25

THE HIERARCHICAL SEMANTICS OF SELF

Georg Northoff and Diego Gorini

Introduction

The self is a basic phenomenon of our life and existence. While the concept of self has been discussed extensively in philosophy over the last three centuries, its presence in psychology and neuroscience is more recent. Various studies have shown complex psychological (Sui & Humphreys, 2015) and neural (Apps & Tsakiris, 2014; Frewen et al., 2020; Northoff et al., 2006; Qin et al., 2020; Tsakiris, 2016, 2017; van den Meer et al., 2010) correlates of the self; yet, a crucial neurosemiotic question remains unanswered: How are these signatures and their psychological manifestation related to semantics? Is there correspondence –if not convergence – between the neural hierarchies of the self and of semantic systems? And if so, what enables such a convergence?

Neurosemiotics investigates signs and their intersubjective manifestations by bridging neurobiological levels with explanations of "psychological" phenomena such as cognition, intentionality, and meaning. Across multiple levels of interactions, semantics is one of the semiotic dimensions more critically related to meaning. In this chapter, we use a tool of neurosemiotics to forge interdisciplinary links between the self with the semantics. We will focus on semantic analogies among neural hierarchies of the self and semantic hierarchies, as well as their respective and convergent timescales. We will enrich our proposals by providing insight from a computational model of vector semantic space operating on a temporal continuum of long and short timescales. Briefly, we aim to address vital questions about the self and semantics, charting their connections with a neurosemiotic approach.

Hierarchical layers of self

Contemporary research on the self spans diverse psychological, neural, and conceptual levels. Psychologically, the self impacts functions such as perception (Sui et al., 2012), reward (deGreck et al., 2008; Yankoskaya et al., 2020), action (Frings & Wentura, 2014), emotion (Northoff et al., 2009), attention (Sui et al. 2013), and decision-making (Nakao et al., 2016, 2019). Neurally, the self (as well as these functions) has been associated with multiple brain regions, including cortical midline structures (Northoff et al., 2006), premotor and motor cortices (Hu et al., 2016; Qin et al., 2020), the insula (Enzi et al., 2009), the temporo-parietal junction (Qin et al., 2016, 2020), and the anterior cingulate cortex (Qin et al., 2016, 2020), as well as subcortical regions such as the amygdala, the ventral tegmental area, and the midbrain (deGreck et al., 2008). These findings have inspired specific conceptualizations of

DOI: 10.4324/9781003051817-30

the self. Following Mesulam (1998) and others, including McLean (1990) and Panksepp (1998), the brain's cortical organization has been conceived in a layer-based way as in terms of radial-concentric organization. The radial-concentric approach is also reflected in the three-layer model by Feinberg and Northoff (Northoff et al., 2011), the assumption of a dynamic core (DePasquale et al., 2018), and the core–periphery hierarchy with a sensory-transmodal gradient (Margulies et al., 2016).

Despite some differences, these radial-concentric models share the assumption of gradient patterns – that is, a continuum of different processes, regions, and functions. Gradient patterns can be conceived as hallmark features of neural hierarchies (Huntenberg et al., 2018; Margulies et al., 2016). Following these developments, we propose an analogous hierarchy of the self, equally typified by gradient patterns (Qin et al., 2020). Such a hierarchical model of the self integrates and unifies different neural, psychological, and conceptual correlates by assuming a gradient pattern of different layers of self-processing – that is, self-related, self-predictive, and self-referential processing (Northoff, 2007, 2016). This view offers fruitful foundations to examine the key question tackled herein.

From the hierarchy of self to semantic similarity and relatedness

Do these layers of self-specificity converge with analogous semantic layers? Our main assumption is that the hierarchy of self converges with a semantic hierarchy of semantic similarity and relatedness. This, as we demonstrate, is possible through shared time windows. While the different layers of self can be distinguished by their different timescales (from short to long), semantic similarity and relatedness (see below for definition) are also mediated by short and long timescales, respectively.

We therefore propose that timescales provide the link, if not the "common currency" (Northoff et al., 2020), of self and semantics, as both are mediated by the co-occurrence of short and long timescales. Our proposal lies at the interface between semantics, semiotics, and neuroscience. It is semantic in that we are concerned with meaning and how it is constructed. For that, we assume that timescales are key in constructing meaning, thus adopting a temporal view of semantics. Such a view invites semiotic insights, as the self deploys acts and objects that take on the role of "signs" through their integration within the timescales of the self. This touches upon neurosemiotics, as we assume these timescales are yielded by and based on the brain, as most basic features of our self.

These three layers of self-specificity constitute the core proposal to trace neurosemiotic links between the self and semantics, as discussed in the first and second parts. Drawing on structuralist research, we distinguish between semantic similarity and relatedness. After establishing these concepts, we derive further insights from a recent computational model of vector semantic space that associates semantic relatedness and similarity with different timescales, long and short. That, as outlined in the third part, provides the link to the self, which also operates on a temporal continuum of long and short timescales. Specifically, we postulate that the three layers of self (self-related, self-predictive, and self-referential) mediate the temporal continuum of semantic similarity and relatedness with their short and long-time windows.

Neural hierarchy of self: Different layers of self-processing

Layers and neural hierarchy of self

A recent large-scale meta-analysis on healthy subjects (Qin et al., 2020) integrated imaging studies that focused on different aspects of the self, namely inner body (interoceptive processes), outer body (extero-proprioceptive processes), and reflections on one's own mental states. Different regions were associated with each of the three layers, and the regions of the lower layer were included within the next upper layer (see below for details). Together, this amounts to a spatial multi-layered nested hierarchical model of self (Qin et al., 2020), including an intero-exteroceptive self, an extero-proprioceptive/cognitive self, and a higher-order reflective or mental self.

The interoceptive self (i.e., how the brain processes and perceives the body's inner organs and their input) was investigated through fMRI studies that measured interoceptive awareness, including cardiorespiratory, urogenital, and gastrointestinal awareness. That was complemented by fMRI studies on the extero-proprioceptive self, focusing on external bodily inputs, such as facial or proprioceptive input signals. Finally, studies were included that tapped on more cognitive aspects of self, employing trait adjectives or other stimuli so that subjects became aware of their own self as distinct from others (Figure 25.1).

The intero-exteroceptive self: Self-related processing

Based on the interoceptive studies, there is a more basic or lower layer of self, an interoceptive self that is related to regions that mostly process interoceptive signals –for example, the bilateral insula, the dorsal anterior cingulate cortex, the thalamus, and the parahippocampus, thus including mainly regions of the salience network (Menon, 2011; Qin et al., 2020). The fact that these regions were shared among the different kinds of interoceptive awareness (i.e., cardiorespiratory, urogenital, and gastrointestinal) speaks to their role in integrating different interoceptive inputs from various inner-body organs (Craig, 2009). One can thus identify an "interoceptive or vegetative self" (Azzalini et al. 2019); Tsakiris, 2017), "bodily self" or "proto-self" (Panksepp, 1998) as the most basic and fundamental layer of self. However, subcortical and cortical regions, such as the thalamus and the insula, don't just process interoceptive inputs form the inner body. They also process exteroceptive stimuli from the outer environment (Craig, 2009). Given such intero-exteroceptive convergence, we characterize this level of self as self-related processing (SLP) and an intero-exteroceptive self.

To better understand SLP, we need to distinguish between intrinsic and extrinsic self-specificity. Self-specificity does not come with the stimuli themselves – it is not an intrinsic feature. For instance, even our name is not intrinsic to our self as our parents could have named us otherwise. The same is true of our face, which due to changes by external events or over the course of life, can change in major ways without affecting our self – it is still my face despite all the changes. Therefore, self-specificity is an extrinsic rather than intrinsic feature of stimuli or contents.

The typical example is the rubber hand illusion. We can conceive another person's hand or arm as our own body's hand or arm – this means that our own body's hand is not an intrinsic feature of our self as my body's own hand can also be related with a hand from another body (Tsakiris, 2017). This may be due mostly to complex processes that integrate the multiple interoceptive input streams with the extero- and proprioceptive input streams (Tsakiris, 2017) in a seemingly purely stochastic way (Apps & Tsakiris, 2014, independent of their respective origin (Northoff, 2016). This results in an intero-exteroceptive self that endows self-specificity to interoceptive and exteroceptive inputs by integrating them in ways that are specific for the respective person (that is, his/her self). The intero-exteroceptive self is thus, in its most basic way, an integrative self (Sui & Humphreys, 2015).

