SISSA
Scuola Internazionale Superiore di Studi Avanzati

Neuroscience Area – PhD course in
Cognitive Neuroscience

Semantic processing with and without awareness.
Insights from computational linguistics and semantic priming

Candidate:
Andrea Nadalini

Advisor:
Prof. Davide Crepaldi

Academic Year 2018-19

Summary

Introduction

Semantic cognition

Semantics–from the ancient Greek σημαντικός–is the branch of linguistics concerned with meaning. Yet, in contemporary neuroscience, semantics rather refers to the cognitive and brain processes due to which we know what the different entities in the world are, and how to interact with them. Crucially, semantic knowledge gives meaning to language, making otherwise meaningless auditory and visual objects appropriate vehicles for a successful communication with our conspecifics.

An adult human brain has a wealth of information about the concepts of the world we live in; this knowledge is acquired progressively throughout life, and it is stored within the semantic memory. A wide variety of human behaviors relies on this conceptual knowledge, such as the recognition and use of objects, the ability to apprehend abstract concepts, to name them and eventually to share them with others. We cannot reason, remember the past or imagine the future without having access to it. All human cultures–whether scientific, literary, religious, artistic–are built around a foundation of conceptual knowledge of this kind. However, despite being involved in almost all human activities, its neurobiological bases are far from being fully understood.

The current thesis aims at exploring the cognitive and brain mechanisms that allow for meaning extraction from a specific type of stimuli: words. This choice reflects the main interest that has driven my PhD, i.e. to understand how lexical–semantic knowledge is organized and accessed, rather than object recognition *per se*. Despite the processes of access to meaning for words and other visual stimuli partially overlap (Shinkareva, Malave, Mason, Mitchell, & Just, 2011; Simanova, Hagoort, Oostenveld, & Van Gerven, 2014), there is also evidence that the two semantic routes are not identical. Several neuroimaging studies reported different patterns of activation elicited by carefully matched words and pictures (Devereux, Clarke, Marouchos, & Tyler, 2013; Gates & Yoon, 2005; Price et al., 2006). Moreover, there are patients who show severe object recognition impairments in spite of a relatively spared word comprehension (Davidoff & De Bleser, 1994; Humphreys & Rumiati, 1998), further suggesting the specificity of semantic access via words.

Theories of semantic cognition

Classic view

Traditionally, semantic memory was thought of as a modular and a–modal system where long–term representations of concepts are stored (Tulving, 1972). Modularity points to a functionally specialized cognitive system, which is different from other memory structures such as episodic memory, which refers instead to the memory of events that took place at a specific time and place. A–modality refers to the independence of the semantic information associated with a given concept from the sensory modality through which it was originally perceived. For example, when reading the word *orange*, we activate its conceptual representation which includes information regarding its shape, color and taste, yet this information is dissociated from the sensory systems used to actually see and taste it.

While Tulving's theoretical framework for semantic memory surely represented the foundation for the scientific study of semantic representations, later research clearly challenged this classic view. Advances in neuroimaging techniques and computational modelling (Jones, Willits, & Dennis, 2015; Martin & Chao, 2001) made it possible to better understand the nature of semantic memory as a part of an integrated structure which is widely distributed across the brain and connected to sensory, perceptual, and motor systems.

Embodied view

Behavioral and neuroimaging experiments have shown that access to word meaning implies to activate sensorimotor information associated with perceiving and interacting with the real–world entities words refer to. That is, unimodal sensory regions – including the visual, auditory and sensorimotor cortex – play an active role in the processing of lexical meaning (Binder & Desai, 2011; Glenberg & Gallese, 2012; Kiefer & Pulvermüller, 2012). For example, comprehending words related to movement, color, sound or emotion activates cortical regions involved in the processing of these specific types of information: lower temporal (motion), fusiform gyrus (color), superior temporal (sound), temporal pole and ventromedial prefrontal cortex (emotion). Similarly, deficits in the comprehension of action verbs have been reported for patients suffering from neurological syndromes that affect motor skills, such as Parkinson's disease (Boulenger et al., 2008) or amyotrophic lateral sclerosis (Grossman et al., 2008). These results have licensed the embodied semantics theory; under the more radical interpretations, this theory posits that understanding concrete words corresponds to activate the sensory–motor representations acquired when making experience with the corresponding referents (Barsalou, 2008). The same process holds for abstract words, whose meanings are constructed as metaphoric extensions from sensory–motor experience (e.g. love is a journey; happiness is up, sadness is down; Lakoff & Johnson, 1980).

Although the activation of sensorimotor information during language understanding is uncontroversial, the question of the causal relation between the two has been the focus of a long–lasting debate. Advocates of strong embodiment have suggested that this activation is not an epiphenomenon, but an essential mechanism of meaning construction, being mandatory (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005), automatic (Ansorge, Kiefer, Khalid, Grassl, & König, 2010; Dudschig, de la Vega, De Filippis, & Kaup, 2014), and attested already at early stages of semantic processing (Boulenger et al., 2006; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008).

Such experiments are surely elegant and highlight intriguing phenomena; yet, whether they truly imply a causal connection between sensorimotor information and meaning, it is far from clear. Upon closer look, other explanations that do not require strong embodiment claims can be licensed. Mahon and Caramazza (2008), for example, pointed out replication issues in the literature, and suggested that most of this evidence could be explained by a disembodied view of cognition that more carefully takes into consideration the dynamics of activation flow between cognitive and brain systems. Other studies directly questioned radical views of embodied semantics. Bottini, Bucur and Crepaldi (2016) found no evidence that words could automatically trigger sensorimotor information outside of awareness, although these words were clearly processed up to the semantic level. Similarly, Miller, Brookie, Wales, Wallace and Kaup (2018) conducted a series of EEG experiments in which participants made hand or foot responses to verbs referring to either hand or foot movements (e.g. punch, kick). While different ERPs were elicited by the specific motor actions required by the task, no such difference was attested for the semantic processing of the hand- versus foot-related target words. These results clearly challenged claims whereby access to the meaning of action verbs would mandatorily recruit motor areas activated when performing the corresponding action. Neuropsychological evidence is also intermixed; for example, there are cases of apraxic patients who were able to name and recognize words referring to objects they could not interact with (Mahon & Caramazza, 2005).

To conclude, a cautious examination of the literature on embodied theory seems to dismiss its more radical versions, and tell us that sensory–motor information is clearly involved in the construction of lexical meaning, but plays a rather secondary and supportive role.

Symbolic view

Semantic representations are thus built upon lifelong verbal and non–verbal experience, and recruit several sensory, linguistic, motor and affective processing systems, which are widely distributed across the brain. Crucially, all the information coming from modality–specific areas eventually converges in regions that act as semantic hubs, and allow perceptual experiences to reach an abstract level of representation (Binder, Desai, Graves, & Conant, 2009; Damasio, 1989). This process seems to capture and incorporate two aspects of word meaning that were heavily studied in

Experimental Psychology and Cognitive Neuroscience: a taxonomic system responsible for assigning categories to lexical meanings, and a thematic system that link them based on frequent co–occurrence of the corresponding referents in events or scenarios (Mirman, Landrigan, & Britt, 2017).

Early studies of semantic memory postulated the existence of taxonomic networks where concepts are stored and connected via parent–child hierarchies (Collins & Quillian, 1969). For example, the node ANIMAL would branch into subordinate nodes REPTILES, BIRDS, MAMMALS, etc., which in turn would branch into their subordinate nodes (e.g. RODENTS, PRIMATES, FELINES, etc.,), and so on. Each node is defined by a set of features (e.g. ANIMAL: breathes, eats, mates, etc.) that are inherited by all the elements at lower levels in the tree. Crucially, this model predicts the existence of a distance effect, so that the farther information is stored in the hierarchy, the longer the processing time; for example, it would be easier to confirm that "dog is a mammal" (one node) than "dog is an animal" (two nodes). While early evidence seemed to confirm such effect, later studies challenged it (Chang, 1986). More recently, taxonomic relationships have been described on the bases of a set of binary features that point to perceptual, functional and encyclopedic aspects of the corresponding entity. This approach relies on the collection of data from human raters in property generation tasks and word meaning, which can be eventually represented by a vector keeping track of such features (Dilkina & Lambon Ralph, 2012; McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008). Featural models does not conceive any distance effect, as they are not hierarchical. Rather, their core predictions stem from the distinctiveness and the overlap of the features associated to the entities, accounting for semantic similarity over and beyond categorical membership. For example, they can explain why an eagle is more similar to a hawk than to a penguin, while all being birds. Yet, they are not perfectly suited to represent the semantic content of abstract entities, whose describing features can be quite difficult to define and seem to be rather situation- and context-specific.

Thematic relationships, instead, reflect association due to contiguity between concepts, which can be represented as nodes within network models (Collins & Loftus, 1975). In these networks, activation would spread from one node to the other, with activation strength proportional to their association. This latter has been typically quantified by asking many subjects to list words brought to mind by a target word. (Nelson, McEvoy, & Schreiber, 2004). This approach has been widely used in psycholinguistic research, particularly to study the dynamics of lexical–semantic access. Yet, association norms represent quite a fuzzy psychological construct; they are not clearly defined and encompass a wide range of rather different types of relationships. For example, category membership (rifle-gun[1]), collocation (macaroni-cheese), synonymy (sofa-couch), meronymy (hammer-tool), antonymy (day-night), scripts (school-student), function (bed-sleep), even proper names of notorious entity (president-Bush).

