



This article is part of the topic “Abstract Concepts: Structure, Processing, and Modeling,” Marianna Bolognesi and Gerard Steen (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview).

## Neurobiological Mechanisms for Semantic Feature Extraction and Conceptual Flexibility

Friedemann Pulvermüller<sup>a,b,c</sup>

<sup>a</sup>*Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin*

<sup>b</sup>*Berlin School of Mind and Brain, Humboldt Universität zu Berlin*

<sup>c</sup>*Einstein Center for Neurosciences Berlin*

Received 9 August 2017; received in revised form 2 May 2018; accepted 9 May 2018

---

### Abstract

Signs and symbols relate to concepts and can be used to speak about objects, actions, and their features. Theories of semantic grounding address the question how the latter two, concepts and real-world entities, come into play and interlink in symbol learning. Here, a neurobiological model is used to spell out concrete mechanisms of symbol grounding, which implicate the “association” of information about sign and referents and, at the same time, the extraction of semantic features and the formation of abstract representations best described as conjoined and disjoined feature sets that may or may not have a real-life equivalent. The mechanistic semantic circuits carrying these feature sets are not static conceptual entries, but exhibit rich activation dynamics related to memory, prediction, and contextual modulation. Four key issues in specifying these activation dynamics will be highlighted: (a) the inner structure of semantic circuits, (b) mechanisms of semantic priming, (c) task specificity in semantic activation, and (d) context-dependent semantic circuit activation in the processing of referential, existential, and universal statements. These linguistic-semantic examples show that specific mechanisms are required to account for context-dependent semantic function or conceptual “flexibility.” Static context-independent concepts as such are insufficient to account for these different semantic functions. Whereas abstract amodal models of concepts did so far not spell out concrete mechanisms for context-dependent semantic function, neuronal assembly mechanisms offer a workable perspective.

---

Correspondence should be sent to Friedemann Pulvermüller, Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin 19145 Berlin, Germany. E-mail: [friedemann.pulvermuller@fu-berlin.de](mailto:friedemann.pulvermuller@fu-berlin.de)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Keywords:* Concept; Context dependence; Embodiment; Grounding; Meaning; Neurobiological mechanism; Quantification; Reference; Existential / universal statement

---

## 1. Introduction

The question about concepts and the brain is an exciting one, although some models addressing this domain have left scholars dissatisfied, especially in answering this question by pointing to rectangles carrying the words “concepts” and “meaning” (see Davis, 1989) or to a similarly labeled brain area. Such “answers” do not address questions about how we acquire concepts, how they are influenced by our genetic endowment and experiences, and how what we learn interacts with our instincts and desires when first grasping the ideas of freedom, justice, or democracy, or when we finally realize that whatever in reality had been labeled by them does not stand up to expectation. I will step into this tradition by not answering the big question here. However, I will make a serious attempt to specify some basic and, as I believe, essential mechanisms necessary for conceptual processing. As all conceptual processes are based on and require brain processes, a significant part of my contribution will be about neurobiology.

This paper will address semantics, the meaning of words, signs, and symbols. While concepts can exist without signs, scientists rarely speak about specific concepts that come without a handy label. Some scholars also believe that the formation of thoughts and concepts is influenced significantly by the available labels and language structures (see, e.g., Boas, 1940; Clark, 1996; Slobin, 1996; Von Humboldt, 1979; Whorf, 1956), and experimental evidence suggests that aspects of this position may in fact be true (Lupyan, Rakison, & McClelland, 2007; Majid, Bowerman, Kita, Haun, & Levinson, 2004; Miller, Schmidt, Blankenburg, & Pulvermüller, 2017).<sup>1</sup> Still, I will not promote a strong opinion on these much debated issues here. Instead, this article will focus on concepts with corresponding word meanings, because the use of word meanings as examples of concepts represents the typically discussed standard. Evidently, the label makes theorizing easier as it adds, so to speak, a handle to the concept by which the latter can be grabbed.

For distinguishing different aspects of semantics and concepts, the tripartite model of word meaning (Fig. 1, left panel, Löbner, 2014; Ogden & Richards, 1923) is useful, although its various implicit assumptions have received ample criticism (see, e.g., Alston, 1964). The model distinguishes the form of a sign or symbol<sup>2</sup> from two aspects of its meaning, labeled the “concept” and the “referents.” If a word is used to speak about a specific object in the world, this object is called the word’s *referent*. As signs can also be used to speak about aspects of objects, for example, color, also these features are sometimes called referential (although in a slightly different sense). One may even allow speaking about referent actions, for example, when the word “grasp” is used to speak about somebody’s grasping of a cup, and about referent interaction types, for example, when the word “democracy” (or “democratic”) is used to speak about the way groups of people decide between alternative options. The second aspect of meaning, the *concept*, designates, in this model’s context, the knowledge about word meaning immanent to the

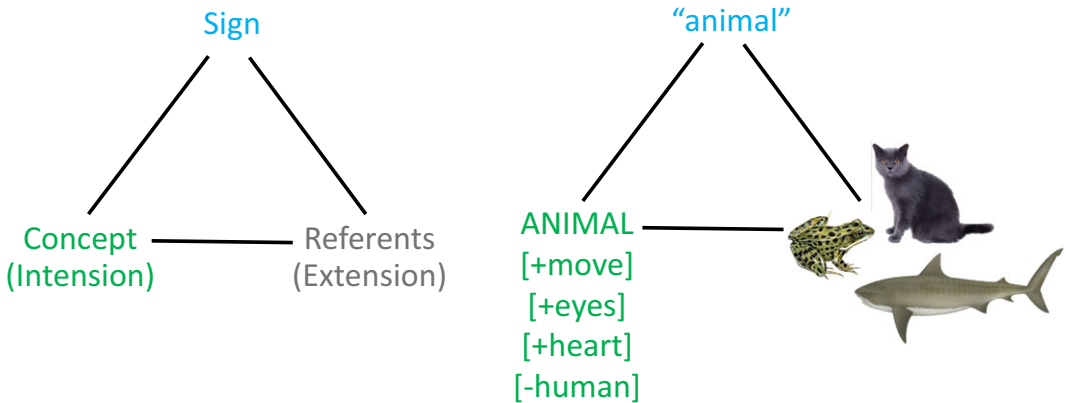


Fig. 1. Meaning, concept, referent. The panel on the left shows the relationship between a sign and its meaning as postulated by the semiotic triangle model (Ogden & Richards, 1923). The form of the sign (in blue) is connected with its related concept (or intension, in green) and the real-world entities the sign is used to speak about, its referents (or extension, in gray). The panel on the right illustrates the three aspects using one specific word form, some of the semantic features of the related concept and concrete example referents.

language user. Importantly, the “concept” understood in this way includes the information about how to use a word or symbol to speak about aspects of the world. Therefore, it is one of the main jobs of the concept—not its only job though—to provide the mental glue between sign and referent (see also Frege, 1892; Löbner, 2014; Quine, 1960). Whereas the concept can be seen as the most essential part of meaning, important criteria for correct application of the sign come from referent contexts. These two aspects of meaning, concept and referents, are sometimes labeled the “intension” and “extension” of the sign.

## 2. The need for semantic grounding

Some semantic theories focus entirely on intension. Following insights from philosophy (Carnap & Bar-Hillel, 1952; Frege, 1918), cognitive and language scientists described meaning in terms of semantic features (Katz & Fodor, 1963; Löbner, 2014), that is, predicates that can be attributed to all elements that fall into the category labeled by a word or symbol. The ANIMAL concept (words in capital letters indicate specific concepts) would thus include features such as “moves,” “has eyes,” “has a heart,” but not the feature “is human”; semantic features of a concept can be indicated in square brackets together with a feature value indicating whether the features applies to a concept [+] or not [–]. The concept ANIMAL would thus include the features plus feature values [+move, +eyes, +heart, –human] (s. right panel, Fig. 1). Some semantic feature theories see the semantic features of a concept as necessary and sufficient for the concept. A different approach is to describe semantics in terms of the relationships between concepts, for example using diagrams with lines of different length symbolizing the strength of the semantic relationship between any two concepts, with shorter lines indexing closer

relations (Collins & Loftus, 1975). The degree of semantic relatedness is obtained from semantic ratings of words (is “beard” semantically related to color?), associative listing tasks (“what comes to mind if you hear “beard”), or from text corpora (from which the co-occurrence probability of the critical word with other words is mapped). The latter strategy rests on the assumption that words with similar meanings appear in similar contexts, that is, together with the same context words, whereas semantically distant words have different contexts. Indeed, the degree of similarity of context vectors of words provides an estimate of their semantic similarity (Burgess & Lund, 1997; Landauer, Foltz, & Laham, 1998; Landauer, McNamara, Dennis, & Kintsch, 2013; Schütze, 1992).

That semantic relationships between symbols can be described and quantified in terms of semantic features, semantic vectors, or other metrics, represents a major achievement. However, this does not answer all semantic questions. In particular, these representations of semantic knowledge leave open the question how symbolic entities (semantic features, dimensions, distances, etc.) relate to the world, the actions humans perform in the world, and their social interactions. Recall that it is one of the main jobs of any semantic theory to explain how signs relate to real-world entities. This job is left undone by these proposals. They specify semantic knowledge in abstract space, but omit the equally crucial semantic knowledge about the link between symbol and what the symbol is used to communicate about (Fig. 1). Such approaches can be called “amodal symbolic” to indicate that they do not include the modality-specific knowledge required to interlink the concept of an EYE with the visual and functional features of typical eyes, and the concept of REDNESS with the perception of the color.

Crucially, one cannot learn the referential symbolic link, and the meaning of symbols in a broader sense, exclusively by relating symbols to other symbols, concepts to other concepts, or concepts to semantic features, because, in these cases, the relationship between the symbols and the meanings these symbols symbolize would remain unclear. It is essential to use symbols in the context of objects and actions in order to set up referential semantic links; and without any referential links, knowledge of meaning is impossible. This is, in a nutshell, the *symbolic grounding problem* (Harnad, 1990; Searle, 1980). A person who does not know Chinese cannot learn this language through Chinese symbol manipulation only. If this person is encapsulated in the “Chinese Room” just receiving Chinese symbols through a mail slot and returning other such symbols to the outside world according to a rule book that tells him which symbol(s) to return upon which stimulus signs, then this person cannot be said to understand the meaning of the symbol forms he is using and processing. He may just look up and follow the combinatorial rules in the book and deliver symbols accordingly, without processing the meaning and content linked to the symbols. Or, putting it differently: If you do not know Chinese, you cannot learn it from looking up Chinese symbols in a Chinese lexicon (if it is a picture-less lexicon). What is required is a link between the word and what it is used to speak about, between “animal,” the ANIMAL concept, and the real-world entity.