Predictive or cognitive self: Self-predictive processing

The next or middle layer of self includes what Qin et al. (2020) describe as proprioceptive or exteroceptive self. This can also be taken as a first instance of a cognitive self. FMRI studies focusing on external bodily-related inputs, such as facial or proprioceptive inputs, implicate regions like the bilateral insula, the inferior frontal gyrus, the premotor cortex, the temporo-parietal junction (TPJ), and the medial prefrontal cortex (MPFC). As these regions process inputs from different sensory modalities, they may be key not only for integrating extero- and proprioceptive modalities, but also different exteroceptive sensory modalities – that is, cross-modal integration. Despite their differences, these regions share the processing of proprioceptive inputs related to one's own body – one can thus speak of a "proprio- or exteroceptive self, or embodied self" (Panksepp, 1998; Tsakiris, 2017). This

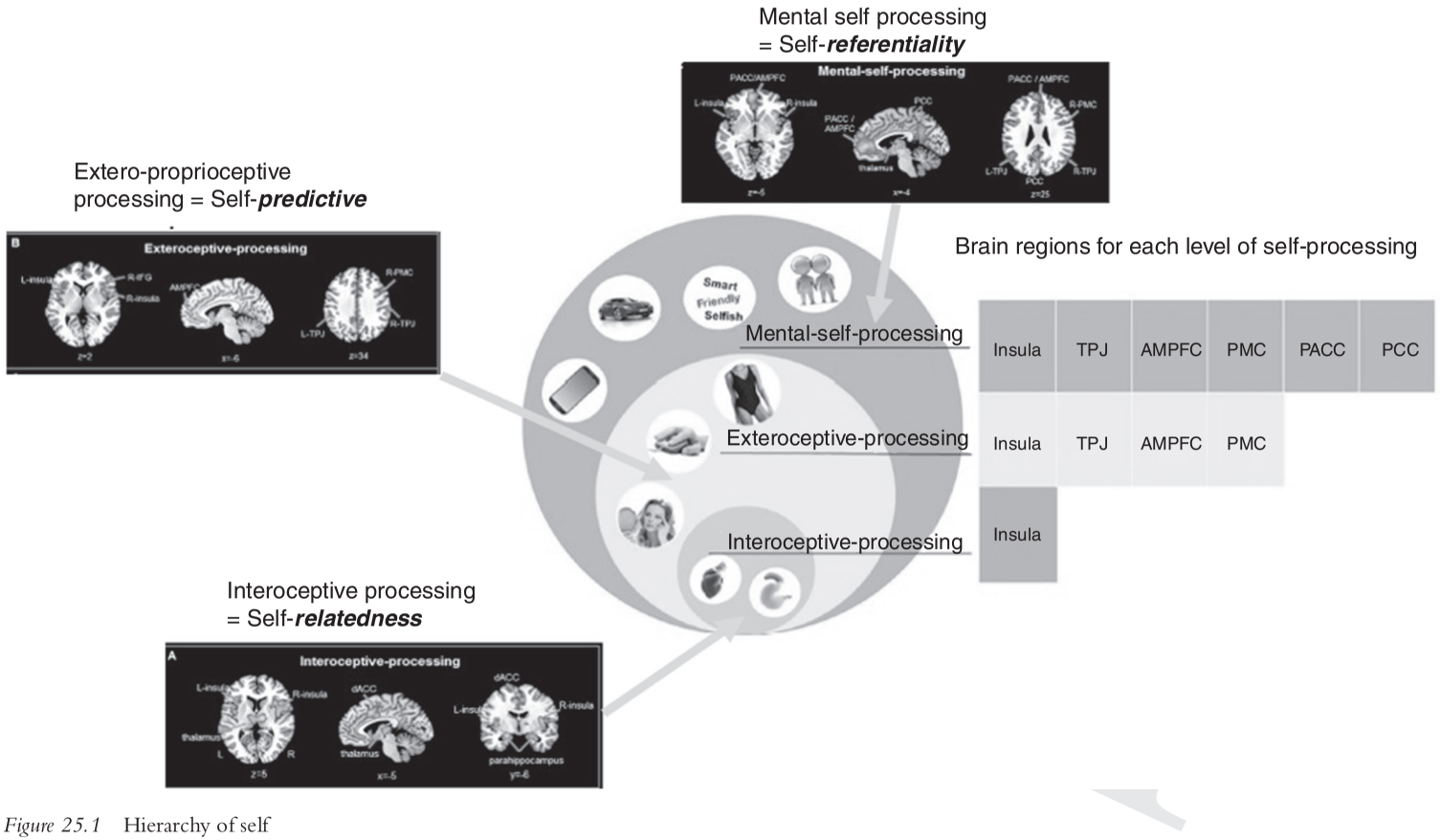

Figure 25.1 Hierarchy of self

This model is presented as obtained in a large-scale meta-analysis of self by Qin et al. (2020). The black boxes show the regions obtained for each layer of self-processing, while the middle blue-free circles show the nested hierarchy with the regions of the lower layer resurfacing and nesting within the next more extended layer and so forth. See e-book for a full-color version of this figure.

may be conceived as the most basic instance of prediction, and thus of cognition. We therefore characterize this level of self as self-predictive processing (SPP) or cognitive self.

SPP implies that self-specific contents can serve to predict or anticipate future stimuli and their respective contents. More specifically, self-specific contents serve as predicted inputs which are then compared with the actual inputs from the body and/or the world. SPP has been mainly examined in paradigms involving either parts of the body, like faces (see Hu et al., 2016, for meta-analysis of self-face paradigms) and arms (as in the rubber hand illusion; Apps & Tsakiris, 2014); or, alternatively, the whole body (Blanke et al., 2015).

This concerns mainly the proprioceptive and exteroceptive stimuli/contents of the body. SPP is based on the theory of predictive coding (Friston, 2010) as applied to the self (Apps & Tsakiris, 2014). Predictive coding claims that the degree of matching or difference between the predicted input and the actual input determines further processing and subsequent behavior (Friston, 2010, 2018; Apps & Tsakiris, 2014).

Applied to the specific case of the self, this results in the following: if the stimulus (i.e., the actual input) is different from the self-specific content that is the predicted input, then there will be a large prediction error indicating a discrepancy between the self-specific content and the stimulus – the actual stimulus will then be endowed with a low degree of self-relatedness. If, in contrast, the self-specific content as predicted input matches strongly with the input stimulus, this results in a small prediction error with a high degree of self-specificity attributed to the stimulus. The underlying mechanism here is that the brain can predict the self based on multisensory integration and a probabilistic representation of the self – that is, self-contents (Apps & Tsakiris, 2014).

The predictive relation in SPP is based on the stochastic properties of the self as predicted input and the actual input. This is so because predictive coding and SPP are based on a stochastically based active inference from its present self-specific content as predicted input to actual inputs – e.g., future stimuli (Seth, 2013; Friston, 2018; Moutoussis et al., 2014; Apps & Tsakiris, 2014). That also entails a rather lower-order processing mode, as such stochastically based inference and matching can well operate in at least a partially implicit and involuntary mode. That distinguishes SPP from the more explicit level of higher-order reflective or mental self-processing, self-referential processing as we will discuss in the next section.

Higher-order reflective or mental self: Self-referential processing

Finally, the highest layer of self (Qin et al., 2020) has been related to typical midline regions of the default mode network (DMN), like the medial prefrontal and the posterior cingulate cortices as well as the regions included in the second level (most notably, the bilateral TPJ) and the first level (the bilateral insula and thalamus). These regions seem to be recruited when one needs to represent one's own self in mental states – one can therefore also speak of a "mental or cognitive self" (Qin et al, 2020) or "extended self" (Damasio, 2010).

Self-specific stimuli in these paradigms usually involve the subjects' own name, his/her face, or personally related trait adjectives (see above). Importantly, the self-specificity of these stimuli is given and presupposed and thus pre-existing (or ready-made): these experimental paradigms do not focus so much on the stimuli themselves but rather on the psychological content of the stimuli and how the subject refers to self- and non-self-specific stimuli. These paradigms thus test for the difference between the subjects' referring to self- and non-self-specific stimuli – the underlying processing mode is therefore described as self-referential processing (SRP) (Sui & Humphreys, 2015; Sui et al., 2012).