Experimental studies of word meaning: the priming paradigm

No matter which specific theory one embraces, each entry in the semantic memory can thus activate a more or less extensive network of knowledge, which is influenced not only by the percept itself (bottom-up processing), but also by information already stored in the brain via previous experience with the stimulus (top-down processing). For example, reading the word *mouse* does not necessarily imply only the activation of a specific piece of encyclopedic and sensory information ("a small rodent that typically has a pointed snout, relatively large ears and eyes, and a long tail"), but possibly a much larger field of knowledge, partly variable from one individual to another. This field of knowledge includes the representations of related entities like cat, cheese, the Speedy Gonzales cartoons, the yard of your grandmother's country house, etc. (Figure 1).

[1] All the examples are taken from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 2004)

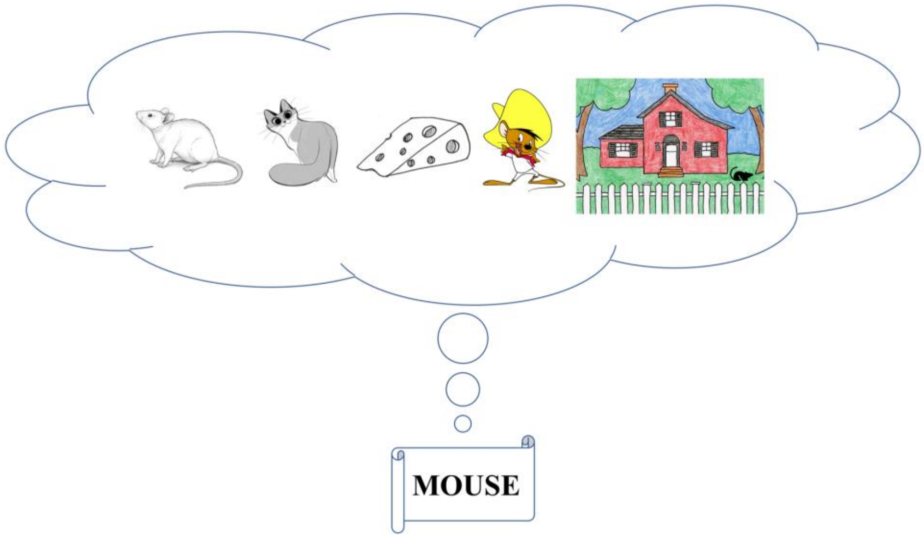

Figure 1. Access to the semantic content of a word (e.g., mouse) can "activate" related conceptual representations (e.g., mouse, cat, cheese, Speedy Gonzales, etc.)

Thus, the activation of a representation stored in semantic memory generally overflows on concepts that are close to it. This phenomenon, called semantic priming, probes access to the meaning of a word by measuring the facilitation it exerts on a neighboring representation (McNamara, 2005). So if you ask a subject to perform a task requiring semantic processing of the word *mouse* – such as, for example, saying whether it is a natural or artificial entity – it will be faster and more accurate if *mouse* had been preceded by the related word *cat*, than if it had been preceded by an unrelated word such as *ship*. This facilitation occurs also when participants are involved in non-semantic task, such as lexical decision and naming; thus, priming seems to be driven by fundamental memory recruitment processes (McNamara, 1992).

Semantic representations are highly complex and multidimensional, and different aspects of word meaning follow different time courses of activation. Thus, a critical factor modulating the emergence and the magnitude of priming is represented by the stimulus onset asynchrony (SOA), i.e., the time passed from the presentation of the prime to the presentation of the target. Longer SOAs are likely to allow for secondary and more effortful aspects of lexical meaning to be processed. For example, Lam, Dijkstra and Rueschemeyer (2015) reported priming for words referring to objects that are manipulated in a similar way (e.g., paper plane-DART) already at a SOA of 100ms, while priming based on visual similarity (e.g., syringe-DART) showed up only at a SOA of 1000ms.

Thus, priming experiments have been fundamental for the study of lexical semantic processing processing; however, the numerous studies using this paradigm that were carried out since the beginning of the 1970s have brought partly contradicting results. Most of the controversy relates to the specific contribution brought by taxonomic (feature-based) vs. thematic (association-based) relationships as described above. Previous studies provided conflicting results, leaving the issue still open and highly debated. For example, Lucas (2000) stated that "pure" feature-based similarity – i.e., in the absence of word association–produces priming, while he found no evidence supporting the opposite claim. Conversely, Hutchison (2003) concluded that both feature overlap and associative relatedness leads to a significant facilitation of related targets. One possibility is that it may not be fruitful to dichotomously differentiate between associative and featural similarity, given that highly associated items in norm production tend to share some form of semantic relationship as well (Brainerd, Yang, Reyna, Howe, & Mills, 2008; Guida & Lenci, 2007). Rather, this distinction points to the extremes of an underlying continuum. A theoretical approach describing meaning-based similarity in continuous terms is represented by distributional semantics.

Distributional semantics

Distributional semantics is a fully symbolic theory defining meaning activation as an a-modal process based on a set of connections linking words to each other. This approach builds upon the theoretical assumption that humans construct semantic representations of lexical items by keeping track of their distribution in language use. If words get their meaning due to the linguistic context they appear in, then words occurring in similar contexts will be similar in meaning. This idea is not new, but dates back at least to the 50s, as we can see from the following quotations:

“*The meaning of a word is its use in the language*” (Ludwig Wittgenstein, *Philosophical Investigation*, 1953)

“*Each language can be described in terms of a distributional structure, i.e. in terms of the occurrence of parts (ultimately sounds) relative to other parts*” (Zellig Harris, *Distributional Structure*, 1954)

“*You shall know a word by the company it keeps*” (John Rupert Firth, *A synopsis of linguistic theory 1930-1955*, 1957)

Nowadays, distributional semantics represents a mainstream research paradigm in Computer Science and Cognitive Neuroscience, mostly due to the great advancements in the development of techniques capable of providing human-like estimates of meaning-based similarity between words. All these procedures are strictly linked to the development of linguistic corpora, large database of text documents made up of billions of words (these models need to be trained on large amounts of material). By looking at their distribution, it is then possible to reveal recurrent patterns that could be eventually used as a proxy to represent lexical meaning, and therefore to account for semantic similarity. One of the major advantages of this approach is that words themselves represent the building blocks of semantic representations, ruling out the weakness of postulating a-priori which “features” constitute the basis for theoretical models of semantics. Moreover, similarity estimates can be automatically obtained for potentially all words attested in a given corpus, while feature-lists and association norms are available only for a limited set of stimuli and require time and resources to recruit participants.

The most immediate way to model semantic relatedness according to word distribution is by looking at surface cooccurrence, based on the assumption that two words that exhibit a tendency to appear near to each other in natural language are likely to be associated in meaning. Typically, co–occurrence is computed within a window comprising from 3 to 5 words, but it may vary according to the specific experimental question being asked. Some studies have been interested into immediately adjacent words, also called bigrams (Pecina, 2010), while others have taken into consideration much wider windows (Vechtomova, Robertson, & Jones, 2003). Moreover, punctuations and function words – those words that convey only little meaning and primarily carry out a syntactic function – are normally excluded before collecting frequency counts, in order to face the data sparsity issue and increase the signal-to-noise ratio. The same reasoning holds for lemmatization, which reconducts all the inflected forms (e.g., speak, speaks, spoke, spoken) to the same abstract representation (e.g., speak).

Mere recurrence is not enough to indicate strong attraction between lexical items, as word pairs may be highly attested due to the individual frequency of the single component. Thus, it is common practice to apply some mathematical transformation to the raw count of co–occurrence. For example, it is possible to estimate joint and conditional probabilities, run statistical tests of independence, compute likelihood and information-based measures (a systematic review can be found in Evert, 2007). Here, we will focus on pointwise mutual information (PMI) between two words, which can be computed via the formula:

$$\mathrm{PMI}(w_1,w_2) = \log_2 \frac{p(w_1,w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ corresponds to the probability of the word pair, while $p(w_1)$ and $p(w_2)$ to the individual probabilities of the two components (Church & Hanks, 1989). PMI expresses how a given word can be used as a proxy for expecting another word, and thus can be rightfully considered as an index of local associative relationship. The metrics found successful applications in psycholinguistic research; for example, it could account for similarity judgements (Recchia & Jones, 2009), reading speed (Ellis, Simpson-Vlach, & Maynard, 2008), and free association and syntactic parsing (Pitler, Louis, & Nenkova, 2010).

More complex methods are based on word embeddings, a set of computational methods that involve the training of distributional semantic models (DSMs) where lexical items are mapped to numerical vectors. Similarity between words is indexed by spatial proximity in the semantic space, and it can then be measured via linear algebra operations, for example, by computing the cosine of the angle formed by two word-vectors:

$$\cos\theta = \frac{a \cdot b}{||a|| \cdot ||b||}$$

Early approaches built word vectors from co-occurrence matrices that kept track of word distribution in a given corpus. These matrices could differ regarding the type of linguistic context taken into consideration. Some models, such as the Hyperspace Analogue to Language (HAL; Lund & Burgess, 1996), relied on word-by-word matrices constructing distributional profiles for words based on which other words surrounded them, via a sliding context window that was normally advanced one word at a time along the entire corpus. Others, such as Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), constructed word-by-documents matrices by counting how many times words appear in broader linguistic contexts like paragraphs or entire text documents. After collecting frequency counts, raw vectors underwent some transformation allowing the model to achieve a better performance. This optimization process could imply reweighting the counts for context informativeness and smoothing them with dimensionality reduction techniques.

