are partially correlated in the stimuli. For example, in the case of our stimuli there was a correlation between semantic category and implied–real world size, in that the implied real world size of the animals was on average larger than that of tools. Using partial correlation as the association metric within RSA (hereafter "partial correlation RSA"), we are robust to the effect of one dimension (e.g. size) while testing for the correlation between the other dimension (e.g. category) and the neural similarity in a given region (Clarke and Tyler, 2014).

Decoding models. We used two different decoding models to solve our four different prediction problems. First, to predict the number of letters composing each word, we applied a regression model in all ROIs. The chosen model was a Ridge regression (linear least squares with l2–norm regularization). The regularization parameter was selected by a nested cross–validation loop. Given the ordinal nature of our problem (i.e. what matters is the rank position, not the absolute value) the metric used to assess the prediction quality was the Kendall rank correlation coefficient (or Kendall tau). The same regression model was used to predict the averaged implied real world size of the objects referred to by the words: all animals and tools where ranked, regardless of their semantic classification, from the smallest (i.e., pencil sharpener) to the biggest (i.e. whale). The ranking scale was devised by the authors considering the average size of the items. When possible, we used information from encyclopedias; when that information was not available, each author gave an approximate estimate and ranked the items independently; it was then

verified that the ranks converged [the rank of the items can be found in the supplementary table 1]. Given that in our set of stimuli the object sizes increased logarithmically, the rank, which we used as our size metric, is equivalent to the logarithm of the sizes (correlation between the ranks and the log of the sizes r2 = 0.98).

To solve the binary classification problem related with the semantic category (i.e. decode whether a given beta map corresponded to an animal or a tool word) we used a support vector machine (SVM) model with linear kernel. The loss function chosen was squared hinge loss with l2–norm regularization and, again, the regularization parameter was selected by a nested cross–validated loop. Finally, the same model was applied to solve the multiclass problem using a one–vs–rest scheme.

For all decoding models, we report the cross–validation scores computed by averaging the scores of 5 folds with a leave–one–run–out scheme: within each subject data from five out of six runs were used to fit the model and data from the held out run were used to test it. The group–level results were then computed averaging the scores obtained by each subject, and their significance was tested against the empirically estimated random distribution. To obtain such a distribution, the procedure used to obtain the group results was repeated 10.000 times randomly permuting the labels.

The same regression and classification models were fed with the stimuli themselves (i.e., the matrices of 0 and 1 representing the physical appearance of the words used during the experiment, averaging across the three fonts used) to rule out that any of our results could be explained by some low–level characteristic of the stimuli. The goal here is to show that in the stimuli themselves there is already enough information to decode the low level physical dimensions (i.e., number of letters), but not higher level semantic dimensions (nor the perceptual one – size, nor the conceptual one – category and cluster), thus showing that what is retrieved from the patterns of brain activity is not due to any low level property of the stimuli used.

All the analyses described in this section were conducted with the machine learning library in Python Scikit–Learn (http://scikit-learn.org).

RSA. The first step of representational similarity analysis was the modeling of predicted similarity matrices corresponding to the different dimensions investigated. Concerning word length the matrix was built computing the pairwise absolute difference in number of letters between every word pair (the simplest measure of visual similarity). For instance, the entry corresponding to *sheep* (n° of letters = 5) vs *cow* (n° of letters = 3) would contain a |5–3| = |2|. The same strategy was applied to the implied real size ranking scale: the entry corresponding to *whale* (position in ranking = 24) vs *pencil sharpener* (position in ranking = 1) would contain a |24–1| = |23|. These first two matrices show distances (i.e. dissimilarity) thus in order to be compared with the neural similarity matrices, their values need to be inverted (similarity = 1 – dissimilarity). As to the conceptual dimensions of our stimuli, two matrices were built: one depicting the two semantic categories and one describing the eight clusters that had emerged from the behavioral study. The first one had 1 for all entries of the same category (i.e. all identical combinations: two animals or two tools) and 0 everywhere else (i.e. all different combinations: an animal and a tool). The semantic cluster matrix was built likewise, thus having 1 for all combinations of items from the same cluster and 0 everywhere else. The four matrices being symmetrical (Fig. 57b), they were vectorized discarding the diagonal and keeping only the upper half, then standardized to have mean 0 and standard deviation 1. It should be noted that there is a significant correlation between the similarity matrix of size and the ones of semantic category ($r = 0.39$, $p<0.001$) and semantic cluster ($r = 0.27$, $p<0.001$), due to the fact that animal–words tend to refer to big items and tool–words tend to refer to small items. There is, clearly, a correlation between the predicted similarity matrix representing the two semantic categories and the one describing the 8 semantic clusters

($r = 0.32$, $p<0.001$). Importantly, there is no significant correlation between the predicted similarity matrix for length and the ones for size ($r = 0.04$, $p = 0.49$), category ($r = 0.06$, $p = 0.32$), or cluster ($r = -0.002$, $p=0.97$).

In order to retrieve the neural similarity matrices, for each subject and in each ROI, we built a vector with all the voxels' values for a given stimulus (i.e. from a given beta map). The six stimulus-specific vectors were averaged and all pairwise correlations between vectors were computed (by means of Pearson's correlation). The 24x24 neural similarity matrix obtained was then vectorized as done for the predicted similarity matrices. We obtain thus four vectors (denoted as X_L, X_S, X_C and X_k) from the predicted similarity matrices and one (denoted as Y) from the neural similarity matrix. In order to directly test our hypothesis, we need to be able to estimate the contribution of each single predicted similarity matrix (e.g., X_k) to the neural one (Y) while controlling for the effect of the other ones (e.g., X_L, X_S, X_C). Expressing the neural similarity vector as a linear combination of the predicted similarity vectors plus a noise term, we are interested in testing the null hypothesis that the partial regression coefficient of a given predicted similarity matrix is not significantly different from zero. That is, given the model $Y = \beta1X_L + \beta2X_S + \beta3X_C + \beta4X_K + \varepsilon$ where ε is a vector of residuals, we would like to test the null hypothesis H_0: $\beta_i \neq 0$ (where i can take the values {1, 2, 3, 4}). The test statistic we used for this hypothesis is the partial correlation between all pairs of Y and X (e.g., Y and X_k), controlling for the remaining variables Z (e.g., X_L, X_S, X_C). The partial correlation of two vectors Y and X while controlling for Z is given as the correlation between the residuals RX and RY resulting from the linear regression of X with Z and of Y with Z, respectively. Since the distribution of this statistic is unknown, we choose to obtain the significance level using a permutation test (Anderson and Robinson, 2001). Thus, for each subject and each ROI, we computed the partial correlation between the neural similarity matrices and each predicted similarity matrix (controlling for all the others). The observed result of

size is thus corrected for the potential residual correlation between the neural signal and length, category and cluster, the one of category is corrected for length, size and cluster, and so on. Then, scores from all the subjects were averaged and the significance of the group–level results was tested against the empirically estimated random distribution similarly to what has been done for the decoding models. Two features of partial correlation RSA should be noted. . First, because it is based on Pearson correlations, partial RSA assumes linear relations between the variables, therefore the inferences might not be valid if a strong non-linearity underlies the relationship between the physical/cognitive variables and the patterns of brain activation. This issue will need to be tackled in the future to further refine this type of RSA analysis. Second, from a neurobiological point of view, the use of partial RSA can elucidate whether multiple (and partially correlated) features of the stimuli can be independently encoded in the same (set of) brain regions. We think that this question is legitimate, especially in light of the fact that pure functional selectivity (i.e., a brain region in which neurons are solely involved in coding one specific stimulus feature) is clearly not a feature of our brain. It is however necessary to remember that the observation of an interaction between brain region and feature would not imply that a given feature (e.g., size) is solely represented in a given brain region (e.g., visual areas). It would only indicate that there is more residual signal related to a given feature in one area compared to the other. Such results could reflect the fact that more neurons code for one feature in one area than in another one. Alternatively, it may suggest that the different features are encoded with a different degree of precision across areas. The current methods do not allow differentiating across these scenarios: detailed electrophysiological studies might be useful to address the question.

All the analyses described in this section were conducted with in–house python scripts.

2.9 Additional Analyses

We performed five supplementary analyses. First, in order to demonstrate that our semantic effects (especially those that we could recover from activity in early visual regions) could not be explained by information present in the physical appearance of the stimuli themselves, we applied all the aforementioned decoding and partial correlation RSA analyses to the images of the stimuli (i.e. the snapshots of the screens with the words we presented to the subjects during the fMRI experiment).

Second, to better qualify the effect of size as separated, thus independent from the effect of length, even though there was no significant correlation between the predicted similarity matrix for length and size ($r = 0.04$, $p = 0.49$), nor between length and size across the stimuli themselves ($r=0.38$, $p.=0.06$), we re-run the partial RSA analyses on a subset of words by removing the two more extreme words length-wise (the shortest and the longest, one animal ("FOCA") and one tool ("TEMPERINO")). This further reduced the already non-significant correlation across Length and Size in our stimuli (down to $R=0.27$ ($p.=0.21$)), and the respective distance matrices (down to $R=-0.03$ ($p.=0.5$)).

Third, to better qualify the presence of different gradients along the ventral stream, we tested for an interaction between the 3 different semantic dimensions (size, category, cluster) and our ROIs by feeding subjects' partial correlation scores (once Fisher r-to-z transformed) into an ANOVA (6 ROIs x 3 dimensions), and then performed trend analyses with SPSS (http://www.ibm.com/analytics/us/en/technology/spss/), testing for a linear, a quadratic, a cubic, a 4-th and a 5-th order term for each of the 3 dimensions.

Forth, to verify the impact of the partial correlation RSA (vs. standard RSA), we also computed, for all predicted matrices and ROIs, standard Pearson correlation (standard RSA), assessing their significance with permutation tests.

Fifth, to investigate whether the effects were lateralized, we run an additional partial correlation analysis on the same ROIs but separately for the right and left hemisphere.

Finally, we attempt to study whole brain activity (via a searchlight with partial correlation RSA) and to determine whether a more fine grained representations of the semantic distance could be appreciated (using the continuous scale obtained from the behavioral experiments instead of the binary classification has belonging to a given semantic cluster or not). These exploratory analyses, whose results overall confirm our general findings, are reported in the Appendix 1.2..

3. Results

In each ROI we applied different MVPA models tailored to our variables and cognitive questions. Firstly, we used decoding to predict: the number of letters composing each word and the relative implied real–size (using the rank from the smallest to the biggest item, approximatively equivalent to the logarithm of the real size), through a regression model; and the conceptual–semantic dimensions at two different scales, that of the semantic category and that of a finer–grained semantic cluster, through a binary classification and a multi–class classification model. We then further qualified the results through partial correlation RSA, and compared the pattern of fMRI activations to words with those predicted by the similarity of the stimulus conditions along the aforementioned dimensions. Extremely low p-value are rounded to $p < 10\text{-}5$ and all p-values inferior to 0. 0083 survive Bonferroni correction for multiple ROIs comparisons ($p = 0.05/6$ areas $= 0.0083$).

3.1 Physical Dimension: number of letters

The number of letters composing each word could be successfully predicted by a regression model in the early visual regions BA17 (mean score = 0.45, $p<10^{-5}$), BA18 (mean score = 0.31, $p<10^{-5}$) and BA19 (mean score = 0.21, $p<10^{-5}$). Likewise, the neural similarity computed from the pattern of activation of these areas significantly correlated with the predicted similarity matrix modelling the difference in number of letters between each word pair: BA17 (mean score = 0.35, $p<10^{-5}$), BA18 (mean score = 0.13, $p<10^{-5}$) and BA19 (mean score = 0.06, $p<10^{-5}$). More anterior temporal regions ceased to reflect such physical dimension of the visual stimulus, in line with the expected increasing invariance to physical dimensions along the ventral stream. These results are therefore a sound sanity check for our models (Fig. 60).

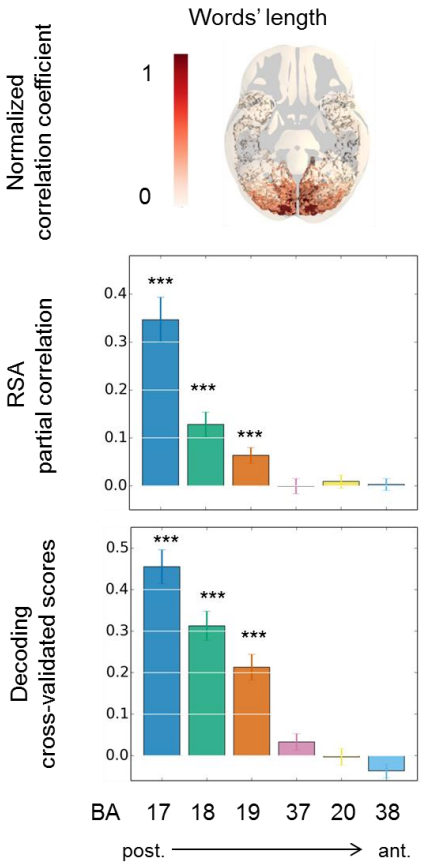


Figure 60 Low level stimuli representation. Results concerning the physical dimension of our stimuli (length of the words). Lowermost: the regression model applied (scoring metric: Kendall tau) was able to predict the number of letter composing each word in primary and secondary visual areas. Middle: the partial correlation between neural similarity matrix and length of words matrix is significant in primary and secondary visual areas (while controlling for the other three dimensions investigated). Uppermost: in a template brain, the six ROIs are colored according to the normalized partial correlation scores, highlighting how the effect of the purely physical dimension is confined in occipital visual areas. We are showing the average scores across subjects (n°=16) and error bars indicate the s.e.m.. Statistical significance (* $p < 0.05$, ** $p < 0.001$, *** $p < 10\text{-}5$) is computed with a permutation test and very low p-value are rounded to $p < 10\text{-}5$. Exact p-values are reported in the text and ** / *** survive Bonferroni correction ($p = 0.05/6$ areas $= 0.0083$).

3.2 Perceptual–Semantic Dimension: implied real word size

We then investigated the brain code for the real–world size of the objects referred to by the words, to which we refer to as a perceptual–semantic dimension (see Fig. 61a). A regression model with the rank of the sizes (equivalent to the log of the sizes) permitted above chance prediction of the relative size in BA17 (mean score = 0.07, p=0.0006), BA18 (mean score = 0.05, p=0.0086), BA19 (mean score = 0.09, p<10–5), and BA37 (mean score = 0.04, p=0.0086). Because in our stimuli implied real–word size and semantic category were correlated (on average, tools were smaller than animals) using decoding we were unable to determine if the source of the information used by the decoder to solve the implied real–world size regression problem was indeed related to the implied–real world size, to the semantic category, or both. The partial correlation RSA, on the contrary, could provide such information. Once we accounted for the conceptual effects (semantic

category and cluster), the similarity in the implied real–world size significantly correlated with the neural similarity observed in primary visual areas (BA17, mean score = 0.06, p<10–5) and then progressively decreased in more anterior areas (BA18, mean score = 0.02, p= 0.0537, and BA19 mean score = 0.03,p= 0.0484) (see Fig. 61a).

3.3 Conceptual–Semantic Dimensions: semantic category and cluster

Next, we tested more conceptual aspects of our stimuli (see Fig. 61b–c): the semantic category (i.e. animals vs tools) and the sub–category semantic clusters (e.g. domesticated animals vs. wild animals). A binary classification model was able to predict above chance the words' semantic category in four occipito–temporal ROIs: BA17 (mean score = 0.54, p=0.0008), BA18 (mean score = 0.53, p=0.0055), BA19 (mean score = 0.57, $p<10^{-5}$), BA37 (mean score = 0.56, $p<10^{-5}$). Again, because of the correlation between semantic category and size, these results were further qualified by partial correlation RSA, which showed that category membership was increasingly correlated with brain activation as we moved along the ventral path from posterior to anterior regions (BA18 mean score = 0.02, p=0.0558, BA 19 mean score = 0.03, p=0.0099), independently from the residual code for size, reaching the peak in BA37 (mean score = 0.05, p=0.0004). Finally, using a multiclass classification model we could decode the subtle semantic clustering of our words in five ROIs: BA17 (mean score = 0.14, p=0.0126), BA18 (mean score = 0.13, p=0.0148), BA19 (mean score = 0.16, $p<10^{-5}$), BA37 (mean score = 0.14, p=0.0295), BA20 (mean score = 0.15, p=0.0001). These results were further qualified by partial correlation RSA, which showed that semantic cluster membership, once accounted for the other dimensions, was represented in the most anterior areas of the temporal lobe (BA19 mean score = 0.04, p= 0.006, BA37 mean score = 0.03, p= 0.0081), peaking in BA20 (mean score = 0.06, p<0.05).

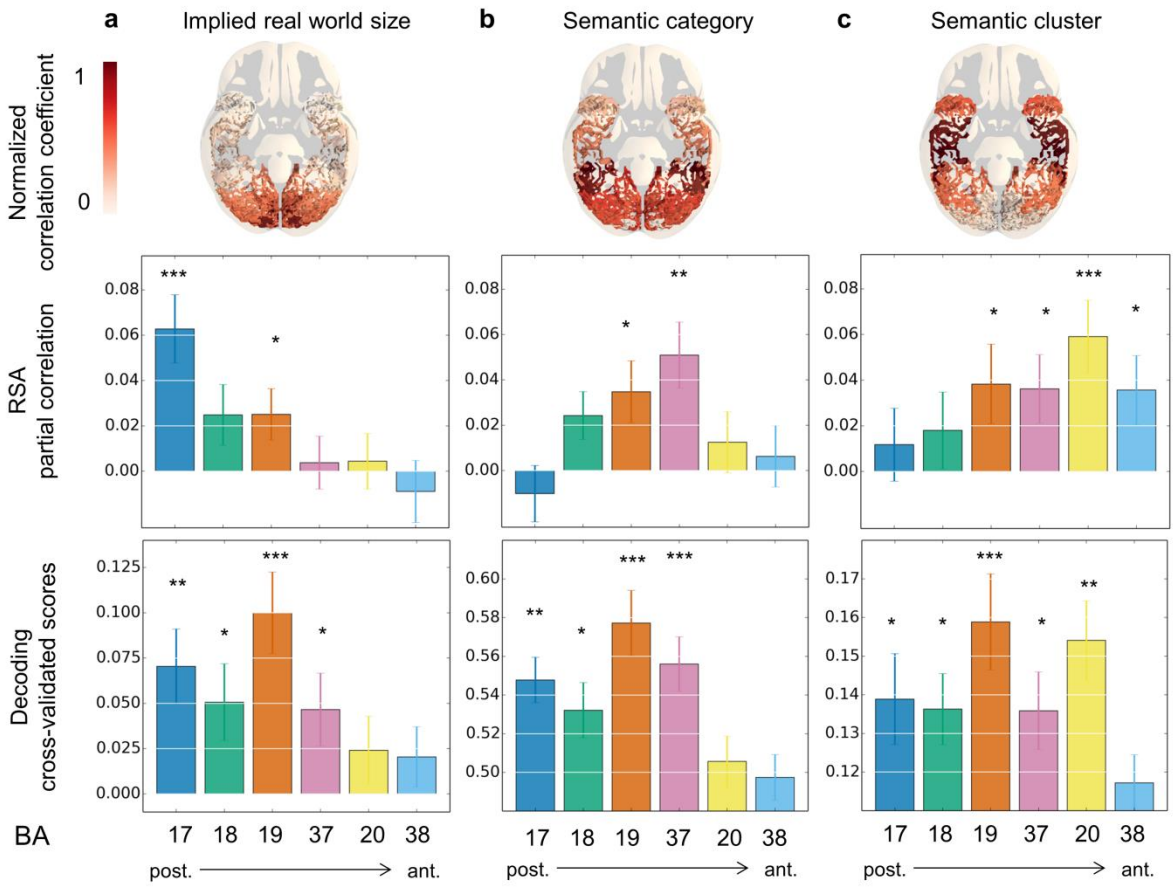


Figure 61 Topography of perceptual and conceptual representations in the ventral path. (a) Lowermost: the regression model (scoring metric: Kendall tau) was able to predict above chance the implied real-world size in four occipito-temporal areas. Middle: the partial correlation between neural similarity matrix and real–world size matrix, while controlling for the other dimensions, is significant in primary visual areas (BA17). Uppermost: the six ROIs are colored according to the normalized partial correlation scores, highlighting how the effect of the perceptual dimension is confined in occipital visual areas. (b) Lowermost: the binary classification model was able to predict above chance the semantic category in four occipito-temporal areas (from BA17 to BA37). Middle: the partial correlation between neural similarity matrix and semantic category matrix is significant in the occipito–temporal cortex (BA19 and BA37). Uppermost: information about semantic category appears to be coded in occipito-temporal areas, anteriorly respect to the implied real-world size and posteriorly respect to the semantic cluster. (c) Lowermost: the multi-classification model was able to predict above chance the semantic cluster in five occipito-temporal areas (from BA17 to BA20). Middle: the partial correlation between neural similarity matrix and semantic cluster matrix is significant in anterior areas, from BA19 to BA38, peaking in BA20. Uppermost: the effect of semantic cluster gets progressively higher the more anterior the areas considered. We are showing the average scores across subjects (n°=16) and error bars indicate the s.e.m.. Statistical significance (* $p < 0.05$, ** $p < 0.001$, *** $p < 10\text{-}5$) is computed with a permutation test and very low p-value are rounded to $p < 10\text{-}5$. Exact p-values are reported in the text and ** / *** survive Bonferroni correction ($p = 0.05/6$ areas $= 0.0083$).

3.4 Controls on Low Level Physical Dimensions

In order to demonstrate that our semantic effects (especially those that we could recover from activity in early visual regions) could not be explained by information present in the physical appearance of the stimuli themselves, we applied all the aforementioned decoding and partial correlation RSA analyses to the images of the stimuli (i.e. the snapshots of the screens with the words we presented to the subjects during the fMRI experiment). Unsurprisingly, the only dimension that this analysis could recover from such input was the number of letters composing each word: decoding score = 0.74, $p<0.001$; RSA score = 0.23, $p<0.001$ (for implied real world size: decoding score = 0.12, $p=0.28$; RSA score = -0.01, $p=0.62$, for semantic category: decoding score = 0.11, $p=0.30$; RSA score = 0.05, $p=0.18$, for cluster category: decoding score = 0.08, $p=0.33$; RSA score = -0.05, $p=0.82$).

We also explored if the variations in word length could explain the effect of size in early visual areas. Although the predicted similarity matrices for length and size were not correlated with each other, because the effect of word length was very strong compared to that of size, as a further control aiming at reducing the variability in length across our stimuli we re-run the partial correlation analyses of size eliminating two stimuli, corresponding to the longest (4 letters) and the shortest (9 letters) words. This partial correlation RSA testing for the effect of size (corrected for length, category and cluster) was smaller compared with the one run on the full set of stimuli, but it remained significant ($p. < 0.05$) in BA17. As for the original analysis, this effect disappeared in more anterior regions.

3.5 Interaction between Semantic Dimensions and ROIs

Our findings illustrate two clear postero–anterior gradients in the neural response profile of the ventral visual path: posterior occipital regions appear as coding for the visuo-perceptual semantic property of the items (the implied average real word size), irrespective to their semantic category, while as we moved anteriorly in the ventral stream, mid-anterior temporal regions discriminate first between semantic categories and further anteriorly between sub-categorical cluster in a way that is insensitive to their visuo-perceptual property of size. Such an interaction between semantic dimensions and our ROIs was explicitly tested with an ANOVA (6 ROIs x 3 dimensions). The results was highly significant: $F(10,150)= 4.48$, $p<0.001$, corroborating the differential contribution of perceptual and conceptual semantic dimensions to the pattern of brain activity in occipital and temporal areas (see Fig. 62). Across the six ROIs, the three effects develop according to different trends: implied real world size shows a significant (decreasing) linear trend ($F(1,15) = 23.92$, $p<0.0001$); semantic category a significant quadratic trend ($F(1,15) = 15.97$, $p=0.001$); semantic cluster a marginal (increasing) linear trend ($F(1,15) = 3.59$, $p=0.07$), not significant likely due to the loss of signal / increased noise in BA38).

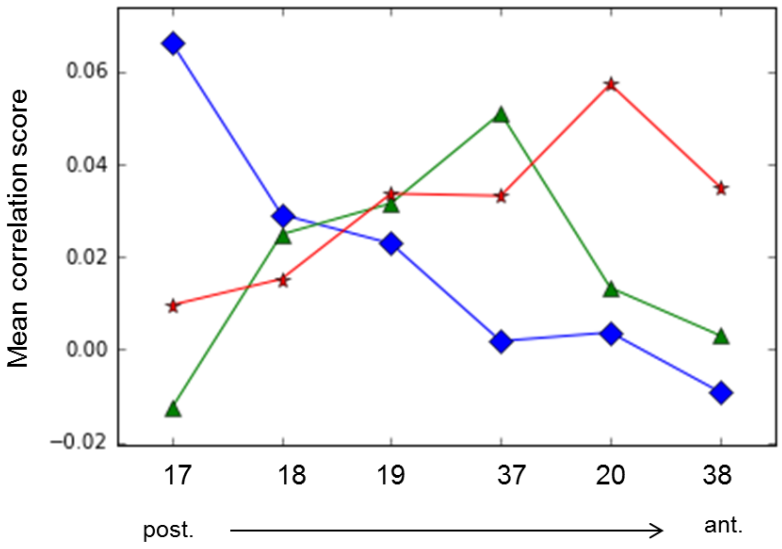


Figure 62 Interaction between the 3 semantic dimensions and the 6 ventral ROIs. For each semantic dimension, the average partial correlation score across subjects (n°=16) is plotted as a function of the different Brodmann areas investigated. Implied real world size (in blue) follows a decreasing trend as one moves from posterior (BA17) to anterior (BA38) areas. Semantic category (in green) and semantic cluster (in red) show the opposite trend.