On the psychological level, SRP involves a variety of different psychological functions. Several self-paradigms require subjects to focus their attention on self-specific stimuli vs. non-self-stimuli (Sui et al., 2012; Sui & Humphreys, 2015). Since self- and non-self-specific stimuli lead to different recruitment of attentional resources (Sui et al. 2013; Sui & Humphreys, 2015), attention may need to be considered yet another important cognitive feature required by self-reference. Moreover, other

cognitive self-paradigms require subjects to imagine self-specific contents versus non-self-specific contents in past, present, and future (d'Argembeau et al., 2010) – self-reference here involves imagination of future and past, described as episodic simulation (Schacter et al., 2012).

Taken together, SRP can be characterized by the subjects' reference to self-specific contents, hence the name self-*referential*. SRP is usually coined self-reflection (Hu et al., 2016;, Sui and Humphreys, 2015, Northoff, 2016), as one reflects upon one's own self and its self-specific contents by recruiting various cognitive functions. Other related concepts include self-certainty, self-recognition, and self-insight (see Philippi & Koenigs, 2014). While somewhat distinct, these terms are nevertheless unified by the fact that they all describe the subjects' reference to self-specific psychological contents in a higher-order reflective or mental way.

Nested hierarchy of self: Different spatial and temporal scales are integrated

Together, these findings point to what Qin et al. (2020) describe as "nested hierarchy of self": regions of the lower level were included in the next higher level, where they were complemented by additional regions, and so forth. For instance, the bilateral insula was present on the most basic level (i.e., the interoceptive self) and resurfaced, in completely independent imaging studies, in both second (i.e., proprioceptive self) and third (i.e., mental self) levels. The same holds true for the bilateral TPJ, which first showed in the intermediate layer of the proprioceptive self and resurfaced again in the third level of the mental self. Accordingly, each of the hierarchical levels of self recruits both overlapping and separate regions compared to other levels amounting to spatial nestedness with a spatially nested hierarchy of self.

This conforms to nestedness as a spatial hierarchy of self: the lower layer regions are integrated within the more extended regions of the next upper layer and so forth. That conforms to an analogous nestedness of different layers of self-processing: self-related processing is integrated and resurfaces within SPP which, in turn, is integrated within SRP. Finally, on the more psychological level, this means that the intero-exteroceptive self is integrated and nested within the predictive self and that, in turn, it resurfaces within the higher-order cognitive reflective or mental self.

It shall also be noted that the spatial hierarchy of self in the brain is accompanied by a corresponding temporal hierarchy. The regions' neural activity displays different timescales – that is, intrinsic neural timescales (INT), which can be measured by the autocorrelation window (ACW). The upper layer regions, like the DMN, and midline structures show the longest ACW with long time windows optimally suited for temporal integration (Golesorkhi et al., 2021). While the intero- and exteroceptive regions exhibit much shorter ACW – the temporal windows of their neural activity are thus better suited for temporal segregation than temporal integration (Golesorkhi et al., 2021).

Recent studies also associate the self with the brain's INT. The higher the degree of self-specificity on the mental level, the longer the ACW and thus the higher the degrees of possible temporal integration (Huang et al., 2016; Kolvoort et al., 2020). This strongly suggests that the three layers of self-specific processing (self-relatedness, self-prediction, and self-reference) can also be distinguished in their temporal hierarchy: self-related processing involves shorter timescales which are nested and contained within the longer ones of SPP and ultimately within the longest ones of SRP. We are now ready to turn to semantics. We will see that semantics may also be viewed in terms of its timescales, which may provide an intrinsic connection between language/meaning, brain, and self.

Self meets language: Semantic similarity and semantic relatedness

Semantic relatedness versus semantic similarity

For creatures living in the midst of a continuous flux of information, categorical perception is a fundamental surviving skill that has been carefully studied by psychology and the cognitive sciences

(Harnad, 2003). In fact, determining that two things are identical, similar, or simply related is a complex cognitive ability that has proven difficult to formalize in semantics (McRae et al., 2012; Schulte im Walde, 2020).

Take, for example, the pairs of words *cup* and *coffee*, and *cup* and *mug*. Which of these pairs has the stronger semantic similarity? We certainly employ the former pair in close proximity much more often than the latter pair, and the word *cup* primes the word *coffee* more strongly than the word *mug*. Thus, we are tempted to say that *cup* and *coffee* are similar words. Yet, cups and coffees are not at all similar, beyond the fact that they are both concrete entities. Cups and mugs, on the other hand, commonly share many attributes, such as shape, material, and function.

More precisely, semantic similarity between words *A* and *B* is commonly defined as having a high degree of substitutability, which usually entails that referents A and B share many attributes (Miller & Charles, 1991; Rapp, 2002; Schütze & Pedersen, 1993). Quite trivially, this explains why one can substitute some words with their synonym: synonyms are very similar words. On the other hand, words *cup* and *coffee* are said to be merely semantically related (rather than being similar), which usually entails that cups and coffees tend to co-occur in space and time. Again, this trivially explains the priming effect that word *cup* has on word *coffee*: one tends to perceive or interact with the referents within the same space-time window (Figure 25.2).

It is tempting to suppose that the distinction between *true similarity* and *simple relatedness* subsists only at the lexical level – that is, as relations between words. However, this semantic distinction actually extrapolates up to the behavioral and neural sciences, in the distinction between taxonomic versus thematic knowledge (Merck et al., 2019; Mirman et al., 2017; Schwartz et al., 2011). For example, given task X, you could determine that tools A and B are similar enough to be substitutable in the performing of that task. Knowing that tool A may substitute tool B maps semantic similarity to taxonomic knowledge, while knowing that both A and B relate to task X maps semantic relatedness to thematic knowledge. Alternatively, on a logical view, one could decide to adopt either a dog or a

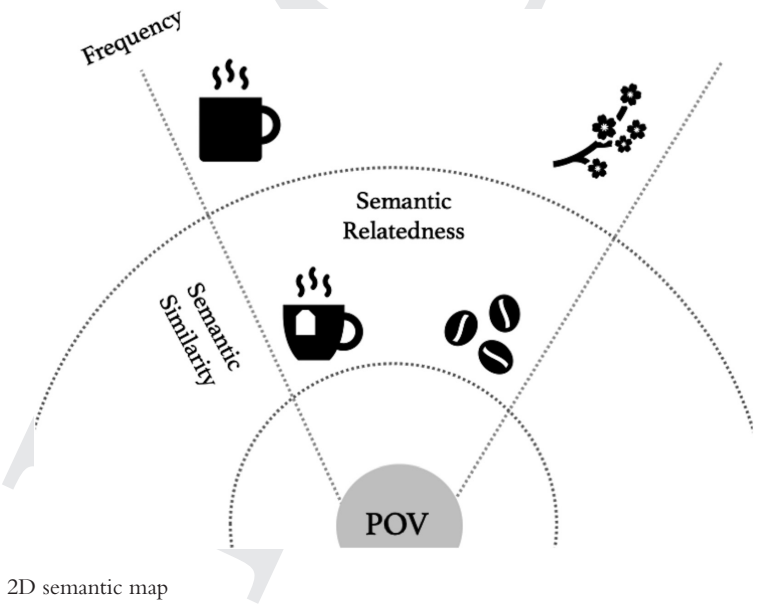

Figure 25.2 2D semantic map

This figure shows four objects (cup, mug, coffee, tea) from a single point of view (POV). Things along the semantic relatedness axis (blue) tend to *co-occur* within the same spatiotemporal window. Things along the semantic similarity axis (orange) tend to share many attributes and functions, hence the possibility of *substitution*. The radial distance from the POV alludes to the likelihood of each object given the respective context. See e-book for a full-color version of this figure.

cat for home; in this case, semantic similarity (and substitutability) maps to the logical operator OR, and semantic relatedness (and co-occurrence) maps to AND, as in $(cat \wedge home) \vee (dog \wedge home)$. A more general cognitive-semiotic process may be found underlying these mechanisms helping us making sense of the world along the distinction of semantic relatedness and similarity.

Finally, the distinction of semantic relatedness versus semantic similarity can also be viewed within the larger context of the self. The self provides references or a standard against or relative to how both semantic relatedness and semantic similarity are perceived. For instance, the self of a coffee addict will certainly draw extremely close semantic relationship of coffee and cup. In contrast, the self of somebody who hates and does not drink coffee at all but tea may rather draw close semantic relation of cup with tea while, at the same time, perceiving large semantic distance – that is, low relatedness of cup and coffee. Hence, the degrees of semantic relatedness and, ultimately, also of semantic similarity, strongly depend upon the context like on the self in which they operate.