More recent models, instead, have tackled vector construction as a supervised task, by implementing neural network architectures that assign weights to the vectors in order to maximize model performance. In particular, the state-of-art model (word2vec; Mikolov, Chen, Corrado, & Dean, 2013) represents a simple neural network consisting of an input, an output and a hidden layer, and is based on a predictive mechanism that allows to infer a target given a cue. There are two different learning architectures that can be implemented: in continuous-bag-of-words (CBOW), a given word is predicted on the basis of the surrounding words, while in skip-gram, the surrounding words are predicted on the basis of a given word. In both cases, learning is performed by adjusting at each training step the weights of the connections between the nodes of the network, based on the difference between the outcome (the target) predicted on the basis of a cue (the context) by the network, and the correct one.

Word2vec – and prediction-based models in general – have been proposed as a psychologically plausible model of learning, such as the Rescorla-Wagner model of classical conditioning (Günther, Rinaldi, & Marelli, 2019; Mandera, Keuleers, & Brysbaert, 2017). Model estimates cover a wide range of classic lexical-semantic relationships, such as synonymy (e.g. king-monarch[2], 0.51) , antonymy (e.g. life-death, 0.42), meronymy (e.g. engine-car, .49). Associative relations as well can be grasped (monkey-banana, .41). Finally, it can account for featural similarity beyond category membership (e.g. shark-dolphin, .46 vs shark-tuna, .24). Experiment evidence has shown that word2vec has been shown to performed better than (or as well as) other DSMs in a variety of task, such as synonym detection, concept categorization, semantic priming (Baroni, Dinu, & Kruszewski, 2014; Mandera et al., 2017; Marelli, 2017).

Despite many DSMs involve the collection of cooccurrence data to construct distributed representations, there is a crucial difference between the two metrics. Spatial proximity in the semantic space reflects overlap in the contexts of use between words that may never cooccur directly. Two synonyms like *car* and *automobile* are not likely to appear in the same sentence; still, they point to the same entity, and are therefore expected to be used with pretty much the same words. Conversely, the fact that two words appear very often close to each other stems from the effective co-presence of the corresponding referents as we experience them in our everyday experience. For example, the words *glove* and *oven* are not strongly related in the semantic memory, but are likely to go together in language due to the fact that every time you need to take out a baking pan from the

[2] All the examples are taken from the CBOW model developed by Mandera, Keuleers, & Brysbaert (2017)

oven, you need a glove for not getting burnt. The two approaches/metrics, therefore, specifically code for different aspects of word associations, even if these different aspects typically correlate.

Conscious and unconscious cognition

Generations of scientists and philosophers have struggled with the uncertainty about how to define consciousness. Traditionally, the conscious state has been defined as a psychological state characterized by a subjective awareness of an experience; thus, a mental representation is described as conscious if and only if it is reportable – "I am aware of seeing this stimulus". The use of this criterion of reportability has been critical in the experimental work aimed at determining the cognitive and brain mechanisms underlying conscious access. Indeed, certain mental representations do not reach the consciousness and are therefore described as unconscious or subliminal.

Over the last decades, the neuroscientific study of consciousness has made significant progress by combining contributions from experimental psychology, functional brain imaging and computational modelling (Dehaene, Charles, King, & Marti, 2014). Due to such improvements, it is now possible to explore the unconscious counterpart of many high-level cognitive functions – such as memory, emotions, executive control, mathematics, language – whose exploration was most often conducted in conscious healthy subjects.

Yet, how to characterize the differences between conscious and unconscious processing is still highly debated. Support for and against a qualitative difference between the two is present in the literature, and such empirical diversity resulted in a rather polarized distinction between firm supporters or deeply skeptics. According to former group, every fundamental high-level function can be carried out by the unconscious mind pretty much as the conscious one does (Hassin, 2013). This position is backed by experimental evidence showing unconscious completion of complex tasks like arithmetic (Karpinski, Briggs, & Yale, 2019), goal setting (Hassin, Bargh, & Zimerman, 2009), sound-symbolism mapping (Hung, Styles, & Hsieh, 2017), syntactic processing (Berkovitch & Dehaene, 2019) or sentence meaning construction (Sklar et al., 2012). Similarly, it has been claimed that working memory includes cognitive processes of which participants are not aware (Logie, 2016).

However, some of these results have failed replication attempts (Mongelli, Meijs, van Gaal, & Hagoort, 2019; Moors & Hesselmann, 2019; Nakamura et al., 2018). These results question the strength of previous claims and rather suggest that conscious and unconscious processing may be qualitatively different. More precisely, it may be possible that the amount of information that can be extracted and processed from a subliminally presented stimulus is reduced and more segregated relative to the conscious counterpart.

This would be in line with the global workspace model (Baars, 2005; Dehaene & Changeux, 2011; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). In this model, unconscious processing is segregated in several modular brain networks. An information represented locally in one of these processors would only access consciousness if it is enhanced by attentive top-down amplification and then spreads, via long–distance connections throughout the cortex, to form a coherent state of activity at the global level in the brain. Such long-distance connectivity allows, at least when it is sufficiently persistent, to make information accessible to high-level processes such as categorization, long-term memorization, emotional evaluation and voluntary manipulation. This global availability of information through this global neuronal workspace would correspond exactly to what we experience in the form of perceptual awareness.

Unconscious semantic processing: the masked priming paradigm

Access to word meaning outside of awareness is generally accepted (Kouider & Dehaene, 2007). Most of the evidence came from masked priming studies in which the prime word is presented very briefly – 50 ms or less – and is embedded between a random sequence of uppercase characters (e.g. XYGDF) and the target word. This procedure, called backward masking, prevents conscious access to the prime, which will still facilitate the processing of a semantically related target.

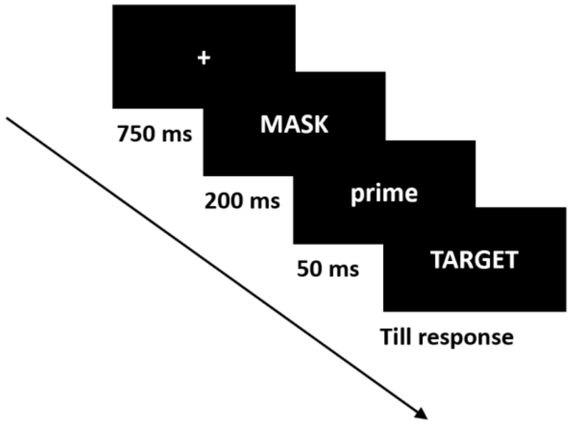

Figure 2. Exemplar trial in a masked priming experiment

The first evidence showing the existence of a subliminal priming effect came in the early 80s. Marcel (1983) found facilitation for related words (child-infant) independently of prime visibility. In another work, the same author followed up results on polysemous word (palm) from Schvaneveldt, Meyer and Becker (1976), who had showed how only one semantic representation at the time could be accessed when the word was processed consciously. Yet, Marcel reported that when the polysemous word was masked, both meanings were activated, suggesting that semantic representations could be richer and independent of executive control in the absence of conscious perception (Marcel, 1980).

These exciting results, however, were widely criticized for their statistical weakness, lack of reproducibility, and also for the dubious effectiveness of the visual masking used, which relied only on the participants subjective report (Holender, 1986; Purcell, Stewart, & Stanovich, 1983). One approach that has been used widely to address this methodological concern is trying the participants with a detection task on the prime itself. Performance is then typically quantified via the Signal Detection Theory sensitivity measure d', which makes possible to assess an objective threshold of conscious perception, now essential in any experiment using subliminal stimuli. Usually at the end of the experiment, participants are asked to perform a forced-choice task directly related to the hidden word, for example a lexical decision task. Results are then analyzed in terms of "hits" and "false alarms", thus making it possible to calculate a detection index, the *d'*. By correlating the priming effect with this index of visibility, it is possible to estimate the priming effect when primes were fully masked, that is, at *d'* equals to 0.

Other criticisms to subliminal semantic priming were raised because of possible stimulus-response mapping mechanisms that could account for the effect. This type of implicit association explained the effect in terms of direct activation by the prime of the response action required by the target, ruling out the semantic processing of the masked stimulus. Stimulus-response associations are likely to emerge when masked prime words are also presented in target position as well. Abrams and Greenwald (2000) neatly showed the non-semantic nature of this mechanism. In their experiment, target words had to be categorized according to their emotional valence, as positive or negative. After having repeatedly categorized *smut* and *bile* as negative words, participants provided faster responses to unpleasant targets when primed with the subliminal word *smile*, which was made up by fragments of the previously seen target words. Similarly, facilitation to pleasant responses was induced by the masked prime *tumor* when *tulip* and *humor* had been previously presented as target words. Indeed, such bias can be easily overcome by ensuring that hidden primes are never presented as visible targets.

All these criticisms allowed for the development of new and stronger paradigms that made the existence of truly subliminal semantic priming no longer a matter of debate (Van den Bussche, Van den Noortgate, & Reynvoet, 2009).

The mechanisms behind masked priming

Traditionally, priming was accounted for via spreading activation mechanisms, both within localist frameworks, where activation spreads among concepts (Neely & Kahan, 2001), and within connectionist frameworks, where activation spreads among features (Plaut, 1995). Crucially, this process has been described as automatic and not liable to strategic control by the reader.