3.6 Standard Pearson Correlation RSA

Second, we verified the impact of the use of partial correlation in RSA, and thus run the "standard" Pearson correlation RSA. This revealed a pattern very close to decoding: due to the relation between implied real world size and semantic category/cluster the three effects are intermingled and result in a less clean gradient from physical (length of words: BA17 mean score = 0.34, p<10–5, BA18 mean score = 0.12, p<10–5, BA19 mean score = 0.06, p=0.0202) and perceptual (implied real world size: BA17 mean score = 0.62, p=0.0133), to conceptual (semantic category: BA37 mean score = 0.05, p=0.0446; semantic cluster: BA20 mean score = 0.05, p=0.0407) (see Fig. 63).

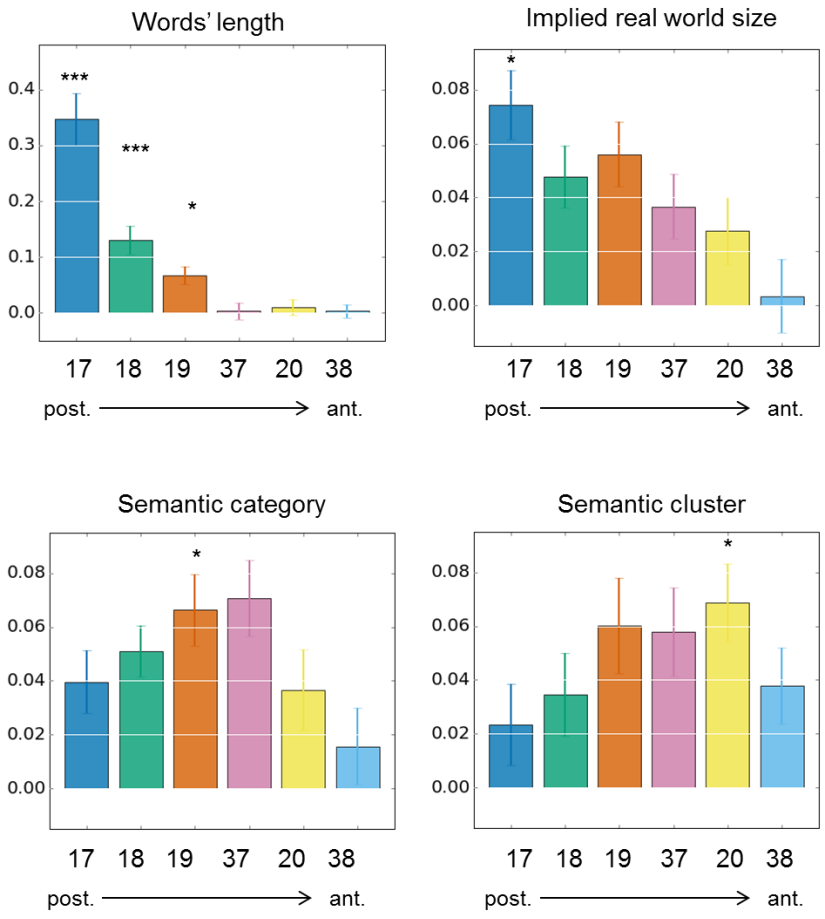


Figure 63 Pearson correlation results. Results for RSA with standard Pearson correlation (not partialling out other variables) for the four dimensions investigated: length of words, implied-real world size, semantic category and semantic cluster. We here show the average scores across subjects (n°=16) and error bars indicate the s.e.m.. Statistical significance (* p < 0.05, ** p < 0.001, *** p < 10-5) is computed with a permutation test and very low p-value are rounded to p < 10-5; ** / *** survive Bonferroni correction (p = 0.05/6 areas = 0.0083).

3.7 Lateralization of the Effects

Finally, when our ROIs were split in left vs right, the profile of the 4 effects followed the same trend bilaterally: moving from posterior to anterior along the ventral path physical (i.e., length of words) and perceptual (e.g., implied real world size) effects decrease, while conceptual ones (i.e., semantic category and cluster) increase (see Fig. 64). On the left hemisphere, length of words: BA17 mean score = 0.33, $p<10^{-5}$, BA18 mean score = 0.14, $p<10^{-5}$, BA19 mean score = 0.05, p=0.0001; implied real world size: BA17 mean score = 0.04, p=0.0018, BA19 mean score = 0.03, p=0.0102; semantic category: BA18 mean score = 0.03, p = 0.0112, BA19 mean score = 0.03, p = 0.012, BA37 mean score = 0.06, p<10–5; semantic cluster: BA19 mean score = 0.03, p =0.0048, BA37 mean score = 0.04, p =0.0029, BA20 mean score = 0.05, p =0.001. On the right hemisphere, length of words: BA17 mean score = 0.25, $p<10^{-5}$, BA18 mean score = 0.09, $p<10^{-5}$, BA19 mean score = 0.06, $p<10^{-5}$; implied real world size: BA17 mean score = 0.06, p<10–5, BA18 mean score = 0.02, p=0.0499; semantic category: BA19 mean score = 0.03, p = 0.0234; semantic cluster: BA19 mean score = 0.03, p = 0.0134, BA20 mean score = 0.05, p = 0.0001, BA38 mean score = 0.04, p =0.0036. It should be noticed that having now 12 ROIs, the Bonferroni correction threshold is now 0.004 (p = 0.05/12 areas = 0.004).

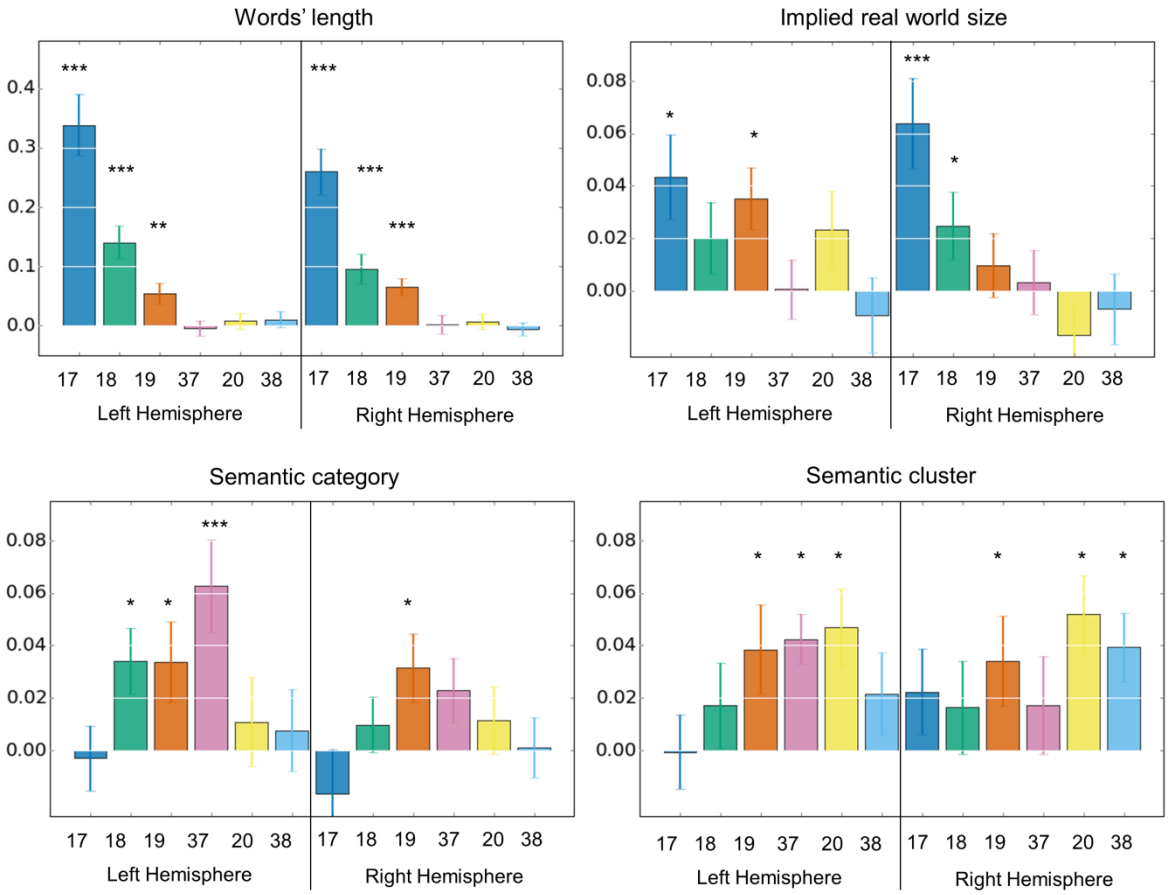


Figure 64 Lateralization of the effects. Results for the four dimensions investigated (length of words, implied-real world size, semantic category and semantic cluster) in the six ROIs of the left and right hemisphere respectevely. We here show the average scores across subjects (n°=16) and error bars indicate the s.e.m.. Statistical significance (* $p < 0.05$, ** $p < 0.001$, *** $p < 10\text{-}5$) is computed with a permutation test and very low p-value are rounded to $p < 10\text{-}5$; ** / *** survive Bonferroni correction ($p = 0.05/6$ areas $= 0.0083$).

4. Discussion

This study investigated the semantic representation of word meaning along the ventral visual path during silent reading and tested the hypothesis that perceptual semantic features of the objects referred to by the words are encoded in brain regions that are partially segregated from those encoding conceptual semantic features. Our task, orthogonal to the dimensions of the semantic space we investigated, ensured that subjects processed the words at an individual level (as opposed to the category or cluster level), and that the representations recovered in the brain activation emerged spontaneously. Furthermore, since we used words instead of pictures as stimuli, our results are free from any possible low-level confound due to visual shape similarity (Rice et al., 2014). We used a combination of multivariate decoding and partial correlation RSA. In

fact, while decoding only tests for the possibility to discriminate classes (without directly assessing in which aspects those classes differ), partial correlation RSA directly tests for the contribution of a given representational geometry onto brain activity.

Implied real–world size information in primary visual areas

One surprising result of this study is that, during reading, early visual areas appear to contain information relative to at least one perceptual–semantic dimension of word meaning: the implied real–world size of the items they refer to (Fig. 61). This information, however, is progressively lost towards anterior temporal regions, which become more progressively involved in encoding more abstract information such as semantic category and sub-categorical cluster. Not surprisingly, if one had to look only at non-partial correlation RSA or decoding, one would have observed much more distributed effects, with size reaching significance also in more anterior areas and category also in more posterior ones. Having run partial RSA, however, we now know that this would have been a spurious effect due to the correlation between size and semantic category and cluster. Partial correlation RSA gives us a cleaner picture of the contribution of this perceptual dimension once accounting for the conceptual ones. In this respect, it is to be noted that the surprising effect of size in early visual areas was also present when we corrected for the effect of word length, which, even though not significantly correlated with size (neither at the level of the raw values nor at the level of the similarity matrices) was not entirely un-related to it. Further, we could retrieve size-related information in BA17 even after we removed from the analyses the two words that were most greatly variable in length. These results suggest that early visual areas play a role in semantics, and not only in low–level vision. They are coherent with recent studies indicating that activity in primary visual cortex contains perceptual information even in the absence of sensory stimulation (e.g., the prototypical color of objects presented as a gray-scale image)

(Bannert and Bartels, 2013) or in presence of ambiguous stimuli (Vandenbroucke et al., 2014). Moreover, calcarine cortex has been shown to allow distinction of words semantically related with visual properties (e.g., "shinny") vs auditory properties (e.g., "loud") (Murphy et al., 2016). Our results also relate to the literature on mental imagery, which indicates commonalities between the neural substrates of perception and of imagery (Farah, 1992; Kosslyn, 2001; Smith and Goodale, 2014). In our experiment we neither explicitly prompted the use of mental imagery nor did we inhibit it, thus we are neutral with respect to the issue of whether the observed effects were related to imagery or not. One way to approach the question in the future would be to directly compare the neural representational geometries in early visual cortices during reading (i.e. reading names of objects of different sizes; the condition we have in the present study), with that elicited during perception (i.e. seeing items of different sizes), and mental imagery (i.e. imaging items of different sizes). The recent success of a voxel-wise encoding model suggests that the same low-level visual features are encoded during visual perception and mental imagery (Naselaris et al., 2015); however, further research is needed to test: (1) whether they differ in representational granularity, as is the case for audition and auditory imagery (Linke and Cusack, 2015); and crucially (2) whether similar results are obtained when subjects are presented with symbolic stimuli, i.e. words, instead of pictures. Despite this open issue, however, our results indicate that activation in primary visual areas contains information related to the real–word size of items even when the items are not physically present but simply evoked by symbols. Interestingly, the results of the preliminary behavioral feature generation task we conducted indicate that subjects spontaneously and consistently report size as a key defining property of both animal and tool words (averaging across items and subjects, size-related features were reported 188 times for animals and 212 times for tools), while color, for example, was reported frequently as a feature defining animals but much less for tools (554 times for animals, 117 times for

tools). Finally, while the scope of the research was not to investigate the internal scale at which object sizes are represented in the brain, because we computed our dissimilarity matrix on the basis of the rank of the sizes, and because the progression in sizes of our stimuli was roughly logarithmic, our results are compatible with the idea that size is encoded in early visual cortex according to a logarithmic scale (Konkle and Oliva, 2011) .

It should be noticed that implied real world size is relatively easily and objectively quantifiable, while other properties, such as color, cannot easily be established for many stimuli. However, in future studies we shall try to parametrize and thus model other visual as well as non-visual sensory properties implied by nouns (e.g., shape, sound) in order to investigate the degree of segregation across sensory regions of these properties. Concerning the anatomy of the real-word size effect, previous literature has shown the implication of lateral-occipital, inferotemporal, and parahippocampal cortices (Konkle and Oliva, 2012; He et al., 2013). The discrepancy between those studies and the current one can be traced down to the numerous methodological differences. Most studies used pictures as stimuli (Konkle and Oliva, 2012 studies 1 and 2), while we used words. Furthermore, when they did not use pictures, but words, as we do, they engaged subjects in tasks involving active size comparison (He et al., 2013) or imagery of objects in their prototypical or atypical size (Konkle and Oliva, 2012 studies 3), thus drawing subject's attention on the size dimension. Instead, in our experiment, subjects were asked to actively think of the whole concept referred to by the words, with no specific focus on the size dimension. Moreover, previous studies compared objects that did not only differ for average size but also belong to largely different semantic categories (animals vs tools vs non-manipulable objects, He et al., 2013), while we present results for the implied real world size effect controlling for categorical differences. Finally, all the aforementioned studies identified the effect of size using univariate analyses, while in our experiment there was no effect, neither in V1 nor in other regions at the univariate level.

Multivariate analyses of those data could reveal if additional information could retrieved from brain activity, and especially from primary visual areas, when the distributed pattern of activity is considered.

Conceptual taxonomic information is mainly encoded in mid and anterior temporal areas

A good number of neuropsychological and neuroimaging findings now converge in indicating a crucial role for ATL in the conceptual semantic processing. Herpes simplex encephalitis with widespread lateral and medial temporal lobe damage is associated with semantic category–specific deficits (Lambon Ralph et al., 2007). Moreover, semantic dementia, a neurodegenerative disorder whose gray and white matter atrophy starts in ATL, shows progressive decline in semantic representations spanning all stimulus presentation modalities (visual, auditory, verbal and pictorial) suggesting a key role of ATL in amodal semantic processing. Neuroimaging studies focusing on regions in anterior temporal cortex which are activated during semantic tasks also show that semantic proximity of words belonging to the same semantic category correlates with the patterns of activity in left perirhinal cortex (Bruffaerts et al., 2013). Virtual lesions through TMS and cortical stimulation also indicate that interfering with ATL generates trouble in a variety of semantic tasks (Pobric et al., 2010; Shimotake et al., 2014). These findings are compatible with the idea that the anterior temporal cortex acts as a hub region where single perceptual semantic features are integrated to give rise to semantic representations. In the current experiment we show that activity in the mid and anterior temporal cortex (but not in more posterior occipito/temporal regions) reflects categorical and sub-categorical conceptual clustering of the words, and is thus in line with the aforementioned literature. However, because in the current study we investigated at the same time conceptual and perceptual semantic dimensions of the words we presented, we could directly demonstrate that the ATL codes for the conceptual dimensions of the semantic

space (category and sub-categorical cluster) in a way that is independent from the single perceptual feature of size. If we had used decoding results only, we would have mistakenly concluded that categorical semantic information is available already in posterior occipital areas. Instead, by partial correlation RSA we can start teasing apart the multiple components of complex representational spaces that characterize word meaning. The finding that even once accounting for the difference across animals and tools in their average size there is enough information in the ATL to discriminate their category and sub-categorical cluster, even if admittedly at a coarse anatomical scale, enriches our understanding of the representational geometry of the anterior part of the temporal lobe. In fact, they complement previous evidence of object category effects in posterior middle/inferior temporal gyrus and ventral temporal cortex (similar to our semantic categories) (Fairhall and Caramazza, 2013), and of semantic similarity effect in left perirhinal cortex (similar to our semantic cluster) (Bruffaerts et al., 2013).

Representational shift along the ventral stream

The third major finding of our study is the observation of two progressive gradients of semantic coding as we move along the ventral stream (Fig. 61 and 62): from perceptual to conceptual and from categorical to sub-categorical.

While visuo–perceptual semantic information appears to be preferentially encoded within occipital visual areas, anterior temporal areas become progressively invariant to such perceptual features, and at the same time progressively more sensitive to the conceptual taxonomic dimensions of the semantic space: the semantic category and the sub-categorical cluster of the words. While a similar posterior–to–anterior gradient of abstraction –from physical to perceptual to conceptual information coding– has been previously reported in the domain of object recognition (Peelen and Caramazza, 2012; Devereux et al., 2013; Carlson et al., 2014; Clarke and Tyler, 2014), to our knowledge no study has previously investigated at the

same time physical, perceptual and conceptual dimensions of word meaning. The presence of a semantic gradient along the occipito–temporal axis was first suggested by clinical data: patients with vascular damage in the territory of the posterior cerebral artery present fine–grained categorical deficits (e.g. disproportionate failures for biological categories) only if their lesion extend to the anterior temporal region, beyond Talairach's y–coordinate –32 (Capitani et al., 2009). We also observed an increasingly fine-grained clusterization of words as we moved along the anterior temporal lobe: while mid-level temporal regions represent the gross semantic category of the words (animals vs. tools), more anterior regions (BA20 and BA38) become progressively sensitive to the sub-categorical clustering, allowing to distinguish words related, for example, to domesticated land animals, wild land animals, sea mammals, and sea non-mammals. A speculative idea is that the nature of the representation in the temporal lobe could be progressively more fine–grained (i.e. reflecting categorical membership in the posterior portion and single item identity in more anterior one). This hypothesis would also fit well with the report of "concept cells", coding for individual items though with a very high degree of invariance (even across symbolic and pictorial presentations) in the medial areas of the human anterior temporal cortex (Quiroga, 2012). This representational shift should be interpreted in light of the coarse anatomical scales we used and better qualified by furthers studies tapping the specific representational granularity (or hierarchy) of the different perceptual and conceptual dimensions involved in word meaning in more precisely defined brain regions.

A multidimensional semantic neural space: theoretical implications

Our ROIs encompass several functionally defined areas responding preferentially to different categories of visual stimuli, such as objects (Lerner, 2001), bodies (Downing et al., 2007), faces (Peelen and Downing, 2005) and words (Dehaene and Cohen, 2011). Beside

this macroscopic parcellation based on categorical preference, other more abstract dimensions, such as animacy (Sha et al., 2014) and real world size (Konkle and Oliva, 2012) have been suggested as additional organizing principles of object processing in the ventral visual path. Recently, it has been proposed that cytoarchitectonic differences underlie the functional segregation observed in the ventral temporal cortex following two computational principles: a lateral-medial axis of specialization (i.e., same computations on different contents), and a posterior-anterior axis of transformation (i.e., different computations on the same content)(Weiner et al., 2016). In our study we could retrieve size and category information from the activity of occipito–temporal areas, but only at the multivariate level, indicating that the activation of this information during passive word reading is more subtle and distributed compared to that directly evoked by looking at the pictures of the stimuli. Moreover, the discrepancy between findings implicating down-stream regions in the processing of size-related information (Konkle and Oliva, 2012) with our observation of an effect already in early, up-stream, regions could tentatively be explained in terms of differences in task requirements between the two studies (Martin, 2015). Generally speaking, the different perceptual and conceptual dimensions characterizing objects (Huth et al., 2012) and words (Just et al., 2010; Huth et al., 2016) semantics appear to be coded in a highly distributed fashion, encompassing visual and nonvisual cortices (Fernandino et al., 2015b; Anderson et al., 2016). All this evidence contributes to the description of a distributed and multidimensional semantic neural space, partially answering the question of how word meaning is encoded in the brain. A current debate, of interest for some, relates to the question of whether the *format* of the representation of the different stimulus features in the various brain regions is abstract or embodied (Glenberg, 2015; Mahon, 2015). Our study, by investigating the representational geometry of word meaning in different brain regions of the ventral stream elucidates where and how, in the brain, semantic information is encoded. However, it remains neutral as to its *format*.

In this respect, we agree with A. Martin (Martin, 2015) that given the absence of a consensus on how to establish the format of a representation, currently no experimental setting seems to be able to actually tackle this problem. Nevertheless, we think that the double dissociation between coded properties and brain regions that we observed is a convincing argument in favor of a distributed theory of semantic processing that accepts the key role of the anterior temporal lobe in conceptual knowledge and that at the same time recognizes an important part played by sensory-motor areas in encoding perceptual components of meaning.

In conclusion, our results indicate that different aspects of word meaning are encoded in a distributed way across different brain areas. Perceptual semantic aspects, such as the implied real word size appear to be encoded, independently from higher order semantic features, primarily in early sensory areas, which represent the aspects of semantic information that are isomorphic with the input they typically process. Conceptual aspects, such as the categorical cluster and sub-clusters, appear encoded primarily in anterior temporal areas, which code taxonomic information in a way that is independent from single perceptual features. Hence, both sensory and association areas appear to play an important role by coding for specific and complementary perceptual and conceptual dimensions of the semantic space.

Acknowledgements:

In addition to my co-authors (F. Pedregosa, M. Buiatti, A. Amadon, E. Eger, and M. Piazza , I would like to thank the LBIOM team of the NeuroSpin center for their help in subjects scanning, C. Pallier for feedback on the first draft of this manuscript and B. Thirion, G. Varoquaux, and S. Dehaene for fruitful discussions. I gratefully acknowledge K. Ugurbil, E. Yacoub, S. Moeller, E. Auerbach and G. Junqian Xu, from the Center for Magnetic Resonance Research, University of Minnesota, for sharing their pulse sequence and reconstruction algorithms. The research was funded by INSERM, CEA, Collège de France, and University Paris VI.

Bibliography

Aldrich MS, Alessi AG, Beck RW, Gilman S (1987) Cortical blindness: etiology, diagnosis, and prognosis. Annals of neurology 21:149-158.

Anderson AJ, Binder JR, Fernandino L, Humphries CJ, Conant LL, Aguilar M, Wang X, Doko D, Raizada RD (2016) Predicting Neural Activity Patterns Associated with Sentences Using a Neurobiologically Motivated Model of Semantic Representation. . Cerebral Cortex.

Bannert MM, Bartels A (2013) Decoding the yellow of a gray banana. Current biology : CB 23:2268-2272.

Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA (1999) An fMRI version of the Farnsworth–Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. Cerebral Cortex 9:257-263.

Bi Y, Wang X, Caramazza A (2016) Object Domain and Modality in the Ventral Visual Pathway. Trends Cogn Sci 20:282-290.

Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Grabowski TJ, Langfitt JT, Loring DW, Lowe MJ, Koenig K, Morgan PS, Ojemann JG, Rorden C, Szaflarski JP, Tivarus ME, Weaver KE (2011) Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. Neuroimage 54:1465-1475.

Binney R, Parker G, Lambon Ralph M (2012) Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. J Cogn Neurosci 24:1998–2014.

Bonner MF, Peelle JE, Cook PA, Grossman M (2013) Heteromodal conceptual processing in the angular gyrus. Neuroimage 71:175-186.

Bruffaerts R, Dupont P, Peeters R, De Deyne S, Storms G, Vandenberghe R (2013) Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:18597-18607.

Capitani E, Laiacona M, Pagani R, Capasso R, Zampetti P, Miceli G (2009) Posterior cerebral artery infarcts and semantic category dissociations: a study of 28 patients. Brain : a journal of neurology 132:965-981.

Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR (2014) The emergence of semantic meaning in the ventral temporal pathway. J Cogn Neurosci 26:120-131.

Clarke A, Tyler LK (2014) Object-specific semantic coding in human perirhinal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:4766-4775.

Cukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nature neuroscience 16:763-770.

Davis T, Poldrack RA (2013) Measuring neural representations with fMRI: practices and pitfalls. Annals of the New York Academy of Sciences 1296:108-134.

Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. Trends Cogn Sci 15:254-262.

Dejerine J (1892) Contribution a l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. Mémoires de la Société de Biologie 4:61-90.

Devereux BJ, Clarke A, Marouchos A, Tyler LK (2013) Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:18906-18916.

Devlin J, Matthews P, Rushworth M (2003) Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. Journal of Cognitive Neuroscience 15:71-84.

Downing PE, Wiggett AJ, Peelen MV (2007) Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:226-233.

Epelbaum S, Pinel P, Gaillard R, Delmaire C, Perrin M, Dupont S, Dehaene S, Cohen L (2008) Pure alexia as a disconnection syndrome: new diffusion imaging evidence for an old concept. Cortex 44:962-974.

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:10552-10558.

Farah MJ, Michael J. Soso, and Richard M. Dasheiff (1992) Visual angle of the mind's eye before and after unilateral occipital lobectomy. Journal of Experimental Psychology: Human Perception and Performance 18.

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PloS one 5:e15710.

Fernandino L, Humphries CJ, Seidenberg MS, Gross WL, Conant LL, Binder JR (2015a) Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. Neuropsychologia 76:17–26.

Fernandino L, Binder JR, Desai RH, Pendl SL, Humphries CJ, Gross WL, Conant LL, Seidenberg MS (2015b) Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. Cerebral cortex.