Isomorphism of structural and semantic relations

Back to lexical semantics, a first step in this direction is to consider similarity and relatedness from the structural point of view (Erk, 2016; Peirsman, 2008; Sahlgren, 2008; Schulte im Walde, 2020). A bit like they are in mathematics, structural relationships in semantics should be understood as those relationships between words that occur by virtue of their *position* in a certain phrase or context, regardless of their content or referent. Studying word meaning without taking into account their referent may seem like a difficult task, but it turns out to be the essential requirement for machines to understand natural language, as we will see later. An example should suffice to clarify this point. A common task in many tests of second-language proficiency is for participants to be asked to guess ("fill in the blank") the most suitable word given a certain context (Landauer & Dumais, 1997). In English, sentences follow the structure subject–verb–object (SVO). So, if given the following sentence:

> "My grandfather ____ a successful lawyer,"
>
> the verbs *was*, *is*, and *will be* would surely come to mind, although, with less probability, other verbs are also possible, such as *met* and *saw*. Notably, this prediction is based entirely on the structural properties of the sentence. In fact, we could replace all of its words and the prediction would remain the same:
>
> "Her grandpa ____ an unsuccessful doctor."
>
> Clearly, the "horizontal" SVO structure allows for this kind of "vertical" substitution. However, finer-grained, nested structures also preserve this lexical property. The words *was*, *is*, and *will be* are merely conjugations of the verb *to be*, and conjugation is a typical example of vertical substitution. Other examples include pronouns (*my*, *her*), and semantic classes such as synonyms (*grandfather*, *grandpa*), co-hyponyms (*lawyer*, *doctor*), and antonyms (*successful*, *unsuccessful*).

More precisely, words having this "vertical" relationship within a linguistic structure are said to be paradigmatic parallels or to have a paradigmatic relation. Conversely, words having a "horizontal" relationship within a linguistic structure are said to be syntagmatic associates or to have syntagmatic relation. And now we can better understand the distinction between true semantic similarity and simple relatedness: Similar words imply an underlying vertical, exclusive, paradigmatic relation, and hence the possibility of substitution. Likewise, related words imply an underlying horizontal, inclusive, syntagmatic relation, and hence the spatiotemporal co-occurrence. Together, one can speak of isomorphism of structural and semantic relations (Figure 25.3).

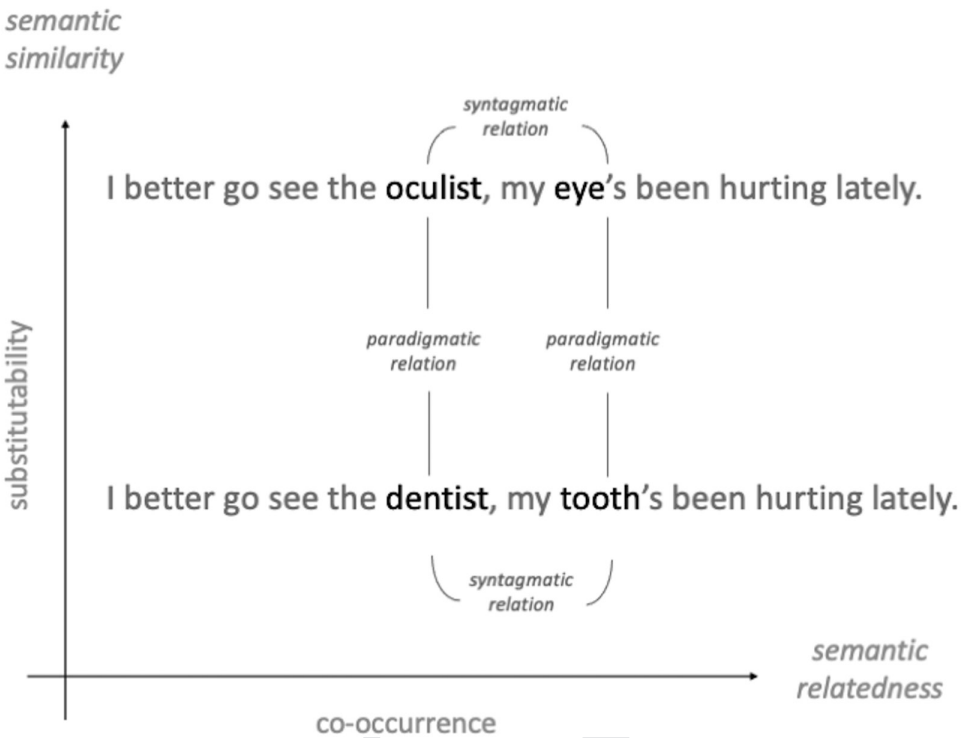

Figure 25.3 Structural and semantic relations along the same linguistic plane

Similar words have a "vertical" relation of substitutability because they are in a paradigmatic relation. Conversely, related words have a "horizontal" relation of co-occurrence because they are in a syntagmatic relation. The words in gray make clear that one and the same linguistic structure can serve as context to several similar *and* related words. See e-book for a full-color version of this figure.

From isomorphism to structuralism: The self is part of sociocultural structure

Isomorphism of meaning and structure has only recently been the subject of study (Erk, 2016; Jezek, 2016; Peirsman, 2008; Sahlgren, 2008; Schulte im Walde, 2020). Its historical origins can be traced to Ferdinand de Saussure's linguistic structuralism (de Saussure, 1922), Wittgenstein's idea of word meaning as usage (Wittgenstein, 1953), and especially to the so-called distributional hypothesis developed by Harris (1954).

In a nutshell, this hypothesis attempts to explain the meaning of words (i.e., semantics) as a function of context (i.e., structure), and has been stated and restated in various ways over time: "Difference of meaning correlates with difference of distribution," "You shall know a word by the company it keeps," "Words which are similar in meaning occur in similar contexts,", and so on. The idea is that you do not necessarily need to know the content or referent of a word to infer its meaning, so as long as you know the context (or "company") it is usually found in. For example, maybe you have never had seen an aardvark, but if you read the following sentences:[1]

> "The long-eared aardvark slept all day and ate insects all night."
>
> "With its rabbit-like ears and pig-like snout, the aardvark is an interesting looking animal."
>
> "Though it is a small mammal, the ant-eating aardvark finds its closest relative in the elephant."

You could probably guess that an aardvark is an animal similar to an anteater but with long ears – example adapted from Lenci (2008). In fact, the distributional hypothesis may explain the fast rate at which we are able to learn new words simply by reading (Landauer & Dumais, 1997). Moreover, acquiring semantic knowledge from distributional information may be particularly important in the case of abstract words, for which no external referent exists (Frassinelli et al., 2017). One can see that linguistic structuralism converges here more or less with what is described as distributional semantics, opening new avenues for mathematical and computational formalization of semantic similarity and relatedness. These will be exploited in the next part.

This view of semantic relatedness and similarity carries important consequences for conceiving the relationship between language and the self. Structuralism conceives semantic relatedness and similarity according to their context, the spatiotemporal features of which are key in especially determining the network of semantic relatedness. The self is, by itself, part of that context; importantly, though, it is not to be identified or equated with the context. The self provides a perspectival view of that context while being itself only part or one aspect of that context. Hence, the above figure may need to be enlarged by drawing the context in which the relation of self to semantic relatedness and similarity occurs.

Back to our example with the coffee and tea drinker, now imagine that the coffee addict is a Chinese man having grown up in a predominant tea culture; this will certainly affect both his self (Han & Northoff, 2008) and how he perceives the semantic relatedness of cup to both coffee and tea. One may guess that this will be different by his counterpart in, for instance, Italy, where coffee is much more often drunk than in China. The same, conversely, applies to the tea lover. Also, an Italian in China may want to drink coffee from a *gaiwan* or a *chawan* (both of which are typical Chinese bowls for preparing and drinking tea) because of the perceived semantic similarity to a coffee cup in Italy. In contrast, one cannot really imagine that a Chinese would associate such semantic similarity with a coffee cup since they relate *gaiwan* and *chawan* exclusively to tea drinking.

