The neuro-cognitive representations of symbols: the case of concrete words

Valentina Borghesani, Manuela Piazza

ABSTRACT

We live our lives surrounded by symbols (e.g., road signs, logos, but especially words and numbers), and throughout our life we use them to evoke, communicate and reflect upon ideas and things that are not currently present to our senses. Symbols are represented in our brains at different levels of complexity: at the first and most simple level, as physical entities, in the corresponding primary and secondary sensory cortices. The crucial property of symbols, however, is that, despite the simplicity of their surface forms, they have the power of evoking higher order multifaceted representations that are implemented in distributed neural networks spanning a large portion of the cortex. The rich internal states that reflect our knowledge of the meaning of symbols are what we call semantic representations.

In this review paper, we summarize our current knowledge of both the cognitive and neural substrates of semantic representations, focusing on concrete words (i.e., nouns or verbs referring to concrete objects and actions), which, together with numbers, are the most-studied and well defined classes of symbols. Following a systematic descriptive approach, we will organize this literature review around two key questions: what is the content of semantic representations? And, how are semantic representations implemented in the brain, in terms of localization and dynamics? While highlighting the main current opposing perspectives on these topics, we propose that a fruitful way to make substantial progress in this domain would be to adopt a geometrical view of semantic representations as points in high dimensional space, and to operationally partition the space of concrete word meaning into motor-perceptual and conceptual dimensions. By giving concrete examples of the kinds of research that can be done within this perspective, we illustrate how we believe this framework will foster theoretical speculations as well as empirical research.

Introduction

You are attending a talk, and suddenly your phone starts ringing. Vibrations propagate in the air and reach your ears, your cochlea, and finally your brain (and the brains of others around you). The auditory information upon which you might act is readily available. Let’s now imagine a different scenario, in which you remembered to put your phone in silent mode. A colleague sees a flickering light on your mobile phone in silent mode. A colleague sees a flickering light on your mobile phone in silent mode. Whispering to you the word instead of writing it down would have led to an equally successful communication. Leaving aside the differences in the pragmatics of the two circumstances, the way information is processed in your brain in the two cases is dramatically different. Yet, they both yield the very same output: you grab the phone and quietly answer the call. The most important difference across these two scenarios is that in the second case you become aware of a physical event occurring in the external world (i.e., your phone is ringing) not by directly perceiving it, but by interpreting a symbol (see Glossary). Symbols are visual or verbal objects that we use to arbitrarily label other (classes of) known objects. These can be put in relation to one another in a generative way, building more and more complex representations that yield different representational geometries (see Glossary). Recollecting our initial example, your attentive colleague wrote five black marks: p h o n e. Whispering to you the word instead of writing it down would have led to an equally successful communication. Indeed, for human adults acquainted with deciphering oral and written language, a symbol can take the physical form of visual (e.g., strokes on a piece of paper) or auditory (i.e., sound waves) inputs, thus generating dramatically different sensory representations (see Glossary). However, the information acquired through these different modalities eventually converges and activates the same semantic representation (see Glossary): one single concept is activated by reading the letters "p h o n e".

Keywords:
- Symbols
- Semantic memory
- Concrete nouns
- Action verbs
- Representational geometry
- Neuropsychology
- Neuroimaging

ARTICLE INFO

Valentina Borghesani, Manuela Piazza

Valentina Borghesani a,b,c,⁎, Manuela Piazza b,c

a École Doctorale Cerveau-Cognition-Comportement, Université Pierre et Marie Curie - Paris 6, 75005 Paris, France
b Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, U992, F-91191 Gif/Yvette, France
c Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

⁎ Correspondence to: UCSF Department of Neurology, 675 Nelson Rising Lane, San Francisco, CA 94158, USA.
E-mail address: valentinaborghesani@gmail.com (V. Borghesani).

The authors declare no competing financial interests.

http://dx.doi.org/10.1016/j.neuropsychologia.2017.06.026
Received 25 October 2016; Received in revised form 18 June 2017; Accepted 21 June 2017
0028-3932/© 2017 Elsevier Ltd. All rights reserved.
or hearing the sound /fʌn/. The first key experimental evidence of the modally invariant nature of semantic representations is the phenomenon of cross-modal semantic priming, where words presented in one modality facilitate the processing of subsequent words when they belong to the same semantic category. This occurs whether they are presented in the same or different modalities (e.g., Holcomb and Anderson, 1993). Second, patients with difficulty retrieving the meaning of words often have deficits in both written and spoken input (e.g., Hodges and Patterson, 2007).

Etymologically, the term semantics (from the ancient Greek σηματικός /σηματικός/), which means to symbolize, to mean) refers to the study of how symbols are represented in our minds and how they represent meaning. Historically, different terms have been used to refer to our ability to store, retrieve, manipulate and share knowledge about objects and concepts: (a) semantic memory, which emphasizes the mnemonic component as dissociated from information about events (episodic memory); (b) semantic knowledge, which emphasizes information that is processed conceptually, as opposed to the perceptual processing that takes place in modality-specific areas after external stimulations; (c) simply semantics, which adopts a linguistic perspective, highlighting a close link with language. Through centuries of scientific exploration, linguists, philosophers, psychologists, computer scientists, neurologists, and cognitive neuroscientists have greatly contributed to the debate on the origin of semantic representations (how much is innate/learned?), their structure (how are they organized in memory?), their grounding (how do they connect with the external reality?), and their neural implementation (where/when/how are they implemented in the brain?).

Semantics emerged as a cognitive reality with the seminal work of cognitive scientists such as Quillian (1967) and psychologists such as Tulving (1972), and the following forty years have witnessed heated debates on its cognitive and neural representations, fueled by an ever-growing body of behavioral, clinical, and neuroimaging data. No theory so far has been able to integrate all the evidence and provide a biologically and computationally sound model that incorporates how symbols in general are acquired, stored, and accessed in the adult brain. However, many hypotheses have been put forward, some relating to specific classes of symbols such as numbers or action verbs, some being broader, and these have been the subject of previous reviews (Caramazza and Mahon, 2003; Binder and Desai, 2011; Pulvermüller, 2013; Eger, 2016; Piazza and Eger, 2016).

Roadmap of this review

In this paper, we review the literature related to a well-defined class of symbols: that of concrete words (e.g., nouns referring to concrete items, such as common tools, animals, objects, and verbs referring to bodily actions). We have chosen this class of words because, together with numerals (which we have recently reviewed elsewhere; e.g., Piazza and Eger, 2016), they are the most frequently studied class, being relatively easy to define in terms of their referents. Moreover, compared to abstract words, concrete words are more frequent, processed faster, learned and remembered better (Paivio and Caspo, 1971; Paivio et al., 1986; Schwanenflugel, 1991). Indeed, their oral forms (i.e., spoken nouns and verbs) are among the first kinds of symbols that children learn, usually in strict temporal association with the actual items (or actions) they refer to. We will structure our review around the following inter-related aspects that are necessary to characterize any representation within a neuroscientific frame: the content (i.e., what is represented in the memory system that holds the meaning of concrete words?), the neural implementation in space and time (i.e., where and when are the semantic representations of concrete words activated in the brain?), and the format (i.e., how does the brain code for that information, i.e., the underlying neural code?). For each of these topics, we will highlight currently unresolved questions and suggest ways in which they could be approached. Crucially, we propose a simple operational definition of semantic representations (see 1.3), and indicate how this can help foster theoretical speculations and empirical investigations on the neural substrate (see 2.4), the temporal dynamics (see 3.3), and the neural implementation (see 4.2) of semantic representations.