Friedman L, Glover GH, Fbirn C (2006) Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. Neuroimage 33:471-481.

Glenberg AM (2015) Few believe the world is flat: How embodiment is changing the scientific understanding of cognition. Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale 69:165-171.

Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK, Weiner MW, Miller BL (2004) Cognition and anatomy in three variants of primary progressive aphasia. Annals of neurology 55:335-346.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nature reviews Neuroscience 15:536-548.

He C, Peelen MV, Han Z, Lin N, Caramazza A, Bi Y (2013) Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. Neuroimage 79:1-9.

Hodges JR, Patterson K (2007) Semantic dementia: a unique clinicopathological syndrome. The Lancet Neurology 6:1004-1014.

Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76:1210-1224.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453-458.

Just MA, Cherkassky VL, Aryal S, Mitchell TM (2010) A neurosemantic theory of concrete noun representation based on the underlying brain codes. PloS one 5:e8622.

Konkle T, Oliva A (2011) Canonical visual size for real-world objects. Journal of experimental psychology Human perception and performance 37:23-37.

Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. Neuron 74:1114-1124.

Kosslyn SM, Giorgio Ganis, and William L. Thompson. (2001) Neural fundation of imagery. Nature Reviews Neuroscience 2:635-642.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60:1126-1141.

Lambon Ralph MA (2014) Neurocognitive insights on conceptual knowledge and its breakdown. Philosophical transactions of the Royal Society of London Series B, Biological sciences 369:20120392.

Lambon Ralph MA, Lowe C, Rogers TT (2007) Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. Brain : a journal of neurology 130:1127-1137.

Lerner Y, Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001) A hierarchical axis of object processing stages in the human visual cortex. Cerebral Cortex 11:287-297.

Linke AC, Cusack R (2015) Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. J Cogn Neurosci 27:1322-1333.

Mahon BZ (2015) The burden of embodied cognition. Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale 69:172-178.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing. Neuron 38:487–497.

Martin A (2015) GRAPES-Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. Psychonomic bulletin & review.

Mion M, Patterson K, Acosta-Cabronero J, Pengas G, Izquierdo-Garcia D, Hong YT, Fryer TD, Williams GB, Hodges JR, Nestor PJ (2010) What the left and right anterior fusiform gyri tell us about semantic memory. Brain : a journal of neurology 133:3256-3268.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191-1195.

Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Ugurbil K (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 63:1144-1153.

Murphy C, Rueschemeyer SA, Watson D, Karapanagiotidis T, Smallwood J, Jefferies E (2016) Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. Neuroimage.

Naselaris T, Kay KN (2015) Resolving Ambiguities of MVPA Using Explicit Models of Representation. Trends in Cognitive Sciences 19:551-554.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63:902-915.

Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL (2015) A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. NeuroImage 105:215-228.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Current biology : CB 21:1641-1646.

Papinutto N, Galantucci S, Mandelli ML, Gesierich B, Jovicich J, Caverzasi E, Henry RG, Seeley WW, Miller BL, Shapiro KA, Gorno-Tempini ML (2016) Structural connectivity of the human anterior temporal lobe: A diffusion magnetic resonance imaging study. Human brain mapping 37:2210-2222.

Pascual B, Masdeu JC, Hollenbeck M, Makris N, Insausti R, Ding SL, Dickerson BC (2015) Large-scale brain networks of the human left temporal pole: a functional connectivity MRI study. Cereb Cortex 25:680-702.

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

Peelen MV, Downing PE (2005) Selectivity for the human body in the fusiform gyrus. Journal of neurophysiology 93:603-608.

Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:15728-15736.

Pobric G, Jefferies E, Ralph MA (2010) Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. Neuropsychologia 48:1336-1342.

Pulvermuller F (2013) How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends in cognitive sciences 17:458-470.

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nature Reviews Neuroscience 11:351-360.

Quiroga RQ (2012) Concept cells: the building blocks of declarative memory functions. Nature reviews Neuroscience 13:587-597.

Rice GE, Hoffman P, Lambon Ralph MA (2015) Graded specialization within and between the anterior temporal lobes. Annals of the New York Academy of Sciences 1359:84-97.

Rice GE, Watson DM, Hartley T, Andrews TJ (2014) Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:8837-8844.

Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K (2004) Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychological review 111:205-235.

Rubinsten O, Henik A (2002) Is an ant larger than a lion? Acta Psychologica 111:141-154.

Setti A, Caramelli N, Borghi AM (2009) Conceptual information about size of objects in nouns. European Journal of Cognitive Psychology 21:1022-1044.

Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, Halchenko YO, Connolly AC (2014) The Animacy Continuum in the Human Ventral Vision Pathway. J Cogn Neurosci:1-14.

Shimotake A, Matsumoto R, Ueno T, Kunieda T, Saito S, Hoffman P, Kikuchi T, Fukuyama H, Miyamoto S, Takahashi R, Ikeda A, Lambon Ralph MA (2014) Direct Exploration of the Role of the Ventral Anterior Temporal Lobe in Semantic Memory: Cortical Stimulation and Local Field Potential Evidence From Subdural Grid Electrodes. Cereb Cortex.

Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. Neuroimage 54:2418-2425.

Simanova I, Hagoort P, Oostenveld R, van Gerven MA (2014) Modality-independent decoding of semantic information from the human brain. Cereb Cortex 24:426-434.

Smith FW, Goodale MA (2014) Decoding Visual Object Categories in Early Somatosensory Cortex. Cereb Cortex.

Tesla N (1919) My Inventions - My Early Life. Electrical Experimenter:696-697.

Vandenbroucke AR, Fahrenfort JJ, Meuwese JD, Scholte HS, Lamme VA (2014) Prior Knowledge about Objects Determines Neural Color Representation in Human Visual Cortex. Cereb Cortex.

Weiner KS, Barnett MA, Lorenz S, Caspers J, Stigliani A, Amunts K, Zilles K, Fischl B, Grill-Spector K (2016) The Cytoarchitecture of Domain-specific Regions in Human High-level Visual Cortex. Cereb Cortex.

Zwaan RA, Stanfield RA, Yaxley RH (2002) Language comprehenders mentally represent the shapes of objects. Psychological science 13:168-171.

CHAPTER 5:
TEMPORAL FEATURES OF SEMANTIC DIMENSIONS

Time is what keeps everything from happening at once.
[Ray Cummings, 1921]

In this chapter I review the work I conducted to investigate the temporal dynamics of the neural representations of different semantic dimensions. Analyses of this dataset are still in progress and a journal paper is in preparation.

Highlights:

- Perceptual dimensions appear to determine early effects (~200ms), mostly in the inter-trial phase coherence.
- Conceptual dimensions appear to determine early effects (~200ms), mostly in the spectral power changes.
- Different dimensions appear to be dissociated in terms of sources and dynamics, more than timing.

1. Introduction

In the previous chapters, we have seen how the many open questions concerning the neural substrate of semantic knowledge (Chap.1) can be investigated by means of different behavioral and imaging techniques (Chap.2). I here present the results of a MEG experiment aiming at corroborating our behavioral and fMRI results (Chap.3 and 4), while adding one crucial piece of information: the timing of semantic knowledge processing.

1. 1 The Temporal Dynamics of Word Reading

As already pointed out, the fine temporal resolution of electroencephalography (EEG) and magnetoencephalography (MEG) have revealed how, during single word reading, brain activation unfolds from occipital areas towards the anterior temporal pole

(Marinkovic et al., 2003; Pammer, 2009). First of all, primary visual cortices host processing of the physical properties of the stimuli presented, and during the first 200 ms, analyses of the visual-orthographic features spreads in a feed-forward wave along the inferior occipital gyrus and fusiform gyrus (Tarkiainen, 1999; Pammer et al., 2004). Second, manipulations of lexicality (words vs non-words) and frequency (high vs low frequency words) influence brain activity in left superior temporal cortex within 200 and 600 ms (Wydell et al., 2003), even if earlier effects have been reported (e.g., 150 ms (Sereno et al., 1998)). Finally, between 300 and 500 ms, activity mainly originating in left fronto-temporal temporal areas, denotes semantic processing (Kutas and Hillyard, 1980). While in the time domain the N400 (a negative evoked related potential peaking at 400 ms), has been long considered the electrophysiological component most revealing of the timing of semantic memory (Kutas and Federmeier, 2000), in the frequency domain the desynchronization of the upper alpha band has been associated with semantic memory retrieval (Klimesch, 1999).

As we have seen in Chap 1, many theories on the neural substrate of semantic representations clash on the hypothesized driving principle: evolutionary relevant domains (Caramazza and Shelton, 1998) or co-occurrences of features (Tyler and Moss, 2001). While great attention has been paid to the investigation on the topographical organization of semantic categories with PET and fMRI, less numerous studies have targeted the timing of their differentiation. The few electrophysiological studies existing tapping this particular question have reported word category effects (e.g., animals vs vegetables, living vs non-living) earlier than the semantic N400 wave, around 250-270 ms (Dehaene, 1995; Hinojosa et al., 2001; Martin-Loeches et al., 2001), while on picture the semantic categories can be dissociated even earlier, around 180-200 ms (Ji et al., 1998; Antal et al., 2000). A handful of studies have used multivariate approaches, and have revealed that it is possible to possible to decode semantic information from M/EEG signal acquired

while subjects are presented with pictures (Carlson et al., 2011), spoken (Correia et al., 2015) and written words (Chan et al., 2011a), with this latter case being the hardest (Simanova, 2010). According to these studies access to the information which is sufficient to discriminate semantic category varies between 250–400 ms (Simanova, 2010) and 550–600ms (Correia et al., 2015). Finally, (Simanova et al., 2015) has shown that it is possible to decode semantic category from MEG data recorded while subjects are instructed to spontaneously generate words corresponding to animals or tools, even without the presentation of any external stimuli. Overall, these findings suggest a broad time window, between 200 and 600 ms, during which semantic effects can be appreciated. However, as previously stressed (Chap 1 and Chap 4), with these studies we cannot test to what extent the observed category-related effects reflect the activation of co-occuring motor-perceptual dimensions of word meaning. In fact, these studies typically manipulate and compare the categorical aspect of words, but not other implied correlated dimensions of the stimuli, such as the shape, color, sound, size, affordance, smell, etc… Moreover, the time window of category effects appears quite broad and variable across studies, possibly as a function of both stimuli and tasks differences. Hence, the process appears inadequately described: it is not clear which kinds of representations are activated at different points in time.

As we will see next, another parallel stream of studies have instead focusses on perceptuo-motor components of word meaning, and have also shown that early motor-perceptual effects can be detected. As in the case of the semantic category, also perceptuo-motor effects seem to emerge much earlier than the timings of the "classical" N400 (e.g., 150 ms in (Kiefer et al., 2008)). These observations, paired with the broad time window reported for categorical effects, open the possibility that semantic content might be recovered not in a unitary fashion, but rather differentially according to the different motor-perceptual or conceptual dimensions considered. Moreover, early detection of differences along motor-

perceptual dimensions has been taken as disproving theories claiming that the sensory-motor activations observed in the neuroimaging literature are post-conceptual, epiphenomenal effects of mental imagery (Mahon and Caramazza, 2008). However, with the data currently available such position cannot be firmly disproved as we lack a direct manipulation of both motor-perceptual and conceptual dimensions within the same experimental paradigm and in the same subjects. The present study approached this question, and investigated how the spatio-temporal dynamics of brain activity during reading reflects access to the perceptual and to the conceptual aspect of word meaning.

1. 2 The Temporal Dynamics of Accessing Semantic Features

A first investigation on the processing of different semantic features in time comes from a seminal paper on the factors influencing the N400. (Federmeier and Kutas, 1999) have shown that the degree of coherence between the unexpected (semantically incongruous) word ending a sentence, and the context of the sentence itself determines the amplitude of the effect (i.e., the more the word is plausible, thus sharing many semantic features with the preceding words in the sentence, the less prominent the N400). More directly, one study attempted at contrasting words with strong visual connotations (i.e., referring to colors and shape, such as "*red*" and "*square*") and strong auditory connotations (i.e., referring to sounds, such as "*whistle*" and "*echo*") (Bastiaansen et al., 2008). By means of a region of interest (ROI) analyses, they observed that in temporal areas the N400 was larger for vision-related words than for audio-related words. Moreover, they found a double dissociation in the frequency spectrum of the left hemisphere: the theta power increase was larger over the temporal ROI for audio-related words and larger over the occipital ROI for vision-related words.

Following the embodied theory of semantic, which postulates a key (therefore early) role for sensory-motor areas in semantics, some authors have attempted to identify the first point in time when these areas are recruited during word processing. Overall, somatotopic coherent semantic differences have been reported at 240 ms (Pulvermüller et al., 2000) and 220 ms (Hauk and Pulvermuller, 2004) after the onset of visual stimuli, and at 172-200 ms (Pulvermüller et al., 2005) after the onset of auditory stimuli during action verb processing.

Capitalizing on multivariate analyses, authors have dissociated the influence of different semantic features on the patterns of brain activity generated by the observation of pictures presented together with the corresponding written names (Sudre et al., 2012). Subjects were asked to answer to specific semantic questions tapping the different perceptual and functional features of the items. Trivially enough, the authors were able to decode low level physical features of the stimuli (e.g., words length) earlier than perceptual-semantic features (e.g., can you pick it up?) and conceptual-semantic features (e.g., is it alive?). Interestingly, ROIs frequently associated with semantic processing (e.g., superior temporal and inferior parietal cortex) did not show the highest decoding performance; rather, the best decoding score were observed in the left lingual gyrus and left latero-occipital complex. With a similar approach, the same group has shown single word meaning decoding through sum of semantic features starting as early as at 100 ms and spreading up to 700 ms, with the best performance being over occipital sensors (Fyshe et al., 2012). More recently, in the domain of object recognition, (Clarke et al., 2015) have deployed an encoding model to show that after 200 ms semantic features significantly increase predictive performance of individual objects' identity over and above visual features.

Finally, it has been suggested that the neural correlate of the integration of unimodal features (e.g., "*red*" + "*big*" = "*bus*", both vision-related) is a sustained increase in high-frequency power (gamma band, 80 –120 Hz). On the contrary, the integration of

multimodal features (e.g., "*red*" + "*loud*" = "*bus*", one vision-related, one audio-related) appears to be associated with enhanced low-frequency power (theta band, 2– 8 Hz) (van Ackeren et al., 2014).

1. 3 Present Study Hypothesis

In this study we are interested in testing the hypotheses that different perceptual and conceptual dimensions of word meaning are supported by the activity of partially distinct brain networks, possibly involving a precise temporal hierarchy. In this respect, we can contrast radically different predictions according to different existing theories. If semantics representations emerge via the reactivation of motor-perceptual features thanks to the converging activity of modality-specific areas (Pulvermüller, 2013), then (a) perceptual effects should be appreciated much earlier than conceptual ones, and (b) their topographies and source reconstruction should indicate an early contribution of early sensory-motor cortices. Alternatively, if semantic information is coded in an trans-modal hub by an abstract code, and post-conceptual mental imagery is responsible for the motor-perceptual effects (Mahon and Caramazza, 2008), we should appreciate an early source of conceptual information, localized in multimodal convergence regions comprising the antero-temporal, infero-parietal, and inferior frontal cortices, followed, later in time, by multiple sensory-specific sources of (epiphenomenal) sensory-specific information. A third option, in line with recent so-called "hybrid models" put forward to reconcile clinical and fMRI data (e.g., Meyer and Damasio, 2009; Lambon Ralph et al., 2017), is that of an integrated and possibly concurrent involvement of both "semantic hubs" (i.e., convergence zones where both conceptual and perceptual information come together) and modality specific "spokes" (where only perceptual information is represented). However, precise predictions of the temporal dynamics underlying the interplay between hub and spokes components of such system have not yet been put forward.

In order to test these hypotheses, in the current work we selected words varying orthogonally along three dimensions: one visuo-perceptual (i.e., the implied real world size), one audio-perceptual (i.e., whether it is associated with a prototypical sound or not), and one conceptual (i.e., the semantic category). Two central questions guided our investigation:

1. Is there a difference in space, time and/or frequency, between the different levels of the three dimensions?
2. Can we establish the temporal hierarchy with which perceptual and conceptual information are activated?

Concerning the timing of our effects, both motor-perceptual (Pulvermüller et al., 2000; Hauk and Pulvermuller, 2004; Pulvermüller et al., 2005) as well as conceptual effects (Dehaene, 1995; Hinojosa et al., 2001; Martin-Loeches et al., 2001) have been reported in early time windows (rarely, but equally so). This suggests that both motor-perceptual and conceptual features might be activated rapidly during word reading, with the great variability possibly due to the material selected and the task assigned to the subjects. As no previous study has directly compared within the same experimental paradigm motor-perceptual and conceptual dimensions, no strong prediction on their relative timing can be put forward.

Concerning the aspect of the signal that might reveal the temporal encoding of semantic representations, based on the literature cited above we predicted that both event-related field potentials (ERFs) (Pulvermüller et al., 2000; Kutas and Federmeier, 2000, Hauk and Pulvermuller, 2004; Kiefer et al., 2008) and brain oscillatory patterns, in particular in the theta (Bastiaansen et al., 2005, van Ackeren et al., 2014) and alpha range (Klimesch, 1999) are potentially relevant.

Finally, concerning the localization of our effects, following our own previous results, we predicted that the perceptual semantic dimension of the semantic space would be primarily encoded in early sensory regions (primary and secondary visual areas for size, primary and secondary auditory areas for sound), while the conceptual

dimensions would be primarily encoded in multimodal associative areas, such as the anterior temporal lobe (Borghesani et al., 2016).

To our knowledge, this is the first attempt to directly compare the spatial, temporal and spectral representations of multiple motor-perceptual and conceptual dimensions within the same subjects, during the same task.

2. Materials and Methods

2.1 Subjects

Fifteen healthy adult volunteers (seven males, mean age 24.57 ± 2.69) participated in the MEG study. Data from two additional subjects were discarded due to magnetic artifacts (the MRI scan suggested the presence of dental implants). All participants were right-handed as measured with the Edinburgh handiness questionnaire, had normal or corrected-to-normal vision, and were Italian native speakers. All experimental procedures were approved by the local ethical committee and each participant provided signed informed consent to take part in the study. Participants received a monetary compensation for their participation.

2.2 Stimuli

As done for the fMRI experiment, target stimuli (i.e. 32 words, 16 names of living items and 16 names of non-living items) underwent both psychological and psycholinguistic validation. First, we ran two preliminary behavioral experiments that involved 130 French native speakers, tested through internet–based questionnaires. These experiments (detailed in Chap. 3.2) suggest an overall distributed semantic space including domesticated (*bull, sheep, cow, chamois, rabbit, rooster, ant, and cricket*) and exotic animals (*elephant, giraffe, gorilla, lama, marmoset, parrot, chameleon, and*

scorpion), house appliances (*fork, wardrobe, sofa, pillow, washing machine, vacuum cleaner, blender, and alarm clock*), and objects linked with means of transportation (*canoe, boots, roller, bike, motorcycle, helicopter, car stereo, and rotating beacon*).

The preselection of the words led to an orthogonal classification of the items as belonging to one or the other hand of the continuum of the two perceptual dimensions (i.e., implied real world size and prototypical sound) (see Fig. 65). To assess the consistency between our predicted classification and that subjectively reported by the participants of the MEG experiment, we implemented two behavioral questionnaires to be administrated after the recordings. In the Visual Task, subjects were asked to rate, on a scale from 1 to 9, the size of the object/animal each word referred to, as compared with a shoe box (i.e. "*could this item fit in a shoe box?*"). In the Auditory Task, subjects were asked to rate, always on a scale from 1 to 9, whether the object/animal was associated with a prototypical sound or not. The order of tasks, and of the categories within each task (i.e. living vs non-living), were randomized across subjects. The results clearly support our initial classification. As far as the Visual task is concerned, across subjects the average score for items categorized as big was 7.84 (± 0.80), while the one for items categorized as small was 3.28 (± 1.27). None of the items categorized as big had a score lower than 6, and none of the items categorized as big higher than 5. Similarly, across subjects the average score for items categorized as having a prototypical sound was 7.67 (± 0.77, none of them having an average score lower than 6), while one for items categorized as silent was 2.43 (± 1.2, none of them having an average score higher than 5).

Finally, we verified that words belonging to the different perceptual and conceptual categories were well matched across several psycholinguistic variables such as number of letters, number of syllables, gender, accent and frequency of use (see Chap. 3.2).

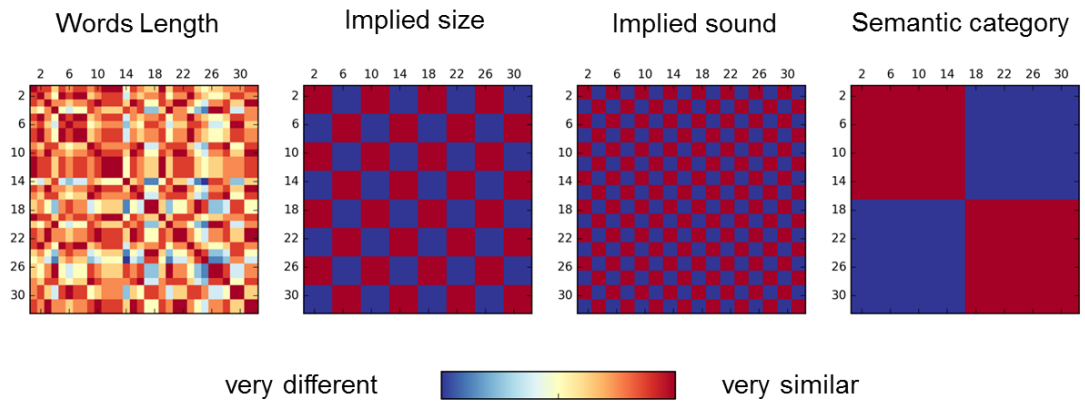


Figure 65 Matrices modeling the similarities across stimuli along the dimensions investigated. The words' length matrix depicts all pairwise differences in terms of number of letters between the stimuli. The implied real–world size matrix indicates whether a given pair of stimuli share the same size (e.g. both big) or not. Similarly, the implied real–world sound matrix illustrates which stimuli share the auditory property of being associated with a prototypical sound. Finally, the semantic category matrix indicates which pairs of stimuli belong to the same category (e.g. both non-living) and which do not.

2.3 Testing Procedures

Subjects were seated in a comfortable armchair in front of the screen (monitor with 60 Hz refresh rate). Subjects were instructed to silently read the target stimuli (i.e. the 32 words referring to living or non-living items) and to make semantic decisions on rare odd stimuli. These odd stimuli appeared on 6% of the trials and consist in a pair of words semantically related to one of the targets (e.g., "*ruminant, wool*" for sheep). The subjects pressed the left or the right hand to indicate whether the odd stimulus was related or not to the previously seen target word (i.e., 1-back task). The hand-answer mapping was counterbalanced within subjects: half of the subjects answered *yes* with the left hand in the first half of the imaging runs and then *yes* with the right hand in the last half; the other half of the subjects follow the reverse order. Importantly, the pairs of words used as odd stimuli did not contain any verb, nor any reference to the dimensions investigated (i.e. implied size or sound). Target stimuli were presented at the center of the screen, printed in Courier New, for 300 ms (18 frames). They were followed by an inter-stimuli-interval that varied randomly between 2167 ms (130 frames) and 3340 ms (200 frames). The odd stimuli were presented for 1670 ms (100 frames) and followed by 1670 ms (100 frames) of blank (see Fig. 66). Within a

given MEG session, the participant underwent 8 runs of ~7 min each. Breaks between runs were tailored on subjects' needs. Each run contained 5 repetitions of each of the 32 target stimuli and 10 odd stimuli, for a total of 170 stimuli per run. Pseudo-randomization ensured that, over the entire experiment, for half of the odd stimuli (i.e., 40 times) a positive answer was expected. Prior to testing the first subject, a photodetector was used to compute the delay between the time at which the triggers were sent to the MEG acquisition computer and the time at which the stimuli actually appeared on the screen. Such delay (50 ms) was corrected during preprocessing. Stimuli were presented with Psychopy.

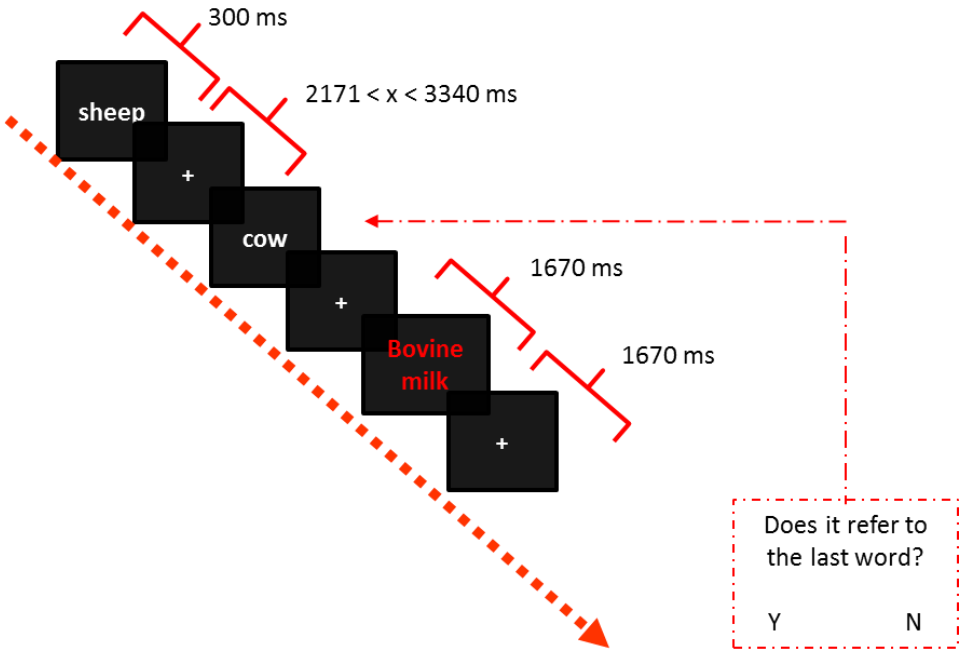


Figure 66 Experimental setting. Example of a sequence of stimuli: during the MEG experiment, subjects were instructed to silently read the target stimuli and to press a button at the presentation of rare odd stimuli. The odd stimuli consist of a two-words definition that could refer to the last seen target word.