Accordingly, in addition to the self, the introduction of structuralism allows consideration of the self by itself as part of the more general and larger structure of the environment, a spatiotemporal structure. The latter, in turn, provides the ground or condition for establishing and constituting different degrees of semantic relatedness and similarity. The self, as discussed in the first part, may then be conceived primarily in terms of those structural relations or networks. It is primarily spatiotemporal, constituting and interweaving itself within the complex network of semantic relatedness and similarity of the respective cultural, social, and environmental context.

Computational mechanisms: Vector semantic space

A most interesting aspect of linguistic structuralism in the distributional hypothesis is that it can be, and has been, very successfully operationalized using relatively modest mathematical and computational means (Erk, 2012; Landauer & Dumais, 1997; Lenci, 2018; Turney & Pantel, 2010). This has made it possible to empirically test the capacity of the distributional hypothesis to model word meaning, and the semantic relations of similarity and relatedness, as well as other finer-grained semantic relations. The main idea behind these models is to map the semantic space of natural language to a mathematical vector space, so that to each word *a* corresponds an *N*-dimensional vector $\vec{a}$ encoding its meaning in a distributed fashion. Ideally, each dimension of the semantic vector space should encode some feature of word meaning (e.g., abstractness, concreteness). Although in practice it is not possible to be sure that that is what actually happens, semantic vector spaces do offer a clear semantic interpretation: vectors representing similar words are close in space. In other words, in a vector space, the semantic similarity of words *i* and *j* is a function of the distance between word vectors $\vec{i}$ and $\vec{j}$:

$$similarity\left(i, j\right) = \frac{1}{distance\left(\vec{i}, \vec{j}\right)}$$

The most commonly used distance metric is cosine, and a typical application consists in finding the *k*-nearest neighbors of a given word. For example, the five nearest neighbors of the noun *lawyer* in this[2] vector semantic space (Kutuzov et al., 2017) are (in order of descending cosine score): *attorney*, *jurist*, *barrister*, *solicitor*, and *politician*. However, keep in mind that the *k*-nearest neighbors of a word are not necessarily those words that could be considered the most similar by human judgment. Rather, the *k*-nearest neighbors are those words with which the target word maintains the strongest paradigmatic relation, and that therefore could reasonably substitute it in a given sentence. This is perfectly in line with the idea that distributional semantics regards words simply as positions in a structure, and thus the vectors encode word meaning (or content) as a function of distributional (or structural) information. But where does that information come from? (Figure 25.4).

Vector semantic models are created from large collections of texts (in the order of billions of words) using either of two methods: by counting co-occurrence frequencies (Landauer & Dumais, 1997; Lenci, 2018; Turney & Pantel, 2010), or by training a neural network to predict masked words in a sentence (Collobert & Weston, 2008; Devlin et al., 2018; Mikolov et al., 2013; Pennington et al., 2014). Despite some negligible differences in performance (Baroni et al., 2014; Lenci, 2018), both methods are based on the same principle (e.g., the distributional hypothesis), and have been proven to be mathematically equivalent (Levy & Goldberg, 2014; Levy et al., 2015) – that is, either an explicit (counting models) or implicit (predictive models) factorization of a matrix of co-occurrence frequencies. For simplicity, here we will only deal with "predictive" models.

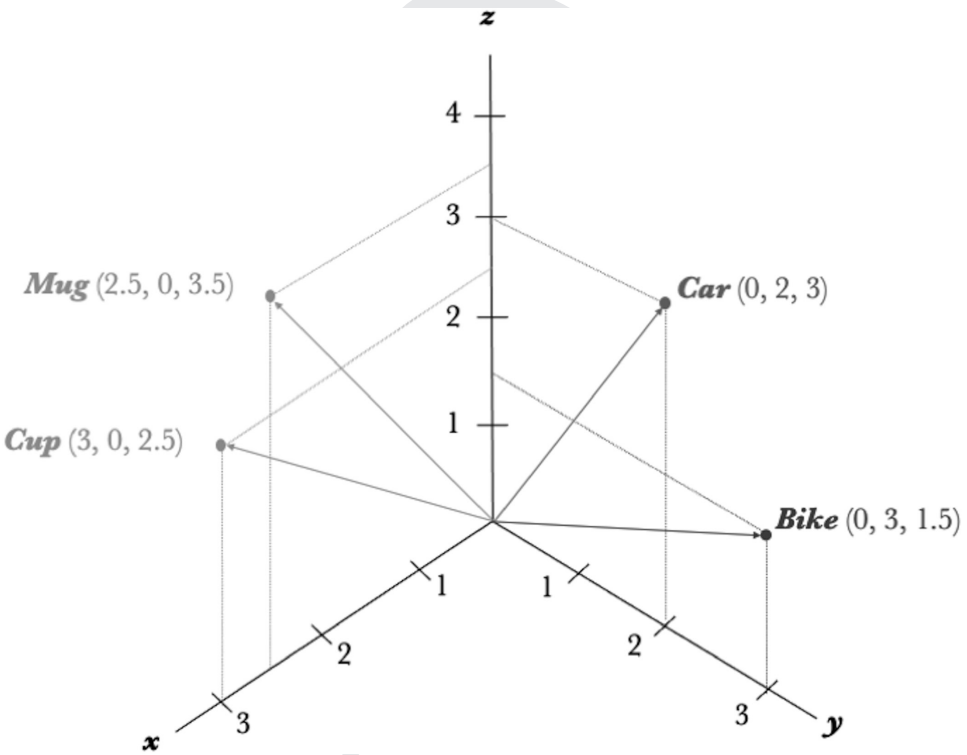

Figure 25.4 3D semantic vector space

Similar words (*mug* and *cup*, *car* and *bike*) tend to be close to each other in space. One may interpret the k-nearest neighbors of a word as a cluster representing a semantic family of some kind. Keep in mind that real semantic vector spaces have hundreds of dimensions (typically 100 or 300), so words can be close and further from each other along dimensions hard to interpret. See e-book for a full-color version of this figure.

Converging semantics, brain, and self through their timescales

Vector semantic space: Short- and long-time relations shape similarity and relatedness

We are now confronting the hardest part: taking into view the link between self and semantics. For that, we first go back to the semantic vector space itself and how it can be trained to yield semantic similarity and relatedness. Training a distributional model is a self-supervised task. Simplifying it quite a bit, a large collection of text is randomly segmented into training examples according to a so-called context window, the size of which ($1 \leq N$) determines the length of the training examples ($2N + 1$). For each training example, the distributional model learns to predict the word at the center, which is masked, based on the surrounding words – that is, based on the context. Essentially, the task is very similar to the "fill in the blank" exercise that was presented above, except that here the model can take into account only N number of words to the left and to the right of the mask (rather than the entire sentence); that is, it can only "see" the context window around the target word (Figure 25.5).

Intuitively, you could think that a longer context window would yield more accurate predictions, and consequently better semantic vector spaces. Note that semantic vector spaces refer primarily to positional distance in space which, as one can assume, translate into temporal distances in speed: the more distant two words are "located" from each in the semantic vector space, the more distant their respective positions in time. Hence, the spatial distances in the semantic vector space may translate into corresponding distances in time – see also Golesorkhi et al. (2021) for an analogous example of temporal and spatial convergence in brain topography. The concept of window, as used in the following sections, can thus be conceived in both spatial and temporal terms.

Extensive experimentation has shown that the size N of the context window actually determines the type of structural information that the model is able to capture (Agirre et al., 2009; Kiela & Clark, 2014; Kiela et al., 2015; Lapesa et al., 2014; Lapesa & Evert, 2017; Levy & Goldberg, 2014; Turney, 2012). Models trained on smaller windows ($N = 1, 2$) produce semantic vector spaces that capture or encode the paradigmatic relations between words, and thus perform best on semantic similarity tasks (in which, clearly, human judgment is the gold standard). Conversely, models trained on larger windows produce semantic vector spaces encoding syntagmatic relations, and thus perform best on semantic relatedness tasks. This result is important not only because it allows to distinctly model

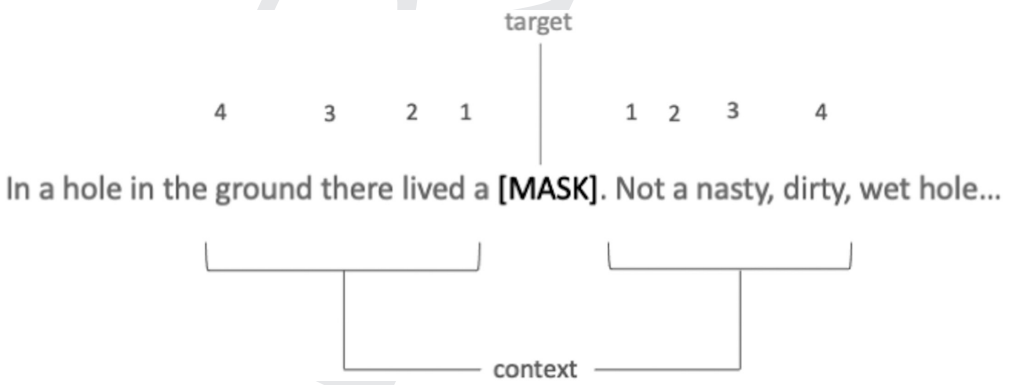

Figure 25.5 Example of a context window

This figure shows a context window of size n equals 4 around a masked word (*hobbit*) in a sentence. Keep in mind, though, that a real training example consists only of the words within the context window, and that training examples are kept within sentence boundaries, such as full stops in English. See e-book for a full-color version of this figure.

semantic similarity and semantic relatedness in a principled manner, but also because it points towards continuity of the semantic space rather than strict polarity, in which semantic similarity appears to be a constrained or "distilled" form of semantic relatedness.