This link can only be learned using real-life criteria shared by speakers of the language (cf. Wittgenstein, 1953). To make meaning learning possible, examples of animals can be shown. And, only after a range of words are already “grounded in” knowledge about the

world, further semantic learning is possible by purely verbal explanation or symbol co-occurrence (Harnad, 1990, 2012; Stramandinoli, Marocco, & Cangelosi, 2017). The verbal explanation “An animal has heart and eyes and moves around” is suitable for semantic learning of the word “animal” only after the terms “heart,” “eye,” and “moves” have been grounded. Likewise, semantic feature theories require grounding of their basic semantic features.

Note that, in normal language acquisition, a majority of vocabulary items are learned from language contexts, without referent objects being present. Still, without a *grounding kernel*, that is, a basic set of symbols learned in the context of semantically relevant objects and actions, the semantic machinery cannot break out of its symbolic circularity (Harnad, 2012). Note also that the grounding kernel needs to be established before learning from context becomes possible, which is plausible, given that it is especially early in life that symbols are acquired in action and object contexts. Importantly, semantic grounding is not restricted to the visual modality (e.g., pointing to animals or showing animal pictures), but it may involve other sensory channels (e.g., hearing and smelling animal sounds and odors) and, crucially, action and interaction (stroking and feeding an animal). Some of this information, which is specific in each modality and still multi-modal in its entirety, is part of referential semantic knowledge.

### 3. Grounding amodal symbols

How can an amodal symbolic semantic model accommodate semantic grounding and symbol reference? As mentioned, standard amodal symbolic models, including abstract semantic feature and vector theories, lack coverage of referential semantic knowledge. Recent “amodal” approaches do not explicitly state the precise format of their postulated symbolic representations (Bedny & Caramazza, 2011; Leshinskaya & Caramazza, 2016; Mahon & Caramazza, 2008); that is, apart from presuming that their representations are semantic or conceptual in nature, it remains unspecified whether semantic relationships may rely on semantic features, vectors or hierarchies. In spite of this lack of specificity, these recent “amodal” approaches, just like the classical ones, are subject to the grounding problem, as they do not explicitly cover the links between symbols and the entities they are used to communicate. Instead, “interface systems” are proposed to make contact with the physical and social world via sensory and motor systems. Note that the real-world interface could indeed open a perspective on incorporating referential semantics and grounding, but this cognitivist type of approach misses this possibility by still restricting all “truly” semantic and conceptual processing to the amodal symbolic system (for detailed discussion, see Pulvermüller, 2013b). The interfaces are, at best, allowed an optional, enriching role in the comprehension process (Mahon & Caramazza, 2008), which therefore cannot be genuinely semantic.<sup>3</sup> How the referential links crucial for semantic grounding are being established remains unclear in this type of approach.

In one view, there is no real grounding problem for amodal symbolic models:

Amodal concepts have to be grounded in the sensory/motor systems. How is that done? Schematically, at least, the answer is not so complex: A line is drawn from the concept to the corresponding sensory/motor information. Grounding solved. (Mahon, 2015, p. 424)

However, a line drawing exercise cannot be the solution if the concept is in the mind and the referent information in the world. A main problem lies in the identification and appropriate selection of the concept: How could one make sure to select the right concept and link it to the perceived instance of an eye? For coming up with a workable mechanism, the concept selection question needs to be answered. In the absence of clear *criteria* for concept selection, amodal symbolic theories attempt at specifying semantic understanding without explaining grounding. Therefore, they miss a major component of semantics. In a sense, their abstract representations remain in a Chinese room of the human mind and brain.

A putative solution of the grounding problem goes back to British empiricism and runs somewhat like this: When people learn the meaning of words, they become able to pick out appropriate referents for given words and find the appropriate verbal labels for objects. Their ability to point to appropriate objects upon being given a word can therefore be a test for their semantic competence (Alston, 1964; Locke, 1909/1847). However, this does not directly solve the grounding of concepts; it restricts itself to the symbolic handles of concepts, words, and symbol forms, and their real-world correlates, the referents. Obvious limitations of this approach come from the facts that many words do not have referents that can easily be pointed to (“democracy,” “causation”) and some even lack any correlate in the world whatsoever (the concept of NEGATION). However, adding word-to-object-correspondences to semantic knowledge may solve the grounding problem for a base vocabulary of symbols and symbolic features (cf. Harnad’s grounding kernel). Further concepts can then be derived from the available grounded items, for example, by building a semantic context vector or by verbal explanation and context (for a related model, see Stramandinoli et al., 2017). However, verbal explanations of highly abstract terms typically use other rather abstract items so that one wonders how abstract semantic features can be transferred from concrete to abstract semantic domains. A major fundamental question remains how words-with-referents bring about conceptual mechanisms that detach from actual perceptions (for discussion, see Pulvermüller, 2018a).

#### 4. Emerging concepts and meanings

In his seminal article, Harnad took up the empiricist tradition suggesting that conceptual representations may emerge from the learning of the relationships between symbol and its referents (Harnad, 1990). He proposes a “hybrid” model, in which information about words and about their referents is associated by way of a neural network; error backpropagation learning provides the mechanism for establishing referential semantic links. Instead of an explicit conceptual representation, the model is assumed to develop

nonsymbolic representations that can pick out the objects to which symbols refer. Connections via a “hidden” middle layer of the hybrid model carry the information about both referential symbolic links and symbolic content—although nothing but referential information enters the network. Detailed investigation of the functional properties of the “hidden” units would be required to link up neural function with cognitive theory (see, e.g., Tomasello, Garagnani, Wennekers, & Pulvermüller, 2017; Wood, 1978).

Although there is broad agreement on the necessity of semantic grounding, there is little knowledge about what grounding actually entails at the conceptual end. To find out, it is necessary to spell out the relevant processes in more detail. The ultimate mechanisms of conceptual thought and semantic processing require and emerge from brain mechanisms instantiated in structural and functional features of neurons and their connections. For this reason, the laws that govern the latter also determine, or at least co-determine, concepts and meanings. A look at the brain and its structural and functional properties may therefore benefit theories about concepts and meaning.

One framework spells out semantic grounding mechanisms at the neurobiological level in terms of distributed semantic neuronal circuits binding together information about words and their meaning (Pulvermüller, 1999, 2013a). Detailed implementation of this model in neural architectures replicating important features of cortical structure and connectivity allows for careful conclusions on the brain localization and, putatively, functionality of the emerging circuits (Garagnani & Pulvermüller, 2016; Tomasello et al., 2017). When such brain-constrained network models were used to imitate learning of semantic relationships between word forms and their referent objects and actions, model neurons were joined together into strongly connected neuronal circuits interlinking word-form related and semantic information. In consequence, these emerging semantic circuits were widely scattered across different cortical regions, including both multimodal and modality preferential areas, for example, visual or motor cortices.

This neurobiological grounding framework predicts a differential involvement of modality preferential areas in the processing of signs of different semantic types. Processing of words used to speak about actions involves neurons situated in areas of the motor system, whereas words with referents normally perceived through sensory modalities involve neurons in the respective sensory systems of the brain, although words used to speak about objects with action affordances produce remarkable activity in cortical sensorimotor cortices too. Apart from sensorimotor activations, the model also predicts and explains multimodal system activations. The model’s brain-constrained simulations provide neural underpinnings for proposals in the semantic grounding framework, which assume that conceptual representations develop during symbol learning in the context of actions and objects (Arbib et al., 2014; Barsalou, 1999; Fischer & Zwaan, 2008; Glenberg & Gallese, 2012; Glenberg & Kaschak, 2002; Kemmerer, 2015a,b; Kiefer & Pulvermüller, 2012; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Pulvermüller, 1999, 2005; Smith & Semin, 2007). There is ample experimental evidence that involvement of modality-preferential areas is present during semantic processing and reflects semantic features of the symbols under processing (Grisoni, Dreyer, & Pulvermüller, 2016; Grisoni, Miller, & Pulvermüller, 2017; Kemmerer, 2015a,b; Kiefer & Pulvermüller, 2012; Klepp,

Niccolai, Buccino, Schnitzler, & Biermann-Ruben, 2015; Klepp et al., 2017; Mollo, Pulvermüller, & Hauk, 2016; Moseley, Pulvermüller, & Shtyrov, 2013; Trumpp, Kliese, Hoenig, Haarmeier, & Kiefer, 2013). Would this type of grounded neurobiological account just imply that associative links are drawn between symbol form and all the memory traces of its previously encountered referents?

## 5. Beyond word-object association: Neurobiological mechanisms of semantic grounding

### 5.1. Interlinking symbols and referent exemplars: Perspectives and limitations

When learning the link between word and object while a teacher speaks out the word and the learner attends to the object, neuronal activity elicited in the visual system will co-occur with activity in those brain regions engaged in speech processing. In the visual system, the neurons of potential interest are at several levels, ranging from primary to higher order visual areas. When specific objects are perceived, the relevant perceptual information is carried by a distributed pattern of neurons spread out across the ventral inferior temporal stream of visual processing (Haxby et al., 2001; Norman, Polyn, Detre, & Haxby, 2006). The individual neurons included in this population respond to perceptual features at different levels of complexity.

A simple learning model of the conceptual-semantic links of the word “eye” may therefore join together the word form representation with perceptual-semantic neurons in the ventral visual stream. A neuronal population processing the word form co-activates with a population of cells activated by the visual stimulus. The former is likely located in perisylvian language cortex (blue dots, left panel of Fig. 2) and the latter in the ventral stream of visual object information processing (green dots, middle panel in Fig. 2). This is indeed very close to drawing a line between word and concept, although the line is realistically a massive set of neuronal connections and the connection is between symbol and referent representations. There is still no correlate of the concept in this mechanism; just associative learning. Firing together yields wiring together (see Hebb, 1949).

What will be associated together? According to exemplar based models, all object or action representations a given word is used to speak about will be associated with its symbol form. Given some referent instances are most typical referents of the symbol and therefore co-occur more frequently with it than others, these will be more strongly interlinked with the word’s representation, thus allowing for modeling more or less prototypical referential-semantic relationships (Rosch & Mervis, 1975). This follows from the Hebbian learning rule of gradually strengthening synapses with increasing numbers of co-occurring activations (for discussion, see Pulvermüller, 2018b).