1. What: the representational content

Interpreting concrete words (as for any other symbol) is a multi-stage cascading process that starts with the reception of the sensory-specific input from the stimulated sensory organ and unfolds in time through two macroscopically separable stages. First, the identity of the symbol is recognized (identification stage). Second, its meaning is activated (semantic stage). There are several pieces of evidence that these two stages are at least partially separable, the most powerful one being the double dissociations observed in patients with certain neuropsychological deficits. Specifically, we find patients for whom impairments in the recognition of visual or spoken words (pure alexia on one side and receptive aphasia without alexia in the other) can be doubly dissociated from impairments in the access to their meaning (semantic dementia) (compare, for example Epelbaum et al., 2008 and Gorno-Tempini et al., 2004). While the representational content of the recognition stage is the identity of the symbols, in the form of a specific sequence of letters or sounds, it is much less straightforward to conceptualize the representational content of the semantic stage. What is “meaning”, the content of semantic representations, and how is it implemented in the brain?

1.1. Unitary interconnected nodes vs. sets of distributed features

Early cognitive models of semantic memory considered semantic representations as nodes in a semantic network (Collins and Quillian, 1969; Collins and Loftus, 1975). In these models, concepts correspond to undivided units connected together via links of asymmetric strength according to their level of essentiality (e.g., it is essential for the definition of the concept ukulele the fact that it is a musical instrument, but not for the concept of musical instrument that one kind is ukulele). The core assumption of these models is that, when a word is processed (e.g., ukulele), not only the node in the semantic network corresponding to the concept is triggered, but activation spreads out along all connected paths in the network (reaching, for instance, lute, banjo, and guitar). Such activation progressively decreases, in a way that is proportional to the strength of the links. According to this view, the semantic space is organized according both to the semantic distance (i.e., the distance between two given concepts along the shortest path), as well as a semantic similarity (i.e., the aggregate measure of all possible paths connecting two given concepts) across words (see How to study semantic distance). This initial geometric model of mental representations has had a long and successful history (Coombs, 1954; Torgerson, 1965), even though it poorly accounted for two known issues. First, asymmetry, in which in a given class of objects, the less frequent concepts, called “variants” (e.g., ukulele), appear to be more associated with the “prototype” of the class (e.g., guitar) than vice versa (Tversky, 1977). Second, contextual effects, which suggests that representational space is probably more flexible than initially conceived - i.e., the specific list of items presented to subjects influences their similarity judgements (e.g., by adding a distal element, such as flute, the reported distance between ukulele and guitar is shrunk, while by adding an intermediate element, such as banjo, it is amplified (e.g., Goldstone et al., 1997). More successful models, which readily accounted for these effects, considered concepts not as unanalyzable units, but rather as a weighted combination of multiple features (Smith et al., 1974). An important novelty of these models is the introduction of the concept of “typicality”, which rests on the distinction between the aspects of word meaning that are essential, called “defining features” (e.g., for the concept bird: being a biped, having wings) and other that are not, called “characteristic features” (e.g., for the concept bird: flying, perching in trees). An
instance of a category will be highly typical if it possesses all the defining features and also most of the characteristic features. For instance, a *canary* is a more typical exemplar of the category *birds* than a *penguin*. Analyses of behaviorally-collected lists of semantic features allowed researchers to notice how different domains (i.e., living vs non-living things) contrast in factors such as feature correlations and distinguishing features (McRae and Cree, 2002). According to these models, the representational geometry of the semantic space is determined by the number of features shared across words and the weights attributed to them: the more features two words share, the closer in semantic space they are (see *How to study semantic features*).

### 1.2. Attribute-based vs company-based models

The fundamental difference between unitary interconnected nodes and sets of distributed features as explanatory theories of semantic space has led to the development of two substantially different kinds of models.

The first set of models can be broadly described as connectionist models, based on feature ensembles (Rumelhart and Todd, 1993; Rogers and McClelland, 2004). They conceive semantic knowledge as emergent properties of units interconnected via weighted links. Sometimes labelled “attribute-based models,” each unit represents specific features, or “attributes” of the referent (e.g., the color, shape, size or affordance). These models rely on the more or less explicit assumption that the semantic space of words is inherently structured by the conceptual space of the objects and actions they refer to, and that this structure is shaped by the similarities and differences of the perceptual and motor features of the objects, in the case of nouns, and of the actions, in the case of verbs. In other words, these models assume that semantic representations reflect object and action representations. Because they directly connect symbolic representations to real object and action representations, these models approach and account for the symbol grounding problem (see Glossary).

Conversely, company-based models emphasize that word meaning can emerge via statistical extrapolation of the relations among words themselves. (Lund and Burgess, 1996; Landauer and Dumais, 1997). This assumes that words co-occurring in text will tend to be strongly linked in memory and close in meaning. Largely based on the statistical analyses of word co-occurrence in large corpora, these models conceive words as points in a multidimensional space where the various dimensions represent a complex combination of other words and the referents they refer to, and that this structure is shaped by the similarities and differences of the conceptual and motor properties of the objects, in the case of nouns, and of the actions, in the case of verbs. In other words, these models assume that semantic representations reflect object and action representations. Because they directly connect symbolic representations to real object and action representations, these models approach and account for the symbol grounding problem (see Glossary).

Company-based models, however, have turned out to be a powerful tool for psychologists because, by estimating to what extent words are used together in language (e.g., the word *dog* is frequently associated with the word *leash*), they provide important measures of associative relatedness. On the contrary, attribute-based models, mainly relying on subjects’ semantic similarity or feature listing judgements provide measures of semantic relatedness (e.g., the word *dog* is semantically close to the word *wolf*). Since early investigations, authors have distinguished between frequently associated and semantically related words: for instance, some found priming effects in the case of semantically similar but not associated words (Fischer, 1977). Others were able to detect automatic priming for associated word pairs, but not for word pairs that were semantically related but not associated (Shelton and Martin, 1992). After 30 years of research, a large meta-analysis indicated that both association strength and feature overlap appear to contribute to automatic priming, stressing the need for further investigations of what appears to be a complex interaction between frequency of co-occurrence and semantics (Hutchison, 2003; see however Lucas, 2000 for evidence of pure semantic priming effects and a refutation of pure associative priming effects).

### 1.3. Future directions - Breaking down the semantic space into interpretable dimensions

Over time, psychologists have tended to embrace a perspective in which words are represented as a sum of features (notwithstanding the recent resurgence of company-based approaches driven by the application of artificial intelligence algorithms to internet-based corpora). Within the attribute-based approach, several models prioritize many (but not all) defining features of concrete words as related to the perceptual and motor properties of objects, such as color, sound, and affordance, etc. (For a recent model based on functional divisions in the human brain see Binder et al., 2016). However, considering that perceptual and motor features are often highly correlated and thus redundant in naturalistic circumstances, we might question whether the set of features that best describes semantic representations truly corresponds to single perceptual-motor properties. Moreover, in those models that also integrate categorical taxonomic information (Rumelhart and Todd, 1993; Rogers and McClelland, 2004), features that are perceivable through the senses (e.g., “being red”, for *tomato*) are not considered as being qualitatively different from features that imply a more complex hierarchical organization (e.g., “being a fruit”, for *tomato*). This appears as a limitation.