However, later studies challenged this view and suggested that access to word meaning without awareness is not automatic; rather, it is prone to top-down influences. More precisely, subliminal semantic priming has been found to depend on the availability of attentional resources. For example, the effect was drastically reduced if, prior to the onset of the prime, participants were engaged in a perceptual task requiring high allocation of attentional resources relative to a task requiring low allocation of attentional resources (Martens & Kiefer, 2009). Similarly, task settings have been shown to moderate the emergence of subliminal priming. While the effect is strongly attested in task tapping semantic properties of the stimuli, it is instead much more fleeting in lexical decision or naming task, where word meaning is de–emphasized (De Wit & Kinoshita, 2015).

These findings have licensed another interpretation: subliminal priming would origin from processes that maximize the uptake of goal-oriented information, via the collecting evidence that is relevant to optimally perform the task. Because of the close contiguity between the prime and the target, evidence is accumulated from both the stimuli, which are effectively confounded (Kinoshita & Norris, 2010). When related prime–target pairs provide converging evidence to accomplish the task, the prime gives a head start to the accumulation process and thus makes the decision to the target easier.

However, the specific information contributing to the such evidence accumulation process has not been fully understood yet. As outlined above, lexical-semantic representations cannot be uniquely defined, as words can be similar under many different aspects. For example, *cat* may prime *dog* due to feature overlap (e.g., they are both furry, have four legs, are kept as pets by humans; Quinn & Kinoshita, 2008), or due to category membership (animals; Abrams, Klinger, & Greenwald, 2002), or due to associative strength (similarly to how *kangaroo* is associated to *Australia*; Anaki & Henik, 2003). All these different aspects of lexical meaning are reflected in words distribution, despite at different levels, from surface cooccurrence to latent language structure. Crucially, while meaning can be processed in all its multidimensional complexity when words are conveyed above the threshold for conscious perception, the unconscious reader may have only a partial access to some specific dimensions. This is exactly the question that has driven my PhD, and that I have tried to address with the experiments that are gathered in this thesis.

Experimental contributions

In the first chapter of this thesis, I tested the idea that conscious and unconscious priming is different in depth of processing. While unconscious semantic representations are built from symbolic information only, conscious representations reflect the contribution of symbolic and situated, extra-linguistic knowledge. Teasing apart these different aspects of word meaning is obviously very difficult, since they overlap in the vast majority of the cases. A very convenient exception to this rule is provided by the mapping between space and time, which can happen along both a vertical and a lateral axis, but only the former is encoded in language use (e.g., "the future is *ahead* of you", not "to your *right*"). We took advantage of this particular feature of the space-time mapping, and tested metaphorical congruity priming along both axes, with primes presented wither masked or visible.

In the second chapter, I tested subliminal and supraliminal priming by modelling semantic similarity as a continuous variable. To better define the symbolic information that is encoded in language, I collected distributional information for a set of prime-target pairs both at the local and at the distributed level, by looking at lexical cooccurrence (PMI; e.g., rubber-penknife) and spatial proximity in a semantic space (cosine similarity; e.g., sofa-hammock) respectively. The two metrics were compared in their capability to predict priming across a series of experiments manipulating prime duration and prime visibility.

In the third chapter, I looked at the electrophysiological correlates of conscious semantic priming, testing the specific contribution of local (PMI) and distributed (cosine similarity) linguistic information to the brain signature of semantic facilitation. More precisely, I recorded EEG signal from participants performing a primed lexical decision, and test for the emergence of the N400 component in word pairs that could be highly co-occurrent but far in the semantic space (e.g., car-tank), or, symmetrically, neighbors in the semantic space but poorly associated locally (e.g., cell-cage).

In the fourth chapter, I tried to explore subliminal priming in a situation where unawareness was not induced by some visual masking technique, but it was rather a stable trait of individuals who have suffered a psychological and/or neurological trauma. Thus, I tested neglect patients, a clinical population that lack attentional resources to consciously report stimuli presented in the affected hemifield (typically, the left one). However, these neglected stimuli are not simply ignored, but they activate cognitive representations that seem to exert an influence upon high-level cognitive processes. In this study I tested semantic priming in lexical decision task using the same set of stimuli as in the previous chapter. Prime visibility was manipulated by presenting the stimuli either on the left (neglected) side of the screen, or on the right one, where they were clearly visible.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science*. https://doi.org/10.1111/1467-9280.00226

Abrams, R. L., Klinger, M. R., & Greenwald, A. G. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196262

Anaki, D., & Henik, A. (2003). Is there a "strength effect" in automatic semantic priming? *Memory and Cognition*. https://doi.org/10.3758/BF03194385

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., & König, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition*. https://doi.org/10.1016/j.cognition.2010.05.010

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progress in Brain Research*. https://doi.org/10.1016/S0079-6123(05)50004-9

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.59.103006.093639

Berkovitch, L., & Dehaene, S. (2019). Subliminal syntactic priming. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2018.12.001

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhp055

Bottini, R., Bucur, M., & Crepaldi, D. (2016). The nature of semantic priming by subliminal spatial words: Embodied or disembodied? *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000197

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2006.18.10.1607

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/PBR.15.6.1035

Chang, T. M. (1986). Semantic Memory. Facts and Models. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.99.2.199

Church, K. W., & Hanks, P. (1989). *Word association norms, mutual information, and lexicography*. https://doi.org/10.3115/981623.981633

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal*

Learning and Verbal Behavior. https://doi.org/10.1016/S0022-5371(69)80069-1

Damasio, A. R. (1989). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation*. https://doi.org/10.1162/neco.1989.1.1.123

Davidoff, J., & De Bleser, R. (1994). Impaired picture recognition with preserved object naming and reading. *Brain and Cognition*. https://doi.org/10.1006/brcg.1994.1001

De Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000074

Dehaene, S., & Changeux, J. P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2006.03.007

Dehaene, S., Charles, L., King, J. R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2013.12.005

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3809-13.2013

Dilkina, K., & Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2012.00333

Dudschig, C., de la Vega, I., De Filippis, M., & Kaup, B. (2014). Language and vertical space: On the automaticity of language action interconnections. *Cortex*. https://doi.org/10.1016/j.cortex.2014.06.003

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Evert, S. (2007). Corpora and collocations. *Corpus Linguistics. An International Handbook*. https://doi.org/10.1515/9783110213881.2.1212

Gates, L., & Yoon, M. G. (2005). Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: An fMRI study. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.08.020

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*. https://doi.org/10.1016/j.cortex.2011.04.010

Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*. https://doi.org/10.1212/01.wnl.0000319701.50168.8c

Guida, A., & Lenci, A. (2007). Semantic properties of word associations to Italian verbs. *Italian Journal of Linguistics*.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619861372

Hassin, R. R. (2013). Yes It Can: On the Functional Abilities of the Human Unconscious. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612460684

Hassin, R. R., Bargh, J. A., & Zimerman, S. (2009). Automatic and flexible: The case of nonconscious goal pursuit. *Social Cognition*. https://doi.org/10.1521/soco.2009.27.1.20

Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2008.20123

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X00021269

Humphreys, G. W., & Rumiati, R. I. (1998). Agnosia without prosopagnosia or Alexia: Evidence for stored visual memories specific to objects. *Cognitive Neuropsychology*. https://doi.org/10.1080/026432998381177

Hung, S. M., Styles, S. J., & Hsieh, P. J. (2017). Can a Word Sound Like a Shape Before You Have Seen It? Sound-Shape Mapping Prior to Conscious Awareness. *Psychological Science*. https://doi.org/10.1177/0956797616677313

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196544

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of Semantic Memory. *Oxford Handbook of Mathematical and Computational Psychology*.

Karpinski, A., Briggs, J. C., & Yale, M. (2019). A direct replication: Unconscious arithmetic processing. *European Journal of Social Psychology*. https://doi.org/10.1002/ejsp.2390

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*.

https://doi.org/10.1016/j.cortex.2011.04.006

Kinoshita, S., & Norris, D. (2010). Masked priming effect reflects evidence accumulated by the prime. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470210902957174

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2007.2093

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*. https://doi.org/10.1016/S0364-0213(80)80017-6

Lam, K. J. Y., Dijkstra, T., & Rueschemeyer, S.-A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00659

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*. https://doi.org/10.1037/0033-295X.104.2.211

Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2015.1136657

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03212999

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03204766

Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643290442000446

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*. https://doi.org/10.1016/j.jphysparis.2008.03.004

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2016.04.001

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*. https://doi.org/10.1016/0010-0285(83)90009-9

Marelli, M. (2017). Word-embeddings Italian Semantic spaces: A semantic model for psycholinguistic research. *Psihologija*. https://doi.org/10.2298/PSI161208011M

Martens, U., & Kiefer, M. (2009). Specifying attentional top-down influences on subsequent unconscious semantic processing. *Advances in Cognitive Psychology*. https://doi.org/10.2478/v10053-008-0067-3

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*. https://doi.org/10.1016/S0959-4388(00)00196-3

McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review*. https://doi.org/10.1037/0033-295X.99.4.650

McNamara, T. P. (2005). Semantic priming: Perspectives from memory and word recognition. In

Semantic Priming: Perspectives from Memory and Word Recognition. https://doi.org/10.4324/9780203338001

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*. https://doi.org/10.3758/BF03192726

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013. *ArXiv Preprint ArXiv:1310.4546*.