2.4 MEG Protocol

Data were collected at Neurospin (CEA-Inserm/Saclay, France) in a dimly illuminated, sound-attenuating, magnetically shielded room. The whole-head Elekta MEG system (Neuromag Elekta LTD, Helsinki) used has 102 magnetometers and 204 orthogonal planar gradiometers. Participants were seated in the upright position, and head positioning was ensured to be in close contact to

the dewar. Subjects were instructed to avoid any head, body, or unnecessary limb movements. At the start of each block, their head position was measured thanks to four head position coils (HPI) placed over the frontal and mastoid areas and compared on-line with the position at the beginning of the recording. To minimize head displacements across the whole recordings, if the head position moved more than 10 mm from the original position in any direction, the subject was assisted to reposition the head closer to the original position. To help the coregistration with the anatomical MRI, prior to the recording, three fiducial points (nasion, left and right pre-auricular areas) and about 100 more supplementary points distributed over the scalp of the subjects were digitalized (3D digitizer, Polhemus Isotrak system). MEG recordings were sampled at 1 kHz, hardware band-pass filtered between 0.03 Hz and 330 Hz, and active compensation of external noise (Maxshield, Neuromag Elekta LTD, Helsinki) was applied. Heartbeats, horizontal and vertical eye movements were recorded simultaneously with the MEG signals thanks to three additional pairs of electrodes for the electrocardiogram (ECG) and the electro-oculograms (EOG) respectively. Right before or immediately after each experiment, empty room recordings of about 2 min were acquired while no subject was sitting under the dewar. These recordings were subsequently used to compute the noise covariance matrix (i.e, the estimation of the noise in the signal needed to estimate a reliable forward model).

2.5 MRI Protocol and Source Reconstruction

Data were collected at Neurospin (CEA-Inserm/Saclay, France) with a 3 Tesla Siemens Magnetom TrioTim scanner using a 32-channel head coil. Anatomical images were acquired using a T1-weighted MP-RAGE sagittal scan (voxels size 1x1x1.1mm, 160 slices, 7 minutes). Volumetric segmentation of participants' anatomical MRI and cortical surface reconstruction was performed with the FreeSurfer software (http://surfer.nmr.mgh.harvard.edu/).

Current source density was estimated with BrainStorm software (http://neuroimage.usc.edu/brainstorm). After cortical and scalp reconstruction, anatomy and MEG signals were coregistered using head position indicators digitized earlier. The forward problem was computed using an overlapping spheres model. Noise covariance was estimated from MEG empty-room recordings. Individual sources were computed with the weighted minimum-norm method (depth weighting factor of 0.5, loosing factor of 0.2 for dipole orientation). They were then projected on a standard anatomic template to perform averages across subjects.

2.6 MEG Data Pre-Processing

After visual inspection to detect bad channels, the first steps of preprocessing included signal space separation (SSS) to suppress external magnetic interference, interpolation of noisy MEG sensors and correction for head movements between data blocks with Maxfilter software application (Elekta Neuromag). Head movement correction was performed with respect to a subject-specific head position, computed as the mean head position across blocks (custom-made software, courtesy of Antoine Ducorps and Denis Schwartz, CENIR, Paris, France), and used afterwards for MEG/MRI co-registration. Data were then visually inspected again to detect bad segments, i.e., segments of recording including clear motor artifacts, or channels jumps/anomalies. Such bad segments were flagged and thus skipped in all the following stages. These raw but cleaned data followed two slightly different preprocessing according to the goal of the analyses: Event-Related Field (ERF) or time-frequency analyses (spectral power and inter-trial phase coherence).

ERF After filtering the data with a low-pass filter at 40 Hz, heartbeat and blinks components were automatically detected (via principle components analysis, PCA), visually checked and removed (by removing the corresponding signal-space projections, SSP). The

stimulus-trigger delay (50 ms) was corrected. Data were then epoched starting 200 ms before and ending 900 ms after the onset of the stimuli. These epochs were downsampled to 250 Hz and baseline corrected using the 200 ms preceding stimuli onset. These preprocessing steps were conducted with Brainstorm.

Time-frequency analyses. After filtering the data with a low-pass filter at 160 Hz, the same artifacts removal and correction for the stimulus trigger time delay implemented in the ERF analysis were applied. Data were then epoched starting 800 ms before and ending 1200 ms after the onset of the stimuli. Epochs were downsampled to 500 Hz and no baseline correction was applied. These preprocessing steps were conducted with Brainstorm.

Spectral power was estimated by computing the time-frequency decomposition with the multi-taper approach implemented in Fieldtrip (http://www.fieldtriptoolbox.org), with parameters adapted to two distinct frequency ranges. For the low-frequency range (4 – 35 Hz in 1 Hz steps), data segments were extracted from sliding time windows with a length of 500 ms between 4 and 10 Hz (frequency resolution = 2 Hz), and with a length equal to 5 oscillation cycles per frequency between 10 (frequency resolution = 2 Hz) and 35 Hz (frequency resolution = 7 Hz), shifted in steps of 40 ms. These parameters were chosen to optimize the frequency resolution for higher frequencies, while keeping a limited time window for lower frequencies in order to test stimulus-related effects. Data segments were tapered with a single Hanning window and Fourier-transformed. Spectral power was computed as the square amplitude of the resulting time-frequency decomposition. The associated time–frequency images had no discontinuities thanks to the continuous frequency resolution function. For the high-frequency range (34 –100 Hz in 2 Hz steps), data segments were extracted from sliding time windows of 200 ms length, shifted in steps of 40 ms. A multitaper approach was applied to each window to optimize spectral concentration over the frequency of interest (Mitra and Pesaran, 1999). Frequency smoothing was set to 20

% of each frequency value. With these settings, the number of tapers used ranged from 2 at 34 Hz (frequency resolution = 7 Hz) to 7 at 100 Hz (frequency resolution = 20 Hz). Spectral power was first estimated per taper and then averaged across tapers.

Inter-trial phase coherence (ITC) was determined for each subject and condition by computing the phase-locking factor (Tallon-Baudry et al., 1996) with the following steps:

- The complex time-frequency decomposition at time *t* and frequency *f* of each single trial (as computed above) is normalized by its absolute value to obtain amplitude-independent unitary vectors in the complex plane;
- These normalized vectors are averaged across single trials to obtain a complex value related to the phase distribution of each time–frequency region around *t* and *f*. The ITC is computed as the modulus of this value.

ITC ranges from 0 (purely non-phase-locked activity) to 1 (strictly phase-locked activity). These analyses steps were conducted with Fieldtrip.

2.7 Univariate Analyses

To fully exploit the temporal richness of MEG data, we took a "data mining" approach as proposed in (Makeig et al., 2004) by evaluating event-related changes in terms of both (1) distribution of the phase of these oscillations across trials by computing the inter-trial phase coherence (ITC), and (2) amplitude of brain oscillations by computing the power of time-frequency representations. ERFs are intrinsically included in these two measures as they are (at least partially) produced by event-related phase-locking (i.e., event-related narrowing of the phase distribution) and may be associated with an increase of oscillatory power. However, ITC has the added advantage of decomposing the ERF into its constituent phase-locked frequency bands (Makeig et al., 2004), facilitating the identification of

experimental effects in specific frequency bands that could be otherwise scrambled by overlapping fluctuations in lower frequency bands. ERFs, while likely being less sensitive, are widely used, thus, for completeness, we report also the ERFs for the three contrast of interest, as they can help the comparison with the previous literature.

Statistical analyses. All statistical analyses aiming at identifying significant differences between experimental conditions were conducted with the non-parametric cluster-based statistical analysis (Maris and Oostenveld, 2007), as implemented in the Fieldtrip toolbox (Oostenveld et al., 2011). This method allows statistical testing on wide time and frequency intervals with no need of a priori selection of spatial ROIs because it effectively controls the type I error rate in a situation involving multiple comparisons by clustering neighboring channel-time-frequency pairs that exhibit statistically significant effects (test used at each channel-time-frequency point: dependent-samples t statistics) and using a permutation test to evaluate the statistical significance at the cluster level (Montecarlo method, 1000 permutations for each test). Results on statistically significant clusters are reported by specifying the polarity of the cluster (positive or negative), its p value, its temporal and spectral extent and the time and frequency of its maximum effect (hereafter indicated as cluster's peak), defined as the time/frequency at which the cluster statistics is maximal. The time course of the cluster statistics is obtained by averaging at each time point the channel-time-frequency point t statistics over all the channels and frequencies belonging to the cluster at that time point. Analogously, the frequency range of the cluster statistics is obtained by averaging at each frequency bin the channel-time-frequency point t statistics over all the channels and time points belonging to the cluster at that frequency bin.

All the statistical tests were performed, with few differences according to the dependent measure investigated, separately for magnetometers and combined gradiometers (i.e., vector sum is used,

the value for each gradiometer pair is equal to the square root of the sum of the squares of the values computed for each gradiometer).

ERF For all contrasts of interest, epochs from the same condition were averaged for each subject and statistical comparisons performed with the cluster-based statistical analysis described above, corrected for multiple comparisons over time and sensor space. To disentangle early and late effects, two time windows were investigated: an early one (from 0 to 300 ms after stimuli onset), and a late one (from 300 to 600 ms after stimuli onset.

Spectral Power and Inter-Trial Phase Coherence. For all contrasts of interest, cluster-based statistical analyses corrected for multiple comparisons over time, frequency and sensor space were applied on the whole time window (from 0 to 600 ms), for three frequency ranges: theta and alfa (4 – 13 Hz), beta (13 - 35 Hz) and gamma (35 – 100 Hz) (the latter for spectral power only).

Source visualization. In order to visualize the anatomical sources of the observed significant effects, spectral power and ITC were estimated at the source level in the time-frequency window of the significant effects observed at the sensor level, with Brainstorm's implementation of Morlet wavelets (same computation as implemented at the sensor level, same frequency resolution). For each subject and condition, the reconstructed sources of both spectral power and inter-trial phase coherence values were smoothed (10 mm kernel) and projected to the default anatomy. Additionally, before smoothing power values were z-scored with respect to [-500 -250] ms baseline (no baseline correction is necessary for ITC since it is already a normalized measure). For each condition of interest, a paired t-test was run and we here report the corresponding significant clusters ($p<0.05$ uncorrected) on a template cortex smoothed at 50%. Importantly, the t-test at the source level is only used to properly describe the source distribution of the statistically significant effect established at the sensor level, and not for a second statistical test at

the source level, therefore no correction for multiple comparison is required (Gross et al., 2013).

2.8 Multivariate Analyses

To test whether distributed patterns of information could distinguish between our conditions of interest, we applied multivariate analyses. Two different analyses were conducted: one on the filtered and time resolved data used for ERF analysis, the other on the time-frequency decomposed data used for power and phase analysis. The first option offers the best time resolution (temporal smoothing is unavoidable when transforming the data from the time-domain to the frequency domain), while the second one offers the possibility to study effects that concern only (or mostly) one specific frequency band.

Time generalization. Three classifiers were trained to discriminate living vs non-living, big vs small, sound-related vs not-sound-related trials respectively. The data fed to the decoders were matrices composed of *n* trials and *f* features (only gradiometers were used), with each feature corresponding to the amplitude of the MEG signal. A linear support vector machine (SVM) with fixed penalization parameter (C = 1) and 5-fold cross-validation was used. All estimators were systematically fitted across trials at each time point (e.g., t) and tested not only on the same time point (t), but also on all others (e.g., t_1, t_2, t_3, t_4, etc…). The resulting matrices have on the y axis the time at which the estimator was fitted, and the x axis the time at which the estimator was evaluated. To summarize estimators' performances and test for their significance, area under the curve (AUC) was computed for each subject and then averaged across subjects. Significance was then tested across-subject using a Wilcoxon signed-rank test. Time generalization was conducted with custom scripts relying on MNE-python (http://martinos.org/mne) and publically released code by Jean-Remi King (https://github.com/kingjr/jr-tools).

Time-frequency-space searchlight. Iteratively, the features fed to the decoder were selected with a sphere of 10 sensors, and a radius of 1 time bin (each time bin is 40 ms) and 1 Hz. In a cross-validated fashion (5 folds), Linear Discriminant Analysis (LDA) classifiers were trained to discriminate the patterns across sensors for our conditions of interest (i.e., living vs non-living, big vs small, sound-related vs not-sound-related). To identify time bins, frequency ranges and sensors yielding above chance classification, a threshold-free cluster-estimation procedure was used, with multiple comparison correction based on a sign-permutation test. Statistical maps were then thresholded at $Z > 1.64$ (i.e., $p < 0.05$, one-tailed) to reveal significant decoding performance. This analysis was implemented in CoSMoMVPA (http://cosmomvpa.org/).

Additional controls. Multivariate analyses were also used to assess that perceptual and conceptual features could not be explained by information present in the physical appearance of the stimuli themselves. We attempted to decode the condition the words belonged to (i.e., big vs small, sound vs no sound, living vs non-living) from the images of the stimuli (i.e. the snapshots of the screens with the words we presented to the subjects during the experiment). Unsurprisingly, the only dimension that this analysis could recover from such input was the number of letters composing each word: decoding score (Ridge regression) = 0.56, p=0.004 (for implied real world size: classification score = 0.38, p=0.78, for implied real world auditory property: classification score = 0.4, p=0.75, for semantic category: classification score = 0.58, p=0.21). These analyses were implemented in ScikitLearn (http://scikit-learn.org/).

3. Results

3.1 Spatio-Temporal Dynamics of Word Processing: basic effects

Averaging across all words and conditions, time-locked evoked general activity indicates three main waves in response to the visual stimulation. The first one, at ~100 ms, in posterior sensors, representing the early processing of the visual stimulus, shortly followed by a second one, more left lateralized, at ~170 ms, typically associated to visual recognition, confined to the occipito-posterior temporal cortex. Finally, a third wave can be appreciated ~450 ms, extending more anteriorly, towards the ATL and the frontal lobe. Note that the Global Field Power (GFP) as well as the corresponding topographies are similar, yet complementary, across sensors type (compare in Fig. 67 magnetometers, gradiometers type 1 and type 2).

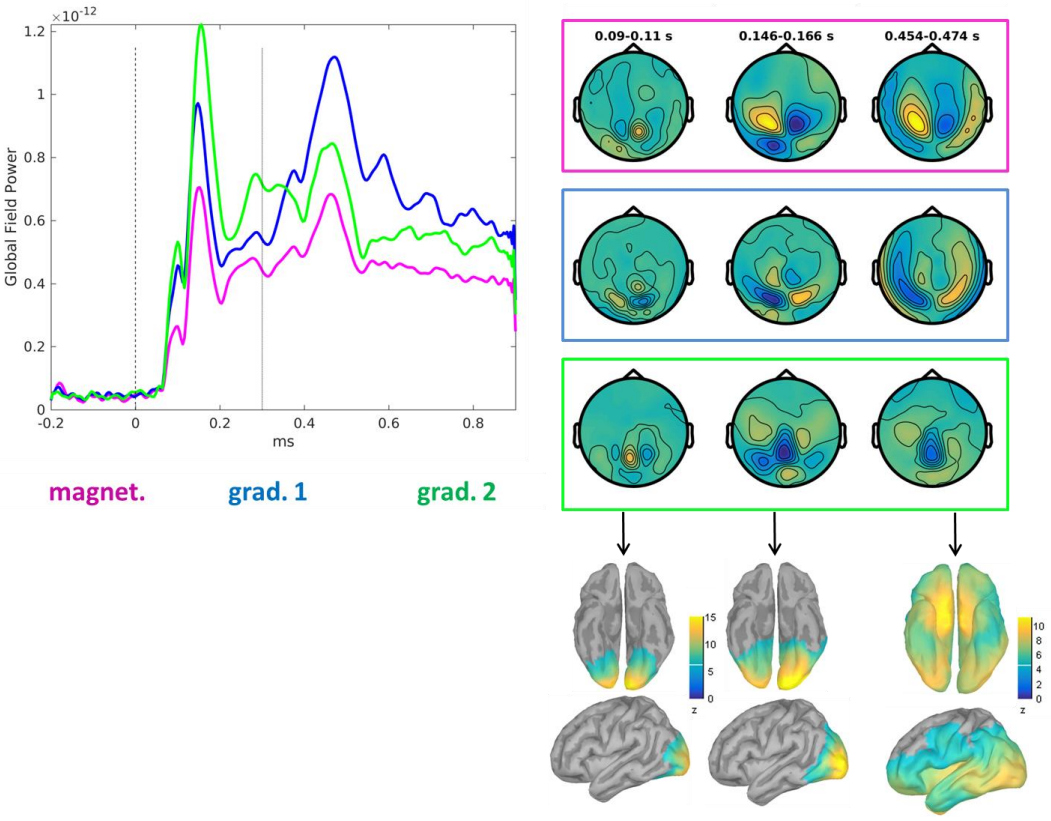


Figure 67 Global evoked activity elicitated by our stimuli. Time course of the evoked response across all stimuli for the three different sensors type (left), together with the corresponding topography (right – upper), and source reconstruction (right - lower). Purple = magnetometers, blue = gradiometers type 1, green = gradiometers type 2.

First of all, as *sanity check* over the quality of our data, we verified whether we could retrieve one basic physical dimension of our stimuli (i.e. word length) both at the univariate and at the multivariate level. The cluster-based permutation statistic of the evoked related response (i.e., univariate level) indicated a significant positive cluster (corrected $p < 0.01$), between 144 and 196 ms (peak at 176 ms). The multivariate decoding approach (Ridge regression) supports this findings by showing how this low level physical dimension of the stimuli is recovered from the distributed pattern of activity starting slightly before 100 ms and peaking twice: ~150 ms after the onset and the offset of the stimuli. Details and figures are reported in the Appendix 1.3.

We then moved to the investigation of the semantic variables of interest: implied real world size, sound, and category. We did so by investigating them in the time-frequency domain: changes in inter-trial phase coherence and spectral power.

3.2 Inter-Trial Phase Coherence

Averaging across all words and conditions, an overall effect of increased inter-trial coherence was observed between 200 and 400 ms, with a bilateral occipito-temporal topography, slightly more left lateralized (see Fig. 68).

Within this time-frequency range, a significant effect of implied real world size was found between 120 and 360 ms (peak at 240 ms), and between 6 and 7 Hz (peak at 6 Hz), in a left occipital cluster (positive polarity, corrected $p = 0.03$): words referring to small items elicit higher ITC than big ones. Fig. 69-upper part illustrates the time-frequency representation of the effect as well as the corresponding sensors topography and source reconstruction. It appears that visual information is confined to the occipital lobe, strongly left-lateralized.

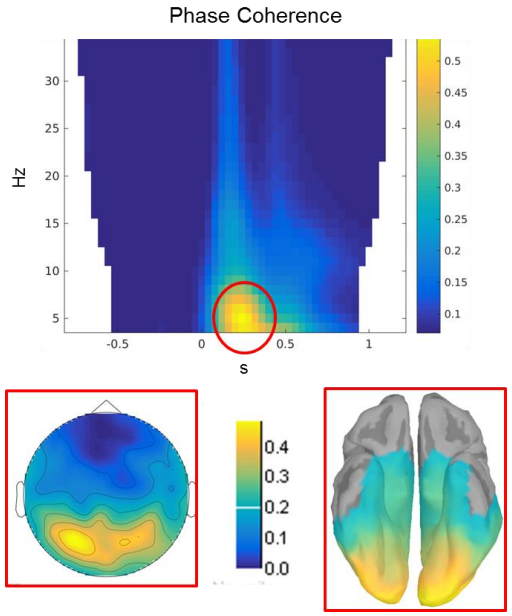


Figure 68 Overall effect of phase coherence. Time-frequency representation of the inter-trial phase coherence across all stimuli. The insert on the left highlights the topography at the sensor level, whose underlying source reconstruction is depicted on the right. Here reported are the results of combined gradiometers, the results for each sensors separately can be found in the Appendix 1.3.

In the same early time window, between 40 and 320 ms (peaking at 200 ms), but in a different frequency band (between 8 and 12 Hz (peak at 10 Hz)), a highly significant cluster for implied real world sound was detected (corrected p = 0.008): words referring to items associated with a prototypical sound elicit higher ITC as compared to those not automatically associated with a specific sounds. The auditory dimension was also recovered in a later time window, between 320 and 520 ms (peak at 400 ms), and between 4 and 6 Hz (peak at 5 Hz) (corrected p = 0.04), where words referring to items not typically associated with sounds elicit higher ITC. As depicted in Fig.69-lower part, source reconstruction suggests that these effects are linked with the activity of occipito-temporal areas, mainly in the left hemisphere, remarkably extending to the superior temporal gyrus of both hemispheres.

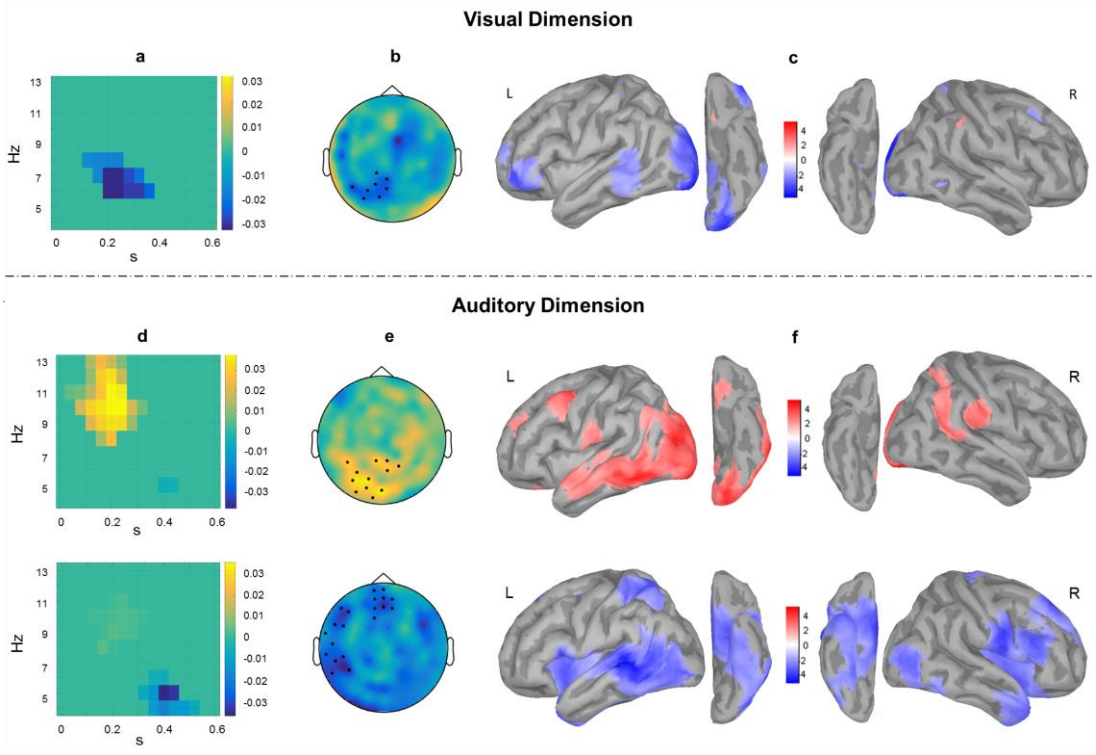


Figure 69 Inter-trial phase coherence effects of the perceptual dimensions. Time-frequency representation (a) and sensors topography (b) of the average difference in inter-trial coherence between the two levels of the visual dimension (i.e., words referring to big vs small items). In the time-frequency plot, non-significant values are masked and in the topography sensors showing a significant difference (Monte-Carlo permutation test) are highlighted. (c) Corresponding source reconstruction (paired ttest, p<0.05). (d-f) Same as in (a-d) but for the two significant clusters of the auditory dimension (i.e., words referring to items with prototypical sounds or not).

All these effects were observed on the combined gradiometers, while only trending (but congruently so) effects could be appreciated in the magnetometers. No effect of semantic category could be appreciated in this aspect of the signal.

3.3 Spectral Power

The power spectrum obtained averaging across all words and conditions illustrate expected results (e.g. Bastiaansen et al., 2005): a left lateralized occipital increase in theta, followed by a bilateral very strong decrease of alpha and beta band, which extends bilaterally along the ventral stream, reaching the anterior temporal areas (see Fig. 70).

When looking at the main contrasts of interest, a strong and long lasting effect was that of the conceptual dimension (i.e., semantic category), which was detected in a left occipital-temporal cluster of gradiometers (corrected p = 0.01) and lasted between 80 and 600 ms (peak at 600 ms), and between 4 and 13 Hz (peak at 9 Hz), where words referring to animals elicit higher theta increase than those referring to tools. When analyses are repeated for two separated time window (0-300 ms and 300-600 ms), two sub clusters can be appreciated within the broad cluster identified: an early one (peaking at 200 ms and 8 Hz, corrected p = 0.02) and a later one (peaking at 600 ms and 10 Hz, corrected p = 0.01). The time-frequency representations of the two clusters and the corresponding sensors topography can be seen in Fig. 71-upper part. The same effect was also appreciated on the magnetometers (positive polarity, corrected p = 0.03, peak at 560 ms, between 8 and 13 Hz). Source reconstruction suggest that the early effect originated primarily from a temporo-parietal network of brain regions in the left hemisphere, including the angular gyrus, while the second effect can be traced back to the activity of bilateral anterior and ventral temporal areas, slightly left lateralized.