We here develop the approach we recently adopted as we believe it could be a heuristically powerful tool for future studies, potentially reconciling the divide between company-based and feature-based models. This approach merges the spatial and hierarchical view of semantic representations (mostly inherited from semantic networks and company-based models), and the notion of word meaning as feature ensembles (mostly inherited from attribute-based models) (Borghesani et al., 2016). The resulting idea is to conceive semantic representations of concrete words as points in a multi-dimensional space where each dimension represents either a single sensory/motor feature of the objects referred to by the words (termed *motor-perceptual features*), or a single higher order descriptor that can be either categorical (for *tomato*, “being a fruit”) or declarative (only learnt through language: “being native to America”) (termed *conceptual features*) (see Glossary). Distances between words in this space can be measured by the pattern of distances along each single dimension. With the umbrella term *motor-perceptual features* we cover all features of objects referred to by words that can be (and typically are) perceived through the senses and/or experienced through action. Motor-perceptual features thus comprise: 1) features of objects that are mainly apprehended through one specific sensory modality, such as color, taste, smell, sound, etc.; 2) features of objects that can be equally resolved via multiple sensory systems, such as average size or prototypical shape, which can be sensed through multiple sensory modalities (such as both vision and touch); 3) the feature of the actions that are typically associated with the objects (e.g., the semantic representation of *tomato* includes the notion that it is typically cut and/or eaten). Because the specific combinations of perceptual and motor features of objects are constrained by the physical laws of the world we live in and by the biomechanical properties of our body, they often correlate with each other (e.g., small objects tend to produce high pitch sounds and tend to be graspable; green food tends to be acidic). According to our proposal, however, each individual feature (which constitutes a single dimension in our putative semantic space) can and should be considered separately, especially when investigating the neural underpinnings of semantic representations (see Section 4). However, semantic representations do not only include motor-perceptual information, but also higher order descriptors, which we term *conceptual features*. With this term we encompass features derived from the integration of multiple, non-correlated perceptual features (e.g., tomato is edible AND it has seeds, thus it is a fruit), thereby requiring non-trivial conjunctive operations of features, or that are derived from...
purely declarative information, i.e. learnt solely through language (e.g., the notion that tomatoes were not cultivated in Europe before the discovery of the Americas). Contrary to the sensory-motor features, which can in principle be resolved through the activation of purely sensory-motor representations, conceptual features are mainly defined by language, and can therefore in principle be resolved linguistically.

**Glossary:**

- **Symbol:** a token, a pointer that stands for a concept. Symbols have three key aspects: (1) arbitrariness: the physical properties of the symbol (the word phone) bear no relation with the semantic content they provide access to (the full concept of telephone, i.e., the device that uses radio frequencies to send sounds over long distances, which has a prototypical form and color, makes a prototypical sound, and is kept in a prototypical place); (2) cultural-dependency: the meaning of symbols is defined only within a given linguistic and cultural milieu: it is a matter of cultural convention that we use a term (e.g., EN: mobile phone) instead of any another one (e.g., IT: cellulare); (3) unbounded combinatorial power: first, symbols can be freely merged, thus creating new symbols-concepts pairings (e.g., a smartphone is the combination of a regular phone with a smart system); second, their reciprocal relations can be analyzed in light of different context and goals, changing the corresponding representational geometry (e.g., the semantic distance between phone, smartphone, and camera will depend on the current need: making a call or taking a picture).

- **Representational geometry:** the pattern of relationships across items, conceived by distances across points in a (multidimensional) space. For instance, consider a bi-dimensional space of edible items where the two dimensions represent the visual features color and shape: in such a space, “carrot” and “banana” would cluster together, as they are both elongated and yellowish, while “raspberry”, which is different in both shape and color, would sit further apart. However, the distance across these items in a higher dimensional space taking into account also a conceptual dimension would be rather different: along the dimension “taxonomy” raspberry and banana cluster together – being fruits – and would be far apart from carrot – a vegetable). Depending on the aspect(s) of the symbols that are relevant for the task at hand, different dimensions of the semantic space can be highlighted and amplified.

- **Sensory neural representation:** the neural activity of primary and secondary sensory areas evoked by the presentation of a stimulus to a given sensory organ. The moment I see a flower, light is transduced by my eyes, information carried by the axons along the optical nerve, projected to my primary visual cortex, and subsequently a coherent representation of its shape, color, size, and position is reconstructed. Similar processes apply to the olfactory sensory representation (i.e., the smell of the flower entering my nostrils). Reading the letters /f l o w e r/ and listening to the sound /flau.w/ will activate different sensory representations.

- **Semantic neural representation:** the neural activity evoked by a symbol beyond its sensory representation; in the case of concrete words, this semantic representations includes both motor-perceptual features of the objects/actions represented by the words, and their more abstract conceptual dimensions. The semantic representation of the word flower (whether perceived in its written or spoken form) is the summation of all the features that define its meaning, both motor-perceptual (e.g., a flower is usually something I can hold with my hand, and with a general overall shape) and conceptual (e.g., a flower is the reproductive structure of angiosperms) ones.

- **Symbol grounding problem:** as first elaborated by Harnad (1990), refers to the problem of how words (i.e., symbols manipulated on the basis of their arbitrary, superficial, shapes) get their meanings, how they connect to the external reality.

**Motor-perceptual semantic features:** aspects of the meaning of a concrete word that under normal circumstances is apprehended through the physical interaction with the object it refers to (e.g., tomato is usually a red, roundish, item, which we usually eat).

**Conceptual semantic features:** aspects of the meaning of a concrete word that can be either (1) derived from the summation of multiple motor-perceptual features of the objects (e.g., a tomato is both edible and has seeds, thus it is a fruit), or (2) learned explicitly in a declarative fashion through language, as they bear no direct link with any combination of motor-perceptual features (e.g., tomatoes where not cultivated in Europe before the discovery of the Americas).

We believe that the general distinction between motor-perceptual vs. conceptual representations is theoretically and methodologically advantageous. First, it bridges the cognitive and neural side of the problem, by forcing researchers to pay attention to the psychologically relevant dissociations, as well as the computational and anatomical constraints that could explain them. Second, while planning empirical testing of different hypotheses for neuro-cognitive substrates, it permits operationalization of the variables at play. Within this framework, different theories would generate different predictions, not only related to the question of the content of semantic representation, but also interrogating the interconnected questions of where, when and how such content is implemented in the brain. This will lead to two main challenges for future research: (1) defining, for each given word, which features are necessary and sufficient to define its meaning, and (2) understanding to which computational mechanisms each feature can be traced back. The first question can be tackled via different empirical methods (e.g. unsupervised clustering of behavioral responses). However, we predict that clinical observations will also be of capital importance. Deficits and degradations of semantic memory have been extensively studied in neurological patients. Our proposal predicts that it should be possible to observe dissociations among degraded motor-perceptual and conceptual features. But many more questions would follow: in degenerative disorders, how do semantic representations of motor-perceptual and conceptual features develop as the disease progresses? In the case of focal lesions, which areas or network(s) are damaged? Additionally, computational mechanisms should aim to identify which particular kinds of neural computations can successfully encode particular pieces of semantic information. This stream of research will require development of computational models able to accommodate behavioral data from healthy subjects as well as patients with disorders in semantic memory.