Miller, J., Brookie, K., Wales, S., Wallace, S., & Kaup, B. (2018). Embodied cognition: Is activation of the motor cortex essential for understanding action verbs? *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000451

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*. https://doi.org/10.1037/bul0000092

Mongelli, V., Meijs, E. L., van Gaal, S., & Hagoort, P. (2019). No language unification without neural feedback: How awareness affects sentence processing. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.116063

Moors, P., & Hesselmann, G. (2019). Unconscious arithmetic: Assessing the robustness of the results reported by Karpinski, Briggs, and Yale (2018). *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2019.01.003

Nakamura, K., Makuuchi, M., Oga, T., Mizuochi-Endo, T., Iwabuchi, T., Nakajima, Y., & Dehaene, S. (2018). Neural capacity limits during unconscious semantic processing. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13890

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In *The nature of remembering: Essays in honor of Robert G. Crowder*. https://doi.org/10.1037/10394-005

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03195588

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-009-9101-4

Pitler, E., Louis, A., & Nenkova, A. (2010). Automatic evaluation of linguistic quality in multi-document summarization. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the 17th Annual Conference of the Cognitive Science Society*.

Price, C. J., McCrory, E., Noppeney, U., Mechelli, A., Moore, C. J., Biggio, N., & Devlin, J. T. (2006). How reading differs from object naming at the neuronal level. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.07.044

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Purcell, D. G., Stewart, A. L., & Stanovich, K. E. (1983). Another look at semantic priming without awareness. *Perception & Psychophysics*. https://doi.org/10.3758/BF03205897

Quinn, W. M., & Kinoshita, S. (2008). Congruence effect in semantic categorization with masked primes with narrow and broad categories. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2007.03.004

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*. https://doi.org/10.3758/BRM.41.3.647

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.2.2.243

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.10.042

Simanova, I., Hagoort, P., Oostenveld, R., & Van Gerven, M. A. J. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhs324

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1211645109

Tulving, E. (1972). Tulving_1985_memory. *Organization of Memory*.

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of Masked Priming: A Meta-Analysis. *Psychological Bulletin*. https://doi.org/10.1037/a0015329

Vechtomova, O., Robertson, S., & Jones, S. (2003). Query expansion with long-span collocates. *Information Retrieval*. https://doi.org/10.1023/A:1023936321956

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*. https://doi.org/10.3758/BRM.40.1.183

Chapter 1. The limits of unconscious semantic processing as revealed by metaphorical priming

Introduction

There has been intense debate on the difference between conscious and unconscious cognition. The human mind was proven able to carry out a variety of tasks outside of awareness (goal setting, Hassin, Bargh, & Zimerman, 2009; arithmetics, Sackur et al., 2008; sentence meanign construction Sklar et al., 2012), to the point that there were suggestions that whatever we can compute consciously, we can also do outside of awareness (Hassin, 2013). However, some of these spectacular unconscious performances were proven difficult to replicate (Rabagliati, Robertson, & Carmel, 2018), and some authors argued that there are both quantitative (Kouider & Dehaene, 2007) and qualitative (Nakamura et al., 2018) differences between conscious and unconscious cognitive processing.

Word meaning is the perfect battle camp for this debate. In fact, the semantic system is highly complex and multidimensional (Borghesani & Piazza, 2017), thus offering wide room for qualitative differences between conscious and unconscious processing to emerge. Words can be semantically related in many different ways. For example, *cat* may be similar to *dog* because these animals share features (e.g., they are both furry, have four legs, are kept as pets by humans; Quinn & Kinoshita, 2008), or because the words belong to the same category (animals; Abrams, Klinger, & Greenwald, 2002), or because they are associated with each other in our experience of the world (e.g., are likely to be primary associates in word association norms; Anaki & Henik, 2003), or again, merely because the words *cat* and *dog* often co-occur with each other in written and spoken language (Brunellière, Perre, Tran, & Bonnotte, 2017). While these different facets of word meaning are obviously all available to the fully aware reader, unawareness may allow only partial access to some of them.

Several studies investigated unconscious semantic processing so far, but the evidence is unclear overall. Priming has been reported for highly associated category coordinates (e.g., table–chair) when prime words were kept unconscious (i.e., presented for a very short time and visually masked), as well as when they were fully visible (Perea & Rosa, 2002). Similarly, semantic facilitation has been observed for word pairs that were related in terms of feature overlap (e.g., goose–turkey), independently of prime visibility (Bueno & Frenck-Mestre, 2008). Conversely, other studies showed different patterns of semantic facilitation depending on whether the prime was available to conscious report. In a lexical decision task, De Wit and Kinoshita (2015) reported priming only when the prime word was fully visible. Bottini, Bucur and Crepaldi (2016) showed that subliminal semantic priming interpreted as the result of unconscious sensorimotor simulations of the words' referents (e.g., simulating an upward movement to understand the word *up*; Ansorge, Kiefer, Khalid, Grassl, & König, 2010) can also be explained by symbolic associations between response labels.

To date, it is still unclear which aspects of word meaning are gathered unconsciously, and which aspects, instead, need conscious access to be retrieved. Indeed, the vast majority of the previous studies focused on whether masked semantic priming happens at all, rather than what kind of information may foster it. This is for a good reason, of course: it is hard to dissociate different aspects of word meaning experimentally, as they (quite unsurprisingly) correlate strongly. For example, associated words (e.g., cat–dog, fork–knife) tend to share semantic features (Brainerd, Yang, Reyna, Howe, & Mills, 2008), and situated knowledge is often encoded symbolically in language use (e.g., the words *red* and *transparent*, which both refer to vision, co–occur more often than words referring to different perceptual modalities, like *red* and *loud*; e.g., Louwerse & Connell, 2011)

A notable exception to this rule, however, is provided by space–time conceptual metaphors. When people talk about time they often use spatial metaphors. In English and many other languages, the future is *ahead* and the past is *behind* (e.g., Clark, 1973). Thus, time flows along a sagittal (front–back) axis. Beyond talking about time using spatial words, it has been shown that people also *think* about temporal sequences using schematic mental representations of physical space. In an experiment using motion capture to assess people's posture, participants were more likely to lean backward when thinking about the past and forward when thinking about the future (Miles, Nind, & Macrae, 2010). Likewise, participants are faster to judge sentences about the future by moving a joystick forward and faster to judge sentences about the past by moving it backward (Ulrich et al., 2012), consistent with expressions like “looking *forward* to retirement” or “thinking *back* on one’s childhood”.

Within some of the same cultures that talk about time as flowing along a sagittal timeline, people also conceptualize time along a lateral timeline, with earlier events on the left and later events on the right. This lateral mental timeline is not encoded in any known spoken language (e.g., Monday comes *before* Tuesday, not *to the left* of Tuesday; Clark, 1973), yet participants are faster to classify words related to the past by pressing a left key and words related to the future by pressing a right key, compared to the opposite arrangement (e.g., Casasanto & Bottini, 2014a). Patients with left hemi-spatial neglect, who ignore objects on the left side of space, also neglect the “left side” of time (i.e., they show better memory performance for events associated with the future than for events associated with the past; Saj, Fuhrman, Vuilleumier, & Boroditsky, 2013). English speakers have been found to gesture according to the lateral mental timeline more often than the sagittal timeline (Casasanto & Jasmin, 2012). Thus, perhaps counterintuitively, the implicit lateral timeline may be activated even more strongly than the sagittal timeline, despite its complete absence from conventional expressions in language.

Both the sagittal and the lateral mental time lines (MTL) have clear sensorimotor origins. For instance, scanning behavior during reading and writing seems to be an important experience to learn and consolidate the horizontal (MTL). In fact, people that read from right to left (e.g., Hebrew speakers) also have a leftward MTL (Fuhrman & Boroditsky, 2010), and the MTL can be transiently reversed by a few minutes of mirror reading (Casasanto & Bottini, 2014a). On the other hand, the sagittal MTL seems to be based on our walking experience in the physical world: as people typically walk in forward direction, they also move forward through both space and time (Clark, 1973).

Accordingly, temporal processing can affect step movements along the sagittal space (Rinaldi, Locati, Parolin, Bernardi, & Girelli, 2016).

Therefore, the metaphorical relationship between space and time appears to be based on the activation of unidimensional spatial schemas that subtend the representation of both spatial and temporal relationships. This hypothesis is further corroborated by neuroimaging experiments that found overlapping activity in the posterior parietal cortex for temporal and spatial conceptual knowledge (Peer, Salomon, Goldberg, Blanke, & Arzy, 2015).

Overall then, time and space are associated along both a sagittal and a lateral timeline. Both schemas are based on sensorimotor, situated experience, but only the sagittal one also emerges in language use (e.g., looking *forward* to retirement), thus creating a further symbolic, associative tie. Taking advantage of this dissociation, we tested the hypothesis that unconscious semantic processing is limited to these symbolic ties and does not allow access to situated spatial representations which are reserved to conscious word processing.

To this aim, we devised a priming paradigm in which sagittal spatial words (*front*, *back*) and lateral spatial words (*left*, *right*) appeared as primes, and temporal words appeared as targets (e.g., *past*, *future*). Primes were presented both above and below the threshold for conscious identification; if our hypothesis is correct, priming should emerge strongly on both axes in the conscious condition, when meaning is fully accessed in all its facets, but should be stronger with sagittal primes in the unconscious condition, when processing would be mostly limited to language–encoded semantic ties.

Experiment 1

Methods

Participants. 120 students at the University of Trieste were recruited into the experiment (30 males, 90 females; mean age=24y, age range=18y-36y). All subjects were right-handed, native Italian speakers, and had normal or corrected-to-normal vision and no history of neurological disorders.