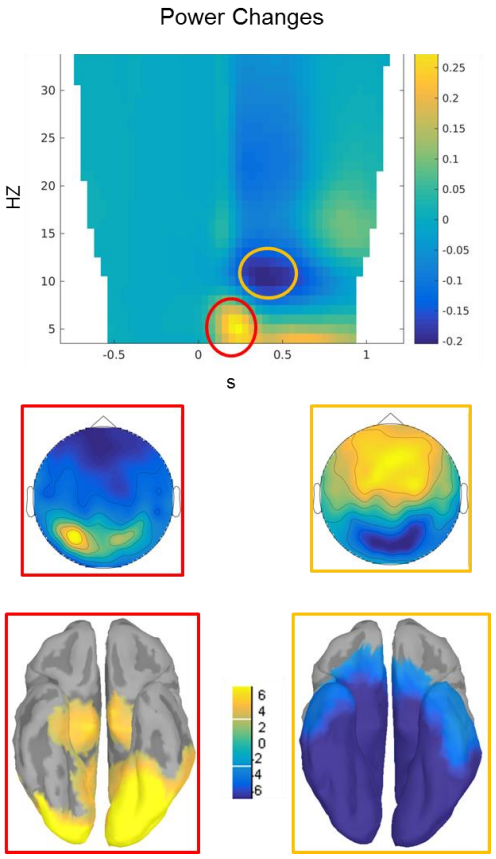


Figure 70 Overall effects of power changes. Time-frequency representation of the changes in power observed across all stimuli. The inserts highlight the topography at the sensor level of the two main effects: bilateral occipital decrease of alpha and beta band, left lateralized occipital increase in theta (source reconstruction is depicted below). Here reported are the results of combined gradiometers, the results for each sensor separately can be found in the Appendix 1.3.

The two perceptual dimensions showed weaker yet significant effects. An implied real world size effect was detected in a left occipital cluster of magnetometers (corrected $p = 0.04$) between 160 and 480 ms (peak at 400 ms), and between 44 and 86 Hz (peak at 74Hz): words referring to small items are associated with an increase in gamma band. This effect was not observed in the gradiometers. Fig. 71-middle part illustrates the time-frequency representation of the effect as well as the corresponding sensors topography and source reconstruction, suggesting the recruitment of left superior-temporal and inferior frontal/parietal regions (~pars opercularis). An implied real world sound effect was also detected, in a bilateral occipital cluster of gradiometers (corrected $p = 0.01$) between 200 and 600 ms (peak at 600 ms), and between 5 and 13 Hz (peak at 10Hz). This last effect was similarly appreciated on the magnetometers (corrected $p=0.008$, peak at 560 ms and 11 Hz). These effects indicate that words referring to items associated with prototypical sounds are associated with a higher decrease in alpha band (i.e., possibly reflecting a higher release from inhibition). As depicted in Fig. 71-lower part, both sensors topography and source reconstruction suggest an involvement of left posterior occipital cortex, bilateral superior-temporal and inferior frontal/parietal regions, yet the auditory dimensions shows an additional cluster in the right superior temporal lobe, which is absent for the visual dimension.

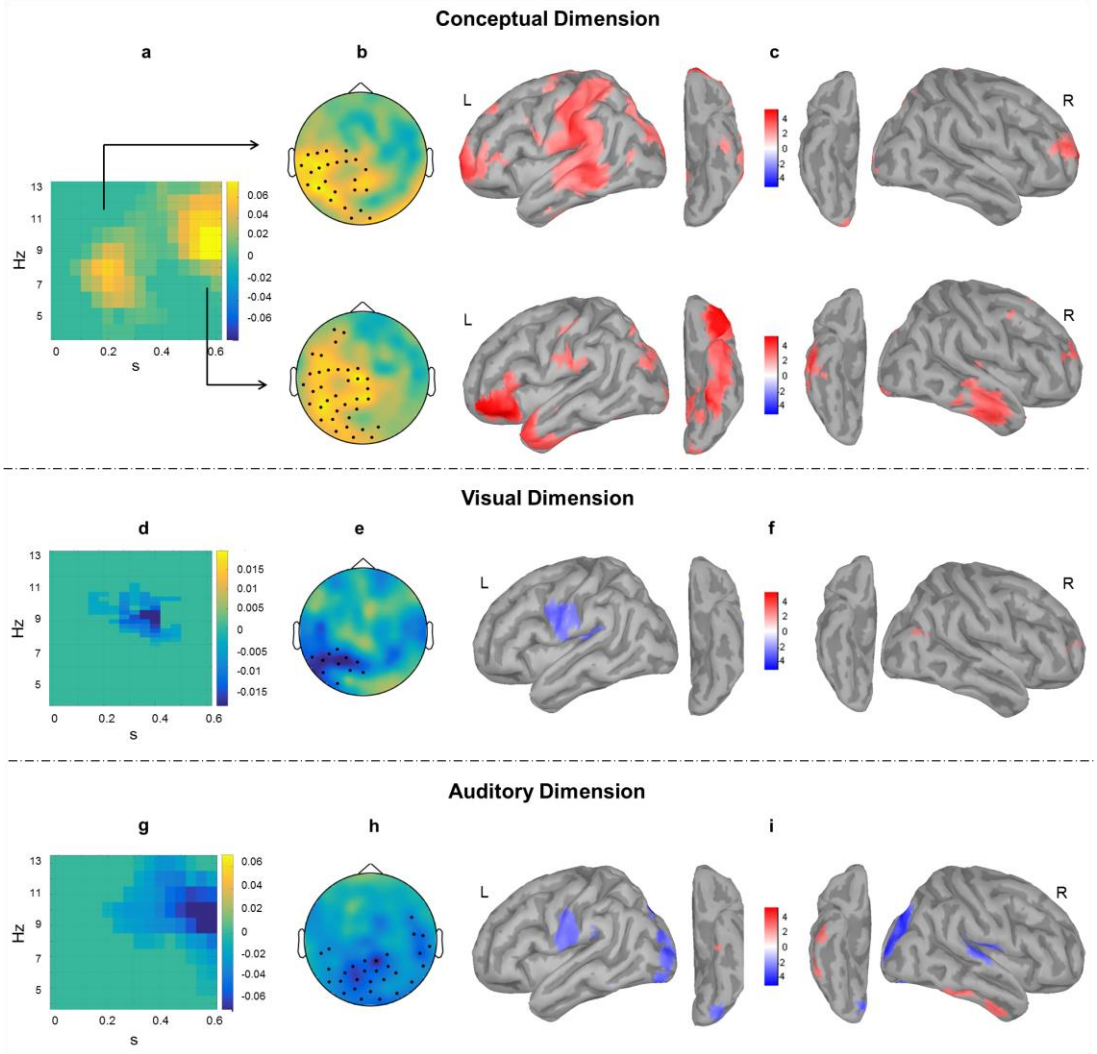


Figure 71 Spectral power effects of the perceptual dimensions. Time-frequency representation (a) and sensors topography (b) of the average difference in spectral power between the two levels of the conceptual dimension (i.e., words referring to animals vs tools). In the time-frequency plot, non-significant values are masked and in the topography sensors showing a significant difference (Monte-Carlo permutation test) are highlighted. (c) Corresponding source reconstruction (paired ttest, $p<0.05$). (d-f) As in (a-c) but for the visual perceptual dimension (implied real world size). (g-i) As in (a-c) but for the auditory perceptual dimension (implied real world sound).

3.4 ERFs

To be able to compare more directly our effects with previous literature, we also explored which of our effects of interest could be recovered from the evoked related fields (ERFs). Two significant clusters were detected on magnetometers combined: a significant effect of implied real world size (a positive cluster over left temporal sensors, between 204 and 232 ms, peak at 212 ms, $p = 0.04$); a significant effect of implied real world sound (a positive cluster over

left occipito-temporal sensors, between 384 and 460 ms, peak at 440 ms, p = 0.009). A cluster of magnetometers approached significance for semantic category (a negative cluster over right fronto-central cluster between 384 and 432 ms, peak at 416 m, p = 0.05). All significant ERF results are reported in the appendix (1.3) together with additional quality checks aiming at excluding possible correlations between our conditions of interest and eye movements.

3.5 MVPA Results

Time generalization. When run on time-resolved data (the same epochs used for the ERFs analyses), the time generalization decoding was very weak (none of the effects survived correction for multiple comparison in time). Overall, the auditory dimension appears to have a stronger and more sustained effect as compared to the visual and conceptual one. All detailed results are reported in the Appendix 1.3.

Time-frequency-space searchlight. For all three categorization problems (sound vs no sound, big vs small, living vs non-living), large portion of occipito-temporal sensors appears to be involved, predominantly in the left hemisphere (see Fig. 72, left and middle). In an attempt to better characterize the results without recurring to ROIs, we examined the time-frequency representation of all posterior sensors (see Fig. 72, right). It can be easily appreciated that information is encoded in theta (and to a less extend alfa) band in all three cases. Moreover, considering the intrinsic smoothing due to the time-frequency decomposition, no strong inference on the relative timing can be drawn as the effects appear to substantially overlap between 200 and 600 ms.

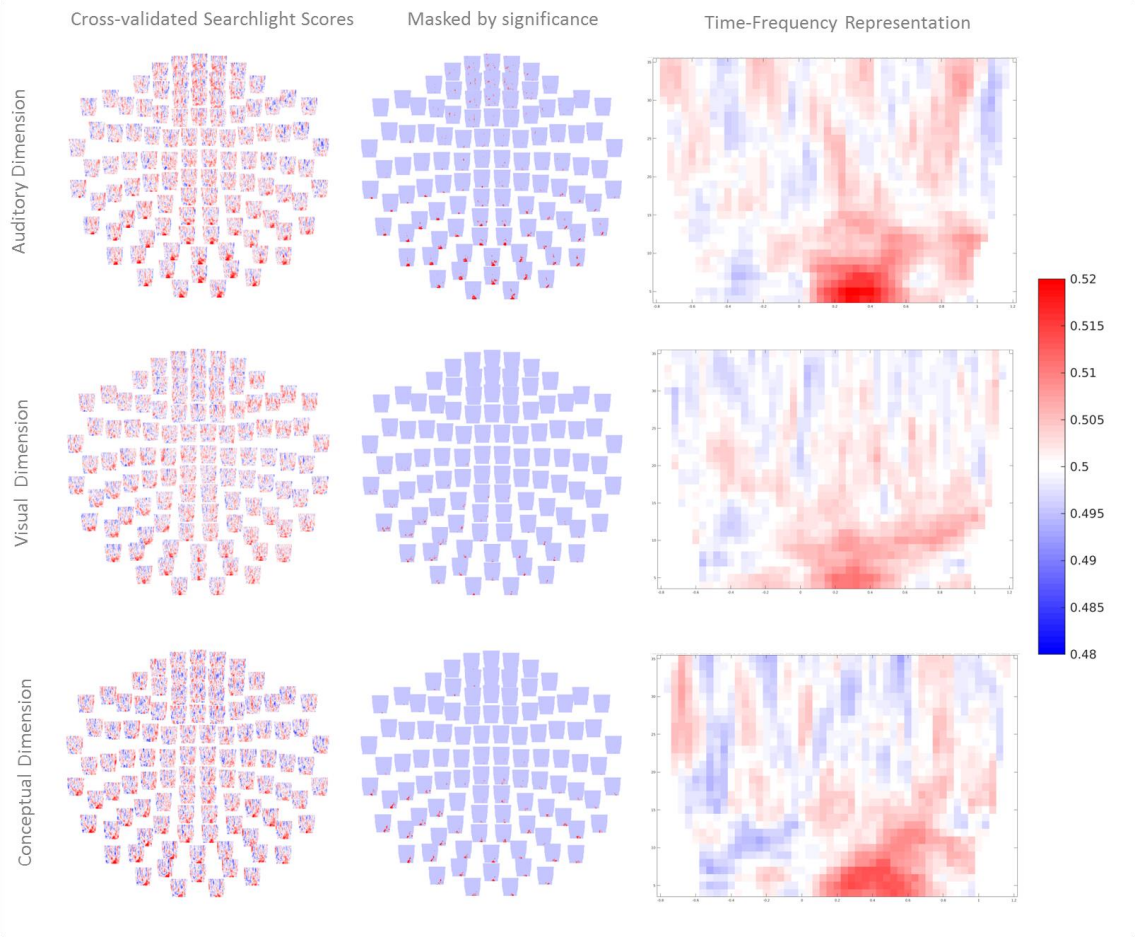


Figure 72 Space-time-frequency searchlight. For the three effects, we report (left) the global topographical representation of the cross-validated classification scores obtained for each time point and frequency bin, (middle) the same scores masked by significance, and (right) the average time-frequency representation in posterior sensors.

An alternative way of comparing the results of the three searchlights, aiming at detecting topographical or time-frequency dissociation, is to compute (1) the average number of bins in frequency and time in which each sensors score reached significance, thus illustrating the topography of significant effects, and (2) the average number of sensors reaching significant score at each time-frequency bin, thus illustrating the time-frequency spectrum of significant effects (see Fig. 73). This visualization helps detecting that the major dissociation between the three dimensions is in terms of the topographical distribution of the corresponding effects: auditory information appears to be more spread and bilateral, visual information more left lateralized and confined to occipital sensors, and

categorical information more spread than the visual one, suggesting the recruitment of more temporo-parietal regions. In terms of the time-frequency distribution, multivariate analyses confirm that retrieval of the three dimensions occurs at low frequencies (mostly in the theta range), rapidly and almost simultaneously. While all effect seem to start around the same timing (ãround 200 ms), their peak seem to differ in time: the conceptual dimension (semantic category) appears to peak (in terms of maximum number of sensors involved) slightly after the perceptual ones: ~250 ms for both visual and auditory dimensions, ~450 ms for conceptual dimension.

Finally, Fig. 74 shows the topography of the three effects in terms of averaged decoding score across the theta range (4-12 Hz) at six representative time points, confirming the observation that the auditory di dimension is the most topographically spread one, while the conceptual dimensions is the one peaking later.

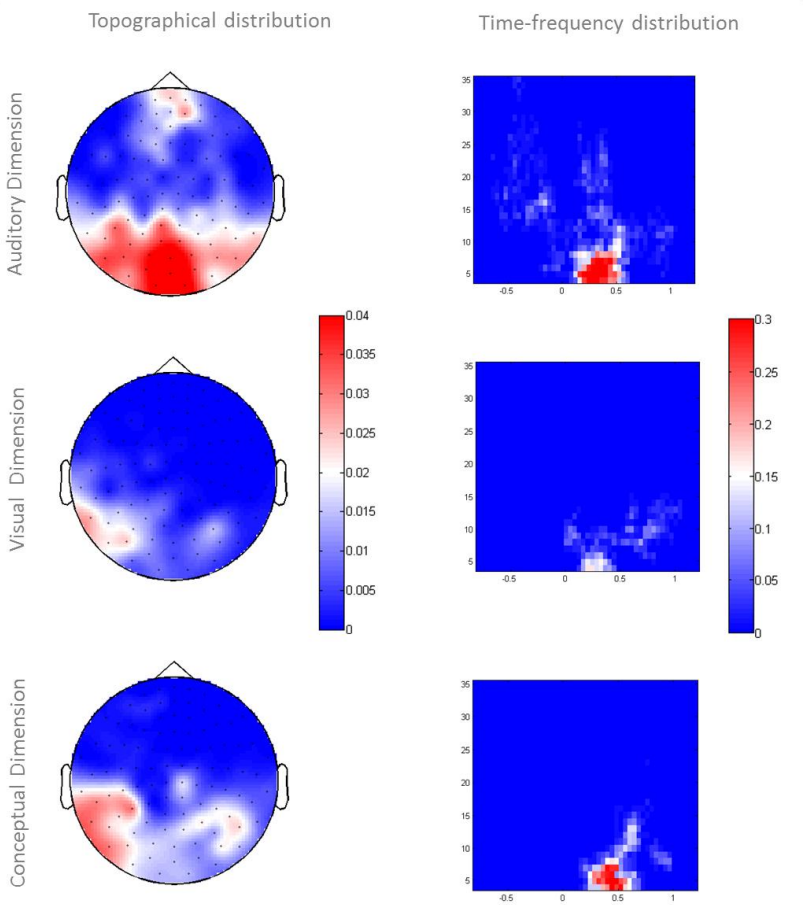


Figure 73 Space-time-frequency searchlight: topographical and time-frequency distribution of the significant effects. For the three classification task, we report (left) the average number of bins time-frequency bins in which each sensors score reached significance, and (right) the average number of sensors reaching significance at each time-frequency bin.

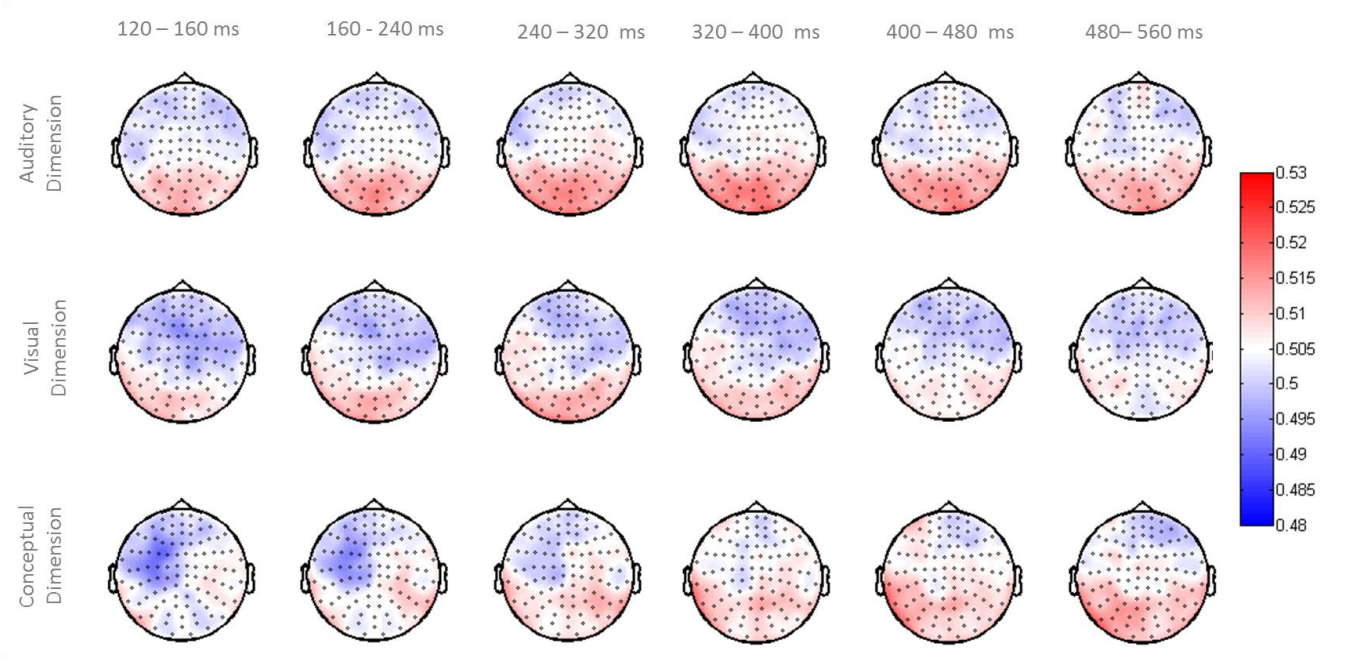


Figure 74 Space-time-frequency searchlight: spatio-temporal evolution. We illustrate the topography of the three effects (auditory dimension, visual dimension and conceptual dimension) in terms of averaged decoding score across the theta range (4-12 Hz) at six representative time points.

4. Discussion

This study investigated the temporal dynamics of different dimensions of word meaning during silent reading. It tested the hypothesis of a temporal hierarchy in the recovery of perceptual and conceptual semantic features of the objects referred to by the words. Our task, orthogonal to all the dimensions of the semantic space we investigated, ensured that the representations recovered in the brain activation emerged spontaneously. To fully capitalize on intrinsically multivariate nature of the MEG signal, we explored not only time-locked changes in the time domain (ERFs), but also phase and power changes in the time-frequency domain.

Overall, the global field power showed the general time course of brain activity classically associated with word reading (e.g. Tarkiainen, 1999; Pylkkänen and Marantz, 2003): activity spreads from posterior occipital areas towards anterior fronto-temporal ones (see Fig. 67). Similarly, the global time-frequency profile was also in line with previous findings (e.g., Klimesch et al., 1997; Bastiaansen et

al., 2005): a general increase in power in the theta band (4–7 Hz) followed by a decrease in the alpha (8–12 Hz) (see Fig. 70). In relation to our specific conditions, while only small effects could be appreciated in the ERFs (see Appendix 1.3), signal power and phase analysis confirmed that all three dimensions of word meaning – i.e., semantic category, as well as implied real world size and sound- are automatically retrieved extremely early in time, and demonstrated that they are coded in partially dissociable sources of the signal.

Early recovery of both perceptual and conceptual dimensions. Both univariate comparison and multivariate decoding analyses converge in demonstrating how the three dimensions of word meaning we investigated, two perceptual (i.e., implied real world size and sound) and one conceptual (i.e. semantic category), can be statistically differentiated from the brain activity in an very early time window (~200 ms). This means that all dimensions of the semantic space are activated in an automatic and possibly parallel fashion extremely early during reading.

Concerning the access to the conceptual dimension, while the majority of the previous studies investigating semantic processes pointed to a differentiation across semantic categories in late time window, at the level of the N400 wave (Lau et al., 2008), we are not the first reporting very early semantic category effect: Dehaene, using EEG, reported a univariate effect of category-selective responses dissociating animal names from proper names, numerals, and verbs within 250 ms after written word onset (Dehaene, 1995), while Chan et al. reported an early (200 ms) multivariate decodability of semantic category (i.e., living vs non-living) of both written and spoken words (Chan et al., 2011a). Direct electrophysiological recordings through microelectrodes in the inferotemporal and perirhinal cortex are able to differentiate semantic categories of words as early as 130 ms (Chan et al., 2011b). One potential reason underlying the lack of early semantic effects in some of the previous studies is that they relied on ERP/ERF, a measure offering lower sensitivity as compared to ITC estimation.

Another, connected, possible explanation is that most of the evoked signal is necessarily dominated by the processing of low-level physical properties of the stimuli, thus possibly washing out the semantic differences unless care is taken to avoid confounds during the design of the experiment. In other words, the selection of highly controlled stimuli appears crucial: as example, consider how in (Dehaene, 1995) length was controlled across the five different semantic categories investigated.

Regarding the two perceptual dimensions, our results are generally in line with data stemming from the investigation of verbs processing and its motor-related "embodied" aspect, reporting somatotopically organized semantic differences across verbs between 150 and 240 ms (Pulvermüller et al., 2000; Hauk and Pulvermuller, 2004; Kiefer et al., 2008). However, to our knowledge, no previous study directly investigated the neurodynamics of the recovery of the perceptuo-semantic features of nouns (see below for a discussion of convergent findings by Sudre and collaborators (2012) of an early effect of implied size using a multivariate approach). Thus, the current study is the first illustrating the finding.

Perceptual and conceptual dimensions are associated with different signal dynamics and cortical sources. One unexpected finding of our work is the dissociation, at early time points, between perceptual and conceptual dimensions in terms of the property of signal that appears to encode them. On one hand, perceptual effects are appreciated in phase-locking changes: around 200 ms after stimulus onset, phase coherence is modulated by the visual dimension in occipital areas (higher for words referring to small rather than big items), and by the auditory dimension in occipito-temporal areas (higher for words referring to items associated with a prototypical sound). On the other hand, the conceptual effect is revealed by power changes: in the same time window (~200 ms) words referring to animals elicit higher theta increase than those referring to tools. One tentative interpretation of this dissociation is that perceptual effects

may involve a partial reinstatement of brain activity elicited by the perception of the real world aspect of interest (sound or size) (Kiefer et al., 2008), a response which appears to be strongly phase-locked. By contrast, the conceptual effect could correspond to a higher-level processing stage, likely encoded in non-phase-locked activity in higher level multimodal regions.

The different dimensions seem also to be partially dissociated in terms of their underlying sources. On one hand, the visual and auditory properties detected in the phase coherence changes are linked predominantly with occipital and posterior-temporal regions, thus mostly involving modality-specific areas. On the other hand, the semantic category effect observed in power changes is linked with posterior parietal, mid-inferior temporal and anterior temporal regions, traditionally associated with multimodal processes and language related functions. This partial dissociation supports hybrid theories on the neural substrate of semantic representations that assign complementary roles to multimodal convergence areas (semantic hub(s)) and modality specific cortices (spokes) (Lambon Ralph et al., 2017).

Source and frequencies dissociation between visual and auditory dimensions. Dissociation across the two different perceptual features in time is not supported by the data: both effects peak between 200 and 250ms. However, they occur at different frequency ranges: visual property at theta, 6-7 Hz, auditory property at alpha, 10-11 Hz. Moreover, both sensors topography and source reconstruction suggest different underlying sources: while both effects involve predominantly the left occipito-temporal cortex, visual information appears to be confined to occipital lobe, while auditory information spreads more temporally and, crucially, appears to involve the superior temporal gyrus of both hemispheres. Our results corroborate, without resorting to ROI analyses, (Bastiaansen et al., 2008) findings of a dissociation within the theta range between temporal and occipital sensors, involved respectively in processing auditory-related vs

vision-related words. It has been suggested that oscillations at low-frequencies (2-8 Hz) might be the neural correlate of semantic feature integration across modalities (van Ackeren et al., 2014), underlying merging of motor-perceptual features into unitary concepts.

Later effects of perceptual and conceptual dimensions. The three effects re-appear in a later time window (400-600). Semantic category and auditory property modulate the power of the oscillations at the 10 Hz. The two aspects (category ad implied sound) do not appear as dissociated in neither time nor frequency (8-12 Hz, ~500 ms); however, they seem to differ in topography. The conceptual dimension, i.e. the semantic category, involves ventro-temporal and anterior temporal regions of the left hemisphere. In contrast, the perceptual dimension, i.e. the implied real world sound, is linked with activity in the left posterior occipital cortex, bilateral superior-temporal and inferior frontal/parietal regions. Similarly, the auditory effect observed in this later time window at the level of phase coherence involves bilateral mid/superior temporal areas.

In the same time window, but at higher frequencies (~70 Hz) an effect of the visual perceptual dimension (implied real world size) is detected. However, contrary to the previous effects, this is an extremely weak one, observed only in the magnetometers (not in the combined gradiometers).