2. Where: the neural topography

The current neuroimaging literature on the neural underpinning of semantic representations can be grouped in two broad classes, according to which of the above reviewed perspectives (attribute-based vs. company-based) is assumed. On one hand, researchers embracing an attribute-based view of semantics use experimental paradigms that allow for isolating the neural substrate of (a limited set of) perceptual or motor features of word semantics (e.g., Fernandino et al., 2015b; see Binder et al., 2016 for a review of this approach). On the other hand, those who resolve semantic representations within linguistic representations themselves, use word embedding to map semantics onto the neural activity of the entire neocortex (e.g., Huth et al., 2016). While this method fails to clarify what is being represented in the different brain regions (Barsalou, 2017), it is effective at modeling behavioral data (see Pereira et al., 2016 for a review).

Despite wide methodological differences, the ever-growing neuroimaging literature on the neural substrate of semantic representations converges in indicating that access to meaning indeed involves a wide
portion of neocortex. When taken together with neuropsychological literature, three clear conclusions emerge: (1) semantic processing encompasses higher order cortices as well as primary and secondary perceptual and motor areas (Fernandino et al., 2016; Huth et al., 2016); (2) observable dissociations suggest that certain brain areas appear to encode preferentially specific semantic categories and/or specific kinds of features (Caramazza and Mahon, 2003; Leschinsky and Caramazza, 2016); (3) generalized, multimodal and pervasive semantic deficits are observed in presence of lesions affecting mainly the anterior temporal lobe (Gorno-Tempini et al., 2004; Lambon Ralph et al., 2017). Accounting for these three observations is the crucial battleground for many diverging theories, revolving around two orthogonal questions: (a) what is the role of the different unsensory and associative areas activated during semantic processing? Does this vast activation reflect a truly distributed semantic network, or is semantic information primarily encoded in a small sub-set of dedicated regions while the activation in the remaining areas (and especially the sensory-motor ones) reflect spurious post-semantic processes, akin to post-semantic mental imagery? (b) are semantic representations structured according to an a-priori built-in set of semantic categories or are categorical representations emergent properties of a system that self-organizes according to feature co-occurrence?

2.1. Localization: distributed network vs. central core

One class of models assumes that semantic memory is widely spread over the cortex in primary, secondary, and associative brain areas, and that each node of the network is essential to recovering a full-blown representation of word meaning. At the opposite extreme, another class of models conceives semantic memory as a dedicated module within the language system, localized in language-specific and circumscribed brain areas dealing with purely symbolic computations.

The most popular kinds of distributed models are the so called embodied (or grounded) ones. Different theories have been put forward spanning a continuum from strong (i.e., perception, action and conceptual processing rest on the same neural systems) to weak (i.e., perception, action and conceptual processing are interdependent, partially sharing the same neural substrates) embodiment. For instance, Barsalou introduced the concept of perceptual symbol systems: according to this proposal the meaning of symbols is grounded in the same neural systems used for imagery and perception, with additional areas outside sensory-motor systems being recruited to support conceptual knowledge (Barsalou, 1999). Similarly, Gallese’s neural parameters simulation affirms that language is an emerging multimodal faculty that exploits pre-existing properties of sensorimotor systems and can be completely resolved in their computations. The sensorimotor systems are thought to provide all elements needed to implement the hierarchical structure that characterizes the mental organization of concepts, eliminating the need for an additional language module (Gallese and Lakoff, 2005; Glenberg and Gallese, 2012). Finally, Pulvermüller’s distributed neuronal assemblies postulates that, because during language learning concrete words are mostly encountered when the objects they refer to are physically present (or the action they refer to is being performed or observed), the perisylvian assembly of neurons dedicated to language establishes long lasting connections between the word forms’ representations and the pre-existing sensory and motor representations of the related objects through the simple Hebbian mechanism of simultaneous firing. Once these connections are established, processing of concrete nouns and action verbs will automatically re-activate the same sensory-motor cortices that were activated during object or action processing, thus grounding the meaning of concrete words (Pulvermüller, 1999, 2013). Evidence in support of these theories (often misinterpreted and distorted, see recent review in Barsalou, 2016) comes from different neuroimaging studies, which show partial overlap between the brain areas activated by elaboration of concrete nouns and action verbs and those activated by object and action processing. The posited neural overlap between word and object processing has been observed in relation to the motor system (i.e., leg, mouth and arm related words elicit somatotopic activations (Hauk et al., 2004; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006; Kemmerer et al., 2008), as well as many sensory systems: reading odor-related terms activates primary olfactory cortex (González et al., 2006), while sound-related ones activate the auditory cortex (Kiefer et al., 2008a), and taste-related ones the gustatory cortex (Barros-Loscertales et al., 2011).

Even though evidence for a distributed sensory-specific network is now compelling, fully embodied models still have some trouble accommodating two important pieces of data. First of all, they struggle to explain how and where nonlinear operations such as feature binding take place: concepts can be filled with properties which are more than the simple sum of their motor-perceptual features. Consider, for instance, composite - often culturally determined - conceptual features such as taxonomic classifications (e.g., tomatoes and cherries are both fruits according to the botanical classification, but only cherries are included in the culinary classification). Similarly, declarative information is acquired linguistically and not through interaction with objects (e.g., tomatoes are native to South and Central Americas). These operations are very difficult to accommodate in purely embodied models. Moreover, different semantic features need to be combined in a nonlinear manner allowing for: (1) appreciation of both superficial (e.g., tomatoes and tennis ball are both rounded) and deep (e.g., tomatoes and bananas are both fruits) similarities, (2) consistency through time and contexts (e.g., a tomato is such irrespective of whether you are reading a recipe for tomato sauce or caprese salad), and (3) adaptation and generalization whenever new information becomes available (e.g., once I learn that tomato can be used in sweet preparations, e.g., tomato sorbet, my concept of tomato, and of desert, will be forever changed) (Lambon Ralph et al., 2010a). Second, while clinical evidence does exist to suggest a crucial role for sensory-motor areas in representing semantic content, it is essentially circumstantial. For instance, evidence suggesting that accessing the meaning of action words depends on the integrity of the cortical (and subcortical) circuits primarily involved in motor control have stemmed from patients with syndromes as different as ideomotor apraxia (Buxbaum and Saffran, 2002; Negri et al., 2007; Pazzaglia et al., 2008b, 2008a; Papeo et al., 2010), corticobasal degeneration (Spatt et al., 2002; Cotelli et al., 2006; Silveri and Ciccarelli, 2007), progressive supranuclear palsy (Daniele et al., 1994, 2013; Bak et al., 2006), Parkinson’s Disease (Pignatti et al., 2006; Cotelli et al., 2007; Rodriguez-Ferreiro et al., 2009), and selective impairments in motor neurons (Bak et al., 2001; Bak and Hodges, 2004; Grossman, 2008; Bak and Chandran, 2012). However, null results have also been reported, such that 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). This contradictory evidence suggests that multiple factors may be at play and that more evidence is needed to conclusively resolve the nature of the observed brain activation in sensory-motor areas during semantic processing in neuroimaging experiments. In particular, it is necessary to gather further evidence to decide whether the sensorimotor regions are truly essential for semantic representations or whether they are simply epiphenomenal.