Participants gave written informed consent for participation, and received 8 Euros in exchange for their time.

Material. All stimuli were Italian words. Primes were 2 spatial words related to the lateral axis (*sinistra*, left, and *destra*, right) and 2 spatial words related to the sagittal axis (*davanti*, front and *dietro*, back). Target stimuli were 8 temporal words. Four of them refers to the past (*prima*, earlier, *ieri*, yesterday, *passato*, past, *scorso*, previous), and four refers to the future (*dopo*, later, *domani*, tomorrow, *futuro*, future, *successivo*, next). Each prime word was coupled with every target item, resulting in 32 different pairs. Each pair was presented 12 times, making up a total of 396 experimental trials.

Procedure. Participants were seated in a comfortable chair and saw the stimuli from a distance of approximately 63 cm. We used a chinrest to keep the distance from the monitor constant and secure a forward orientation. All stimuli were shown in Arial font 32, in white against a black background, displayed on a 22'' monitor with a refresh rate of 120 Hz, using MatLab Psychtoolbox (Kleiner et al., 2007). Responses were collected with an external CEDRUS RB-740 response pad. Each trial started with a fixation point (+) displayed for 750 ms. In the unmasked condition, a blank screen was shown for 200 ms, followed by the prime and by another blank screen, both lasting 50 ms. In the masked condition, where participants were not informed about the presence of the prime, the blank screens were replaced with two visual masks (10 random uppercase letters, e.g. XCBFTYUOIM). Finally, in both conditions, the target word was presented for 1500 ms, or until a response was provided (see Figure 1). Prime visibility was manipulated between subjects, i.e., half of the participants were assigned to the masked condition and half to the unmasked condition.

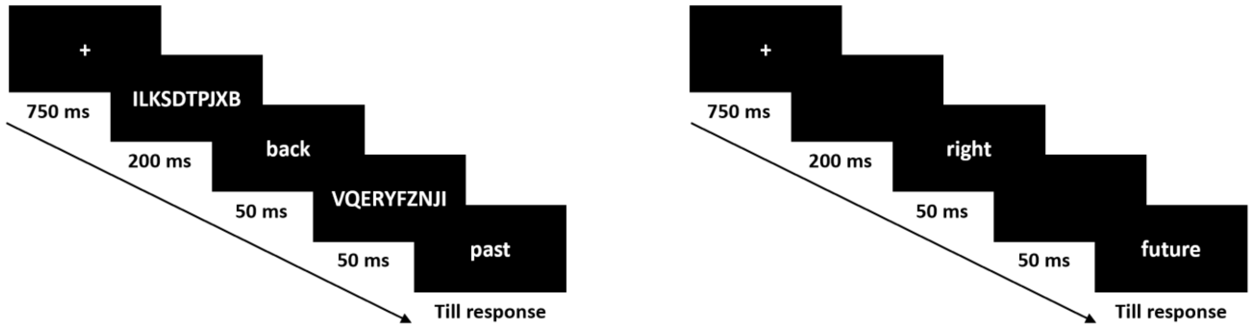

Figure 1. Trial timeline in the visible (left) and masked (right) conditions.

The 396 experimental trials which were divided in two blocks. In one block, participants were instructed to press the central button of the response box when target words were related to the past, whereas in the other they were told to press the same button when target words were related to the future—a go–no go task. The order of the two blocks (go-Past, go-Future) was counterbalanced across subjects. Twelve practice trials were presented before each block. In addition to the main break between blocks, participants took one further period of rest half way through each block.

We stress three important aspects of our design, which guarantee a fair assessment of semantic priming and overcome some limitations in the previous literature. First, the trial timeline was identical in the sub–liminal and supra–liminal conditions: as primes were presented for the same exact amount of time, we ensured that any difference would only depend on awareness, not on prime presentation time (Kanwisher, 2001). Second, target words never appeared as primes; this excluded the possibility that a priming effect could be due to (non semantic) stimulus–response associations (Damian, 2001) or action-trigger conditions (Kiesel, Hoffmann, & Kunde, 2003). Moreover, the go–no go paradigm allowed us to avoid lateralized responses, i.e., left or right button key presses; this excludes that participants' behavior was influenced by any spatial coding of the response (Bottini et al., 2016).

Prime visibility. After the main task, participants in the masked condition were informed about the presence of the prime, and were tested for their ability to perceive it consciously in a prime visibility task (Reingold & Merikle, 1988). More precisely, they were asked to assess whether the masked stimulus was a real word (vs. a string of identical lowercase letters, e.g., aaaaaaaaaa, xxxxxxxxxx).

As real words, we used the same four spatial words that we employed as primes in the main task. Participants were instructed to press either a left or a right key to provide their response. In order to make sure that participants knew where the prime was within the trial, they saw two examples where prime duration was increased to 150 ms before starting the task, so that the prime became visible even with the visual masks. The prime visibility task included 10 practice and 128 experimental trials.

Statistical analyses. Statistical analyses of the reaction times were conducted via mixed–effects linear regression, which is most appropriate when the design includes crossed random effects for both subjects and items (Baayen, Davidson, & Bates, 2008). Following the principles of the New Statistics (Cumming, 2014), we based our analyses on confidence intervals and did not rely on null–hypothesis significance testing. Models were fitted using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in the statistical software *R*. We had fixed effects for Congruity (prime–target congruent vs. prime–target incongruent), Axis (lateral vs. sagittal), Prime Visibility (masked vs. unmasked), and their interactions. We additionally included random intercepts for Subject and Target Word. We modelled the fixed effects in order to expose the parameters that are most relevant to our predictions (Meteyard & Davies, 2019), that is, (i) the contrast between congruent and incongruent primes in the sagittal, masked condition (sagittal masked priming); (ii) how much more (or less) effective are congruent primes in the lateral, masked condition, as compared with (i) (the contrast between sagittal and lateral masked priming); and (iii) how much more (or less) effective are congruent primes in the lateral, unmasked condition, as compared with (ii) (how the difference between sagittal and lateral primes changes in the unmasked, compared to the masked condition). Model–based estimated of response times in each design cell were obtained via the R package *emmeans* (Lenth, 2018).

Open practices statement. This experiment was not formally pre–registered. All data and analysis code are available at https://osf.io/wc7by/, and can be accessed independently from the authors.

Results

Overall accuracy in the experiment was 98%. The mean RT on accurate trials was 550 ms. RT analyses were conducted only on accurate go trials. One participant was excluded because of a particularly anomalous performance (mean accuracy= 88.8%, while every other participant was above 93.7%). In order to reduce the effect of extremely long and short RTs, those individual data points standing at more than 2 standard deviations from each participant's mean were also removed from the analyses. This reduced the analysis set to 21648 data points, which corresponds to a loss of ~4.5% of the potentially available dataset.

Sagittal congruent primes determined quicker RTs than incongruent primes in the masked condition, β = -9.89 [-14.63 – -5.15]. This facilitation shrank substantially with lateral primes in the masked condition, β = +8.13 [+1.43 – +14.84]. In the unmasked condition instead, congruent lateral primes were again effective, β = –10.54 [-20.02 – -1.07].

This pattern of results is represented in the model estimates illustrated in Figure 2. In the masked condition, RTs for congruent prime–target pairs were quicker than for incongruent pairs on the sagittal axis, 505 ms [489 – 522] vs. 516 ms [499 – 535], but much less so (if not at all) on the lateral axis, 510 ms [494 – 528] vs. 513 ms [496 – 531]. Supraliminally instead, facilitation was similar with sagittal, 509 ms [492 – 527] vs 516 ms [499 –534], and lateral primes, 511 ms [494 – 529] vs 517 ms [500 –535].

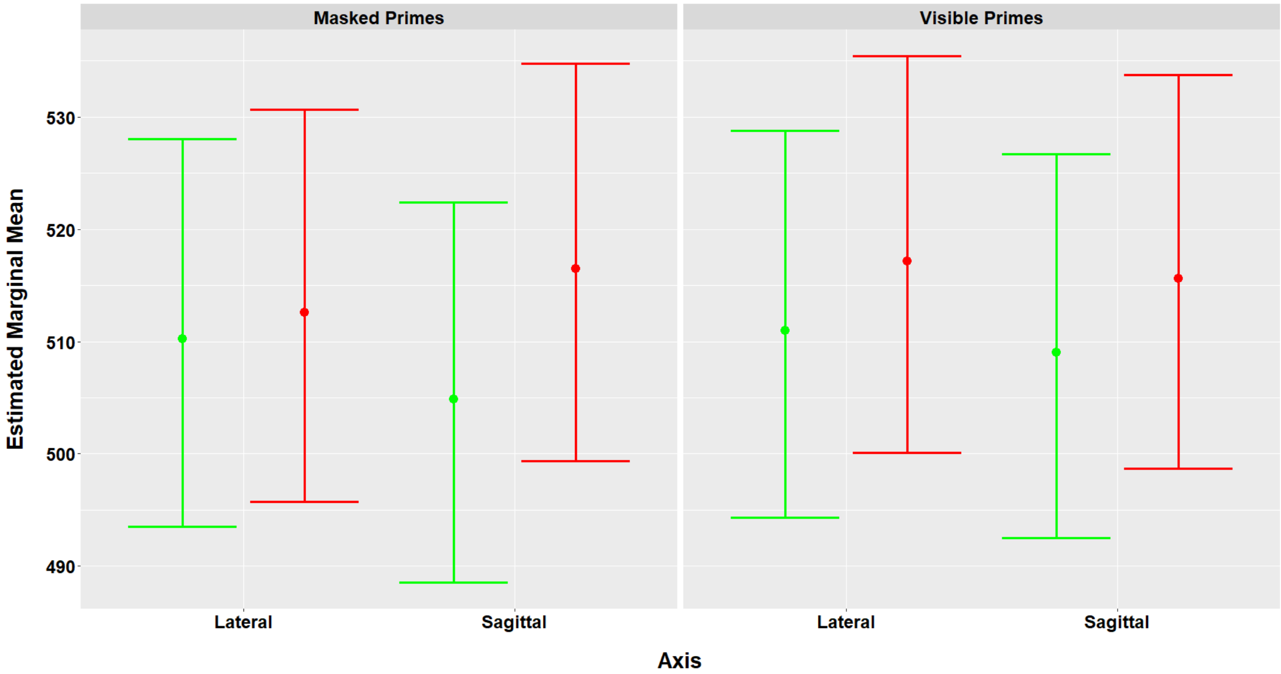

Figure 2. Estimated priming effect on the masked (left panel) and the visible condition (right panel). The congruent condition is plotted in green, and the incongruent condition in red. Error bars refer to the 95% confidence intervals.