However, this whole-representation centered position does not capture important properties of semantic mechanisms. What is co-activated at the neuronal level are not “exemplars,” but sets of different nerve cells, each of which has its own specific connection pattern and resultant response characteristics. As a lot is known about the response



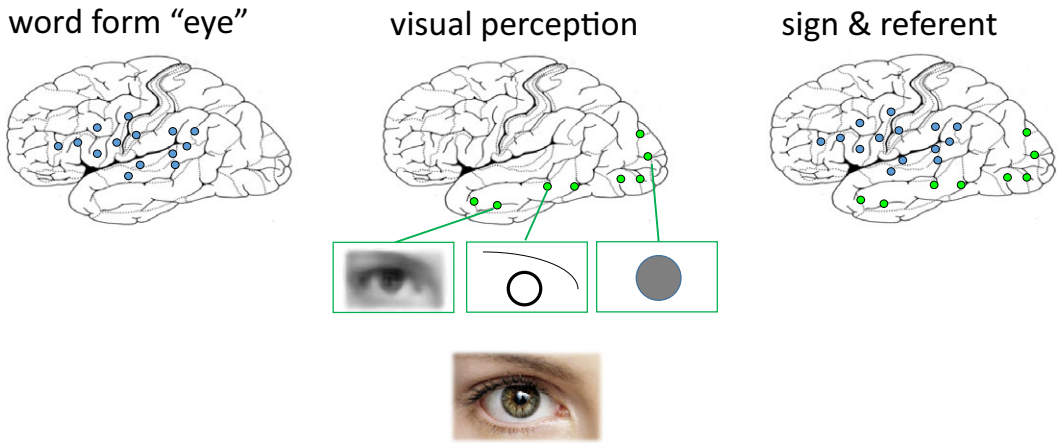


Fig. 2. Illustration of semantic grounding mechanisms for one specific word form in the perception of one single object. Left panel: The processing of the word form “eye” is neurobiologically underpinned by the activation of a neuronal circuit distributed across the perisylvian language areas of the cortex. Middle panel: The perception of an eye—a possible referent of the word—activates a neuronal circuit in the ventral visual stream. Right panel: Correlated activation of these circuits, for example, when the word is used in context of the referent, leads to synaptic strengthening between neurons of the word and object-related circuits and thus emergence of a strongly connected higher order *semantic circuit* (or *lexicosemantic circuit*). Bottom panels: At different levels of the visual processing hierarchy, neurons respond to visual features of different complexity. In early visual areas, neurons respond to elementary features (e.g., round shape on a background), and, further up in the visual processing stream, feature conjunctions (e.g., circular shape plus curved line) and even quite complex visual patterns can be represented (e.g., a blurred general eye shape).

properties of visual neurons and distributed neural activation patterns in the ventral stream (Hubel, 1995; Kriegeskorte & Kievit, 2013; Norman et al., 2006; Perrett, Mistlin, & Chitty, 1987), more specific statements and educated guesses about the response characteristics of the neurons contributing to semantic learning are well motivated.

When an eye stimulus appears, some of the activated neurons respond to elementary visual features, such as the center-surround feature of the colored pupil on its white background of the sclera. Such neurons are particularly common in primary visual cortex and other early visual areas at the bottom of the processing hierarchy. Feature conjunctions and more complex stimulus constellations are coded at higher levels of the visual processing stream. For example, there may be neurons responding to the complex arrangement of the round iris on the background of the sclera together with a finely curved line above or below. A variety of other arrangements of basic shapes, lines, circles, and other features may be coded similarly. There will also be cells that respond to the specific color, the particularly bushy type of eyebrow, or the watery look of a particular eye. Possibly, there are even neurons of a “grandmother” or “cardinal cell” type specific to a particular object or object type as a whole (Barlow, 1972; Fuster, 1995; Kanwisher, 2000; Perrett et al., 1987); those seem more likely in anterior parts of temporal cortex. All these neurons together would code for the particular eye at which the learner may look when hearing the word in a particular situation. An exemplar of the object, neuronally indexed

by an assembly of visual neurons spread out between primary visual cortex and “higher up” anterior temporal cortex of the ventral visual stream, would provide a possible neural basis for grounding of the word form in the visual feature information provided by the referent. And, as mentioned, with further learning, more and more perceptual representations of exemplars may be bound to the word form circuit.

So far so good, but is this still not just associative learning between a word and exemplars of referent stimuli, similar to what has once been proposed in the behaviorist tradition (Skinner, 1957; Watson, 1925)? Explaining semantic learning in terms of connecting words with all of their referents—or as “the sum total of (...) sensory and motor experiences” (Hickok, 2014, p. 130)—is not sufficient or feasible. Apart from language-theoretical caveats (Alston, 1964), there are neurobiological reasons why the exemplar association perspective is unrealistic (see Keysers, Perrett, & Gazzola, 2014; Pulvermüller, 1999).

### 5.2. Semantic feature extraction by correlation

When encountering different instances of eyes, varying, for example, in shape and color, different neuron sets will activate. However, because eyes resemble each other and share perceptual features, and because the majority of neurons in the visual system are sensitive to such features, there is an overlap in the activated feature neurons when different instances of eyes are being encountered. The *overlap neurons* will link to the word representation most strongly and will provide the word’s connection with the typical features of the EYE concept (Fig. 3).

One may request detailed justification of these claims, so let me elaborate on the underlying mechanisms: The reason for this preference of some neurons lies in the correlation structure of language and perception—and an analog argument applies to action too. Strengthening of neuronal connections is induced by neuronal co-activation when the pre-synaptic neuron activates the post-synaptic one. On the other hand, weakening of synaptic strength results from pre- and post-synaptic activations happening independently, or when the latter precedes the former, thus leading to the gradual *delinking* of neurons. Therefore, realistic neuronal learning implies association of co-occurring neuronal patterns, but delinking and in fact dissociation for uncorrelated or anti-correlated ones. Both mechanisms working together have important implications for language (Pulvermüller, 2018b).

Please consider the case of the word form “eye” being used alternately to refer to three different eyes, each characterized by distinct sets of perceptual features, which, however, share some of these features (see Fig. 3). Take eye colors as examples of not-shared features and the center-surround-plus-curved-line pattern as case of a shared one. Now, whenever the word form is used while the subject either looks at, or mentally imagines (“simulates”), one of the eyes, the shared perceptual neurons will co-activate with the word form representation. This implies strong links between the shared neurons and the word form circuit. This contrasts with the mechanisms applying to visual neurons that process not-shared or even idiosyncratic perceptual features of individual referent

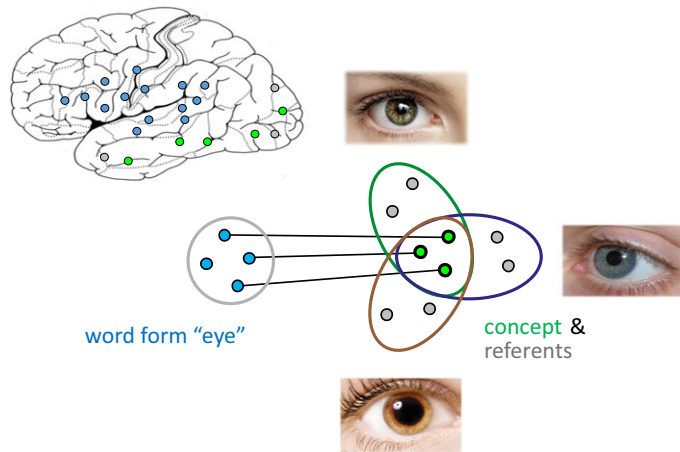


Fig. 3. Perceptual semantic feature extraction in semantic grounding. If a word is used to refer to different objects and the word and referent representations are co-active in the brain, not all of the information about the referent exemplars will be associated with the word form. Only neurons that process visual features shared by referent exemplars will show a high correlation with the word form and will thus be linked into semantic circuits. For example, neuronal units for certain form features (e.g., “circular shape with curved line”) may be included (green mini-circles), but not more variable features (e.g., the color of the skin around the eye; gray mini-circles). Perceptual-visual semantic feature processing can be assumed to be brain-based at different levels of the ventral visual stream of object processing (as indicated by green mini-circles on the brain diagram).

exemplars. For these neurons, there will be synaptic strengthening only when the respective exemplar and the word form circuits co-activate, but synaptic weakening will be in effect in all the other cases (in the other two-thirds of the cases of the current basic example, but much more frequently in more realistic scenarios). This follows from the “anti-Hebbian” part of the synaptic learning rule, which implies that neurons that fire independently from each other weaken their links and functionally dissociate. Because low correlation of neuronal firing leads to delinking, a neurobiological perspective implies a main function of perceptual and action-related feature overlap in grounding symbolic form representations in semantic information about perceptions and actions.<sup>4</sup> The strong links of the perceptual feature overlap ground the word form in perceptual knowledge about typical features of referents, and, therefore, this overlap serves as a primary structural basis for semantic knowledge.

Note again the relevance of the neurobiological argument in the context of the cognitivist versus behaviorist tradition: The reason why the “semantic overlap” develops lies in the correlation structure of the visual features across EYE exemplars and the word form, and, hence, the correlation structure of the firing of the neurons involved. Therefore, rather than storing each pair of word-object exemplars, the brain tends to bind together those neurons that frequently fire together and exhibit a high degree of correlation in their firing. Consequently, the word form is not just associated with all referents encountered. Instead, the typical features shared by many referents of the words, or instances of the

concept, will be emphasized and bound to the word form representation. In contrast, the more idiosyncratic, not-typical and rarely occurring features will not be bound into the semantic circuit if neurobiological learning principles apply. In this way, neurobiological principles may explain how some of the *semantic features* postulated by cognitive theories of meaning can result from the interplay between perceptions, actions, and language use.

In essence, the proposal is as follows: Due to properties of the neurobiological machinery—the mapping of correlations and the responsiveness of neurons in the visual system to features of shape, color, motion, and complex combinations thereof—some of the perceptual features of referent objects are extracted and bound to word form/symbolic knowledge (Fig. 3). Crucially, the neurobiological mechanism delivers a representation, which filters the frequently occurring and prototypical features from the available information about referents. These features are semantically relevant. In this way, the grounding process yields structured semantic representations. The perceptual semantic features correspond to some (but not all) features of semantic feature theories, for example, [+round] or [+curved], although relevant complex perceptual (or action-related) semantic features lack correlates in these theories (e.g., round-shape-plus-curved-line).