The radically opposing perspective on the neural correlates of semantic representations is that of completely disembodied theories that assign the burden of performing all semantic-related computations to specific and circumscribed brain areas. These models usually highlight how the meaning of concepts can be resolved in purely symbolic operations, for instance linguistic co-occurrences in a corpus (Landauer and Dumais, 1997). Emerging properties and non-linear summation of features are operations that are readily resolved by abstract symbols manipulations. However, as reviewed above, the symbol grounding problem is not (explicitly) faced and the details of the neural
implementation of how different aspects of semantic representations are bonded are largely underspecified. The existence of a core node for semantic processing is supported by the observation that generalized, multimodal and pervasive semantic deficits are detected in patients only in the presence of lesions or degenerations that affect the anterior temporal lobe as illustrated by the clinical profile of semantic variant Primary Progressive Aphasia (svPPA) (Warrington, 1975; Mesulam, 1982, 1987; Snowden et al., 1989; Hodges, 1992; Gorno-Tempini et al., 2004; Gorno-Tempini et al., 2011). svPPA is a specific type of frontal-limbic lobar degeneration in which deterioration starts in the anterior temporal lobe (ATL) and then progressively spreads posteriorly towards the parietal lobe, as revealed by in vivo anatomical imaging (Galton et al., 2001; Rosen et al., 2002; Davies et al., 2006; Brambati et al., 2009) and metabolic functional imaging (Diehl et al., 2004; Nestor et al., 2006; Desgranges et al., 2007). The key role of the anterior temporal cortex in semantic representations is also supported by converging findings coming from the analyses of white matter abnormalities (Agosta et al., 2009; Galantucci et al., 2011), and is corroborated by post-mortem pathological findings (Davies et al., 2005).

Given that the bulk of findings reported cannot be accommodated by an entirely distributed/embodied or entirely localized/symbolic perspective, several more recent proposals have attempted to solve this rigid dichotomy. First, it has been suggested that while semantic representations are distributed across the cortex, the integration of these representations takes place in dedicated multimodal hubs, so-called convergence zones (Damasio et al., 1996, 2004; Tranel et al., 1997a), distributed within convergence regions (e.g. temporal pole, anterior infero-temporal cortex, frontal operculum). These convergence regions are thought to be innately dedicated to performing conjunctive operations (i.e., available prior to any individual experience), but their content is shaped by learning and experience. Anatomical constraints are imposed on the location of convergence regions, for instance due to white matter tracts connecting different areas. A similar model, termed hub-and-spokes, posits that the integration needed to give rise to coherent, generalizable concepts takes place in a “transmodal semantic hub”, which is the site where crossmodal conjunctive representations (Binder, 2016) are stored (Rogers et al., 2004, 2006; Lambon Ralph et al., 2007, 2010b; Patterson et al., 2007; Lambon Ralph, 2014). According to the original proposal, the region believed to correspond to the semantic transmodal hub is the ATL, however, other potential candidate hubs have been proposed including the dorsomedial prefrontal cortex, the fusiform gyrus, the inferior frontal gyrus, the inferior parietal cortex, the posterior cingulate cortex, and the precuneus. Recently formalized, Reilly’s dynamic multilevel reactivation framework sees semantic memory as the emergent property of a series of local hubs that re-engage sensorimotor spokes during online reconstruction of object concepts (Reilly et al., 2016). Two kinds of hubs are described: (a) low-order hubs (such as the angular gyrus), where features coming from various modality-specific regions are bound together; (b) high-order hubs in the anterolateral temporal lobes, where symbolic transformations occur upon these integrated features.

2.2. Organization: categorical vs featural clustering

The second ongoing debate concerns whether the organization of semantic knowledge in the brain is based on categorical constraints (i.e., domain-specific clusters dedicated to evolutionary relevant semantic categories (Caramazza and Shelton, 1998)) or, rather, on featural constraints (i.e., certain concepts, and thus the words referring to them, cluster together in the representational space due to the correlations of sensory-motor and functional features that define them (Warrington and Shallice, 1984; Warrington and McCarthy, 1987; Tyler et al., 2000)). Semantic dissociations such as category specific semantic deficits can be elicited by appropriate testing when damage is confined to specific components of the semantic network (Warrington and Shallice, 1984; Capitani et al., 2003; Caramazza and Mahon, 2003). For instance, patients suffering from herpes simplex virus encephalitis (HSVE) often show category-specific semantic deficits: performance can be disrupted for living things, but spared for non-living items or artifacts (Warrington and Shallice, 1984; Pietrini et al., 1998; Sartori et al., 1993; Laiacona et al., 2003). Another source of data are patients with ischemic or hemorrhagic strokes who present with focal lesions along the ventral visual path, who sometimes demonstrate selective and domain-specific semantic impairments (e.g., a selective deficit for fruits and vegetables (Hart et al., 1985; Samson and Pillion, 2003) or for conspecifics (Ellis et al., 1989; Miceli et al., 2000), sometimes resulting in sharp double dissociations: 1) between living vs. non-living items (Warrington and Shallice, 1984; Pietrini et al., 1988; Sartori et al., 1993; Caramazza and Shelton, 1998; Laiacona et al., 2003; Blundo et al., 2006; Sacchetti and Humphreys, 1992; Laiacona and Capitani, 2001); 2) between inanimate living things (e.g., vegetables) vs. animate living things (i.e., animals) (Hart et al., 1985; Hillis and Caramazza, 1991; Farah and Wallace, 1992; Crutch and Warrington, 2003; Samson and Pillion, 2003; Hart and Gordon, 1992; Caramazza and Shelton, 1998). Some authors have re-interpreted these findings in light of statistical differences in the distribution of motor-visual features across different semantic categories (Humphreys and Forde, 2001). However, blind subjects, who have not acquired any visual experience with animals or tools, present a similar categorical organization in the ventral occipito-temporal path (e.g., Mahon et al., 2009; He et al., 2013, for a review see Bi et al., 2016), suggesting that the main bias underlying the categorical organization of the information in the ventral stream is not necessarily driven by visual experience.

2.3. A recent shift in the debate: which kind of information is encoded where?

Recently, especially due to the introduction of multivariate techniques for the analyses of neuroimaging data (so called MVPA, see How to study representational geometries), the literature has seen a shift away from the aforementioned debates towards the characterization of semantic information that can be read-out from the activity of different brain regions. One perspective, called encoding, convincingly showed how distributed (Huth et al., 2016), yet specialized (Fernando et al., 2015a, 2015b) the organization of the neural substrate of semantic memory is: while most of the neocortex is involved in word meaning representations, sensory-motor areas are preferentially recruited to process words that are strongly associated with the sensory-motor attributes they refer to (see Fig. 1a). The complementary approach, called decoding (Haynes, 2015), along with representational similarity analyses (RSA, Kriegeskorte et al., 2008), also provides important and converging insights. With respect to categorical representations, the pattern of activation of the whole brain contains information which allows for distinguishing words on the basis of their broad semantic category (e.g., living vs. non-living item; tool vs. dwelling, (Shinkareva et al., 2011), and that locally, such broad distinctions are especially supported by the mid-inferior temporal cortex (Fairhall and Caramazza, 2013; Simonova et al., 2014), and the mid-anterior temporal cortex (Borghesani et al., 2016). Beyond the distinction into broad semantic categories, other studies investigated the neural substrate of more specific semantic features, testing the hypothesis that they are each encoded in specific brain areas. Towards this end, a seminal work implementing factor analysis revealed that three different semantic features (Can it be used for shelter? Can it be manipulated? Is it food-related?) have differential loadings across the cortex (Just et al., 2010). Recently it was shown that even lower-level perceptual features of the semantic space (e.g., the implied real word size of the item referred to), can be traced back to specific brain regions, and appear to be preferentially encoded in early visual areas (Borghesani et al., 2016). These findings corroborate and expand previous observations of a perceptual-to-conceptual gradient of object representations along the ventral visual path (Peelen and Caramazza, 2012). Finally, semantic similarity -
reflecting overall conceptual proximity in terms of semantic features, both perceptual and conceptual - correlates with patterns of activity in the left perirhinal cortex (Bruffaerts et al., 2013; Liuzzi et al., 2015).