Prime visibility task. No participant reported having noticed the prime. Data in the prime visibility task were analyzed in terms of d–prime, which is based on the ratio between correct YES response (hits) and incorrect YES responses (false alarms) for each participants. The d–prime distribution is shown in Figure 3; the average value was 0.35 [0.25 – 0.44]. These values are widely taken to indicate that primes were effectively masked from perceivers' awareness (Kouider & Dupoux, 2005).

In order to conclusively exclude that prime visibility was an important driver of the facilitation in the sagittal primes condition, we further analyzed the data by regressing the amount of priming against d–prime values (Greenwald, Klinger, & Schuh, 1995). With this linear model, we can estimate facilitation when the d–prime is zero, that is, when prime visibility is null. As illustrated in Figure 3, the 95% CI at the intercept lies entirely above the origin, indicating that priming is indeed estimated to be higher than zero even when primes are completely outside of awareness. According to the model prediction, we would observe a sagittal priming effect of 10 ms [5 – 15] when the d-prime is zero. Finally, the individual d–prime values did not correlate with the size of the sagittal masked priming effect, r= 0.022 [-0.24 – +0.28], further confirming that facilitation does not depend on prime visibility.

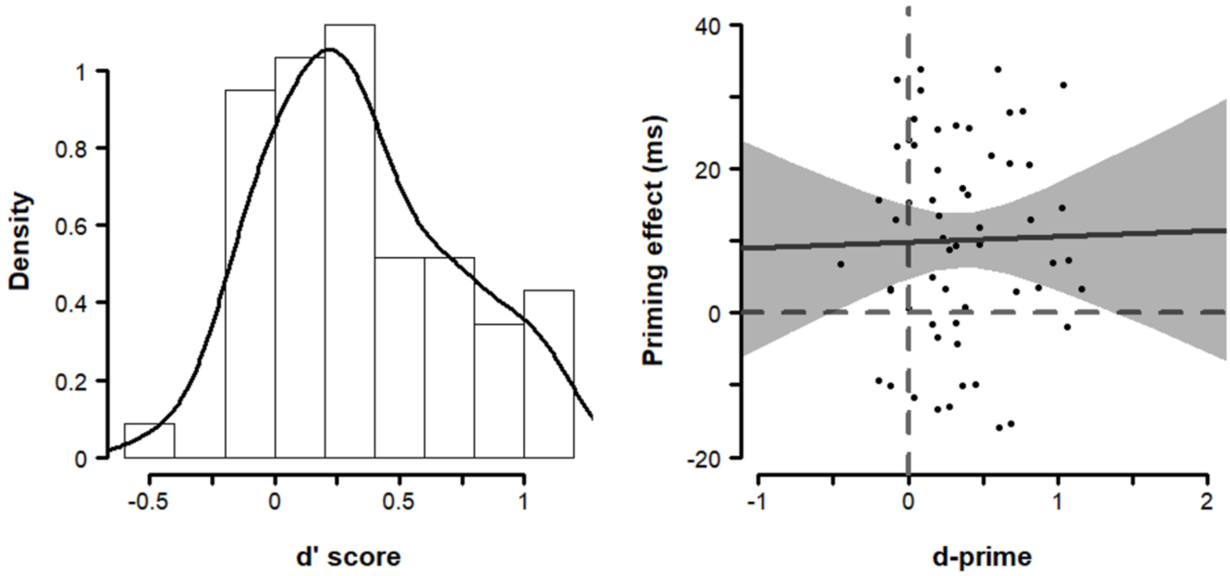

Figure 3. Density plot representing the distribution of the participants' d-prime in the prime visibility task (left panel). Relationship between priming and prime visibility (right panel). Points represent individual participants, and the shaded area indicates the 95% confidence interval of the regression line. Note that priming is measured by subtracting mean RTs on congruent trials from mean RTs on incongruent trials, that is, positive values indicate facilitation.

Discussion

Supraliminal primes generate significant congruity effects on both the sagittal and the lateral axis, consistent with previous studies that provide evidence for sagittal and lateral mental timelines (Casasanto & Bottini, 2014b; Clark, 1973). Furthermore, priming does not differ across axes. The pattern of results is clearly different with masked primes, which yield substantial priming only on the sagittal axis; facilitation on the lateral axis is very small, and substantially smaller than with sagittal primes. These findings comply with the hypothesis under scrutiny—subliminal priming shows little or no sensitivity to semantic ties that are not represented in language use.

Moreover, any role for some residual visibility of the masked primes was ruled out here, in four ways: (i) none of the participants reported noticing any of the masked primes; (ii) the d–prime analysis indicated that primes were effectively kept outside of participants' awareness, consistent with previous work on unconscious word processing (e.g., Kouider & Dupoux, 2005); (iii) the correlation analysis between prime visibility and the size of the facilitation effect showed no relationship between the two and estimated priming to be significantly above zero when d–prime is zero (i.e., there is no prime visibility whatsoever); (iv) it is unclear why residual visibility would selectively affect lateral, but not sagittal primes.

In order to ensure that these results are solid, and in the light of the recent challenges to reproducibility in Experimental Psychology (Open Science Collobaration, 2015), we carried out a replication study. In this replication, we also improved the design by varying prime visibility within subjects, that is, all participants took part both in the masked and unmasked conditions, thus reducing spurious variance in the comparison between sub–liminal and supra–liminal priming due to individual variability.

Replication Experiment

Method

Participants. 56 students at the University of Trieste were recruited into the experiment (18 males, 38 females; mean age=23y, age range=19y-30y). None of them took part in Experiment 1. All subjects were right-handed, and they all stated being native Italian speakers, with normal or corrected-to-normal vision and no history of neurological disorders. Subjects gave written informed consent for participation, and received 15 Euros in exchange for their time.

Material, Procedure and Analyses were the same as in Experiment 1, with the only difference that the same participants took up both the masked and unmasked tasks, that is, we adopted a within-subject design for prime visibility too. This required splitting the experimental sessions in two blocks. In the first block, participants underwent the masked priming and prime visibility tasks, while in the second, which took place 3 to 5 days later, they concluded the study with the visible priming condition.

Results

The overall accuracy in the experiment was 98%. The mean RT on accurate trials was 539ms. Both metrics are very similar to the previous experiment. Data trimming led to the exclusion of ~5% of the total observations, resulting in 20374 datapoints available for the mixed–effects linear regression; again, these figures are very similar to the original experiment.

Linear mixed models reveal again that congruent trials yielded faster RTs than incongruent trials in the masked, sagittal condition (β = -7.69 [-12.63 – -2.74]). With lateral primes, again in the masked condition, this facilitation was reduced (β = +4.42 [-2.57 – +11.41]). Although both parameters shrink towards zero as compared to the previous experiment (see Figure 4), they seem to confirm the original pattern. The highest–level parameter, which tracks the difference between masked and overt priming, varies more substantially as compared to the previous study, and is now close to zero (β = -0.10 [-9.98 – +9.78]).

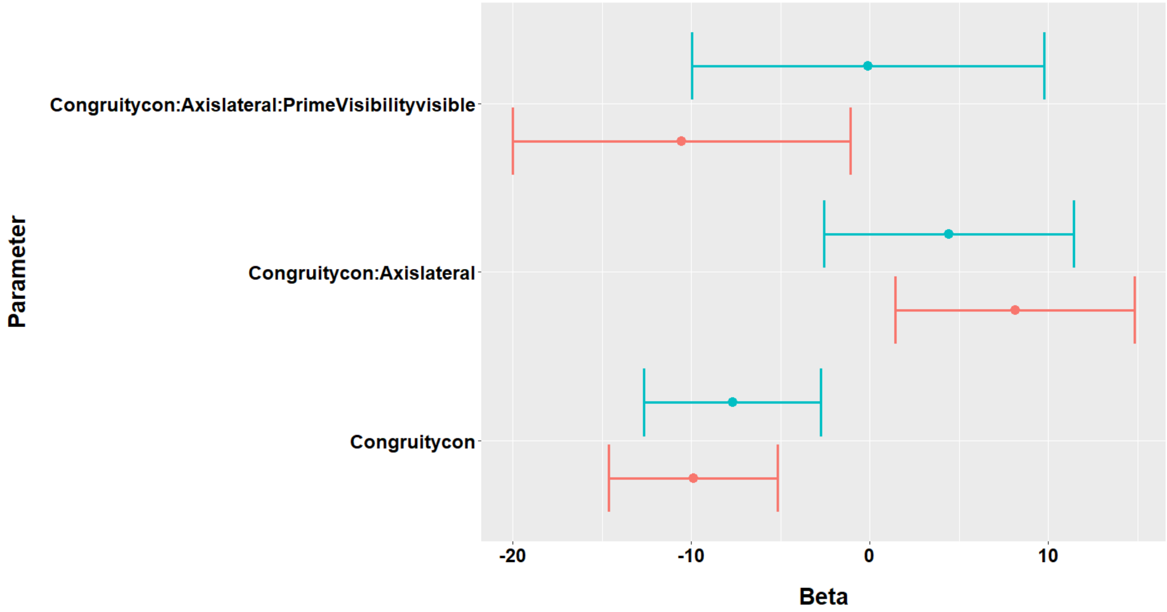

Figure 4. Model betas for the parameters of interest in the analysis. Values from the original experiment are shown in red, and values from the replication experiment are shown in blue. Error bars refer to the 95% confidence intervals.