The emergence of semantic features was exemplified in the visual domain, but the same type of mechanism applies to auditory, somatosensory, olfactory, or gustatory information processing, and to action-related and motor processes as well. And nothing prevents semantic neurons from linking up across modalities. The auditory feature neuron indexing high pitch, the visual neuron for high towers and the motor neuron contributing to high-effort action may well each link up with the word form “high” so that they would be stimulated each time the word form activates. The cross-modal link in conceptual information processing can be by way of neurons in multimodal areas and by direct connections between perceptual and motor system neurons.

## 6. The structure of semantic circuits

Upon close examination, sets with necessary and sufficient semantic features of all possible referents do not exist for most concepts and word meanings (see, e.g., Jaszcolt, 2016; Lakoff, 1987; Lebois et al., 2015; Löbner, 2014; Rosch & Mervis, 1975; Wittgenstein, 1953). This calls into question strong forms of semantic feature theories and, likewise, the role of a strict semantic overlap of all instances of a concept as unique conceptual kernel. Therefore, the model proposed in Fig. 3 can only serve as a basic suggestion, which needs elaboration and modification.

Some features are gradually more characteristic of the members of the semantic category, while others are less typical but still far from idiosyncratic. The neurobiological model implies that the degree of correlation between symbol form and referent activations gradually determines the strength of semantic links. Factors that determine this symbolic link between a word form circuit and a semantic feature neuron include (a) the degree to which the feature is shared by different referents, (b) the frequency with which

the symbol is used to speak about referents exhibiting the feature, and (c) the degree to which the feature correlates with other semantic features of the same symbol. This does not imply nor require a core set of semantic features shared by *all* instances of a concept. A “relatively more typical” feature neuron can therefore become part of the *semantic kernel* and, in the case of words without shared characteristic features of their instances, a discontinuous set of semantic features shared only by different subsets of instances is feasible. The latter property of “family resemblance” may be of particular relevance for the mechanisms underlying abstract concepts (for discussion, see Pulvermüller, 2013a). As prototypical features tend to be shared by more referents than less prototypical ones, it makes sense to distinguish the semantic kernel, where a relatively large number of the prototypical instances of the concept overlap, from a *semantic halo* containing semantic feature neurons shared by only a few category members or by members that are not so common (Fig. 4); this distinction only approximates what can be assumed to be a gradual contribution of underlying neural elements to semantics. The links of the kernel to the word form circuit are strongest, those of the halo are weaker, but still not without semantic function (Fig. 4). Entirely non-prototypical idiosyncratic and rare features of referent representations are not part of the semantic circuit. In this view, the semantic-conceptual circuit with its halo and kernel is embedded in the larger set of referent feature representations, some of which are semantically irrelevant. The neuronal circuits of new referent representations overlap with the semantic kernel and halo, dependent on the typicality of the new referent.

## 7. Dynamics of activation and control

The described neuronal ensembles for conceptual and semantic processing have complex internal structure in which more central and more peripheral semantic features are gradually linked, yielding a circuit with kernel and halo along with a periphery to which activity may spread. How would such a neuronal network activate and how would it behave after activation? In what way would activation be influenced and which dynamics might emerge in different contexts?

A zero assumption may seem to be that concept processing is static and, whenever a word is used, “the same “abstract” concept [...] is retrieved” (Mahon & Caramazza, 2008, p. 69). Such an inflexible perspective on conceptual processing (also suggested by Fig. 3) contrasts with a main stream position in linguistic semantics and pragmatics, where, within the meaning range of a linguistic form, different *uses*, *understandings*, or *senses* are distinguished (Fillmore, 1975; Frege, 1892; Wittgenstein, 1953). A structural linguistic unit (element of the “*langue*,” De Saussure, 1916) has a meaning range, but, as a concrete utterance (of language use, “*parole*”), its meanings-in-context, sometimes called *senses*, are not always the same. The context-relevant senses of “dog” differ when speaking about toy poodles and Great Danes, in particular when discussing the option to fit them under a chair, in which case size aspects will be emphasized. Even more obviously, when a deictic word such as “this” is used together with a noun, one specific

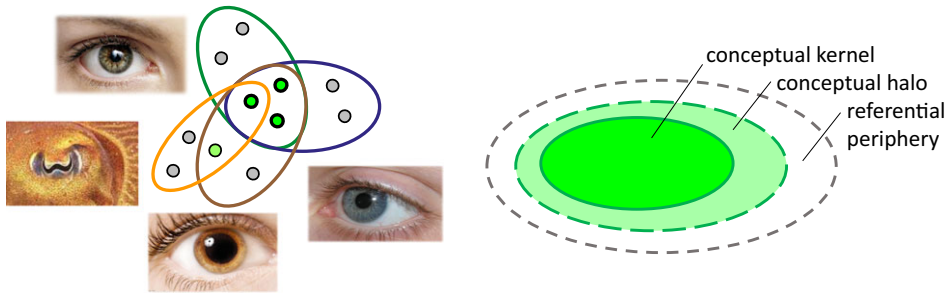


Fig. 4. Internal structure of a semantic circuit. Between the conceptual or semantic kernel, which includes the neural representations of typical semantic features shared by all or most of the category members, and the referential periphery, where idiosyncratic referent features are processed, there is an intermediate “conceptual/semantic halo” with feature neurons just applying to some or a few referent exemplars. Corresponding to the correlation structure of the neuronal activations, the connections within the kernel are strong, whereas their strengths fall off toward the periphery. The bottom diagram repeats the example of overlapping referent circuits (Fig. 3) to illustrate a case of a semantic halo neuron (indexing the eye color feature “brown” shared by the human eye and octopus eye exemplars).

object (or scene, action, interaction) is typically referred to, and other aspects of the entire possible meaning range of the expression may not be relevant. Likewise, the word “eye” may be used to speak about one specific eye, or in a general sense, with different perceptual-semantic features being relevant, or “foregrounded,” depending on context; and different relevant action-semantic aspects of the word “hammer” come into play when talking about hammering or stumbling over something (cf. Glenberg & Robertson, 2000). These *differences in sense* need to be captured by semantic models. In essence, a degree of “flexibility” in conceptual and semantic processing needs to be modeled. Recent claims and observations in the cognitive neuroscience of language fit with this well-established position in semantics (Chen, Davis, Pulvermüller, & Hauk, 2013; Hauk & Tschentscher, 2013; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; Kemmerer, 2015a,b, 2016; Kiefer & Pulvermüller, 2012; Van Dam, Brazil, Bekkering, & Rueschmeyer, 2014; Willems & Casasanto, 2011).

Still, if features of perceptions, actions, and internal states are firmly linked to symbolic forms, how would such flexibility be possible? Some researchers have suggested that neurobiological language models cannot capture context-dependent meaning and related brain activity (Raposo, Moss, Stamatakis, & Tyler, 2009; Willems, Toni, Hagoort, & Casasanto, 2010). On the other hand, a too “flexible” semantic theory may be seen as a weak theory—especially if any strong predictions are meant to be treated “flexibly.” It is therefore important to consider the dynamics of activation of the proposed neuronal circuits to determine which precise processes of “flexible” activation are predicted and explained, and which could falsify the model.

In the following sub-section 7.1, I will outline basic features of neuronal circuit dynamics in order to provide the necessary theoretical underpinnings for a model of conceptual flexibility. Sub-section 7.2 will then treat one basic case of conceptual/semantic

flexibility, the difference between the recognition of a semantically primed symbol and that of an unprimed one. After brief comments on the necessity of regulation and control in neuronal systems (7.3), sub-section 7.4 will address three cases of more sophisticated conceptual flexibility, those of symbol processing in referential, existential, and universal contexts.

### 7.1. Ignition, reverberation, and priming

Brain-constrained network simulations in conjunction with invasive and non-invasive neurophysiological studies have revealed aspects of the dynamics in time when neuronal assemblies activate and maintain activity for some time (Fuster & Bressler, 2012; Pulvermüller & Garagnani, 2014; Pulvermüller, Garagnani, & Wennekers, 2014; Verduzco-Flores, Bodner, Ermentrout, Fuster, & Zhou, 2009; Zipser, Kehoe, Littlewort, & Fuster, 1993). When a neuronal assembly is activated, its strongly connected kernel and its less strongly interlinked halo behave differently. After stimulation, activation spreads from the assembly periphery to its kernel, from where a rapid full activation process originates (latency ca. 100–250 ms). This full activation process, called “ignition,” includes both the kernel and the halo of the circuit. Regulation processes and neuronal fatigue lead to rapid termination of the ignition process and, because neuronal connections are strong in the kernel but weak in the halo, activation starts to vanish from the halo. Thereafter, the halo has lost its activity entirely, whereas the kernel still holds above-baseline reverberating activity, owing to its strong connections, which lasts for tens of seconds if not minutes (Fuster, 2015). Likewise, if the circuit is activated only slightly so that no ignition is induced, activity will primarily be maintained in the circuit kernel.

These mechanistic neuronal processes have implications at the cognitive level. Full ignition of a semantic circuit provides a mechanistic basis for the *recognition* of the concept, whereas its lasting reverberation underpins *working memory*. Slight circuit stimulation only partly activating a circuit, without leading to an ignition (“pre-activation”), provides the correlate of *priming*. Circuit stimulation can be due to activation of a different overlapping circuit, or to activity flow from an entirely different but strongly connected circuit.

### 7.2. Semantic priming as an example of context-related semantic processing

A basic example of the flexibility of neurocognitive activation is semantic priming. Compared with unprimed presentation, the same word is processed more reliably and faster if the previous context includes an item that semantically relates to it. The brain response to the primed word is typically reduced compared with the unprimed case, and this reduction is particularly obvious in specific brain indicators of semantic processing (Bentin, McCarthy & Wood, 1985). How would the neurobiological model accommodate this priming-related flexibility in the brain processes underlying semantics?