2.4. Future directions – Mapping different dimensions of the semantic space onto the cortex

We are far from a comprehensive answer to the question of the neural substrate of semantic representations. Taken together, the existing literature suggests that the multidimensional space of word meaning is distributed throughout the neocortex, requiring the interplay of both modality-specific and supramodal areas (see Fig. 1a). Moreover, this distribution is not random, but rather determined by the nature of the features encoded. However, we are just beginning to scratch the surface of the precise anatomical and functional constraints underlying this organization. In the framework we have proposed (see Section 1.3, and Borghesani et al., 2016), word meaning, considered as points in a multi-dimensional space, where each point represents one word and each dimension represents one feature, we intend to punctually disentangle the contribution of different dimensions to the representational geometry that can be inferred from different brain areas (see Fig. 1b). According to our proposal, the first distinction could be made between those features that relate to motor-perceptual aspects of the objects and actions referred to by the words, and those that do not directly relate to unisensory perception or action, including information emerging from the integration of multiple, non-correlated perceptual and motor features, as well as purely declarative knowledge.

According to our prediction, simple motor-perceptual features could be independently encoded in the brain, possibly through the activity of the same neural populations that encode them during real world interactions with the objects, such as during perception and action. On the other hand, conceptual features, which emerge due to conjunctive operations across simple features and/or via verbal definitions, could be encoded in high-order associative areas of the brain. Thus, despite word meaning being defined by both motor-perceptual and conceptual features, we predict a neuroanatomical segregation between motor-perceptual dimensions and conceptual dimensions. Reviewing the few neuroimaging studies of word processing that addressed this dimension of the problem, there seems to be initial evidence that (1) motor-perceptual features are instantiated in the primary and secondary cortices, where the same features are encoded during perception and action (Kiefer et al., 2008b; Just et al., 2010; Borghesani et al., 2016); (2) the various conceptual groupings of words, which emerge from their combined set of distinguishing features (or “semantic categories”) are encoded in the mid-anterior temporal lobe (Bruffaerts et al., 2013; Fairhall and Caramazza, 2013; Liuzzi et al., 2015; Borghesani et al., 2016). Future research in this direction could address some pivotal questions:

- What are the principles organizing the neural representation of semantic knowledge? Are they domains, features or a combination of the two? Precise analyses of how representational geometries change across the cortex, describing spaces dominated by one or more motor-perceptual or conceptual features, will help us understand where different dimensions are (preferentially) encoded and where they overlap.
- How many convergence hubs are there? What is their specific contribution to semantic processing? We can presume that different hubs have different roles, for instance in integrating information from different sources. If serving different kinds of integrations, they are likely 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 most likely to answer this kind of question. For instance, further exploration of functional and anatomical connectivity constraints can help determine which features can be integrated where
(see Fig. 2). Different cortical regions have already been shown to manifest preferential anatomical connections to distinct sub-regions within the ATL (Papinutto et al., 2016), and this has been used to support the proposed graded specialization of the putative semantic hub (Rice et al., 2015). Similar questions can be asked with respect to the description of the localization and specific computations carried out to support representations of the different motor-perceptual dimensions defining concrete word meaning.

- **Which of the involved areas are actually necessary (and not just accessory) components of the semantic network?** In order to gather data to support causal inference, we need to expand efforts in clinical and virtual lesion studies. Specifically, it is crucial to compare patients with congenital abnormalities of sensory-motor systems with those presenting with acute brain lesions or cortical degeneration in order to determine if those systems are involved in building and/or retrieving conceptual knowledge, and to assess possible intervening compensatory mechanisms. Critically, perceptual features outside the motor domain should be investigated, to appreciate the involvement of sensory-motor cortices at large. Finally, patients with lesions involving the putative hub(s) should be examined to understand how their performance is affected in terms of learning, manipulating and retrieving semantic features.

3. **When: the dynamics of processing**

The two stages of symbolic processing, the *identification* and the *semantic stage*, can be dissociated not only in terms of content (see 1) and localization (see 2), but also with respect to their temporal dynamics. Overall, during single word reading, brain activation unfolds from occipital areas towards the anterior temporal pole (Marinkovic et al., 2003; Pammer, 2009). Similarly, listening first elicits 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 (*identification stage*) are resolved within the first 200 ms. During this time, analysis of the visual-orthographic features starts in the primary visual cortex and then spreads in a feed-forward wave along the inferior occipital gyrus and fusiform gyrus (Tarkainen, 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). Finally, between 300 and 500 ms, activity mainly originating in left fronto-temporal areas has been historically related to semantic processing (Kutas and Hillyard, 1980). In particular, the N400 (a negative event-related potential peaking at 400 ms), has been long considered the electrophysiological component of accessing semantic memory (Kutas and Federmeier, 2000), because it appears to reflect facilitated access to given words due to the activation of features in long term memory (for a seminal review elucidating a possible interpretation of this effect see Lau et al., 2008). Recent findings, however, question this rigid view on the timing of semantic access, suggesting 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 hand). Hence, current theories on semantic representations present diverging hypotheses on two crucial aspects of the temporal dynamic of the semantic stage: (1) the timing of the earliest semantic activation and its characteristics: does it indicate an involvement of sensory-motor cortices? (2) the often over-looked long range aspect of the temporal dimension: how do current task goals and previous experience shape semantic representations?

3.1. **Early sensory motor involvement?**

Traditionally, research in language processing has accepted that semantic processes (i.e., those processes supporting the retrieval of word meaning) are reflected by late components of the event-related potentials (ERPs) (Holcomb and Neville, 1990). However, behavioral evidence has led some authors to suggest the interplay of two distinct mechanisms: a purely linguistic, co-occurrences-based representation, and a truly conceptual representation, possibly including perceptual simulations (Ostarek and Vigliocco, 2016; Zwaan, 2016). Moreover, a few electrophysiological studies have reported word category effects
(e.g., animals vs vegetables, living vs non-living) much earlier than the semantic N400 wave, at around 250–270 ms (Dehaene, 1995; Hinojosa et al., 2001; Martin-Looches et al., 2001), and even earlier, at around 180–200 ms for pictures, (Ji et al., 1998; Antal et al., 2000). The handful of studies that have used multivariate approaches have indeed revealed that it is possible to decode the semantic category from M/EEG signal acquired while subjects are presented with both pictures (Carlson et al., 2011), spoken words (Correia et al., 2015) and written words (Chan et al., 2011) with this latter case being the hardest (Simanova, 2010), as early as 250 ms and throughout the time between 200 and 400 ms.