The precise reason for this strong correlation of size of the context window with the type of structural (and semantic) information encoded in the vector spaces remains unknown. It could be partly explained by the fact that small windows best capture the immediate syntactic function of a word in a certain linguistic context, whilst larger windows tend to associate words by topic (Erk, 2016).

However, even if distributional models were able to overcome the noise produced by the long-range syntactic dependencies commonly found in language, this explanation induces a gratuitous linguistic bias into the hypothesis, which does not allow to extrapolate the structural approach beyond word meaning. After all, semantic knowledge in humans is grounded on much more than just syntactic, combinatorial, or distributional schemes (Andrews et al., 2009; Carota et al., 2017; Pulvermüller, 2018; Vigliocco et al., 2004, 2009). So, the fact that vector spaces are able to model semantic knowledge only from a structural proxy of language demands deeper analysis and suggests that meaning may actually emerge from structure itself, while the strong correlation of window size with type of structural relation suggests that detecting nested structures requires a variety of information processing windows.

Self, brain, and environment are predispositions of semantics

Together, the semantic vector space seems to provide a continuum of semantic relatedness and similarity (as distinguished from their polarity or dualism). Yet, where and how is such continuum coming from? What is the context for such continuum? This leads us back to the self and its respective environmental context.

The self provides a perspectival context for semantic relatedness and similarity: the continuum of semantic relatedness and similarity is viewed from and within a particular perspective, the perspective of the self as a particular person. That introduces differences in how the continuum of semantic relatedness and similarity is viewed: depending on the angle or perspective, the same two words, like *cup* and *coffee*, can be viewed to stand in different degrees of semantic closeness or distance in their semantic relatedness.

The self thus provides a perspectival context for the continuum of semantic relatedness and similarity. Yet where is the context itself, independent of its perspectival nature, coming from? This leads us to the environment – the cultural, ecological, and social context. We already argued that the coffee and tea lovers may view the degree of semantic relatedness of cup, tea, and coffee depending on their respective cultural context. Hence, it is ultimately the relation of the self to its respective environmental context that determines the perspectival context as a setting for the continuum of semantic relatedness and similarity.

The semantic structure (i.e., the continuum of semantic relatedness and similarity) is in itself deeply embedded and anchored within the structural relation of self and environment. That structural relation of self–environment can, by itself, be traced to the relation of the brain and environment as the brain provides the basis for the self. The brain–environment relation as a basis for the self–environment relation is determined in spatiotemporal terms: the timescales of the world/environment relate to the timescales of the brain and ultimately the self. The self's timescales, in turn, may provide the temporal windows that map semantic relatedness and similarity on a temporal continuum of short and long windows.

In sum, we can now see how the vector- and window-based temporal determination of the continuum of semantic relatedness and similarity is related to the self and its own brain-based intrinsic temporal relationship with the environmental context. Like the self and the brain by themselves

(Northoff et al., 2020), semantic relatedness and similarity are intrinsically linked to the timescales of the environment as mediated through self and brain. It is because of such primarily temporal nature of our semantics that we can ascertain meaning (i.e., semantics) to words in terms of their similarity and relatedness.

Losing such context (i.e., our sense of self or the environment) means losing the temporal structure of semantics. This is what we mean when we say that the intrinsically temporal nature of self, brain, and environment provides a necessary condition or predisposition for semantics featured by the long and short windows of semantic relatedness and similarity.

From the hierarchy of self to the temporal continuum of semantic similarity and relatedness

We have so far treated the self as homogenous in the context of semantics. Yet that is not supported by the empirical data described in the first section. We differentiated self-related, self-predictive, and self-referential processes, each associated with different functions (intero-exteroception, cognitive, and higher-order-reflective dimensions, respectively). This raises the question of how these layers of self are related to the temporal continuum of semantic relatedness and similarity.

The three layers of self can be distinguished on spatial and temporal grounds. Following the meta-analysis by Qin et al. (2020), the lowest layer, interoceptive and self-related processing, is spatially restricted to subcortical regions and insula as cortical region; this means the spatial extension of the neural processing of interoceptive inputs is rather restricted to small set of regions. The same applies temporally as the timescales are rather short here – that is short-time windows as reflected in the short ACW of these regions.

What does the spatial and temporal restriction of the lowest layer of self imply for its semantic processing? We saw that semantic similarity is related mainly to short spatial, and thereby temporal, windows in computational models. Short-time windows allow for temporal precision, which is exactly what is needed when determining semantic similarity: two different words need to be precisely compared and mapped onto each other, which requires the temporal precision provided by short-time windows. Albeit tentatively, we can see that short-time windows of the intero-exteroceptive layer of self-related processing may be ideally suited to allow for semantic similarity.

How about semantic relatedness? This requires longer-time windows. Those longer-time windows, as well as more spatial extension, are more prevalent at the opposite end of the hierarchy of self, the mental self with its self-referential processing recruiting longer timescales and large parts of cortical regions, the DMN. Given such especially temporal expansion, we assume that the longer timescales of the mental self and its self-referential processing are ideally suited to mediate semantic relatedness as that, per vector space, require longer-time windows.

Together, albeit tentatively, we assume that the brain's temporal (and spatial) hierarchy of self is related to the structural relation of semantic relatedness and similarity. Lower layers of the hierarchy of self with this more bodily-based self are more suited to process semantic similarity, while an increasing degree of semantic relatedness may be taken into view in the higher layers of the hierarchy of self, the mental self.

Importantly, what mediates the relationship of self and semantics is the temporal scale (and spatial ranges) with the transition of the self's short and long timescales providing the brain-based temporal continuum of semantic similarity and relatedness. The different durations of the timescales on the three layers of self allow for integration of different signs by means of which distinct meanings are constructed. We can thus see how semantics, which focuses on meaning, converges here with neurosemiotics, the neural basis of signs, through the timescales of our brain-based self (Figure 25.6).