As mentioned, priming leads to an enhanced level of activity in the circuit. This can result from previous ignition of that same circuit (as in repetition priming) or activation of an overlapping or connected different circuit (as in phonological and semantic priming). Thus, primed and unprimed conditions differ in so far as, when the critical, to-be-recognized stimulus appears, its corresponding circuit is at a low level of resting activity without priming, but at a moderately enhanced level after priming. Because of this relatively enhanced activity level, a full ignition happening after priming leads to a smaller activity increase than an ignition of an unprimed, previously inactive circuit. Thus, the *activation difference* brought about by the critical stimulus is small in the priming case and large in the case of unprimed ignition, due to the difference in the baseline before critical stimulus appearance. In effect, the model provides a mechanistic explanation for *reduced activity changes* in the primed recognition process (ignition). In semantic priming, the prime and critical stimuli are meaningful symbols so that a semantic circuit is subject to priming and ignition. In this case, it is primarily the kernel and, to a lesser extent, the halo of the semantic circuit that provide the basis for priming (for illustration, see Fig. 5, panels on the left).<sup>5</sup>

The mechanisms of semantic priming between circuits and activity reverberation in the semantic kernel can accommodate a range of experimental results about “flexible” semantic or conceptual brain activations. In semantic priming experiments, Ulrich and colleagues showed priming-related modulation of neurophysiological activity in a range of areas known to contribute to semantic processing (Ulrich, Adams, & Kiefer, 2014), and Grisoni et al. found that semantic priming between meaningful sounds and action-related words significantly reduced semantic activation in similar areas. These authors also reported very specific semantic priming effects in the motor system, where the priming of specific semantic features was reflected at different locations. Words used to speak about actions typically performed with the face or leg gave rise to semantic priming effects in face- or leg-related motor cortices, respectively (see Fig. 6 middle and right panels, Grisoni et al., 2016).

In their seminal study on semantic task and context effects, Hoenig and colleagues showed that action- and visually related nouns designating artifacts and natural “things” activated the brain differently in tasks emphasizing action- and visually related information processing (Hoenig et al., 2008). They used an attribute verification task, where subjects had to indicate whether visual (e.g., elongated) or action-related features (e.g., cutting) could be attributed to typical referents of target words. The results showed that in congruent task contexts, there was activity reduction in those areas related to the type of semantic processing. For example, inferior-frontal activity close to the motor system was only significant for artifact nouns when presented in the visual task, whereas living-things nouns only activated the inferior temporal cortex when presented in the action task. Target words presented in congruent task contexts led to reduced activity. The authors interpret their results in terms of semantic priming. This is consistent with the model in Fig. 4. Semantic priming effects play an important role in explaining the “flexibility” of conceptual semantic priming and its brain correlates, and especially in cases



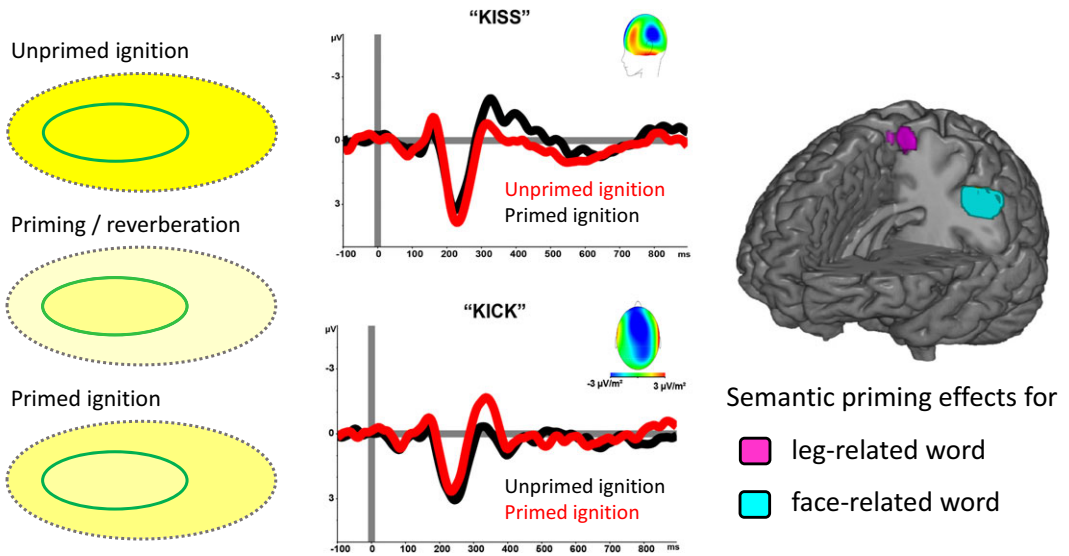


Fig. 5. Flexibility of semantic circuit activation: semantic priming. Left panels: After previous activation of a semantic circuit (after an initial ignition or semantic priming input, top panel), its kernel remains partly active due to reverberating activity supported by strong connections (middle panel). Due to this reverberating activity, a subsequent circuit ignition will be reduced compared with the first (bottom panel). Degree of yellow indicates degrees of activation. Middle panels: Brain activation to semantically primed and unprimed action-related words (“kick” and “kiss”). Note the smaller responses to the primed items: in priming context, the word-elicited brain activity reduces substantially. Right panel: The brain regions where this priming effect—the difference between unprimed and primed activations—is manifest include motor areas. The loci of the priming effects in motor areas indicate the body parts with which the words’ referent actions are typically executed (face for the word “kiss,” which shows semantic priming effects in the lateral face representation in motor cortex, as shown in light blue; leg for the word “kick,” which yields semantic priming effects in the dorsal leg representation of the motor cortex, as shown in purple). Artwork for middle and right panels adopted from Grisoni et al., 2016, reprinted with permission.

where words from a specific semantic category did not produce the behavioral and brain correlates of some of their semantic features (see Kiefer & Pulvermüller, 2012; Pulvermüller, 2013a). A recent study related physiological priming effects to predictive brain activity in action-semantic processing and found that the anticipatory cortical excitation just preceding a predictable target word helped explain the subsequent word-elicited event-related brain response (Grisoni et al., 2017). The observed reduced activations in primed contexts can be explained by the model’s enhanced baseline activity entailing reduced subsequent activation increase during ignition.

In sum, the semantic circuit model provides an explanation for the brain dynamics underlying semantic priming, which represents a key case of flexibility in neurocognitive processing, and appears to be in good agreement with related experimental results. According to this model, the brain correlates of concept processing are not static, but modulated by well-defined features of the context.

### 7.3. *The need for regulation and its implication for task-related flexibility*

It is generally agreed that a complex system such as the human cortex requires regulation and control of activity in order to be functional (Braitenberg, 1978; Braitenberg & Schüz, 1998; Milner, 1996; Palm, Knoblauch, Hauser, & Schuz, 2014). How should it otherwise be possible to ascertain that the cortical activity does not cease, and does not go out-of-bounds as, for example, in epileptic attacks? Gain control of cortical activity can be provided by a feedback loop which controls the general level of excitation/inhibition within specific areas (Bienenstock, 1994; Elbert & Rockstroh, 1987; Knoblauch & Palm, 2002; Palm et al., 2014; Wennekers, Garagnani, & Pulvermüller, 2006). Without such regulation, realistic neuronal circuit models are difficult to construct. A range of neurocomputational simulations combined background activity (noise), which prevents extinguishing of network activity, with area-specific inhibitory feedback regulation, which prevents over-activations, thus controlling the degree of competition between partly active reverberating cell assemblies and allowing only one circuit ignition at a time (Bibbig, Wennekers, & Palm, 1995; Garagnani, Wennekers, & Pulvermüller, 2008; Knoblauch & Palm, 2001; Palm & Sommer, 1995; Wennekers et al., 2006). The gain or amplification factor of the regulation function can differ across areas or wider regions of cortex.

Regulation and gain control are essential for modeling neurocognitive flexibility across different task conditions. For example, to model a perceptual or cognitive task, the gain of inhibitory feedback regulation in motor area (M1) is adjusted to a higher value, to reduce motor activity for preventing overt action. Likewise, the inhibitory gain is specifically increased in perceptual areas for modeling motor tasks, to limit perceptual input. The difference between a language task and one in which attention is directed away from language is implemented by high versus low gains (weak vs. strong mutual inhibition) within the perisylvian language cortex (Garagnani, Wennekers, & Pulvermüller, 2008). For implementing semantic tasks, feedback-inhibitory gains are set low in some or all semantic areas. More generally, any task can be modeled as a gain function across cortical areas. Rather than viewing gain control and regulation as amodal processes with unclear origin, it is conceivable that specific linguistic/cognitive operations are realized, in part, as a change in gain control functions. For example, the warnings “look!” or “listen!” may well be grounded in gain reduction mechanisms for visual or auditory areas specifically, so that, when understanding these words in context, the linguistic machinery would adjust perceptual gains. The psychological correlate of such inhibitory gain reduction would be increased attention to a specific sensory modality or cognitive domain (Garagnani et al., 2008), or to parts of the receptive field and even to specific semantic domains.

It is clear that a task characterized by low (strong) inhibitory gain and thus weak (strong) neuronal competition in a given region will emphasize (reduce) cortical activity in this region. In this perspective, a task not requiring semantic processing (e.g., a phonological one) may reduce or abolish measurable semantic processes in relevant areas (see Chen, Davis, Pulvermüller, & Hauk, 2015; Chen et al., 2013; Tomasino, Fabbro, & Brambilla, 2014; Tomasino, Werner, Weiss, & Fink, 2007) by way of increased inhibitory

gain. In semantic tasks, when the gain in areas relevant for semantics is low, semantic activity is normally evident (Kiefer & Pulvermüller, 2012). The strongest physiological evidence for semantic brain models claiming an involvement of specific areas in specific meaning-related processes comes from neutral tasks or even experiments where subjects are distracted from language processing, so that any activity in the putative meaning-processing area is likely automatic in the sense of “independent of attention being directed to semantics.” For example, a body of research using such distraction paradigms supports the involvement of motor areas in the processing of specific action semantic information (Grisoni et al., 2016; Pulvermüller, Shtyrov, & Ilmoniemi, 2005; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014; Shtyrov, Hauk, & Pulvermüller, 2004). Task-specific adjustment of gain control mechanisms can explain effects of task-related modulations of semantic brain activation, which can be understood in the sense of a tuning of attention toward or away from meaning.