Beside the dissociations between broad semantic categories, many studies in the last twenty years searched for the electrophysiological signatures of the recruitment of sensory-motor areas during semantic processing. Most of the findings come from studies that attempt to elucidate the first point in time when motor areas are recruited during conceptual processing by comparing verbs which are related with different body parts (e.g., kiss, a mouth-related verb vs kick, a foot-related verb). First, visually presented verbs can be dissociated from visually presented nouns as early as 150 ms after word onset (Moseley et al., 2013). This difference, however, is not necessarily semantic: according to linguists, it is far from clear if the verb-noun distinction is semantic, syntactic, or morphological (Kemmerer and Eggleston, 2010). Second, among visually presented body-related action-words (i.e., verbs), somatotopic differences can be observed already around 200 ms (Pulvermüller et al., 2000; Hauk and Pulvermüller, 2004). Similarly, when spoken action-words are presented, this somatotopic effect is reported around 170–200 ms (Pulvermüller et al., 2005) after the words can be disambiguated (recognized as unique lexical items). More recently, the same group has shown that when presenting spoken nouns and verbs, the characteristic somatotopic activations in motor cortex can be detected as early as ~80 ms after the acoustic disambiguation point (Shtyrov et al., 2014). However, the same six words were presented throughout the experiment (each word was seen 180 times), and the somatotopic distinction 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 reflected ultra-rapid semantic activity due to the particular and restricted experimental set of stimuli used.

3.2. The long range: the role of context and experience

Far from being resolved, the question “when?” is further complicated in light of several considerations. First of all, the depth, thoroughness, and therefore the timing of semantic processing likely depends on the task that subjects are performing. It is conceivable that different circumstances will lead to a different load on embodied vs. abstract representations. It is likely that different motor-perceptual and conceptual features are evoked only when task relevant, in a fast, automatic yet task-conditional fashion: for instance, reading “a p a 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). 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 shaped by the different interactions people have with the items their 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). Because it appears that concepts are less static than most existing theories can account for, we are in great need of explicit and testable predictions on when and why in certain situations different aspects of word meaning (more motor-perceptual, more higher level conceptual, or more declarative) are expected to be activated.

How to study...

Semantic Distance: One frequently used way to access subjects’ representation of the semantic distance between different words was to directly ask them to rate their pairwise similarity. Then, dimensionality reduction techniques such as multidimensional scaling (MDS) and hierarchical clustering are used to obtain a graphical representation of the reconstructed representational spaces (Shepard, 1980). When attempting to retrieve represented distances of a large set of items, the number of word pairs to be evaluated quickly becomes prohibitive. In those cases, alternative procedures such as so-called inverse-MDS can be used (Kriegeskorte and Mur, 2012).

Semantic Features: One way to investigate which features determine the representational geometry is that of directly asking volunteers to list a number of features that would describe a set of words. This method (Feature Listing) has been extensively used (Garrard et al., 2003; Cree and McRae, 2003; McRae et al., 2005) and shown to have fair correspondence with the complementary feature rating procedure, where subjects are presented with different features and are asked to rate how much that feature applies to each item (Tranel et al., 1997b; Gainotti et al., 2009). A direct comparison of the two approaches has revealed that ratings, not relying exclusively on participant’s verbal ability to describe the items, 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). However, both approaches (features listing and feature ratings) are limited (and biased) by psychologically relevant yet complex concepts: e.g., a tiger’s typical features (tail, stripes, paws, etc...) are concepts themselves. Recently, attention has turned to the definition (and validation) of set of features that are biologically sound, i.e., based on known modalities of neural information processing: vision, touch, audition, etc... (Binder et al., 2016). Beside these explicit comparisons or ratings, another paradigm adopted in numerous studies on language (and semantic in particular), is that of priming. A priming effect is observed when the presentation of a prior stimulus (prime) facilitates the processing of a second one (target). Traditionally, studies have focused on a conceptual relation between prime and target, such as belonging to the same semantic category (e.g., bread-cake (Fischler, 1977)). Subsequently, some authors have attempted to explore the effects of semantic features overlap, such as in the case of a shared visual shape (e.g. apple-ball (Schreuder et al., 1984)), similar associated movement (e.g., piano-typewriter (Myung et al., 2006)), or same color (e.g., emerald-cucumber, (Yee et al., 2012)). However, motor-perceptual priming appears to be possible only under specific circumstances, for instance it seems that attention of the subjects must have been directed to the targeted feature beforehand (Pecher et al., 1998; Yee et al., 2012).

Representational geometries: multivariate pattern analysis of neuroimaging data allows for new neorocognitive hypotheses to be put to test (Davis and Poldrack, 2013). In particular, thanks to representational similarity analyses (RSA, (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013) it is possible to directly and explicitly assess the representational geometries encoded in different brain areas at different time points, and to compare it across a variety of tasks, populations of subjects and with respect to computational models.

3.3. Future directions – Mapping different dimensions of semantic space as they unfold in time

The traditional semantic category effects, often but not solely linked
to the N400, combined with recent observations of early activity in the motor-perceptual circuits during action verb processing, might suggest that high-order conceptual representations are accessed later than motor-perceptual ones. However, given the type of data available to date, this interpretation is not certain: no single study has directly compared the timing of activation of both motor-perceptual and conceptual dimensions of word meaning within the same subjects, and using the same stimuli and task. This kind of comparison is necessary if one wishes to compare the temporal dynamics of different components of the semantic space. Furthermore, insufficient attention has been paid to the interaction between different tasks and how specific features are recovered. Overall, the role of timing (at different scales) appears crucial in enabling conclusions on the role of different distributed networks while distinguishing bottom-up and feed-forward processes from top-down and feedback ones (Hauk, 2016). In this respect, critical questions include:

- **Can we recover a temporal hierarchy across motor-perceptual and conceptual features?** If so, to what extent is it determined, or at least influenced, by the task that subjects are performing? One possibility is that all features of the semantic space are retrieved in parallel. Alternatively, a specific temporal hierarchy could be observed: purely linguistic, co-occurrence-based representations could be followed by purely semantic ones (Louvee and Connell, 2011), and among those, different timings could characterize motor-perceptual and conceptual features.

- **What are the mechanisms underlying the integration of semantic features when subjects read or hear object-related words?** If different dimensions of the semantic space are coded in different brain areas, what are the mechanisms through which the brain binds them during word processing, supporting our feeling of a unitary (instead of a scattered) representation of word meaning? Attempts in this direction might also disentangle possible differences between the integration mechanisms acting at the level of motor-perceptual features (e.g., how the feature “shape” is built via visual and tactile information) and conceptual ones (e.g., how the feature “being a fruit” is built via declarative knowledge of what seeds are).

- **How does such integration occur during the course of learning new word meaning?** Motor-perceptual and conceptual features are normally acquired in different ways. How does the absence of direct motor-perceptual experiences shape the semantic space (e.g., blind subjects learning the concept of color)? Being taught about a new concept’s features in a declarative or experiential way potentially impacts the neural underpinnings of learning. How is this different for motor-perceptual and conceptual aspects? Training experiments will deepen our understanding of the impact of structural-anatomical factors vs experiential ones, while controlling for stimuli features and their correlations.