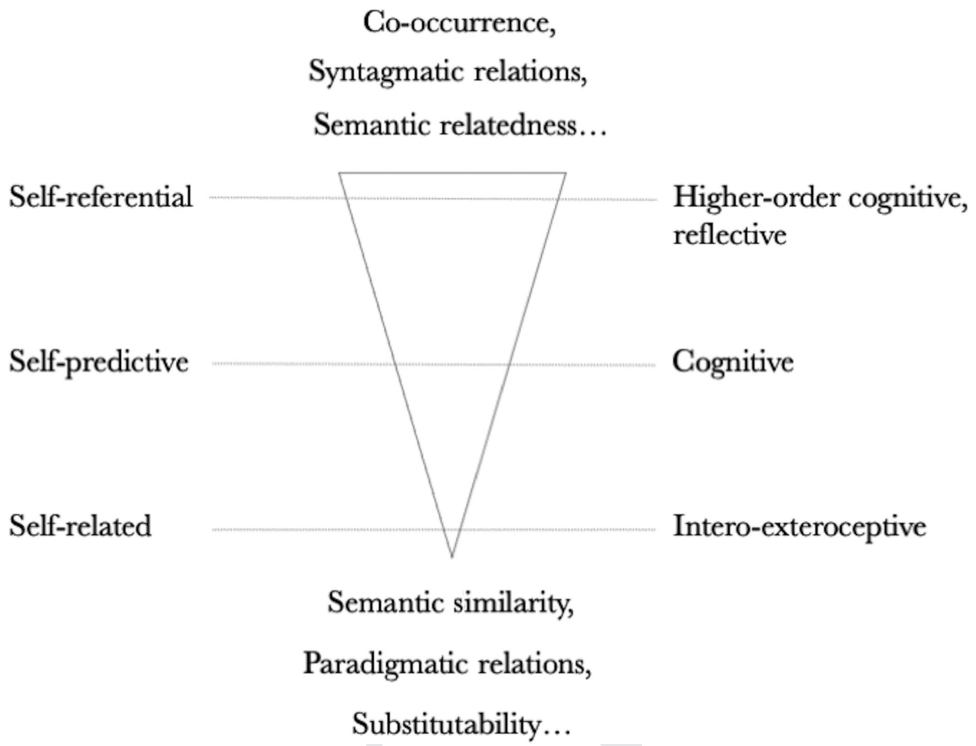

Figure 25.6 Layers of self

Layers of self (left and right) are related to the temporal continuum of semantic similarity and relatedness (middle cone). The conical shape recalls the radial perception in Figure 25.2. Semantically, it represents the fact that as the context window increases, the perception of semantic relatedness also increases. In relation to the self, it represents the fact that longer timescales mediate the more mental self, which is semantically related to semantic relatedness rather than semantic similarity.

Conclusion

We are confronted with an enormous psychological, neural, and conceptual diversity of self in current neuroscience and psychology. We here extend the previously proposed "Basis Model of Self-specificity" (Northoff, 2016) in a hierarchical way by assuming different layers of self-specific processing including self-related, self-predictive, and self-referential processing. Considering recent results in semantics, we postulate that the hierarchy of self is related to a more or less corresponding hierarchy of semantics along the lines of semantic similarity and relatedness.

Importantly, what mediates the self-semantic relation, as we postulate, are temporal features: lower layers of self are mediated by the brain's short timescales, which are ideal for yielding semantic similarity, while higher layers of self exhibit longer timescales that process semantic relatedness. Accordingly, we assume convergence of the brain-based self with semantics: through the active deployment of the self's neural timescales, different linguistic inputs are integrated into different meaningful wholes. What connects brain, self, and semantics/semiotics are, thus, their timescales –a potential "common currency" between them (Northoff et al., 2020).

Notes

1 https://wordsinasentence.com/aardvark-in-a-sentence
2 http://vectors.nlpl.eu/explore/embeddings/en

References

Agirre, E., Alfonseca, E., Hall, K., Kravalová, J., Pasca, M., & Soroa, A. (2009). A study on similarity and relatedness using distributional and wordnet-based approaches. In M. Ostendorf, M. Collins, S. Narayanan, D. W. Oard, & L. Vanderwende (Eds.), *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 19–27). Boulder, CO: Association for Computational Linguistics.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review, 116*(3), 463–498. doi: 10.1037/a0016261

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral signals shape brain dynamics and cognition. *Trends in Cognitive* Sciences, 23(6), 488–509. doi: 10.1016/j.tics.2019.03.007

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In K. Toutanova, & H. Wu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long papers)* (pp. 238–247). Baltimore, MD: Association for Computational Linguistics.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, *88*(1), 145–166. doi: 10.1016/j.neuron.2015.09.029

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

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In A. McCallum, & S. Roweis (Eds.), *Proceedings of the 25th international conference on Machine learning* (pp. 160–167). New York: Association for Computing Machinery.

Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. doi: 10.1038/nrn2555

D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van der Linden, M., Feyers, D., Maquet, P., & Salmon, E. (2010). The neural basis of personal goal processing when envisioning future events. *Journal of Cognitive Neuroscience*, *22*(8), 1701–1713. doi: 10.1162/jocn.2009.21314.

Damasio, A. (2010). *Y el cerebro creó al hombre*. Madrid: Destino.

de Greck, M., Rotte, M., Paus, R., Moritz, D., Thiemann, R., Proesch, U., Bruer, U., Moerth, S., Tempelmann, C., Bogerts, B., & Northoff, G. (2008). Is our self based on reward? Self-relatedness recruits neural activity in the reward system. *NeuroImage*, *39*(4), 2066–2075. doi: 10.1016/j.neuroimage.2007.11.006

de Pasquale, F., Corbetta, M., Betti, V., & Della Penna, S. (2018). Cortical cores in network dynamics. *NeuroImage*, *180*(Part B), 370–382. doi: 10.1016/j.neuroimage.2017.09.063

de Saussure, F. (1922). *Cours de linguistique generale* (C. Bally, A. Sechelaye, & A. Riedlinger, Eds., W. Baskin, Trans.). New York: McGraw-Hill.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 arXiv*, https://aclanthology.org/N19-1423

Enzi, B., de Greck, M., Prösch, U., Tempelmann, C., Northoff, G. (2009). Is our self nothing but reward? Neuronal overlap and distinction between reward and personal relevance and its relation to human personality. *PLoS One*, *4*(12), e8429. doi: 10.1371/journal.pone.0008429

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, *6*(10), 635–653. doi: 10.1002/lnco.362

Erk, K. (2016). What do you know about an alligator when you know the company it keeps? *Semantics and Pragmatics*, *9*, 1–63. doi: 10.3765/sp.9.17

Frassinelli, D., Naumann, D., Utt, J., & m Walde, S. S. (2017). Contextual characteristics of concrete and abstract words. In C. Gardent, & C. Retoré (Eds.), *IWCS 2017 – 12th International Conference on Computational Semantics – Short papers* (Paper 10). Stroudsburg: Association for Computational Linguistics.

Frewen, P., Schroeter, M. L., Riva, G., Cipresso, P., Fairfield, B., Padulo, C., Kemp, A. H., Palaniyappan, L., Owolabi, M., Kusi-Mensah, K., Polyakova, M., Fehertoi, N., D'Andrea, W., Lowe, L., & Northoff, G. (2020). Neuroimaging the consciousness of self: Review, and conceptual-methodological framework. *Neuroscience & Biobehavioral Reviews*, *112*, 164–212. doi: 10.1016/j.neubiorev.2020.01.023

Frings, C., & Wentura, D. (2014). Self-prioritization processes in action and perception. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1737–1740. doi: 10.1037/a0037376

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. doi: 10.1038/nrn2787

Friston, K. (2018). Am I self-conscious? (Or does self-organization entail self-consciousness?). *Frontiers in Psychology*, April 24. doi: 10.3389/fpsyg.2018.00579

Golesorkhi, M., Gomez-Pilar, J., Zilio, F., Berberian, N., Wolff, A., Yagoub, M. C. E., & Northoff, G. (2021). The brain and its time: Intrinsic neural timescales are key for input processing. *Communications Biology*, *470*, 970. doi: 10.1038/s42003-021-02483-6.

Han, S., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience*, *9*, 646–654. doi: 10.1038/nrn2456

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162. doi: 10.1080/00437956.1954.11659520

Harnad, S. (2003). Categorical perception. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (pp. 169–177). London: Nature.

Hu, C., Di, X., Eickhoff, S. B., Zhang, M., Peng, K., Guo, H., & Sui, J. (2016). Distinct and common aspects of physical and psychological self-representation in the brain: A meta-analysis of self-bias in facial and self-referential judgements. *Neuroscience & Biobehavioral Reviews*, *61*, 197–207. doi: 10.1016/j.neubiorev.2015.12.003

Huang, Z., Obara, N., Davis, H. H., Pokorny, J., & Northoff, G. (2016). The temporal structure of resting-state brain activity in the medial prefrontal cortex predicts self-consciousness. *Neuropsychologia*, *82*, 161–170. doi: 10.1016/j.neuropsychologia.2016.01.025

Huntenburg, J. M., Bazin, P. L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cognitive Sciences*, *22*(1), 21–31. doi: 10.1016/j.tics.2017.11.002

Ježek, E. (2016). *The Lexicon: An Introduction*. Oxford: Oxford University Press.

Kiela, D., & Clark, S. (2014). A systematic study of semantic vector space model parameters. In A. Allauzen, R. Bernardi, E. Grefenstette, H. Larochelle, C. Manning, & S. Wen-tau Yi (Eds.), *Proceedings of the 2nd Workshop on Continuous Vector Space Models and Their Compositionality (CVSC)* (pp. 21–30). Gothenburg: Association for Computational Linguistics.