#### 7.4. Context-related semantic changes: “This”—“all”—“there is”?

In linguistics and logic, different usages of concrete nouns are commonly distinguished. Among these is the referential use of a noun (“this eye . . .” with a pointing gesture), the use in an existential (“there is”) statement (“there is at least one eye for which it applies that it . . .”), and that in a universal (“all”) statement (“for all eyes, it applies that they . . .”). Covering the mechanisms underlying these differences in use or sense of a symbol is key to the neurocognitive and neurolinguistic agenda. The neurobiological model of semantic processing outlined in the previous paragraphs can model principal differences between these usages. Here is an outline how, in these three cases, the structured semantic circuit would activate and interact with its referential periphery.

When applying the sentence “Your eyes are so beautiful” to make a compliment, the word “eyes” is used together with a pronoun to speak about specific referents. Similarly, a deictic expression, such as “these” or “this,” can be used together with a pointing gesture to establish a unique reference relationship in a given context. Typically, there is visual input in such a situation, so that the semantic system is activated twice, from its linguistic and visual/perceptual ends. In such deictic referential context, the word “eye” activates a full referent representation. This includes specific neuronal elements in the semantic circuit’s kernel and halo (the relative extent varies between prototypical and more exceptional referents). Other possible competitor representations need to be inhibited, which requires local competition in the cortical regions involved in conceptual semantic processing (moderate inhibitory gain). The upper panel in Fig. 6 illustrates this situation: The word form circuit for “eye” activates its connected semantic kernel and halo; at the same time, perceptual information activates one referent representation, which leads to further activity in specific parts of semantic halo and kernel. After the word form has been recognized, the pre-activated perceptual circuit will ignite so that a link is made between the word and the perceived referent. Such flexible linkage is frequently assumed to rely on neuronal synchrony and/or oscillation.

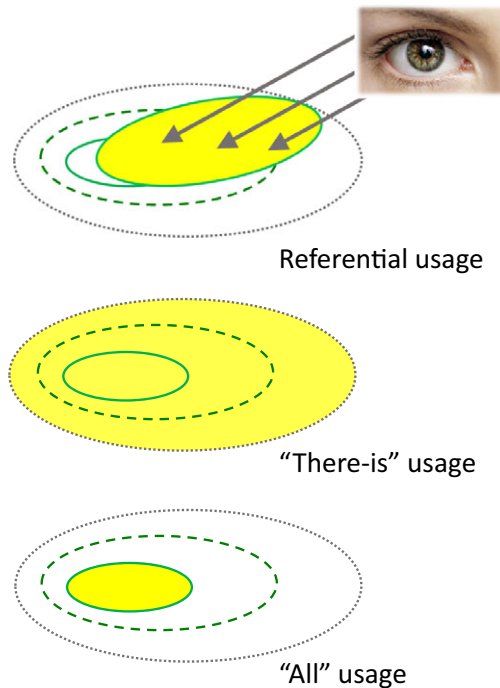


Fig. 6. Flexibility of semantic circuit activation: referential, existential, and universal usage. Top panel: Semantic circuit activation (active parts in yellow) when speaking about a perceived animal in referential context (“this animal ...”); in this case, one specific referential representation is active, which is also stimulated by perceptual input. Middle panel: Existential statements when speaking about any item falling into the animal category (“there is an animal that ...”) are modeled by broad activation of semantic kernel, halo, and referential periphery (low gain). Bottom panel: The proposed neurobiological correlate of universal usage when speaking about animals in a general sense (“all animals ...”) is the activation of the semantic kernel (and possibly part of the semantic halo; high gain).

The existential statement requires a quite different neuronal environment. If someone claims that “there is at least one eye whose natural iris color is magenta,” I may go through all my memories of “eye” referents in order to verify or falsify this claim. This may be most efficient when search happens in parallel, so that all stored EYE instances are allowed to be active simultaneously to some degree. To this end, the model would need to implement gain reduction in the semantic and perceptual processing areas (including the ventral visual stream). Because the level of inhibition is low in this case, recognition of the word form “eye” can activate (prime) its entire semantic kernel and halo, including even the referential periphery, thus priming various different perceptual representations in parallel (middle panel in Fig. 6). Only if one of the stored perceptual eye representations is substantially more strongly activated than all its competitors, an ignition will take place. This can be driven by the context word “magenta” activating, by way of its perceptual-semantic kernel and halo, one of the perceptual eye representations (i.e., the memory of a previously encountered magenta eye). In this case, activations

induced by the words “eye” and “magenta” accumulate in the respective referent’s representation and ignite it, so that the existential statement can be verified. The lack of full activation of a perceptual representation and ongoing priming of the referential periphery provides the basis for falsifying the existential statement.

Once again a quite different activation dynamic is required for modeling a general “all” statement. The claim that “All eyes are blue” can be assessed if consideration is restricted to the typical features. Since blue is common as an eye color but not among the most typical features of an eye, with many eyes being brown, black, or green, the claim will be rejected. The basis for this can be the assessment of features represented in the semantic circuit’s kernel. To restrict activity to the kernel, the inhibitory gain in semantic areas is adjusted to a very high level, so that the activation of any specific referent circuit is blocked and activity is restricted to most strongly connected kernel neurons that code most typical semantic features of the concept or word. This comes at a risk of falsely accepting statements by ignoring the non-prototypical members of the category (e.g., that all eyes are round in the center), but such errors seem quite common.<sup>6</sup>

In this perspective, adjustment of cortical regulation parameters and gain control underlies the contextual flexibility of the different understandings of a given word, typically a noun, in referential, existential, and universal contexts. Highest gain control and levels of inhibition are essential for the “all” sense, intermediate gain characterizes referential usage, and quite low levels of inhibition allow a wide range of otherwise competing perceptual representations to become active in parallel in the existential context. One may claim that a downside of this model is lack of an instance that adjusts the regulation parameters. But, in fact, an important point made above is that feedback parameters and the magnitude of the gain can systematically be linked to context, even to single words and longer constructions. Therefore, the proposal is that an “all”-assertive context is the factor that upregulates gain and the “there is” context downregulates it in wide cortical areas processing sensorimotor and semantic information, with “this” contexts leaving gain control at intermediate levels. In other words, specific words - including so-called function words like “all” and “this” - and grammatical constructions, and the communicative function they carry are being semantically grounded in the machinery for cortical activity regulation in specific areas. For allowing the language learner to tune her or his vocabulary to cortical activity regulation mechanisms, learning situations are necessary in which language use and the child’s attentional focusing are intimately related to each other (Carpenter, Nagell, & Tomasello, 1998).

## 8. Summary and outlook

The simulation of referential, existential and universal statements by focally adjusting, lowering and raising the gain of an activation regulation and control mechanism has two important implications. First, the grounded neurobiological framework sometimes labeled *action perception theory*—is flexible enough to model important semantic distinctions in the processing of concepts. It is not clear how classic semantic feature theories or amodal

concept models would capture these differences. Second, a good deal of supplementary semantic machinery over and above sensory and motor neurons is required for the semantic circuit mechanism to function properly. The latter has probably the most important consequences for the way we need to think about semantics: Distinctions such as that between the meanings of “this,” “there-is,” and “all” seem to relate to elementary brain mechanisms, as they are involved in the regulation of neuronal activation. Likewise, the logical concepts “and,” “or,” and “not” have an obvious neuronal correlate in neuronal circuits (Kleene, 1956; McCulloch & Pitts, 1943; Pulvermüller, 2008). The neuronal implementation of excitatory and inhibitory strong and weak connections, of the activation function of neurons and of the regulation mechanisms governing neuronal functionality, offers a basis for modeling important semantic and logical distinctions. It comes without question that there is still much work to do in spelling out further details of such abstract grounding mechanisms.

This article had two main goals: to show how the neurobiological mechanisms of semantic grounding give rise to semantic features and to highlight mechanisms for conceptual flexibility. The first point goes further than previous claims that *implicit* conceptual codes may arise in the semantic grounding process. The proposed model may be helpful in spelling out the putative neurobiological basis of specific semantic features. This is certainly easier for some features ([red], [elongated], [upward], [move], etc.) than for others ([cause], [make], [free], [beautiful], etc.), but given first steps have been taken, a broader coverage of a rich set of semantic features and their specification in terms of putative brain mechanisms appears as a realistic goal for the future. As neurobiological models including at least a starter set of semantic features are available, experimental work can address specific model predictions. On the other hand, some abstract terms may seem outside the reach of semantic grounding in action and perception (democracy, causation), although future research may show otherwise. In the previous paragraphs, I argued that semantic grounding can be achieved by extracting perceptual and action-related semantic features in a neurobiological machinery based on linguistic input and referent information. Given such grounding of an elementary vocabulary is achieved, contextual learning can build upon it. In contrast, semantic grounding of symbols and constructions indexing referential, existential and universal statements was proposed to be based on attention mechanisms.

Regarding the second goal, different neurobiological mechanisms for conceptual-semantic “flexibility” were highlighted. These may be key to understanding context effects and the different “senses” with which the same symbol can be used in different physical and social environments, interaction types and communicative settings, which have constantly been highlighted by many scholars across decades (see, e.g., Barsalou, 1993; Hoenig et al., 2008; Jaszczolt, 2016; Rosch & Mervis, 1975; Wittgenstein, 1953). For modeling context-dependent semantics, three different mechanisms have been discussed: (a) semantic priming, implemented as pre-activation of semantic circuits due to activity reverberation in the most strongly connected circuit parts and due to input from overlapping or strongly connected circuits, which influences semantic processing and cortical activation dynamics upon presentation of a meaningful target symbol, (b)

task-related modulation of local cortical activity by gain control, which influences the excitability of those parts of distributed neuronal circuits that reach into the areas subject to excitability modulation, and (c) language-related modulation of gain control in areas relevant for semantic processing driven by the use of specific linguistic forms, in particular deictic and logical particles. The semantic circuit model overcomes the context inflexibility of many amodal symbolic theories (for discussion, see Kiefer & Pulvermüller, 2012) and accommodates the frequently emphasized need for concrete mechanisms for conceptual flexibility (Barsalou, 2016; Hauk & Tschentscher, 2013), which may spark future exciting experiments.

## **Acknowledgments**

For comments and discussion of previous version of this text and the related talk, I would like to warmly thank Art Glenberg, Vivian Dittmer, Felix Dreyer, Martin Fischer, Ken McRae, Michael Pauen, Diane Pecher, Gün Semin and Gabriella Vigliocco, Liliya Starodubtseva. This work was supported by the German Research Foundation (DFG, Pu 97/15-1 and -2, Pu 97/22-1, Pu 97/23-1) and by the Freie Universität Berlin.