### 4. How: the implementation of semantic representations

Generally speaking, how semantic information is represented in the brain is still an open question, which is hard to tackle because (1) we lack agreed-upon definitions of the variables at play (i.e., the content, the format and the neural code of a representation), and (2) we face technical limitations concerning what we can investigate (rarely single neurons, more often the combined activity of widely spread populations, or indirect measures of the cortical hemodynamics).

#### 4.1. Defining and investigating the elements of the implementation

We have seen that the content of semantic representation can be described in terms of the relationship between elements in the semantic space (i.e., the representational geometry that is encoded) and that based on which dimensions are considered, a specific pattern of distances/similarities will emerge. However, one must acknowledge that a given geometry (i.e., a given content) can in principle be instantiated through different formats (different terms have been used such as “symbolic” or “pictorial” (Pearson and Kosslyn, 2015)), in the same way that one figure could be represented on a computer by vector or pixel graphics, differing only by the kind of operations that they can support (for instance, only vector graphics can be stretched limitlessly). One possibility is that we could assume an isomorphism between the localization of the activation (i.e., the brain region where the neural activity is observed), and its representational content and format (Mahon and Hickok, 2016). For example, observing activation of the motor cortex during verb processing can be taken to indicate that the content of the representation underlying the activation is motor, and that the format of the representation is the same during semantic processing and action execution. Alternatively, format and content of the representations can be thought as entirely dissociated not only from each other but also, crucially, from the brain region involved (Mahon and Caramazza, 2008). According to this perspective, the brain imposes little or no constraints on either the kind of content it can represent or on its format. According to this view, motor and visual information conveyed by words can be equally well represented in the motor and in the visual cortex, and, irrespective of the brain region, in an entirely abstract format. Finally, the neural code is 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). While most of the currently available neuroimaging methods such as fMRI and MEG lack adequate spatiotemporal resolution to directly determine this level of neural implementation, intracranial recordings can indeed attempt to study the pattern of brain activity that carries information. For instance, in vivo cellular recordings could be used to compare the pattern of neuronal spiking rates between the motor cortex, the visual cortex, and the ATL during the reading of action verbs with those during actual action execution.

#### 4.2. The need to focus on representational geometries

We believe the most fruitful direction is to systematically investigate representational geometries through multivariate pattern analysis (or adaptation) across different brain areas. 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). For instance, in a recent experiment, we have used this approach to investigate and compare the representational geometries of different regions along the ventral stream while subjects read words. Results indicated that a size-based representational geometry was predominant in early visual areas, while a category and cluster based geometry characterized anterior temporal areas (Borghesani et al., 2016). One could take this evidence of a representational space dominated by a visual geometry (e.g., differences in size), detected in an occipital area (traditionally associated with visual processing), as an indication of a visual, depictive format. However, this would require the adoption of the above mentioned a priori assumption of an isomorphism between content/geometry, localization, and format. We agree with Martin (2015) that the format debate, as currently formulated, should be put aside as an ill-posed problem. The focus should be on questions that can be tackled empirically: how do representational geometries differ across brain areas, time points, and tasks? Different theories on the organization of semantic knowledge should make precise predictions:

- **Where can we observe representational geometries dominated by motor-perceptual vs. conceptual dimensions?** For instance, we could expect clear topographical dissociations where each sensory-motor area shows the appropriate unidimensional geometry (e.g., size in visual
areas, sound in auditory areas, etc.), while semantic hubs would present either multimodal geometries combining two or more motor-perceptual features (e.g. size and sound) or higher order conceptual dimensions (e.g., taxonomic categories).

- **When can we observe representational geometries dominated by motor-perceptual dimensions?** One hypothesis is that this would be possible each and every time symbols imbedd with those dimensions are read/heard. Conversely, it could be the case that those geometries are recovered only if the task performed explicitly requires access to those aspects of the concepts.

### Conclusions

The importance of words in our daily life cannot be overstated. Something as apparently trivial as reacting to a colleague handing us a note with five black marks, \( p h o n e \), necessitates the interplay of multiple cognitive and neural mechanisms, crucially revolving around our ability to access the semantic representation we associate with that symbol. The neuropsychological and neuroimaging evidence here reviewed supports the view of semantic representation of concrete words as points in a multidimensional space where each dimension likely corresponds to different features (and/or combination of features), defining its meaning. We propose an operational distinction contrasting two different kinds of dimensions: motor-perceptual (i.e., the aspects of the objects/actions that can be apprehended through the senses) and conceptual ones (i.e., the product of the integration of multiple uncorrelated motor-perceptual features as well as verbal declarative knowledge). Brain activity supporting different semantic features seems to cover a large portion of the cortex. Motor-perceptual features are likely encoded in a distributed set of brain regions including the sensory-motor cortices activated during real world interaction with the objects referred to by the words. On the other hand, integration of non-correlated motor-perceptual features (including their nonlinear summation which allows for the emergence of higher order properties and the creation of non-trivial categories), as well as the encoding of other purely declarative conceptual features likely takes place in higher order cortices, also previously referred to as “convergence zones”. Among the candidate locations for these areas, clinical and neuroimaging evidence suggest the ATL as the ideal cortical area where information from different sensory-motor cortices converge and be merged, and interface with language (see Fig. 2). The description of the precise computational flow that allows the recovery of the meaning of words is still extremely poorly understood. In particular, we lack a description of which aspects of the semantic space are accessed at any given time, and whether and how this may be modulated by the task at hand. Finally, while studies with greater spatio-temporal resolution (e.g., direct electrophysiological recording studies) will deepen our understanding of the neural code, neuroimaging data can approximate the question of whether and how this may be modulated by the task at hand. Finally, device from category-specific semantic deficits. Trends Cogn. Sci. 7, 354–361.

### Acknowledgments

We gratefully thank Ariane Welch for helpful comments on the manuscript.

### References


Corcilius, C.J., see also (2) and (3), flexible and dynamic, to accommodate changes necessitated by the specific task, context, and prior experience. Moreover, given the multidimensional nature of the questions left unanswered and of the hypotheses put forward, the empirical efforts should integrate multiple methods (i.e., neuroimaging as well as cognitive neuropsychology), and a combination of perspectives (i.e., data-driven approaches with naturalistic stimuli as well as hypotheses-driven approaches with controlled stimuli). We believe that great steps forward in the investigation of the neuro-cognitive correlates of symbolic representations will be taken in the next few years, by tackling the questions we have here described, and by comparing predictions from the opposing views we have here reviewed.

Cree, G.S., McRae, K., 2003. Analyzing the factors underlying the structure and composition of the meaning of chipmunk, cherry, chisel, and cello (and many other such concrete nouns). J. Exp. Psychol.: General. 132, 163–201.


Ellis, A.W., Young, A.W., Critchley, E.M.R., 2003. Analyzing the factors underlying the structure and composition of the meaning of chipmunk, cherry, chisel, and cello (and many other such concrete nouns). J. Exp. Psychol.: General. 132, 32, 906


McRae, K., Cree, G.S., Seidenberg, M.S., McNorgan, C., 2005. Semantic feature produc-
featural model for semantic decisions. Psychol. Rev. 81, 214.