Multivariate pattern analyses corroborate univariate results. The multivariate spatio-temporal-spectral searchlight analyses, integrating pattern of brain activity extending in time, space and frequency, confirm the general picture observed with univariate statistics: both perceptual and conceptual dimensions are recovered early in time (since ~200ms) and at low frequencies (theta, to a less extent alfa) over occipital sensors. Moreover, the auditory effect appears as more broad, and the conceptual appears slightly later in time and is appreciated in slightly more anterior sensors. These results are in line with previous findings indicating the possibility recovery

the meaning of words in a large window between 150 and 600 ms thanks to information in distributed patterns at delta, theta and, to a less extend, alpha frequencies (Fyshe et al., 2012). The only previous finding that attempted to dissociate the contribution of different semantic features does not report the timing information for the auditory properties tested *("does it make a sound?*", appears to be decodable but not clear when), while it describes an effect of size information ("*is it bigger than a car*?") around 200 ms and related manipulation information ("*can you hold it*?") even earlier, at 150 ms (Sudre et al., 2012). The same group has recently showed that the meaning of both adjectives and nouns can be predicted on the basis of MEG signal around 100 ms after their onset, with the best performance being reached in occipital areas (Fyshe et al., 2016). These results are coherent with our observation of an early window for word meaning decoding and for a key contribution of posterior areas, even though in their case semantic differences are associated with grammatical-syntactic ones, while our results are purer in that we only use one grammatical class of words (nouns).

In conclusion, our results indicate that different aspects of noun meaning are retrieved automatically, rapidly and simultaneously, yet thanks to different underlying sources and signals. Visual and auditory perceptual semantic aspects (i.e., the implied real word size and sound) are best appreciated in terms of phase coherence changes over occipital and temporal regions respectively. Conversely, conceptual aspects (i.e., the semantic category) are best retrieved in power changes over superior temporal cortices at early time points, and anterior-temporal and ventro-temporal cortices at later time points. Hence, specific perceptual and conceptual dimensions of the semantic space appear to be accessed concurrently yet differentially already within the first 300 ms of reading. Hence, specific perceptual and conceptual dimensions of the semantic space appear to be accessed concurrently yet differentially already within the first 300 ms of reading. The early contribution of sensory-motor cortices to the

retrieval of motor-perceptual dimensions was predicted by embodied views on semantics, however, such theories would not be able to explain the almost simultaneous retrieval of the conceptual dimension in associative areas. On the other hand, both the timing and the reconstructed sources of the effects cannot be accomodated by an abstract theory on semantics, which would consider post-conceptual mental imagery responsible for the motor-perceptual effects. Thus, these results speaks against a purely embodied model or purely amodal perspective on the neural substrate of semantic dimensions, calling for hybrid model where symbolic input are followed by a rapid activation on both a trans-modal hub (dedicated to the processing of conceptual dimensions) and associated modality-specific spokes (dedicated to the processing of motor-perceptual dimensions).

Acknowledgements:

In addition to my co-authors (M. Buiatti and M. Piazza), I would like to thank the LBIOM team of the NeuroSpin center for their help in subjects recruiting. The research was funded by INSERM, CEA, Collège de France, and University Paris VI. I gratefully acknowledge D. Trübutschek and V. van Wassenhove for helpful exchanges on the steps of the MEG analyses.

Bibliography

Antal A, Keri S, Kovacs G, Janka Z, Benedek G (2000) Early and late components of visual categorization: an event-related potential study. Cognitive brain research 9:117-119.

Bastiaansen MC, Oostenveld R, Jensen O, Hagoort P (2008) I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. Brain Lang 106:15-28.

Bastiaansen MC, Van Der Linden M, Ter Keurs M, Dijkstra T, Hagoort P (2005) Theta responses are involved in lexical—Semantic retrieval during language processing. Journal of cognitive neuroscience 17:530-541.

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

Caramazza A, Shelton JR (1998) Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Journal of Cognitive Neuroscience 10:1–34.

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. Journal of vision 11.

Chan AM, Halgren E, Marinkovic K, Cash SS (2011a) Decoding word and category-specific spatiotemporal representations from MEG and EEG. NeuroImage 54:3028-3039.

Chan AM, Baker JM, Eskandar E, Schomer D, Ulbert I, Marinkovic K, Cash SS, Halgren E (2011b) First-Pass Selectivity for Semantic Categories in Human Anteroventral Temporal Lobe. Journal of Neuroscience 31:18119-18129.

Clarke A, Devereux BJ, Randall B, Tyler LK (2015) Predicting the Time Course of Individual Objects with MEG. Cereb Cortex 25:3602-3612.

Correia JM, Jansma B, Hausfeld L, Kikkert S, Bonte M (2015) EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. Frontiers in psychology 6:71.

Dehaene S (1995) Electrophysiological evidence for category-specific word processing in the normal human brain. Neuroreport 6:2153–2157.

Federmeier KD, Kutas M (1999) A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. Journal of Memory and Language 41:469-495.

Fyshe A, Sudre G, Wehbe L, Murphy B, Mitchell T (2012) Decoding word semantics from magnetoencephalography time series transformations. In: 2nd NIPS Workshop on Machine Learning and Interpretation in NeuroImaging.

Fyshe A, Sudre G, Wehbe L, Rafidi N, Mitchell TM (2016) The Semantics of Adjective Noun Phrases in the Human Brain. bioRxiv.

Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R, Parkkonen L, Taylor JR, van Wassenhove V, Wibral M, Schoffelen JM (2013) Good practice for conducting and reporting MEG research. Neuroimage 65:349-363.

Hauk O, Pulvermuller F (2004) Neurophysiological distinction of action words in the fronto-central cortex. Human brain mapping 21:191-201.

Hinojosa JA, Martín-Loeches M, Muñoz F, Casado, P., Fernández-Frías C, Pozo MA (2001) Electrophysiological evidence of a semantic system commonly accessed by animals and tools categories. . Cognitive Brain Research 12:321-328.

Ji J, Porjesz B, Begleiter H (1998) ERP components in category matching tasks. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 108:380-389.

Kiefer M, Sim EJ, Herrnberger B, Grothe J, Hoenig K (2008) The sound of concepts: four markers for a link between auditory and conceptual brain systems. The Journal of neuroscience : the official journal of the Society for Neuroscience 28:12224-12230.

Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain research reviews 29:169-195.

Klimesch W, Doppelmayr M, Pachinger T, Russegger H (1997) Event-related desynchronization in the alpha band and the processing of semantic information. Cognitive Brain Research 6:83-94.

Kutas M, Hillyard SA (1980) Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological psychology 11:99-116.

Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. Trends in cognitive sciences 4:463-470.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. Nature reviews Neuroscience 18:42-55.

Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of physiology, Paris 102:59-70.

Makeig S, Debener S, Onton J, Delorme A (2004) Mining event-related brain dynamics. Trends Cogn Sci 8:204-210.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing. Neuron 38:487–497.

Martin-Loeches M, Hinojosa JA, Gomez-Jarabo G, Rubia FJ (2001) An early electrophysiological sign of semantic processing in basal extrastriate areas. Psychophysiology 38:114-124.

Meyer K, Damasio A (2009) Convergence and divergence in a neural architecture for recognition and memory. Trends in neurosciences 32:376-382.

Mitra PP, Pesaran B (1999) Analysis of dynamic brain imaging data. Biophysical journal 76:691-708.

Pammer K (2009) What can MEG neuroimaging tell us about reading? Journal of Neurolinguistics 22:266-280.

Pammer K, Hansen PC, Kringelbach ML, Holliday I, Barnes G, Hillebrand A, Singh KD, Cornelissen PL (2004) Visual word recognition: the first half second. Neuroimage 22:1819-1825.

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

Pulvermüller F, Härle M, Hummel F (2000) Neurophysiological distinction of verb categories. . Neuroreport 11:2789-2793.

Pulvermüller F, Shtyrov Y, Ilmoniemi R (2005) Brain Signatures of Meaning Access in Action Word Recognition. Journal of Cognitive Neuroscience 17:884-892.

Pylkkänen L, Marantz A (2003) Tracking the time course of word recognition with MEG. Trends in cognitive sciences 7:187-189.

Sereno SC, Rayner K, Posner MI (1998) Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. Neuroreport 9:2195-2200.

Simanova I, van Gerven MA, Oostenveld R, Hagoort P (2015) Predicting the semantic category of internally generated words from neuromagnetic recordings. J Cogn Neurosci 27:35-45.

Simanova I, Van Gerven, M., Oostenveld, R. and Hagoort, P., (2010) Identifying object categories from event-related EEG: toward decoding of conceptual representations. PloS one 5:e14465.

Sudre G, Pomerleau D, Palatucci M, Wehbe L, Fyshe A, Salmelin R, Mitchell T (2012) Tracking neural coding of perceptual and semantic features of concrete nouns. Neuroimage 62:451-463.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. The Journal of Neuroscience 16:4240-4249.

Tarkiainen A, Helenius, P., Hansen, P.C., Cornelissen, P.L. and Salmelin, R., (1999) Dynamics of letter string perception in the human occipitotemporal cortex. Brain : a journal of neurology 122:2119-2132.

Tyler LK, Moss HE (2001) Towards a distributed account of conceptual knowledge. Trends in cognitive sciences 5:244-252.

van Ackeren MJ, Schneider TR, Musch K, Rueschemeyer SA (2014) Oscillatory neuronal activity reflects lexical-semantic feature integration within and across sensory modalities in distributed cortical networks. The Journal of neuroscience : the official journal of the Society for Neuroscience 34:14318-14323.

Wydell TN, Vuorinen T, Helenius P, Salmelin R (2003) Neural correlates of letter-string length and lexicality during reading in a regular orthography. J Cogn Neurosci 15:1052-1062.

CHAPTER 6:
CONCLUSIONS AND PERSPECTIVES

Se Dio esistesse, sarebbe una biblioteca.

[If God existed, He would be a library. Umberto Eco]

In this chapter I summarize the contributions of the theoretical and experimental work conducted during this thesis. I stress which questions are left answered and suggest how future investigations could tackled them.

Highlights:

- The cognitive semantic space is organized around multiple perceptual and conceptual dimensions
- They are distributed yet partially dissociated way across the cortex.
- They recovered automatically, rapidly, and simultaneously during reading.

1. Main Empirical Results

Different terms have been used to refer to our ability to store, retrieve, manipulate and share knowledge about objects and concepts: *semantic memory*, when the accent is on the mnestic component and on the dissociation with respect to information on events (episodic memory); *semantic knowledge*, when the stress is on the conceptual nature of the information that is processed, as opposed to the perceptual processing that takes place in modality specific areas after external stimulations; simply *semantics*, when a linguistic perspective is sought and the tight link with language is highlighted.

Cognitive and neural correlates of semantic representations have long been investigated with multiple neuroimaging techniques as well as thanks to neuropsychological evidences (Lambon Ralph et al., 2017). Having reviewed the main open questions and major

theoretical positions (Chap. 1), the thesis here presented sheds some light onto the neuro-cognitive correlates of semantic representations by means of behavioral, fMRI and MEG experiments. In this work, I narrowed my exploration (1) by focusing on semantic representations as static entities, without dwelling on the processes acting on them, (2) by choosing symbol meaning retrieval (i.e. word reading) as proxy for all other components of semantic knowledge.

1.1 Motor-Perceptual vs Conceptual Dimensions

Through the behavioral experiments conducted (Chap. 3), I observed that the cognitive semantic space is organized around multiple motor-perceptual and conceptual dimensions, not easily isolated. Moreover, I showed that the representational spaces retrieved with different behavioral and corpora-based methods (i.e., Semantic Distance Judgment, Semantic Feature Listing, WordNet) appear to be highly correlated and overall consistent within and across subjects.

In this thesis I propose a heuristic distinction between motor-perceptual features (i.e., those attributes of the objects and actions referred to by the words that are perceived through the senses) and conceptual features (i.e., the information emerging via the integration of multiple, non-correlated motor-perceptual features (e.g., *tomato is edible and it has seeds, thus it is a fruit*). This general distinction is both theoretically and methodologically advantageous. First, it bridges the cognitive and neural side of the problem, by forcing researchers to pay attention to both the psychologically relevant dissociations, and the computational and anatomical constraints that could explain them. Second, while planning empirical testing of the different hypothesis on the neuro-cognitive substrate, it helps the operationalization of the variables at play.

1.2 Topographical Dissociations

Following the proposed distinction between motor-perceptual and conceptual dimensions, and capitalizing from recent advance in data analyses (Chap. 2), first I investigated the neural substrate of perceptual and conceptual dimensions with an fMRI experiment (Chap. 4). I focused on one visuo-perceptual dimension (i.e., implied real world size) and two conceptual ones (i.e., semantic category and cluster). I have been able to show a representational shift along the ventral visual path: from perceptual features, preferentially encoded in primary visual areas, to conceptual ones, preferentially encoded in in mid and anterior temporal areas (Borghesani et al., 2016).

A follow-up MEG experiment supports the observed topographical partial dissociation (Chap. 5): both sensors topography and source reconstruction suggest that different cortical areas are responsible for the perceptual and conceptual effects detected. The two perceptual dimensions investigated (i.e., implied real world size and sound) appear to be encoded in modality specific areas, while the conceptual one (i.e., semantic category) in multimodal, associative areas.

Together, these results indicate that complementary dimensions of the semantic space are encoded in a distributed yet partially dissociated way across the cortex.

1.3 Temporal Dynamics

I investigated the temporal dynamics of semantic representations thanks to four priming experiments as well as with an MEG study. Two important conclusions stemmed from the evidence collected with the priming experiments (Chap. 3). First, perceptual dimensions of word meaning (implied real world size and, especially, implied sound) are recovered during reading in an automatic way. Indeed, it appears as they are retrieved even for words that are not the target of the task at hand (i.e., the prime stimuli), and even when the

task does not explicitly requires it (as it focuses on other aspects of the stimulus semantics). Second, such recovery of perceptual features greatly interacts with the task performed by the readers.

In addition, our MEG results suggest that perceptual and conceptual dimensions, while sharing a similar temporal evolution, can be dissociated both in terms of the features of the brain signal encoding them and their sources (Chap. 3). Inter-trial phase coherence appears to be key for the encoding of perceptual features (i.e. the implied real world size and sound). Conversely, spectral power changes appear to support encoding of conceptual dimensions such as semantic category. Crucially, differences along both perceptual and conceptual dimensions are detected, almost simultaneously, extremely early in time: around 200 ms after stimulus onset.

Together, these results suggest that motor-perceptual and conceptual dimensions of the semantic space are recovered automatically, rapidly, and simultaneously during reading.

2. Implications for the Neuro-Cognitive Representation of Word Meaning

While writing this dissertation, as well as while performing the associated research, I have been guided by four core questions on the neuro-cognitive correlates of semantic representations: *what* is represented, *where* and *when* in the brain, and *how* so. As often the case in science, simple questions can lead to complicated answers, however, the empirical results obtained over the course of the present thesis can be used to provide partial solutions.

2.1 What is the Content of Semantic Representations?

Our results support a distributed, feature-based perspective on the content of semantic representations. In particular, I proposed and adopted an operational definition of motor-perceptual and conceptual features, which has proven to be instrumental to the observation of dissociations on both the topographical (see fMRI results) and the temporal-spectral (see MEG findings) dimensions of the problem. While acknowledging the fact that such distinction largely oversimplifies the conceptualization of the semantic space, I believe it can act as useful framework in guiding future research from different perspectives:

- *Learning in natural circumstances.* By definition motor-perceptual and conceptual features are acquired in very different ways, thus it would be interesting to see what happens when motor-perceptual experiences are lacking (e.g., blind subjects learning the concept of color).
- *Learning in controlled training situations.* When teaching about new items, their features can be taught in a declarative or experiential way. How does this impact the speed (and the outcome) of the learning process? How differentially so for motor-perceptual and conceptual features?
- *Clinical observations.* Deficits and degradations of semantic memory have been extensively studied in neurological patients. Is it possible to observe dissociations among motor-perceptual and conceptual features? Do they enlighten us on the brain topography of different dimensions of word meaning?
- *Dynamics of featural integration.* How are different features integrated? Are there differences between integration mechanisms acting at the level of motor-perceptual features and conceptual ones? Is integration a mechanism that occurs only during learning or is it re-instated every time during semantic processing?

2.2 Where are Semantic Representations Encoded in the Brain?

Our results support two overreaching conclusions on the localization of semantic representations and its driving principles.

Need for a hybrid model. According to our theoretical proposal and empirical findings, the cognitive divide across motor-perceptual and conceptual features seem to map onto the neural level. Motor-perceptual features are represented primarily in the early sensor-motor areas that encode those same features during perception and action. In contrast, conceptual features are encoded in the mid-anterior temporal lobe (and possibly other semantic hubs). Complementary dimensions of the semantic space appear thus to rely on different cortical areas, calling for hybrid models postulating the interplay of both modality specific and supramodal areas.

Representational geometry as empirical asset. While the multidimensional semantic space of word meaning appears to be spread throughout the neocortex, such distribution does not seem to be random. However, we just begun to scratch the surface of the precise anatomical and functional constrains that determine it. One way to pursue this path is by punctually disentangling the contribution of different motor-perceputal and conceptual dimensions to the representational geometry that can be read out of a given area. Precise analyses of how the representational geometries changes across the cortex, describing spaces dominated by one or more motor-perceptual or conceptual features, will help understanding where different dimensions are (preferentially) encoded and where they overlap.

2.3 What are the Temporal Dynamics of Semantic Representations?

With the priming and the MEG experiments conducted, I gathered insight on two levels of the temporal dynamics of word meaning representations.

The timing of semantic dimensions retrieval. Overall, our results indicate that both motor-perceptual and conceptual dimensions of the semantic space can be recovered automatically and rapidly during word reading. This suggests that retrieval of these dimensions is not simply epiphenomenal (i.e., due to mental imagery) as it is observed even during orthogonal tasks and it occurs at extremely short latencies.

Contextual effects: the role of the task. We observed how the requests of the task and the underlying dimensions of the semantic space interact determining either a facilitation or an inhibition of behavioral performance. This indicates that even simple processed such as single word reading are greatly influenced by the goal of the subjects, hence supporting a more dynamic view on semantic representations that does not assume a constant and univocal retrieval processes.

2.4 How are Semantic Representations Implemented in the Brain?

As for many other kinds of representations, we are far from understanding the neural code underlying semantic knowledge encoding and retrieval. Moreover, the techniques described and deployed in this thesis cannot provide a conclusive answer as they open a too indirect window on neural operations.

However, I exemplified how both fMRI and MEG data, especially thanks to the multivariate analyses recently developed, can

be used to investigate representational geometries changes across both space and time. The geometry of a representation can be used as an indirect descriptor of the kind of operations that can be performed on it, thus acting as a proxy for the format of the representation. As illustrated in my fMRI work, double dissociations can be appreciated: one area might appear to code (over and above categorical differences) for perceptual ones (e.g., V1 coding for implied size once the other effects are accounted for), while another area might show the reverse pattern (e.g., ATL coding for semantic cluster).

In terms of reaching an understanding of the underlying neural code, the closest finding comes from single cell recordings, thus this thesis does not speak to this question. Supporting a distributed feature-based neural code, it has been shown that neurons firing at the presentation of a specific concept (e.g., Yoda), show comparable response to semantically related concepts (e.g., Darth Vader), thus suggesting that concepts (and the events they are linked to) are not coded by the activity of single units but rather via partially overlapping assemblies firing for objects and concepts sharing certain features (Quiroga, 2016).

3. General Discussion

While each single study potential criticisms have been covered in the discussion of the corresponding chapter, I would like to highlight here the general limitations and open questions left unanswered from my studies. In brief, I explored semantic features in terms of neural topographical dissociations and promptness of recovery, yet (1) interpretable features are not the only viable organization of the semantic space, (2) even if dissociated, the different dimensions of such space interacts in a dynamic way, (3) early automatic effects do not imply that retrieval of those dimensions is necessary to te understanding symbols.

3.1 Interpretable Features, Evolutionary Domains or Latent Dimensions?

In the neuroimaging literature, the correlate of numerous semantic features have been investigated, notably color (e.g., Simmons et al., 2007), shape (e.g., Wheatley et al., 2005), motor attributes (e.g., Hoenig et al., 2008), auditory properties (e.g., Kiefer et al., 2008), as well as taste (e.g., Goldberg et al., 2006) and smell (e.g., González et al., 2006). However, it is still a matter of debate whether correlation within and across these different features are sufficient to explain domain-specific categorical effects. Such effects have been observed in neuropsychological patients as well as detected with imaging techniques. Domains that have been extensively studied and shown to potentially dissociate include living entities such as animals, fruits and vegetables (e.g., Gainotti, 2010), as well as non-living items, such as tools (e.g., Campanella et al., 2010). Likely, the neural organization observed in healthy adults is the outcome of a complex interplay between:

- biologically determined computational constrains, i.e., which material is best processed in each brain area (e.g., when details need to be decipher, high spatial accuracy is sought, thus areas receiving input from foveal region of the visual field are best),
- evolutionary relevant categories, i.e., classes of stimuli that call for specific behavioral responses such as conspecific (e.g., one can interact with them), dangerous animals (e.g., one should run away), and comestible plants (e.g., one can eat them);
- and statistical associations learned during life-long experiences with the external world, e.g., most of the tools that operate mechanically produce noise.

More studies are needed to dissociate purely categorical effects (if they exist) from the results of feature associations. The ideal setting is the combination of experiments with carefully selected stimuli - as to control for possible confounds-, and with naturalistic stimuli (e.g., texts encountered in daily life), as to cover the broad spectrum of

semantic features that organize complex semantic spaces. Training experiments will also be crucial in determining how different features/domains are assimilated during learning (e.g. Bauer and Just, 2015; Malone et al., 2016): are there differences between being taught that one item belongs to a certain domain in an explicit, declarative way (e.g., "*a X is an mammals*"), and learning it by (direct) observation of its properties?

A second open debate concerns whether semantic features are better understood in terms of interpretable attributes (Rogers and McClelland, 2004), or if the semantic space can be described by vector spaces devoid of meaning (Lund and Burgess, 1996). The long-standing tradition aiming at detecting explicit and directly accessible attributes culminates with the attempt by (Binder et al., 2016) to define biologically sound features, starting from those aspects of the physical and mental world we know to be encoded in specialized brain regions: e.g., texture, temperature, smell, harm, fear (total of 65 experiential attributes). The complementary perspective considers concepts as the outcome of the latent distribution of attributes in the real world, and has led to the development of several methods to extract the underlying vector spaces. These distributional semantic models differ in terms of the free parameters that need to be tuned and in their ability to generalize to new settings (Rogers and Wolmetz, 2016). Even if they appear to be able to account for many behavioral data (Pereira et al., 2016), we are far from an omni-comprehensive model capable, on its own, to cover all manifestations of semantic knowledge, including clinical data. Further comparisons are need: are models based on biologically inspired features (Anderson et al., 2016) better than those based on co-occurrence statistics (Huth et al., 2016) in predicting behavioral, clinical, and imaging data?

I believe that in the long term the problem will be better cast as a matter of localizing devices supporting specific kinds of computation, rather than semantic features per se. This will require investigation of functional and anatomical connectivity constrains that

determine *which* features are best processed and integrated *where*. Ultimately, the brain optimizes computational costs recycling hardwired brain maps to new tasks and materials (Dehaene and Cohen, 2007). In other words, the anatomy compels the global organization, while experience molds the content: concepts are not innate, but brains are constrained. The aim of future work should thus be the exploration of possible interesting dissociations between different features (or, better, computational demands), in order to understand the evolution and organization of semantic representations.

3.4 Dissociated, yet Interacting

In parallel with the debates on which are the features organizing the semantic space (and whether they are interpretable or not), the quest is open as for their neural substrates. The exploration follows three interconnected paths.

Which are the "spokes"? In other words, which are the peripheral centers where modality-specific dimensions are processed? Are all the primary and secondary sensory-motor cortices involved? Particularly interesting is the case of those motor-perceptual attributes which can be experienced via multiple senses (e.g., implied real world shape, a feature that can be perceived when seeing and object but also when touching it). Are these kinds of features coded in both primary visual and sensory areas? What are the differences, if any, between the two representations?

Which are the "hubs"? Convergence zones, likely located in multimodal cortices, are needed in order to integrate information about different motor-perceptual features. Moreover, they presumably support encoding of conceptual attributes learned in a purely declarative fashion (e.g., tomatoes came from the Americas). Clinical and neuroimaging data highlighted the role of the ATL, but is it the only semantic hub? Another likely candidate for cross-modal convergence of multisensory information is the angular gyrus (AG). As ATL, AG appears to be internally specialized (Seghier, 2013) and

future research should target how the two areas differ (in anatomical and functional terms) and interact (both online – i.e., when engaged in semantic tasks- and offline).

How is division of labor implemented? The definition of *hub* and *spokes* implies a hierarchy across regions which is yet to be confirmed and appropriately described. One hypothesis is that the central hub, according to the requests of the cognitive operation being performed, actively integrates sensory-motor information coming from the spokes, thus requiring their activation only if and when necessary. An alternative view considers the retrieval of sensory-motor information in specific areas sufficient to give access to the semantic representation, with the intervention of the amodal hub being necessary only under certain conditions. Predictions made by the different perspectives can (and should be) empirically tested.

3.5 Early, yet Superfluous?

The evidence coming from perceptual semantic priming experiments can be used to support a sensory-motor view of the cognitive (and neural) semantic system. However, priming effects can be interpreted as fast spreading of activation in a purely symbolic system capable of sensorimotor representations (Mahon and Caramazza, 2008): they do not necessarily entails the activation of sensory-motor representations/areas/formats. Similarly, the observation of early effects with time-resolved imaging techniques does not necessarily imply that those activations are central to the recovery of word meaning.