## **Notes**

1. Furthermore, verbal labels may play a special mechanistic role in forming abstract concepts (see Pulvermüller, 2018a).
2. To distinguish the form from the meaning of signs, the terms “word form” or “symbol form” will be used. Word forms are spoken phoneme or written grapheme sequences. Symbol forms include word forms along with facial and manual gestures and written and drawn symbolic forms.
3. In a previous publication, I pointed out some inconsistencies within and between these proposals (Pulvermüller, 2013b).
4. This article focuses on visual semantic features, but it is assumed tacitly that the very same mechanisms apply for semantic information related to other perceptual modalities and to action-related meaning.
5. As a further factor in priming, the reverberating pre-activated neurons may also show fatigue or adaptation, thus generally reducing their activity. This means that, when the critical symbolic stimulus arrives after priming, semantic feature neurons may reach a lower level of activity (due to fatigue) compared with the unprimed case. Differential fatigue and pre-activation may both contribute to the reduced activation difference brought about by a symbol after it has been semantically primed.
6. Note that even linguistic textbooks made and still make obviously incorrect statements about shared semantic features of semantic categories. For example, the meaning description of “bachelor” as [+male, +adult, –married] has been

perpetuated even though obvious counterexamples (polygamous young urban professionals, the pope, etc.) have been pointed out decades ago (Lakoff, 1987).

## References

- Alston, W. P. (1964). *Philosophy of language*. Englewood Cliffs, NJ: Prentice-Hall.
- Arbib, M. A., Bonaiuto, J. J., Bornkessel-Schlesewsky, I., Kemmerer, D., MacWhinney, B., Nielsen, F. A., & Oztop, E. (2014). Action and language mechanisms in the brain: Data, models and neuroinformatics. *Neuroinformatics*, 12(1), 209–225. <https://doi.org/10.1007/s12021-013-9210-5>.
- Barlow, H. (1972). Single units and cognition: A neurone doctrine for perceptual psychology. *Perception*, 1, 371–394.
- Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Lawrence Erlbaum Associates.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–609.
- Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, 23(4), 1122–1142. <https://doi.org/10.3758/s13423-016-1028-3>.
- Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: The case from action verbs. *Annals of the New York Academy of Sciences*, 1224, 81–95. <https://doi.org/10.1111/j.1749-6632.2011.06013.x>.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343–355.
- Bibbig, A., Wenekers, T., & Palm, G. (1995). A neural network model of the cortico-hippocampal interplay and the representation of contexts. *Behavioral Brain Research*, 66(1–2), 169–175.
- Bienenstock, E. (1994). A model of the neocortex. *Network: Computation in Neural Systems*, 6, 179–224.
- Boas, F. (1940). *Race, language, and culture*. Vol. 90449. Chicago, IL: University of Chicago Press.
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim & G. Palm (Eds.), *Theoretical approaches to complex systems*. (Lecture notes in biomathematics, vol. 21) (pp. 171–188). Berlin: Springer.
- Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity* (2nd ed.). Berlin: Springer.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177–210.
- Carnap, R., & Bar-Hillel, Y. (1952). *An outline of a theory of semantic information*. Boston, MA: Research Laboratory of Electronics, MIT.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–vi, 1–143.
- Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2013). Task modulation of brain responses in visual word recognition as studied using EEG/MEG and fMRI. *Frontiers in Human Neuroscience*, 7, 376. <https://doi.org/10.3389/fnhum.2013.00376>.
- Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: Evidence from spatiotemporal brain dynamics. *Journal of Cognitive Neuroscience*, 27(9), 1738–1751. [https://doi.org/10.1162/jocn\\_a\\_00815](https://doi.org/10.1162/jocn_a_00815).
- Clark, A. (1996). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82(6), 407–428.
- Davis, G. A. (1989). The cognitive cloud and language disorders. *Aphasiology*, 3, 723–733.
- de Saussure, F. (1916). *Cours de linguistique generale*. Paris: Payot.



- Elbert, T., & Rockstroh, B. (1987). Threshold regulation—a key to the understanding of the combined dynamics of EEG and event related potentials. *Journal of Psychophysiology*, 4, 317–333.
- Fillmore, C. J. (1975). An alternative to checklist theories of meaning. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 1, 123–131.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, 61(6), 825–850.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, 100, 25–50.
- Frege, G. (1918). Der Gedanke: Eine logische Untersuchung. *Beiträge zur Philosophie des Deutschen Idealismus*, 1, 58–77.
- Fuster, J. M. (1995). *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.
- Fuster, J. M. (2015). *The prefrontal cortex* (5th ed.). London: Elsevier.
- Fuster, J. M., & Bressler, S. L. (2012). Cognit activation: A mechanism enabling temporal integration in working memory. *Trends in Cognitive Sciences*, 16(4), 207–218. <https://doi.org/10.1016/j.tics.2012.03.005>.
- Garagnani, M., & Pulvermüller, F. (2016). Conceptual grounding of language in action and perception: A neurocomputational model of the emergence of category specificity and semantic hubs. *European Journal of Neuroscience*, 43(6), 721–737. <https://doi.org/10.1111/ejn.13145>.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically-grounded Hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27(2), 492–513.
- Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7), 905–922. <https://doi.org/10.1016/j.cortex.2011.04.010>.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401.
- Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cerebral Cortex*, 26, 2353–2366. <https://doi.org/10.1093/cercor/bhw026>.
- Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience*, 37(18), 4848–4858. <https://doi.org/10.1523/JNEUROSCI.2800-16.2017>.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42, 335–346.
- Harnad, S. (2012). From sensorimotor categories and pantomime to grounded symbols and propositions. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford handbook of language evolution* (pp. 387–392). Oxford, UK: Oxford University Press.
- Hauk, O., & Tschentscher, N. (2013). The body of evidence: What can neuroscience tell us about embodied semantics? *Frontiers in Psychology*, 4, 50. <https://doi.org/10.3389/fpsyg.2013.00050>.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>.
- Hebb, D. O. (1949). *The organization of behavior. A neuropsychological theory*. New York: John Wiley.
- Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. New York: WW Norton & Company.
- Hoening, K., Sim, E. J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*, 20(10), 1799–1814.
- Hubel, D. (1995). *Eye, brain, and vision* (2nd ed). New York: Scientific American Library.
- Jaszczolt, K. M. (2016). *Meaning in linguistic interaction: Semantics, metaseantics, philosophy of language*. Oxford, UK: Oxford University Press.

- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3(8), 759–763.
- Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, 39, 170–210.
- Kemmerer, D. (2015a). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, 22(4), 1068–1075. <https://doi.org/10.3758/s13423-014-0784-1>.
- Kemmerer, D. (2015b). Does the motor system contribute to the perception and understanding of actions? Reflections on Gregory Hickok's The myth of mirror neurons: The real neuroscience of communication and cognition. *Language and Cognition*, 7(3), 1068–1075.
- Kemmerer, D. (2016). Erratum to: Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, 23(4), 1143. <https://doi.org/10.3758/s13423-016-1031-8>.
- Keysers, C., Perrett, D. I., & Gazzola, V. (2014). Hebbian learning is about contingency, not contiguity, and explains the emergence of predictive mirror neurons. *Behavioral and Brain Sciences*, 37(2), 205–206. <https://doi.org/10.1017/S0140525X13002343>.
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>.
- Kleene, S. C. (1956). Representation of events in nerve nets and finite automata. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies* (pp. 3–41). Princeton, NJ: Princeton University Press.
- Klepp, A., Nicolai, V., Buccino, G., Schnitzler, A., & Biermann-Ruben, K. (2015). Language-motor interference reflected in MEG beta oscillations. *NeuroImage*, 109, 438–448. <https://doi.org/10.1016/j.neuroimage.2014.12.077>.
- Klepp, A., Nicolai, V., Sieksmeyer, J., Arnzen, S., Indefrey, P., Schnitzler, A., & Biermann-Ruben, K. (2017). Body-part specific interactions of action verb processing with motor behaviour. *Behavioral Brain Research*, 328, 149–158. <https://doi.org/10.1016/j.bbr.2017.04.002>.
- Knoblauch, A., & Palm, G. (2001). Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Netw*, 14(6–7), 763–780.
- Knoblauch, A., & Palm, G. (2002). Scene segmentation by spike synchronization in reciprocally connected visual areas. I. Local effects of cortical feedback. *Biological Cybernetics*, 87(3), 151–167.
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412. <https://doi.org/10.1016/j.tics.2013.06.007>.
- Lakoff, G. (1987). *Women, fire, and dangerous things. What categories reveal about the mind*. Chicago, IL: University of Chicago Press.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2013). *Handbook of latent semantic analysis*. Hove, UK: Psychology Press.
- Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cogn Sci*, 39(8), 1764–1801. <https://doi.org/10.1111/cogs.12174>.
- Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review*, 23(4), 991–1001. <https://doi.org/10.3758/s13423-015-0870-z>.
- Löbner, S. (2014). *Understanding semantics* (2nd ed.). London: Routledge.
- Locke, J. (1909/1847). *An essay concerning human understanding, or, the conduct of the understanding*. Philadelphia, PA: Kay and Troutman.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083.
- Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, 30(4), 420–429. <https://doi.org/10.1080/23273798.2014.987791>.

- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology—Paris*, 102(1–3), 59–70.
- Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3), 108–114.
- McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48(7), 788–804. <https://doi.org/10.1016/j.cortex.2010.11.002>.
- Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2017). Verbal labels facilitate tactile perception. *Cognition*, 171, 172–179. <https://doi.org/10.1016/j.cognition.2017.10.010>.
- Milner, P. M. (1996). Neural representation: Some old problems revisited. *Journal of Cognitive Neuroscience*, 8, 69–77.
- Mollo, G., Pulvermüller, F., & Hauk, O. (2016). Movement priming of EEG/MEG brain responses for action-words characterizes the link between language and action. *Cortex*, 74, 262–276. <https://doi.org/10.1016/j.cortex.2015.10.021>.
- Moseley, R. L., Pulvermüller, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: Brain activity dissociates between conceptual categories within 150 ms. *Scientific Reports*, 3, 1928.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>.
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism*. New York: Harcourt, Brace & World.
- Palm, G., Knoblauch, A., Hauser, F., & Schuz, A. (2014). Cell assemblies in the cerebral cortex. *Biological Cybernetics*, 108, 559–572. <https://doi.org/10.1007/s00422-014-0596-4>.
- Palm, G., & Sommer, F. T. (1995). Associative data storage and retrieval in neural networks. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of neural networks III* (pp. 79–118). New York: Springer Verlag.
- Perrett, D. J., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurons responsive to faces. *Trends in Neurosciences*, 10, 358–364.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–336.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582.
- Pulvermüller, F. (2008). Grounding language in the brain. In M. de Vega, A. Graesser, & A. M. Glenberg (Eds.), *Symbols, embodiment, and meaning* (pp. 85–116). Oxford, UK: Oxford University Press.
- Pulvermüller, F. (2013a). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9), 458–470. <https://doi.org/10.1016/j.tics.2013.06.004>.
- Pulvermüller, F. (2013b). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, 127(1), 86–103. <https://doi.org/10.1016/j.bandl.2013.05.015>.
- Pulvermüller, F. (2018a). The case of CAUSE: Neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 373, 20170129. <http://dx.doi.org/10.1098/rstb.2017.0129>
- Pulvermüller, F. (2018b). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, 160, 1–44. <https://doi.org/10.1016/j.pneurobio.2017.07.001>.
- Pulvermüller, F., & Garagnani, M. (2014). From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment. *Cortex*, 57, 1–21.

- Pulvermüller, F., Garagnani, M., & Wennekers, T. (2014). Thinking in circuits: Towards neurobiological explanation in cognitive neuroscience. *Biological Cybernetics*, 108(5), 573–593. <https://doi.org/10.1007/s00422-014-0603-9>.
- Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. J. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17(6), 884–892.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, 47(2), 388–396.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.
- Schütze, H. (1992). *Dimensions of meaning*. Washington, DC: IEEE Computer Society Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457.
- 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(18), E1918–E1923. <https://doi.org/10.1073/pnas.1323158111>.
- Shtyrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. *European Journal of Neuroscience*, 19(4), 1083–1092.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.
- Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science*, 16(3), 132–135.
- Stramandinoli, F., Marocco, D., & Cangelosi, A. (2017). Making sense of words: A robotic model for language abstraction. *Autonomous Robots*, 41(2), 367–383.
- Tomasello, R., Garagnani, M., Wennekers, T., & Pulvermüller, F. (2017). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*, 98(4), 111–129. <https://doi.org/10.1016/j.neuropsychologia.2016.07.004>.
- Tomasino, B., Fabbro, F., & Brambilla, P. (2014). How do conceptual representations interact with processing demands: An fMRI study on action- and abstract-related words. *Brain Research*, 1591, 38–52. <https://doi.org/10.1016/j.brainres.2014.10.008>.
- Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fMRI study of mental imagery and silent reading of action phrases. *NeuroImage*, 36 (Suppl 2), T128–T141.
- Trumpp, N. M., Kliese, D., Hoenig, K., Haarmeier, T., & Kiefer, M. (2013). Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex*, 49(2), 474–486. <https://doi.org/10.1016/j.cortex.2012.02.002>.
- Ulrich, M., Adams, S. C., & Kiefer, M. (2014). Flexible establishment of functional brain networks supports attentional modulation of unconscious cognition. *Human Brain Mapping*, 35(11), 5500–5516. <https://doi.org/10.1002/hbm.22566>.
- van Dam, W. O., Brazil, I. A., Bekkering, H., & Rueschemeyer, S. A. (2014). Flexibility in embodied language processing: Context effects in lexical access. *Topics in Cognitive Science*, 6(3), 407–424. <https://doi.org/10.1111/tops.12100>.
- Verduzco-Flores, S., Bodner, M., Ermentrout, B., Fuster, J. M., & Zhou, Y. (2009). Working memory cells’ behavior may be explained by cross-regional networks with synaptic facilitation. *PLoS ONE*, 4(8), e6399.
- von Humboldt, W. (1979). *Sprache als Weltansicht Schriften zur Sprachphilosophie* (pp. 10–20 (Auszüge)). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Watson, J. B. (1925). *Behaviorism*. New York: Transaction.

- Wennekers, T., Garagnani, M., & Pulvermüller, F. (2006). Language models based on Hebbian cell assemblies. *Journal of Physiology—Paris*, 100, 16–30.
- Whorf, B. L. (1956). *Language, thought and reality: Selected writings*. Cambridge, MA: Technology Press of Massachusetts Institute of Technology.
- Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2, 116.
- Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience*, 22(10), 2387–2400.
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, UK: Blackwell.
- Wood, C. C. (1978). Variations on a theme by Lashley: Lesion experiments on the neural model of Anderson, Silverstein, Ritz, and Jones. *Psychological Review*, 85(6), 582–591.
- Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. M. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience*, 13(8), 3406–3420.

**Q/A commentaries related to Friedemann Pulvermüller’s paper, presented at the symposium “Abstract concepts: debating their structure, processing and modeling” (Amsterdam, 18 November 2016)**

KM: Thank you for the talk, very interesting. You have this idea of “conceptual kernel” or “kernel neurons,” and it seems to be related to ideas that have been around for a long time, such as “core of a concept,” “context-independent features,” “prototype.” Theoretically, how does the idea of “conceptual kernel” or “kernel neurons” relate to those concepts and what does it add?

FP: The *conceptual or semantic kernel* is different from “the core of the concept” as conceived by standard cognitive theories. The classic idea is that all semantic features included in the “core” are always relevant and processed when the word is used (necessary and sufficient semantic features). This idea is inadequate for theoretical reasons, as pointed out, and empirically unsupported. As I tried to show, kernel activation can be flexible (for example, in “this” and “there-is” contexts, as I illustrated). The kernel, halo, and referential periphery distinction of semantic circuits makes it possible to precisely model conceptual flexibility. In contrast, the semantic core idea contradicts conceptual flexibility.

----

GS: Thank you for the inspiring talk. When you talk about abstract concepts such as *beautiful*, well, these concepts have a special status compared to other abstract concepts, such as *time*... Concepts such as *beautiful* reflect interoceptive experiences, emotional experiences. *Beautiful* represents a sensation that we have, for which we learnt a linguistic label, the word *beautiful*, and it is a sensation that reappears across a number of different experiences (music, art, landscapes), but it does not fit in necessarily into any semantically decipherable features.

FP: Indeed, this is an emotionally loaded concept. But the idea that there is an internal state that only I can access is problematic. If I go to an art exhibition with my wife, and look at an art piece, my wife is able to tell whether I like it or not, because I somehow

express this through my body. Affective experiences, emotion concepts, as well as the meanings of emotion words, *do have* a manifestation in motor areas. And these emotion-expressing action schemas play a crucial role in learning the meaning of words for internal states and feeling, including not only “beautiful” but also “love,” “joy,” and “hope.” Rachel Moseley argues that autistic persons, who have a problem in expressing emotions in their non-linguistic behavior, show concordant deviance in their processing of emotion language (Moseley et al., 2015).

GS: Thank you. My other question relates to the influence of context. Think about the sentence “the sun is rising.” I give you three possible contexts, in which the meaning of that sentence changes completely. (i) Two people in Watergate, in a chamber. And one man says to the other “the sun is rising,” implies that they have to get out, because people are arriving in the office soon. (ii) Two lovers, both married, but to other people. In this case “the sun is rising” implies that at least one of them has to leave, but for other reasons. (iii) A married couple, both farmers, early in the morning, the wife says “the sun is rising,” which means it’s time to go to the field to work. So, that particular sentence is modified in different ways, as a function of the context. My question is: would the same neural activity be observed?

FP: So your question is whether and how the social context contributes. The answer is of course a clear yes, it contributes, a lot. We have run some experiments on similar examples. Consider the word “water.” In some contexts, it may be used to just name a thing, for example to teach the word *water* to a child, while pointing at a glass of water. In other contexts, it can be used for example to request a glass of water from a waiter, and this use implies more social and motoric interaction. My colleague Natalia Egorova measured the neural activation in these two different contexts (naming and requesting) and found indeed that different cortical areas were activated (Egorova, Shtyrov, & Pulvermüller, 2016). In the request context, the predictive action sequence structure that characterizes linguistic and pragmatic interactions was reflected in specific neural activity, which included motor and theory of mind systems.

----

AU: Thank you, great talk. What if for some concepts the kernel is made of related words. So, think about a word such as *freedom*. What if the features that are often activated, the kernel, are related words, rather than grounded features? And we know that there are related words that are often activated, because there is semantic priming. So, in this case how would you define the semantic kernel of *freedom*?

FP: In the neurobiological language model I am proposing, grounded referential semantic features are implemented by kernel and halo of semantic circuits. These circuits include neurons in widespread modality-specific and multimodal areas of cortex. In contrast, combinatorial or distributional semantics, which is captured by semantic vectors describing the co-occurrence probabilities of words in texts, has a different mechanism: the neuronal connections between circuits for stored linguistic forms in perisylvian language cortex. However, the two types of information, grounded and combinatorial, may not be easily kept apart: When learning from context “that strawberries are red” after knowing RED from experience, how can we exclude that grounded

semantic features are incorporated into the (combinatorially manufactured) STRAW-BERRY representation? Francesca Carota recently explored the brain correlates of distributional semantic similarity (using a technique called latent semantic analysis, LSA) and found that the *representational similarity of neuronal activation patterns* reflected LSA similarities in areas including prefrontal, premotor, and inferior-frontal areas—all “extrasyllvian” areas likely involved in semantic grounding (Carota, Kriegeskorte, Nili, & Pulvermüller, 2017).

**Abbreviations:**

KM: Ken McRae

FP: Friedemann Pulvermüller

GS: Gun Semin

AU: Audience

**Additional references, mentioned during the debate:**

Carota, F., Kriegeskorte, N., Nili, H., & Pulvermüller, F. (2017). Representational similarity mapping of distributional semantics in left inferior frontal, middle temporal, and motor cortex. *Cereb Cortex*, 27(1), 294–309. <https://doi.org/10.1093/cercor/bhw379>

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *Neuroimage*, 125, 857–867. <https://doi.org/10.1016/j.neuroimage.2015.10.055>

Moseley, R. L., Shtyrov, Y., Mohr, B., Lombardo, M. V., Baron-Cohen, S., & Pulvermüller, F. (2015). Lost for emotion words: What motor and limbic brain activity reveals about autism and semantic theory. *Neuroimage*, 104, 413–422. <https://doi.org/10.1016/j.neuroimage.2014.09.046>