Kiela, D., Hill, F., & Clark, S. (2015). Specializing word embeddings for similarity or relatedness. In Ll. Màrquez, C. Callison-Burch, & J. Su (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2044–2048). Lisbon: Association for Computational Linguistics.

Kolvoort, I. R., Wainio-Theberge, S., Wolff, A., & Northoff, G. (2020). Temporal integration as "common currency" of brain and self-scale-free activity in resting-state EEG correlates with temporal delay effects on self-relatedness. *Human Brain Mapping*, *41*(15), 4355–4374. doi: 10.1002/hbm.25129

Kutuzov, A., Fares, M., Oepen, S., & Velldal, E. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In M. Jonsson (Ed.), *Proceedings of the 58th Conference on Simulation and Modelling* (pp. 271–276). Reykjavik: Linköping University Electronic Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi: 10.1037/0033-295X.104.2.211

Lapesa, G., & Evert, S. (2014). A large scale evaluation of distributional semantic models: Parameters, interactions and model selection. *Transactions of the Association for Computational Linguistics*, *2*, 531–546. doi: 10.1162/tacl_a_00201

Lapesa, G., & Evert, S. (2017). Large-scale evaluation of dependency-based DSMs: Are they worth the effort? In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th conference of the European chapter of the Association for computational linguistics: Volume 2, short papers* (pp. 394–400). Valencia: Association for Computational Linguistics.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171. doi: 10.1146/annurev-linguistics-030514-125254

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. *Advances in Neural Information Processing Systems*, *27*, 2177–2185.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225. doi: 10.1162/tacl_a_00134

MacLean, P. (1990). *The Triune Brain in Evolution*. New York: Plenum Press.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12574–12579. doi: 10.1073/pnas.1608282113

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The Adolescent Brain: Learning, Reasoning, and Decision Making* (pp. 39–66). Washington, DC: American Psychological Association.

Merck, C., Noël, A., Jamet, E., Robert, M., Hou, C., Salmon, A., Belliard, S., & Kalénine, S. (2019). Identification of taxonomic and thematic relationships: Do the two semantic systems have the same status in semantic dementia. *Journal of Clinical and Experimental Neuropsychology*, *41*(9), 946–964. doi: 10.1080/13803395.2019.1641186

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506. doi: 10.1016/j.tics.2011.08.003

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013–1052. doi: 10.1093/brain/121.6.1013

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv.* doi: https://doi.org/10.48550/arXiv.1310.4546

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28. doi: 10.1080/01690969108406936

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. doi: 10.1037/bul0000092

Morel, N., Villain, N., Rauchs, G., Gaubert, M., Piolino, P., Landeau, B., Mézenge, F., Desgranges, B., Eustache, F., & Chételat, G. (2014). Brain activity and functional coupling changes associated with self-reference effect during both encoding and retrieval. *PLoS One*, *9*(3), e90488.

Moutoussid, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Fiston, K. (2014). Bayesian inferences about the self (and others): A review. *Conscious Cognition*, 25(100), 67–76. doi: 10.1016/j.concog.2014.01.00910.1371/journal.pone.0090488

Nakao, T., Miyagi, M., Hiramoto, R., Wolff, A., Gomez-Pilar, J., Miyatani, M., & Northoff, G. (2019). From neuronal to psychological noise: Long-range temporal correlations in EEG intrinsic activity reduce noise in internally-guided decision making. *NeuroImage*, *201*, 116015. doi: 10.1016/j.neuroimage.2019.116015

Northoff, G. (2007). Psychopathology and pathophysiology of the self in depression – neuropsychiatric hypothesis. *Journal of Affective Disorders*, *104*(1–3), 1–14. doi: 10.1016/j.jad.2007.02.012

Northoff, G. (2009). What are the subjective processes in our brain? Empirical and ethical implications of a relational concept of the brain. *American Journal of Bioethics*, *9*(9), 27–28. doi: 10.1080/15265160903098390

Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. *Cognitive Neuroscience*, 7(1–4), 203–222. doi: 10.1080/17588928.2015.1111868

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain: A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. doi: 10.1016/j.neuroimage.2005.12.002

Northoff, G., Wainio-Theberge, S., & Evers, K. (2020). Is temporo-spatial dynamics the "common currency" of brain and mind? In Quest of "Spatiotemporal Neuroscience." *Physics of Life Reviews*, *33*, 34–54. doi: 10.1016/j.plrev.2019.05.002

Northoff, G., Wiebking, C., Feinberg, T., & Panksepp, J. (2011). The "resting-state hypothesis" of major depressive disorder: A translational subcortical-cortical framework for a system disorder. *Neuroscience & Biobehavioral Reviews*, *35*(9), 1929–1945. doi: 10.1016/j.neubiorev.2010.12.007

Panksepp, J. (1998). *Affective Neuroscience*. Oxford: Oxford University Press.

Peirsman, Y. (2008). Word space models of semantic similarity and relatedness. In K. Balogh (Ed.), *Proceedings of the 13th ESSLLI student session* (pp. 143–152). Hamburg: Freie und Hansestadt Hamburg.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Doha: Association for Computational Linguistics.

Philippi, C. L., & Koenigs, M. (2014). The neuropsychology of self-reflection in psychiatric illness. *Journal of Psychiatric Research*, *54*, 55–63. doi: 0.1016/j.jpsychires.2014.03.004

Pulvermüller, F. (2018). The case of CAUSE: neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170129. doi: 10.1098/rstb.2017.0129

Qin, P., Grimm, S., Duncan, N. W., Fan, Y., Huang, Z., Lane, T., Weng, X., Bajbouj, M., & Northoff, G. (2016). Spontaneous activity in default-mode network predicts ascription of self-relatedness to stimuli. *Social Cognitive and Affective Neuroscience*, *11*(4), 693–702. doi: 10.1093/scan/nsw008

Qin, P., Wang, M., & Northoff, G. (2020). Linking bodily, environmental and mental states in the self-A three-level model based on a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *115*, 77–95. doi: 10.1016/j.neubiorev.2020.05.004

Rapp, R. (2002). The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *COLING 2002: The 19th International Conference on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Disability Studies*, *20*, 33–54.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*(4), 677–694. doi: 10.1016/j.neuron.2012.11.001

Schulte Im Walde, S. (2020). Distinguishing between paradigmatic semantic relations across word classes: human ratings and distributional similarity. *Journal of Language Modelling*, *8*(1), 53–101. doi: 10.15398/jlm.v8i1.199

Schütze, H., & Pedersen, J. (1993). A vector model for syntagmatic and paradigmatic relatedness. In *Proceedings of the 9th annual conference of the UW Centre for the new OED and text research* (pp. 104–113). Oxford: Oxford University Press.

Schwartz, M., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., Mirman, D., & Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Science of the United States of America*, *108*(20), 8520–8524. doi: 10.1073/pnas.1014935108

Sui, J., He, X., Humphreys G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1105–1117. doi: 10.1037/a0029792

Sui, J., & Humphreys, G. W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in Cognitive Sciences*, *19*(12), 719–728. doi: 10.1016/j.tics.2015.08.015

Sui, J., Rotshtein, P., & Humphreys, G. W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7607–7612. doi: 10.1073/pnas.1221862110

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology*, *70*(4), 597–609. doi: 10.1080/17470218.2016.1181768

Tsakiris, M., & Critchley, H. (2016). Interoception beyond homeostasis: Affect, cognition and mental health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160002. doi: 10.1098/rstb.2016.0002

Turney, P. D. (2012). Domain and function: A dual-space model of semantic relations and compositions. *Journal of Artificial Intelligence Research*, *44*, 533–585. doi: 10.1613/jair.3640

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. doi: 10.1613/jair.2934

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: A theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience & Biobehavioral Reviews*, *34*(6), 935–946. doi: 10.1016/j.neubiorev.2009.12.004

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

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*(4), 422–488. doi: 10.1016/j.cogpsych.2003.09.001

Wittgenstein, L. (1953). *Investigaciones filosóficas (Philosophische Untersuchungen)*. London: Kegan Paul.

Yankouskaya, A., Bührle, R., Lugt, E., Stolte, M., & Sui, J. (2020). Intertwining personal and reward relevance: Evidence from the drift-diffusion model. *Psychological Research*, *84*, 32–50. doi: 10.1007/s00426-018-0979-6