Recently, the interest has shifted towards interference paradigms which can have stronger implications for the causal role played by sensory-motor representations in semantics. The reasoning is as follow: if two tasks engage the same neural substrate, then performance should suffer (in terms of RTs and/or errors). Thus, if accessing meaning of words requires retrieval of perceptual features,

concomitant tasks should interfere with subjects' performance proportionally to the involvement of related sensory-motor features. For instance, understanding words with a strong auditory component should be affected by concomitant auditory tasks, while performance with words with strong visual components by a visual task. Ultimately however, only clinical data and evidence from virtual lesion studies (e.g., conducted with TMS) will be able to provide causal inference.

3.6 Effect of Context and Experience

Questions on the temporal dynamics of semantic representations have two ranges of implications: short and long term. In this work, I mainly focused on the short time frame investigating how rapidly and automatically different perceptual and conceptual dimensions are recovered during single word reading. The longer time frame of semantic processing has been partially considered by the priming experiments, where different tasks where compared. Further investigations are needed in order to establish (1) which perceptual features are consistently retrieved, and (2) which factors determine whether their retrieval will interact in a positive (priming) or negative (interference) way with the task. Moreover, there are much broader implications that deserve further examination. Words are never processed in a vacuum: they are usually heard/read in sentences, always in a communicative context (even if only with oneself), by individuals with given experiences and certain goals. Progressively more attention is being paid to naturalistic stimuli, individual differences and contextual variables (e.g., Hsu et al., 2011; Wehbe et al., 2014; Huth et al., 2016). Finally, cross-disciplinary research comparing different languages and culture can help defining how much of the content of semantic representation is culture-dependent and how different languages determine the mapping between semantic and conceptual spaces (for instance, given the same set of household containers, English speakers would produce 7 different names –and

corresponding classifications-, while Spanish speakers 15 (Malt and Majid, 2013)).

4. General Perspectives

Aiming at answering some of the above mentioned open questions and criticisms, the works presented in this manuscript can be directly extended along three, strongly interconnected, axes (not exhaustive of all the potential follow-ups called for by our finding).

4.1 Clinical Relevance of Features Dissociation

The topographical dissociation between perceptual (posterior) and conceptual (anterior) semantic dimensions I reported needs to be replicated and corroborated by evidences coming from the clinical population. Do the deficits of SD patients follow a similar (inverse) progression as the disease spread from anterior to posterior? An additional hypothesis is that this anterior-to-posterior gradient can be detected not only in the spokes (primary and secondary modality-specific areas), but also within the semantic hub (the ATL) where specific areas are responsible for coding of combined motor-perceptual features building up the higher level conceptual ones. These hypotheses could be tested with behavioral experiments aiming at triggering dissociation among motor-perceptual and conceptual features, comparing semantic dementia patients at different stages of progression of the neurodegeneration (where one would expect conceptual deficits to be more prominent than motor-perceptual ones), patients with posterior lesions (where the opposite pattern should be observed) and healthy controls.

4.2 Role of Hub(s) and Spokes

If both the integrative hub (ATL) and the spokes (modality-specific areas) are necessary every time concepts are retrieved, semantic deficit can arise because of (1) degraded representation at the level of the hub (central representational deficiency), (2) ineffective connections between the hub and the spokes (impaired access), (3) degraded representation at the level of the spokes (peripheral representational deficiency). While multivariate analyses of fMRI data can shed light onto the state of the representations in both the hub and the spokes, functional and structural connectivity data are needed in order to understand if/how communication flows within the system. Hence, patients manifesting key dissociations between different motor-perceptual and conceptual dimensions should be compared with protocols including functional neuroimaging as well as connectivity measures.

4.3 Dynamic Emergence of Semantic Representations

The ultimate goal of a theory of the dynamics of semantic representation should be the description of how the semantic (neural) code emerges (i.e., evolution during infancy), changes (by virtue of learning and training), and degenerates (e.g., in patients with semantic dementia). Synchronization, proposed as a mechanism for establishing communication between brain areas (Palva and Palva, 2012), likely plays a key role in binding the distributed codes of properties defining a given concept, to form a coherent semantic representation. Time-resolved techniques can be used to investigate how information is integrated thanks to the interplay between the hub(s) and the spokes. Interestingly, this could be tested not only during semantic processing of consolidated materials (e.g., during reading of words or sentences), but also before, during and after learning of new concepts.

Hopefully, researchers in the field will strike a balance between the questions we can answer (given the ever-improving methodological techniques available) and the questions we should ask (given the burning cognitive conundrums to be solved). We are in need of a fruitful integration of cognitive theories, clinical data, and cutting edge methodological practices, mutually pushing each other forward towards the resolution of the mysteries of semantic representations.

Bibliography

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

Dehaene S, Cohen L (2007) Cultural recycling of cortical maps. Neuron 56:384-398.

Hsu NS, Kraemer DJ, Oliver RT, Schlichting ML, Thompson-Schill SL (2011) Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. Journal of Cognitive Neuroscience 23:2544-2557.

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532:453-458.

Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of physiology, Paris 102:59-70.

Malt BC, Majid A (2013) How thought is mapped into words. Wiley interdisciplinary reviews Cognitive science 4:583-597.

Palva S, Palva JM (2012) Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs. Trends Cogn Sci 16:219-230.

Quiroga RQ (2016) Neuronal codes for visual perception and memory. Neuropsychologia 83:227-241.

Seghier ML (2013) The angular gyrus: multiple functions and multiple subdivisions. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry 19:43-61.

Wehbe L, Murphy B, Talukdar P, Fyshe A, Ramdas A, Mitchell T (2014) Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. PloS one 9:e112575.

APPENDIX:
VERBA VOLANT, SCRIPTA MANENT

Doctoral theses are changing, no one can deny it (Gould, 2016). Yet I do not believe we know exactly what we would like them to become. I confess, I wrote this manuscript to (1) test myself with the challenges represented by such an achivement, and (2) leave something behind, or better, have something to carry with me in my future career. Here are some additional informations/details of the analyses that could not fit the main chapters, yet are worth mentioning. I close by setting in stone (so to speak) some *dos* and *donts*, which I hope will be useful, at the very least, to my future self.

1. Supplementary Materials

1.1 Behavioral Experiments

One of the main aims of the analyses we perfomed in Chap. 3 was to obtain a low dimensional represenation of the semantic space we could (1) visualize and (2) compare across different sources of distance measures (SDJ, SFL, WordNet). Multidimensional scaling (MDS) is a set of methods that, given a matrix of pairwise distances or dissimilarities, permits the visualization of how near or far items are. In this dimensionality reduction technique, users need to set the goodness-of-fit criterion to be minimized (so called stress), and the number of dimensions they desire. Once data are arranged as a dissimilarity matrix, and one has an intuition on the number of clusters to be expected in a low dimensional representation (e.g., k clusters), a data-partitioning algorithm such as k-means can be used to assigns different observations to exactly one of the k clusters.

As our goal was to obtain the best representation in the lowest dimension possible, for each set of data we computed the vector of minimized stress for four different criterions: normalized by the sum of squares of the inter-point distances, normalized with the sum of 4th powers of the inter-point distances, normalized with the sum of squares of the dissimilarities, and normalized with the sum of 4th powers of the dissimilarities. We repeated the process iteratively chosing with all possible number of dimensions from 1 to n-1 (where n is equal to the number of items in each data set). It was then possible to plot value of the minimized

stress as a function of the number of dimension for each criterio. Applying the heuristic known as the "elbow rule", we identified the best criterion-dimension, thus attempting to strike a balance between maximum compression of the data and maximum accuracy. It is important to stress that these methods only offer qualitative ways to analyze distances data by visualizing one of their possible low dimensional representations. The interpretation of the results is thus instrecly ambiguous. For instance, k-means algorithm will always converge and provide a partion of the data into the desired k clusters, but the results can be very unstable as, while being deterministic, they rely on a random initialization (i.e., would not replicate unless the random seed generator is fixed). Similar observations can be made for the orientation of the dimensions in the MDS.

We here present, as example, the results for the SDJ of both set of stimuli (Study 1 and Study 2). We report the stress profile of the chosen criterion over the n-1 possible dimensions to appreciate the presence of an elbow, not always uniquely identifible but always lighing between 2 and 3 dimensions. Moreover, we report the corresponding Shepard plot, the scatterplot of the distances between points in the MDS plot against the observed dissimilarities (i.e., the closer to a perfect diagonal the better). Finally we show the position of the centroides as identified by the K-means algorithm with respectively 4 (Study 1, Fig. 75) and 2 (Study 2, Fig. 76) dimensions.

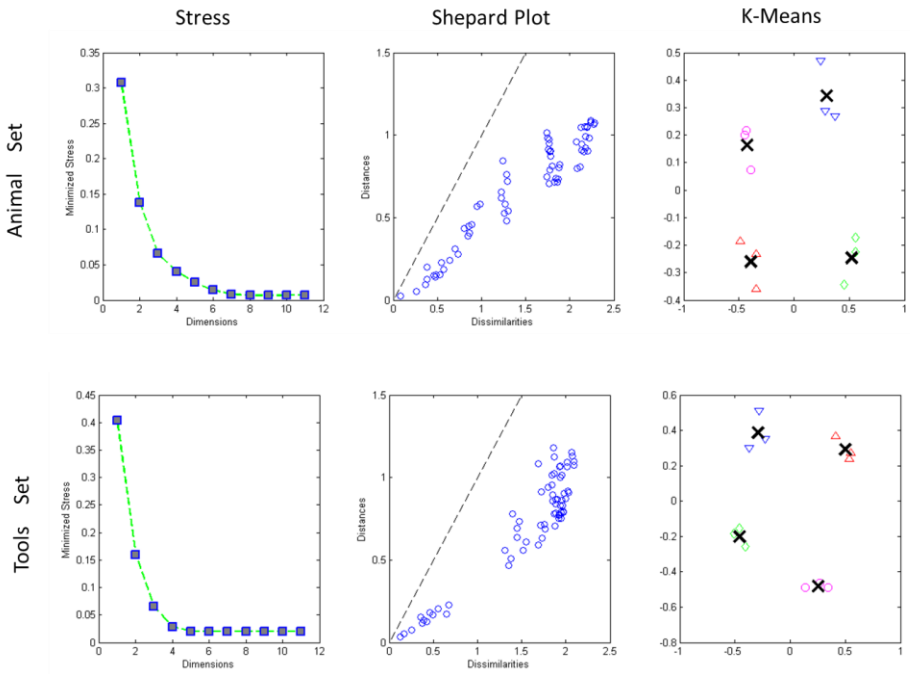


Figure 75 Preliminary analyses on the set of stimuli for Study 1. In an effort to visualize the semantic spaces retrieved by the two behavioral experiments conducted, we performed a series of stress analyses and K-means clustering. For both the animal (upper) and tool (lower) set of stimuli, we report the stress profile of the chosen criterion over the n-1 possible dimensions (left), the corresponding Shepard plot (central), and the position of the K-means centroides (right).

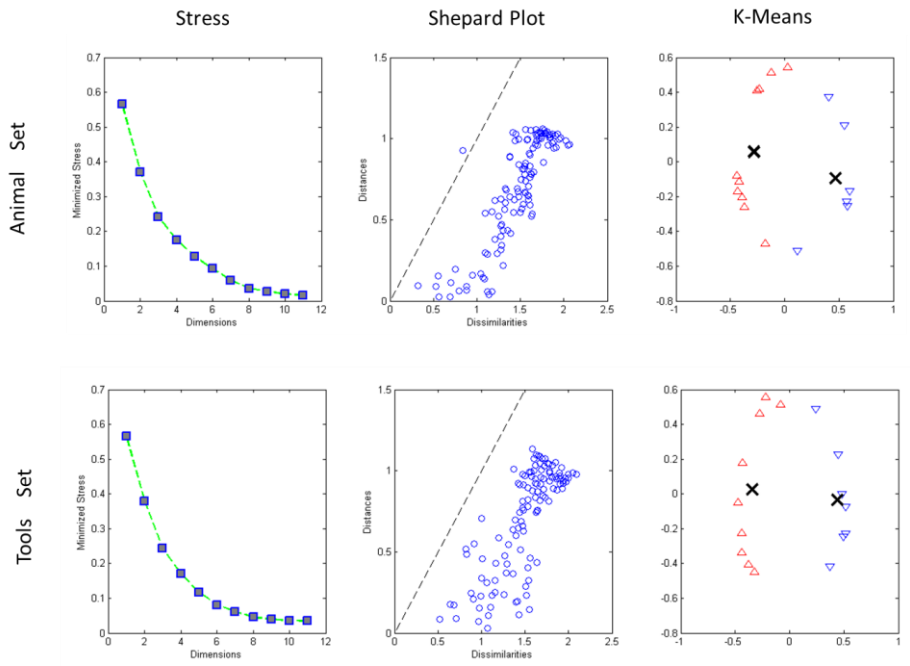


Figure 76 Preliminary analyses on the set of stimuli for Study 2. As in Fig. 1, but for the 32 French words used in Study 2.

1.2 fMRI Experiment

Searchlight. We run an exploratory searchlight with partial correlation RSA. For each subject and each predicted effect, we obtained a map depicting for each voxel the partial correlation score obtained by in a surroding sphere of 8 mm. The resulting maps were entered into a one-sample t-test with subjects as random factor in SPM. Fig. 77-a shows the temporal clusters surviving FWER correction ($p<0.005$) Length, Size and Cluster. Those shown for Category are uncorrected as none survived otherwise (in interpreting these results please bear in mind that there are partial correlations RSA, thus the effect of category was partialled out from the one of cluster). This picture is fairly coherent with the results we obtain from the ROIs analyses in pointing towards a postero-antero gradient from perceptual to conceptual dimensions of word meaning. Fig. 77-b shows corresponding whole brain results.

Fine grained description of the semantic space. In an attempt to test whether a richer, fine-grained semantic space (as the one described by our behavioral data) would fit better brain responses, we run two additional versions of the partial correlation RSA. Aiming at observing dissociations at different level of granularity (i.e., if some regions represent general superordinate category while others make finer grained distinctions across sub-categorical clusters), we kept the two different hierarchical levels of the semantic space

separated in both analyses. First, we tested whether the use of a continuous distance inside each cluster, as a replacement with our binary cluster matrix (see Fig 78-b) would improve the results. Such predicted cluster matrix is extremely correlated with the binary cluster measure (R = 0.85) and the results we obtain are virtually identical to our original results (reported for comparison in Fig. 78-a). Next, we attempted to model the category matrix as a continuous one as well (see Fig 78-c). Now, the two matrices are even more correlated with each other (r = 0.85 across the continuous matrices vs. r = 0.32 across the binary matrices). Moreover, it should be noticed that as we did not ask subjects to rate across-category pairs, the proximity of the within category items is here under-estimated compared to the across category ones. It is thus not surprising that the results show how most of the variance in the data is accounted for by the categorical one (even though a trend for the partial effect of cluster in the two most ATL regions can still be appreciated).

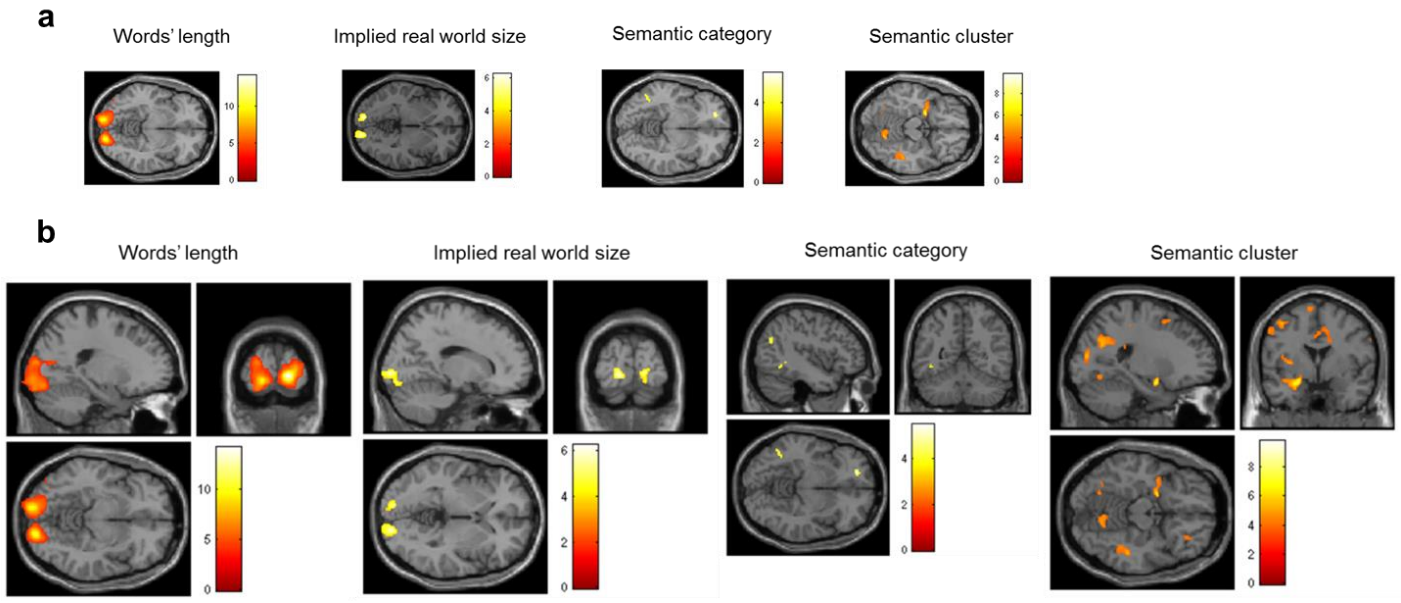


Figure 77 Results of the searchlight analysis run with partial correlation RSA. The map of partial correlation scores (spheres of 8 mm), where entered into a one-sample t-test with subjects as random factor in SPM. We here show the clusters surviving FWER correction ($p<0.005$), except for semantic cluster where uncorrected clusters are illustrated as no voxel survived correction.

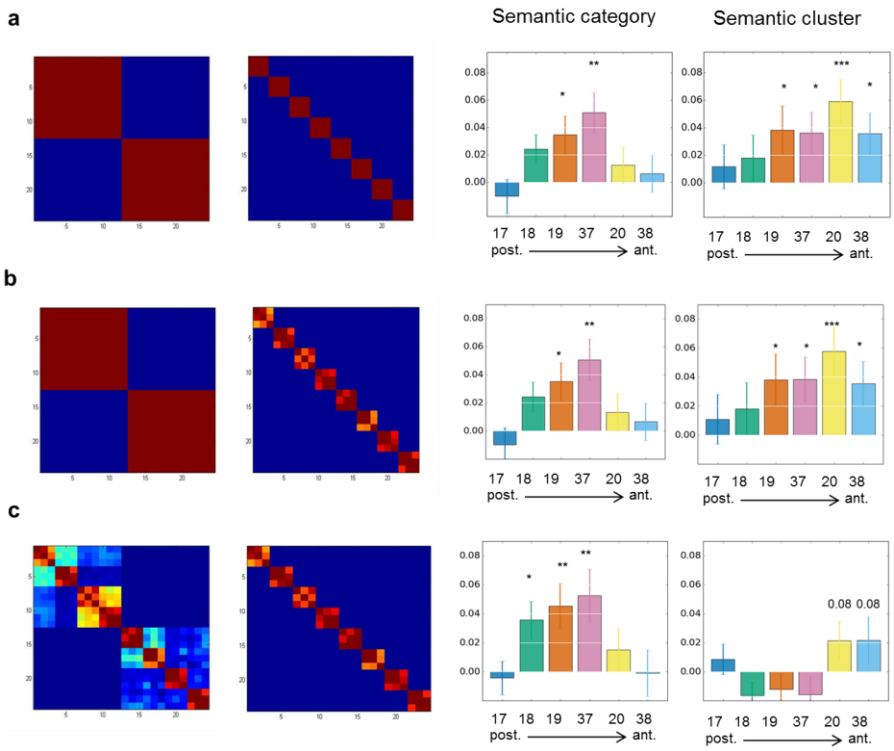


Figure 78 Results with more fine grained predicted matrices. Results of the partial correlation between the neural similarity matrix and the one model (a) binary for both semantic category and cluster, (b) binary for semantic category and continuous for semantic cluster, (c) continuous for both. We are showing the average scores across subjects (n°=16) and error bars indicate the standard error of the mean (SEM) across subjects.Statistical significance (* p < 0.05, ** p < 0.001, *** p < 10-5) is computed with a permutation test and very low p-value are rounded to p < 10-5. Exact p-values are reported in the text and ** / *** survive Bonferroni correction (p = 0.05/6 areas = 0.0083).

1.3 MEG Experiment

Overall inter-trial phase coherence. Figure 79 describes the overall pattern of inter-trial phase-locking (or phase coherence) for the three sensors type separately. It appears clear that for all sensors type the main effect is a theta increase in inter-trial phase coherence around 200 ms after the stimuli onset. The effect is confined over occipital sensors, with a more marked left lateralization in the gradiometers.

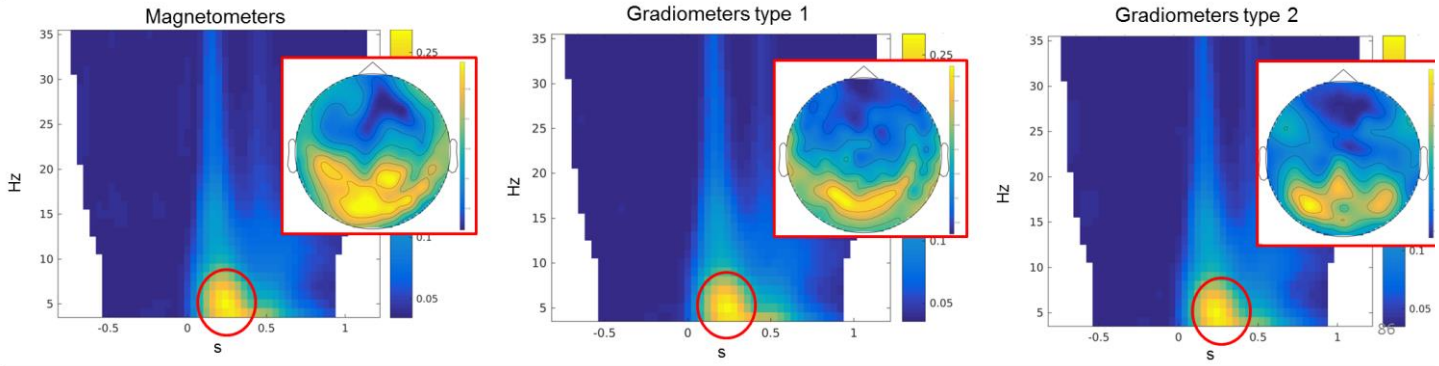


Figure 79 Time-frequency representation of the inter-trial phase coherence (ITC) across all stimuli. The insert highlights the topography at the sensor level of the main effect: increase in ITC in theta frequency range over occipital sensors, slightly left lateralized.

Overall power changes. Figure 80 illustrates the overall pattern of power changes for the three sensors type separately. Again, the two main effects are comparable across sensors type: theta increase around 200 ms, and beta/alfa decrease around 400 ms. As for ITC effect, the left lateralization is clearer on the gradiometers.

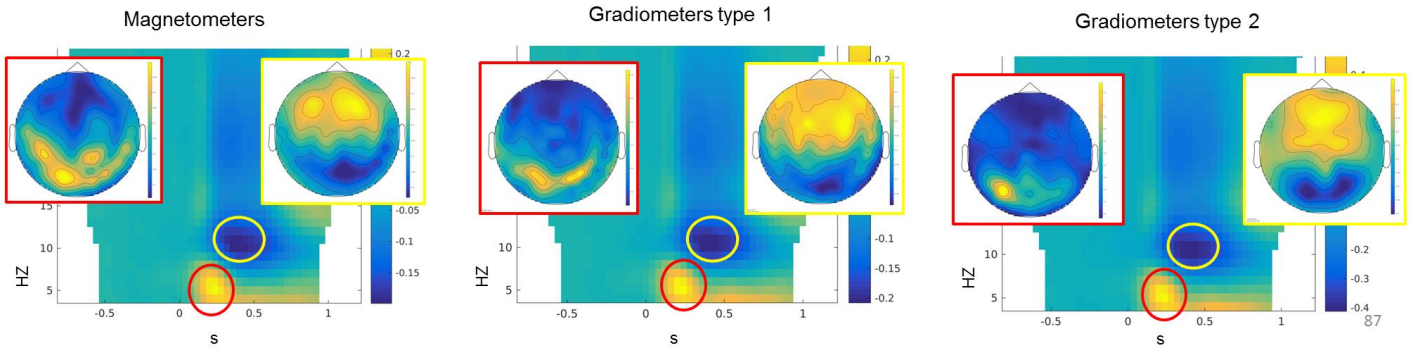


Figure 80 Time-frequency representation of the changes in power observed across all stimuli. The inserts highlight the topography at the sensor level of the two main effects: bilateral occipital decrease of alpha and beta bands, left lateralized occipital increase in theta.

Additional quality check. As we were observing relatively early semantic effects, we took care to control for possible confunds due to eye movements. If not properly taken into account during the preprocessing stage, differences across conditions in terms of eye arificats could potentially contaminate our results. We thus computed for each participant and each condition, the average vertical and orizontal electrooculogram and tested, time point by time point, whethe r any on the three effect of interest (i.e., words referring to noisy items or not, big items vs small ones, living vs non-living). No significant differences were observed (Figure 81).

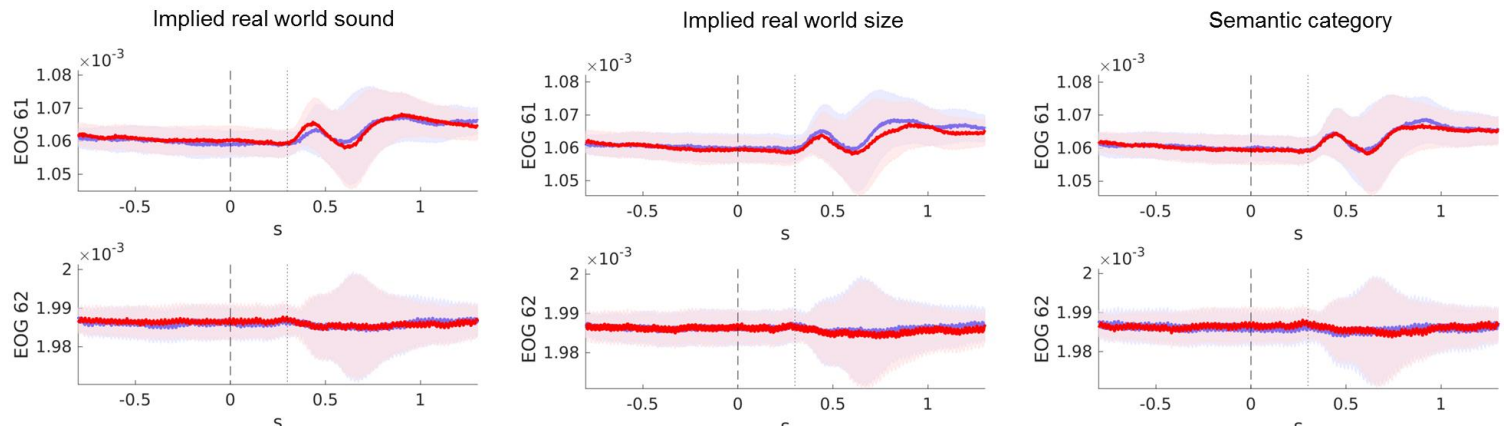


Figure 81 Additional controls for eye movements differences across semantic contrasts. We report for vertical, upper, and horizontal, lower, EOG the average and the standard deviation across subjects condition by condition. (left) Blue = words that refer to items associated with a prototypical sound, red = = words that refer to items not associated with a sound (middle) Blue = words that refer to big items, red = words that refer to small items (right) Blue = words that refer to animals, red = words that refer to tools.

Low level physical dimension of the stimuli. We assessed the possibility of retrieving the physical dimension of our stimuli (i.e., the number of letters composing each word) both at the univariate and the multivariate level. The pipeline of analyses used for the main contrasts of interest was slighlty modified to accommodate the continuous variable. For the cluster-based univariate analyses, we implemented the same ERF procedure as the one described for the main effects (Montecarlo method, 1000 permutations, see Chap. 5.2.7), but substituing the T-statistic with a regression statistic as implemented in Fieldtrip. Results indicate a significant cluster (corrected $p < 0.01$), between 144 and 196 ms (peak at 176 ms) (see Fig. 82).

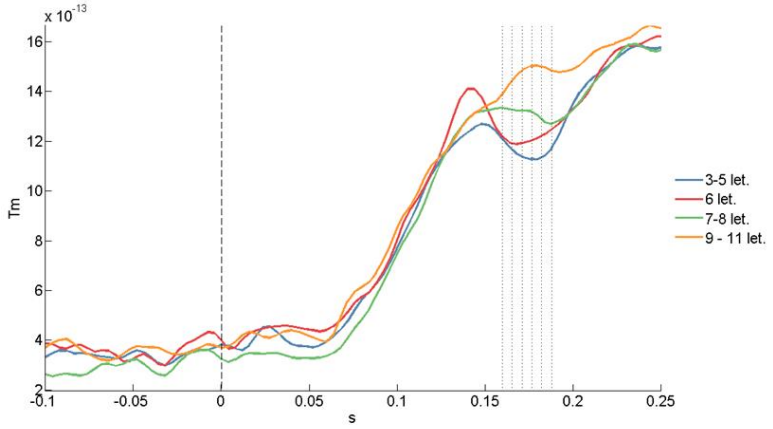


Figure 82 Event-related effect of word length. For visualization purposes only, the time course of the 32 words is here illustrated clustered in 4 groups of different length: 3-to-5 letters (8 words), 6 letters (7 words), 7-8 letters (10 words), 9-11 letters (7 words) Cluster-based significant effect ($P < 0.05$, corrected) is highlighted by vertical dashed lines. We are showing the average scores across subjects (n°=15). Stimulus onset is marked by a vertical dotted line.

For the multivariate analyses, a Ridge regression (default parameters as implemented in Scikit-learn) was used to decode the number of letter composing each word from the distributed pattern of brain activity in the time domain. This sanity check was used to test the effect on the decoding performance of some of the techniques of feature selection mentioned in Chap. 2. Fig. 83 reports the results of this exploratory analysis.

Event-related fields (ERF) Using univariate sensor-level statistics (see Chap. 5.2.7), we examined whether any significant difference along our dimensions of interest could be detected in the time-locked evoked activity (see Fig. 84). A significant effect of implied real world size (i.e., whether words referred to big or small items) was found in a left temporal cluster between 204 and 232 ms (peak at 212 ms, $p = 0.04$). Moreover, a significant effect of implied real world sound (i.e., whether words referred to items producing a prototypical sound or not) was detected between 384 and 460 ms (peak at 440 ms, $p = 0.009$) in a left

occipito-temporal cluster. These effects were observed on the combined gradiometers, while only trending effects could be appreciated in the magnetometers.

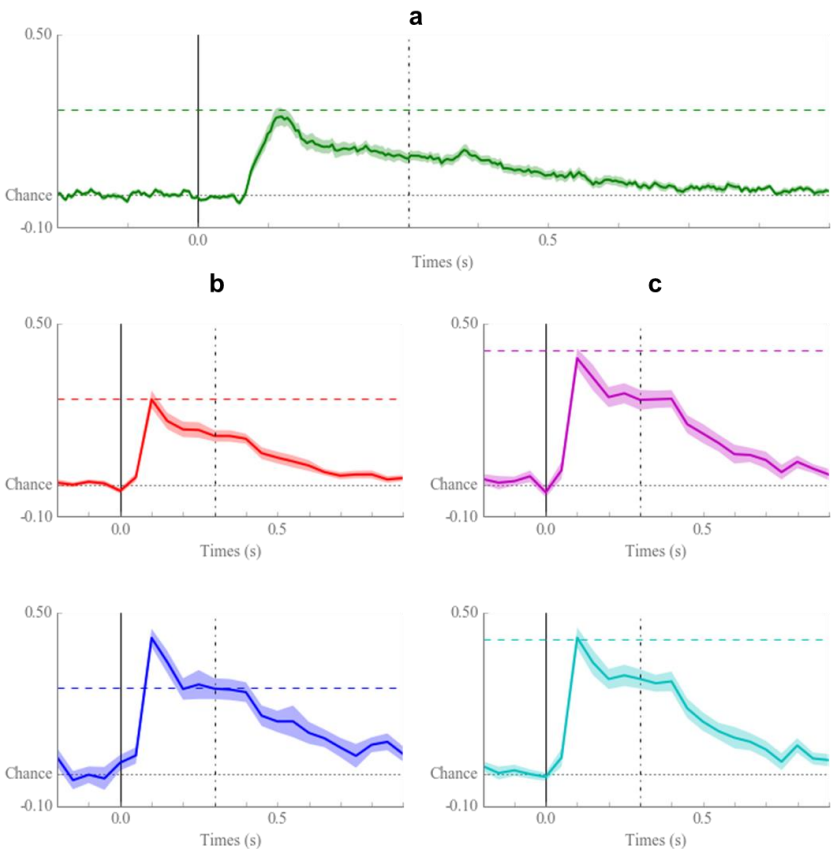


Figure 83 Decoding of the physical feature of the stimuli. Ridge regression was trained to classify words according to the number of letters composing them. We are showing the average cross-validated scores across subjects (n°=15) and shaded area indicates standard error of the mean (SEM) across subjects. Different features preprocessing pipeline and cross-validation scheme have been compared. (a) the classifier is trained and test on single trial epochs at each time point following a 5-fold stratified shuffle split cross-validation scheme. (b) the classifier is trained and test on single trial epochs concatenating 20ms second (red) and additionally averaging 5 trials (blue), same cross-validation as above (c) same as (b-blue) but leave-one-run-out cross-validation was implemented (purple) and feature selection (best 100 features) added (light blue). Horizontal black continuous lines indicates stimulus onset, black dotted lines stimulus offset. Vertical colored dotted lines indicate performance reached using all available information, i.e. all data points, trials and sensors.

Finally, the conceptual effect (i.e., difference between words denoting living vs non-living items) approached significance in a right fronto-central cluster between 384 and 432 ms (peak at 416 m, p = 0.05). Having appreciated earlier and stronger perceptual effects in terms of inter-trial phase coherence (ITC), we concluded that ERF might not be the most appropriate measure to detect small effects confined to a certain frequency band. When epochs are band-pass filtered to include only the frequency bands of interest (i.e., were significant clusters were detected in the ITC analyses), the effects of both visual and auditory dimensions are recovered (see Fig. 85).

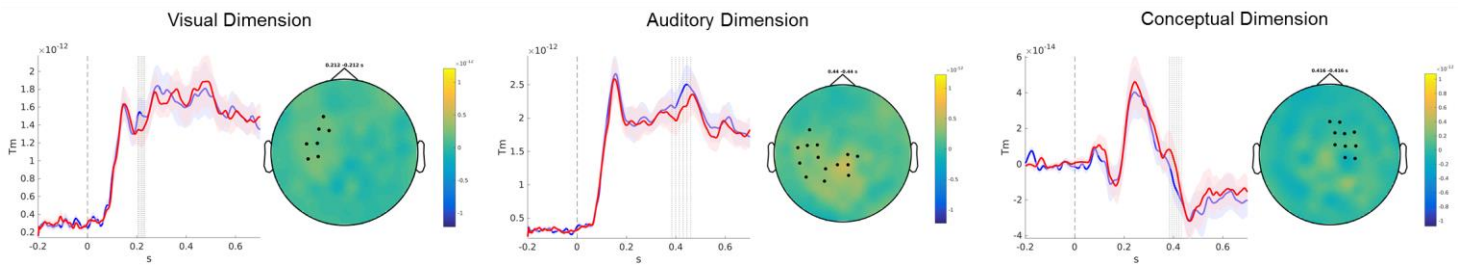


Figure 84 Event-related effects of perceptual and conceptual dimensions. (left) Visual dimensions effect (words referring to big items in blues, small items in red). (center) Auditory dimensions effect (audio-related in blue, not audio-related in red). (right) Semantinc dimensions effect (living in blue, non-living in red). We are showing the average scores across subjects (n°=15) and the shaded area indicates standard error of the mean (SEM) across subjects. Cluster-based significant effects (P < 0.05, corrected) are highlighted by vertical dashed lines (time-course) and * (topography). Stimulus onset is marked by a vertical dotted line.

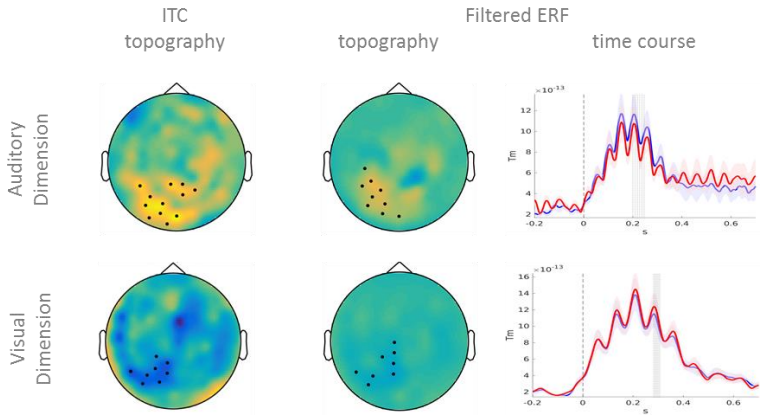


Figure 85 Perceptual dimensions effect as retrieved with ITC and ad evoke-related response on bandpass filtered data. We here show the comparison of the topographies of the significant clusters recovered with ICT and ERF on bandpass filtered data analyses. For the auditory dimension, a significant ITC effect (corrected p = 0.008) was observed peaking in the alfa band (10 Hz) at 200 ms (higher ITC for words referring to items associated with prototypical sound). In the ERF-bandpassed data, the same cluster shows a significant effect (corrected p = 0.01) peaking at 234 ms. For the auditory dimension, a significant ITC effect (corrected p = 0.039) was observed peaking in the theta band (6 Hz) at 240 ms (lower ITC for words referring to big items). In the ERF-bandpassed data, the same cluster shows a trending effect (corrected p = 0.07) peaking at 290 ms.

As described in the main chapter, we attempt to retrieve semantic information corresponding to the three dimension in the multivariate pattern of evoked activity. Fig. 86 illustrates the results of the time generalization technique adopted.

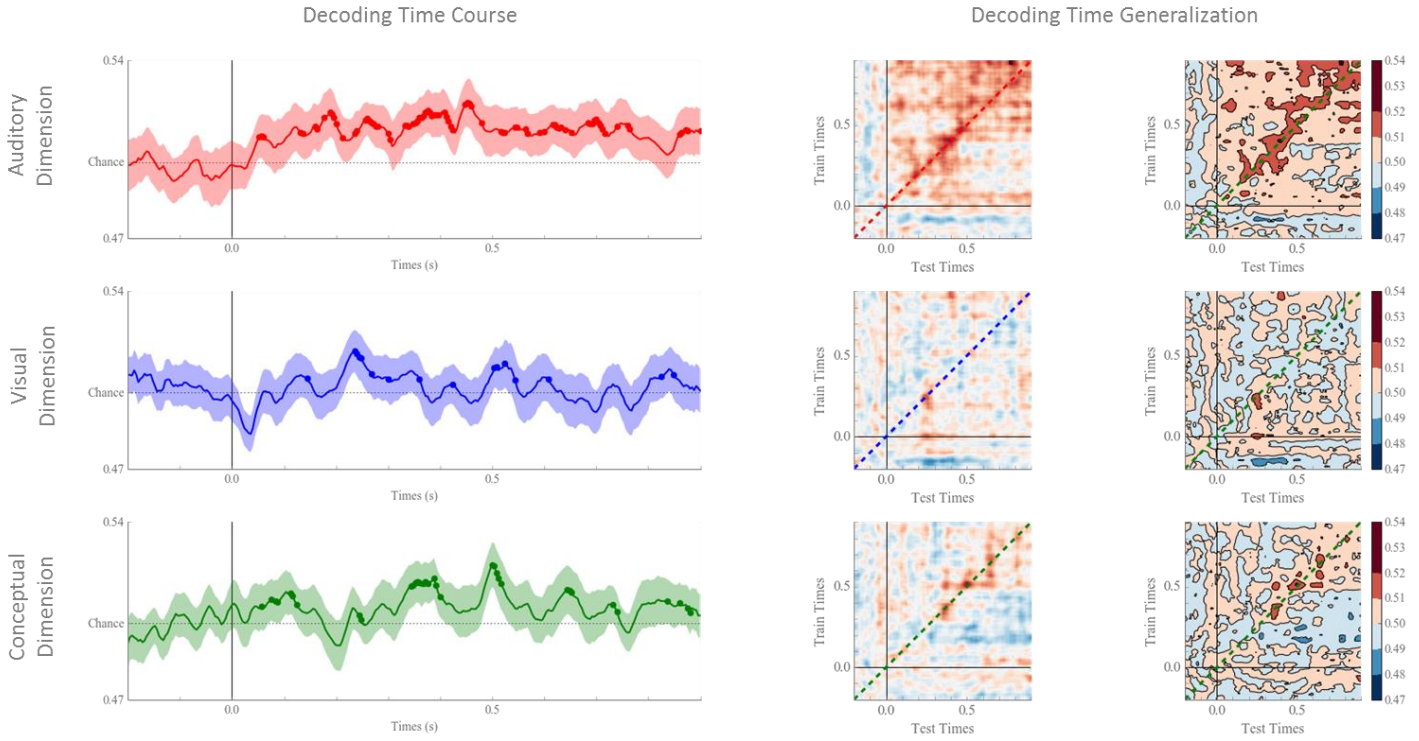


Figure 86 Decoding of the three semantic dimensions. The time course and the time generalization of the cross-validated decoding scores for the three effects: whether the words referred to items associated with a prototypical sound or not (upper, in red), whether the items are big or small (middle, in blue), and whether they are living of non-living entities (lower, green). Dots on the lines indicate significant decoding (Wilcoxon signed-rank test across subjects, uncorrected). For display purposes, data were smoothed using a moving average with a window of five samples. We are showing the average scores (and matrices) across subjects (n°=15). The shaded area indicates standard error of the mean (SEM) across subjects.

2. Software

When I started my doctoral studies, I had no previous experience with the analyses of neuroimaging data nor any knowledge of programming. Throught this manuscript I have highlighted the software I used, firmly believing software is a central part of modern scientific discovery too often underestimated (Pradal et al., 2013).

Psychotoolbox - (http://psychtoolbox.org/) is a free set of Matlab/Octave that can be used to design and control experiments (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). With version PTB-3 it moved to an open source development model. Great tutorial, demos and support (via a lively forum) are available. I used it to program the experiments presented in Chap. 3 and 4.

Psychopy - (http://www.psychopy.org/) is an open-source package for running experiments in Python. It offers the choice between two interfaces (Builder vs Coder) to allow design of rich experiments irrespective of the coding proficiency (Peirce, 2007, 2009). I used it to program the experiment presented in Chap. 5.

SPM - (http://www.fil.ion.ucl.ac.uk/spm/) is a software for analyses of fMRI, PET, and M/EEG data that runs in MATLAB. It is freely distributed and widely spread, being one of the most frequently adopted tools in neuroimaging (Penny et al., 2011). I used SPM for the preprocessing and univariate analyes of my fMRI data (see Chap. 4).

Scikit-learn - (http://scikit-learn.org/) open source, easily accessible, machine learning library in Python (Pedregosa et al., 2011). It is constantly growing thatnks to an international community effort. I used Scikit-learn for most of my decoding analyses (see Chap. 4 and 5), in fact, the best definition I can provide of machine learning is still "*what you need to import scikit-learn for*".

Nilearn - (https://nilearn.github.io/) is a open source library for machine learning on neuroimaging data in Python. It provides ready to ues advanced statistical techniques (heavily relies on Scikit-learn) (Abraham et al., 2014). I used Nilearn to plot many of the brain images

presented in this thesis, but it can do much more: e.g., predictive modelling, functional connectivity, brain parcellations.

Brainstorm - (http://neuroimage.usc.edu/brainstorm/) is a collaborative, open-source application dedicated to the analysis of brain recordings: MEG, EEG, fNIRS, ECoG, depth electrodes and animal electrophysiology (Tadel et al., 2011). I used Brainstorm for the preprocessing and source analyses of the MEG data (Chap. 5).

Fieldtrip - http://www.fieldtriptoolbox.org/ is a MATLAB software toolbox for MEG and EEG analysis (Oostenveld et al., 2011). I used Fiedltrip for the univariate statistical analyses of the MEG data in Chap. 5, and relied on its plotting function for the time-frequency searchligh described in the same chapter.

MNE-python - http://martinos.org/mne MNE is a software package for processing electroencephalography (EEG) and magnetoencephalography (MEG) data (Gramfort et al., 2013; Gramfort et al., 2014). It is the freely distributed output of a community-driven effort. I used MNE-python for the multivariate statistical analyses of the MEG data (Chap. 5), with the exeption of the time-frequency searchligh.

CoSMoMVPA - http://cosmomvpa.org/ is a open source library for MVPA implementations in Matlab/Octave. Handles fMRI volumetric, fMRI surface-based, and MEEG data through a uniform data structure across a variety of data formats (Oosterhof et al., 2016). I used CoSMoMVPA for time-frequency searchligh described in Chap. 5.

3. Dos and Donts

You need more data. Cutting edge methods developed for genetics, vision, etc…will always require a bigger sample, i.e. more data point than the one traditionally acquired for an fMRI/MEG experiment. Much more. Around 20% of the dataset will be need just to validate whatever your model tries to learn/predict/estimate, while a good percentage of what is left will be needed to tune parameters (or other sorceries).

You need better data. First of all, data needs to be meaningful: you won't find many potatoes in a cornfield. Second, data needs to be clean. In the case of fMRI data, the issues concern overlapping HRFs and subjects' movements inside the scanner. As for MEG, problematic factors are, again, subjects' head movements but any other source of magnetic artifacts such as blinks. Hence, general tips include: in fMRI, aim for longer ISI (ideally >3s), otherwise the HRF overlap, with no need to appeal to non-linearties, will killing your chance of detecting a meaningful signal; for both techniques, test well trained subjects, who will keep their head perfectly still and will blink only between trials.

Start from 0. By repeting this mantra, I try to remind myself fo two important lessons I learned during this four years of reasearch:

- At any point in your srudy (design, implementation, data analyses), invest time in the very first steps (e.g., checking data quality before data analyses), few things are as frustrating as wasting time second guessing yourself.
- Always try the simplest model/analysis first. There's plenty of time to complicate things, but only once chance to keep it simple and clean.

Try your pipeline on pure noise. As we do not usually do reasearch in a double blind fashion (i.e., we have the hypothesis, we design the experiment, we analyze the data, we interpret the results), any small choice is influenced by our goals. You find what you look for, and you look for what you want to find. A *better-than-nothing* procedure to test for misleading pipelines is to check which results would be obtained repeating exactly the same analyses (from preprocessing to statistics) on *completely random* data, pure *noise*. I am aware that the definition of noise/random itself poses some challenges, but "*it has to start somewhere*" (Guerrilla Radio, Rage Against the Machine).

Additional References

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Muller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G (2014) Machine learning for neuroimaging with scikit-learn. Frontiers in neuroinformatics 8.

Brainard DH (1997) The Psychophysics Toolbox. Spatial Vision 10:433-436.

Gould J (2016) What's the point of the PhD thesis? Nature 535:26-28.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. Neuroimage 86:446-460.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. Frontiers in neuroscience 7.

Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci:1-9.

Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Frontiers in neuroinformatics 10.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research:2825-2830.

Peirce J (2007) PsychoPy - Psychophysics software in Python. J Neurosci Methods 162:8-13.

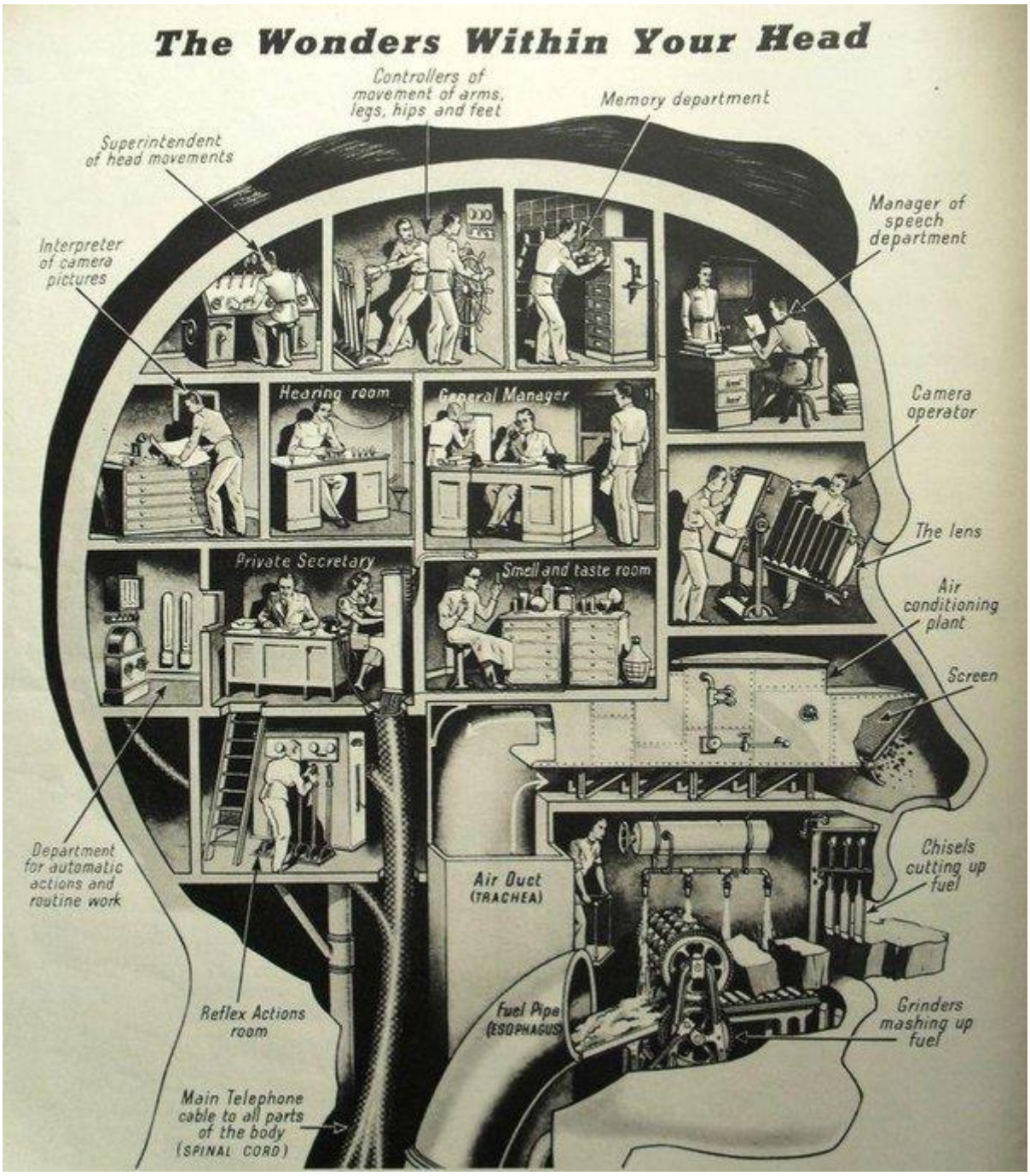
Peirce J (2009) Generating stimuli for neuroscience using PsychoPy. Front Neuroinform 2.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision 10.

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE (2011) Statistical parametric mapping: the analysis of functional brain images.: Academic press.

Pradal C, Varoquaux G, Langtangen HP (2013) Publishing scientific software matters. Journal of Computational Science 4:311-312.

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly application for MEG/EEG analysis. Computational intelligence and neuroscience 8.



December 6, 1938 issue of LOOK magazine