Model estimates of the RTs per condition are represented in Figure 5. Overall, the pattern is very similar to the original experiment (see Figure 2 for comparison) and show stronger priming for sagittal than lateral primes in the masked condition, and similar facilitation on the two axes with visible primes[3].

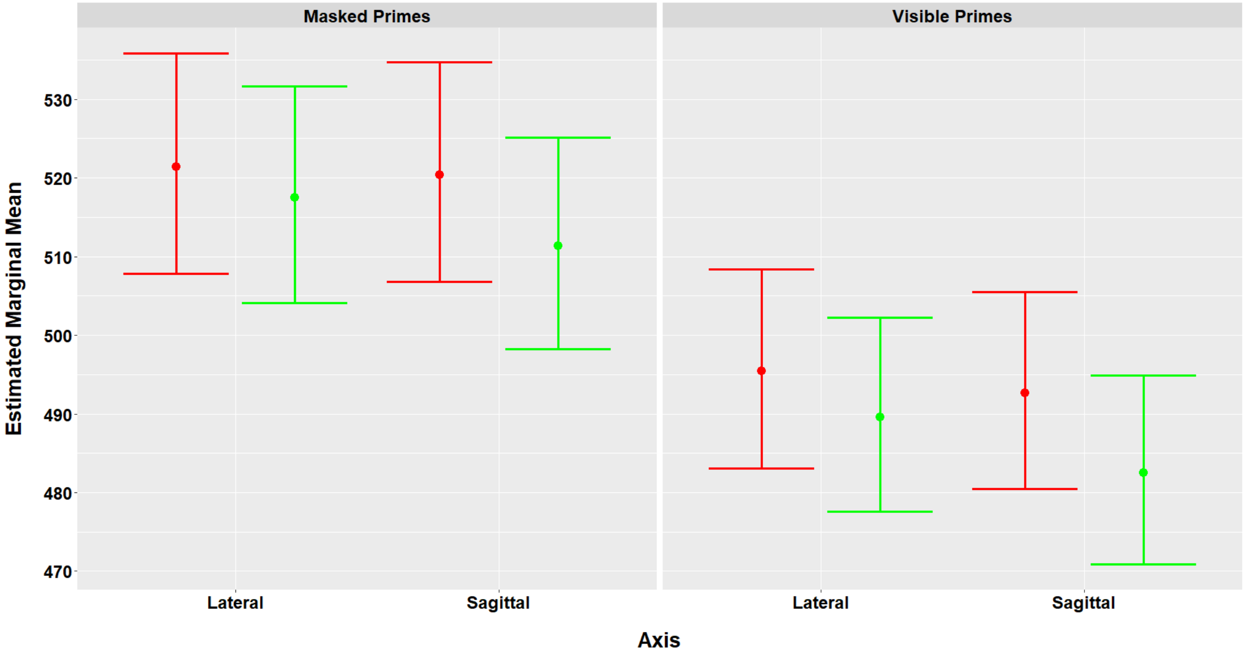

Figure 5. Estimated priming effect on the masked (left panel) and the visible condition (right panel). Congruent condition is plotted in green, and incongruent condition in red. Error bars refer to the 95% confidence interval.

Prime visibility task. As in the original study, no participant reported having noticed the primes. The d–prime distribution is shown in Figure 6; the average value is 0.39 [0.29 – 0.49], very similarly to the original experiment. The correlation between d–prime and amount of priming turned out to be slightly stronger in this experiment than in the original one, r= 0.191 [-0.08 – +0.43] (see Figure 6). The estimated priming when the d–prime is null is still a rather substantial 5 ms [-1 – 11], suggesting again the presence of sagittal masked priming outside of awareness.

[3] RTs are now generally shorter in the visible than in the masked condition. This is probably due to the within–subject design, which required participants to be tested twice on the same material. Because subjects needed to be unaware of the presence of the primes in the masked condition, the corresponding session took place first for all participants. As a result, visible primes may have benefitted from an increased familiarity with the task and the testing materials.

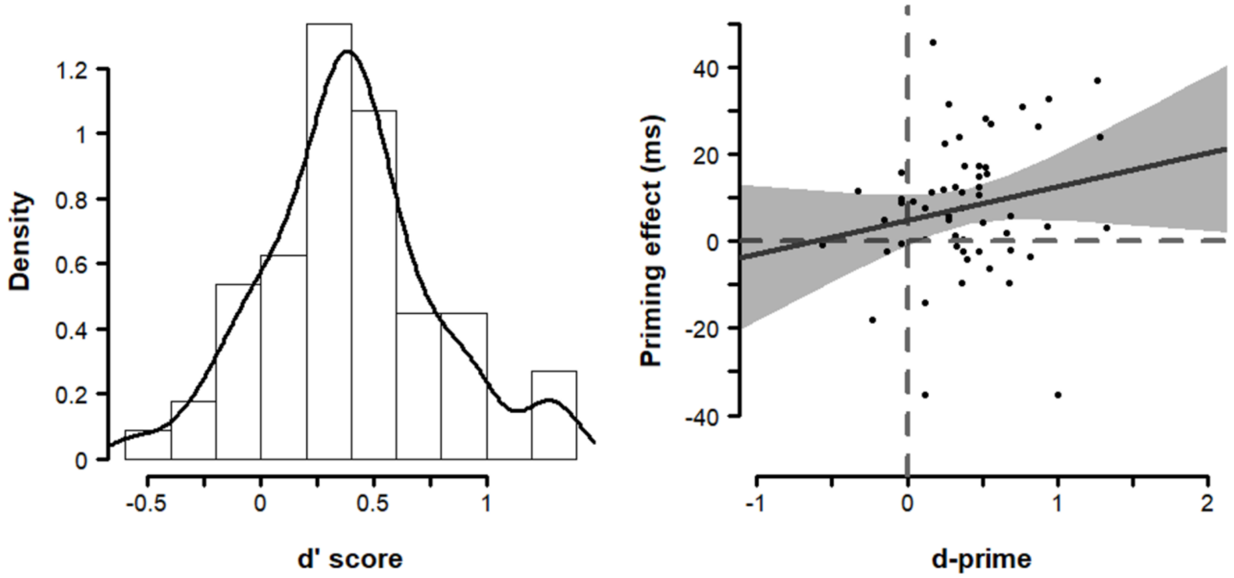

Figure 6. Density plot representing the distribution of the participants' d-prime in the prime visibility task (left panel) . Relationship between priming and prime visibility (right panel). Points represent individual participants, and the shaded area indicates the 95% confidence interval of the regression line.

Meta–analysis of the original and replication experiments

To deliver the full potential of the data collected in this work, we merged the original experiment and its replication in a meta–analysis. The Bayesian approach is particularly suitable here, as it allows to build cumulatively on previously acquired knowledge, i.e., the posterior of the original experiment becomes the prior for the replication (Kline, 2013). Following this approach, we computed a mean RT for each subject in each design cell (i.e., congruent, sagittal, masked primes; incongruent, sagittal, masked primes; congruent, lateral, masked primes; and so on), and then carried out a Bayesian t test for each congruent–incongruent contrast; this procedure allowed us to estimate facilitation for sagittal and lateral primes, in the masked and unmasked condition. For the original experiment, we used an uninformative Cauchy prior (scale parameter=.707; Strachan & Van Dijk, 2003), with a directional hypothesis (we hypothesized that congruent primes could only determine quicker response times); the posterior distribution in the original experiment then became the prior for the replication. This analysis was carried out with JASP (Wagenmakers et al., 2018), while the posterior distribution in the original experiment was estimated in R.

Results are illustrated in Figure 7. Bayes factors in favor of the alternative hypothesis that congruent primes yield quicker RTs than incongruent primes are 16.46, .86, 20.30 and 20.66 for masked sagittal, masked lateral, unmasked sagittal and unmasked lateral primes, respectively. There is thus strong evidence for priming in all conditions, expect for the lateral primes, sub–liminal one. Importantly, while the 95% credible intervals are very similar for sagittal and lateral primes in the unmasked condition, [-.497, -.099] vs. [-.508, -.152][4], they are very different outside of awareness, where sagittal prime generate a strong effect, [-.758, -.362], while lateral primes only yield very weak (if any) facilitation, [-.263, -.009].

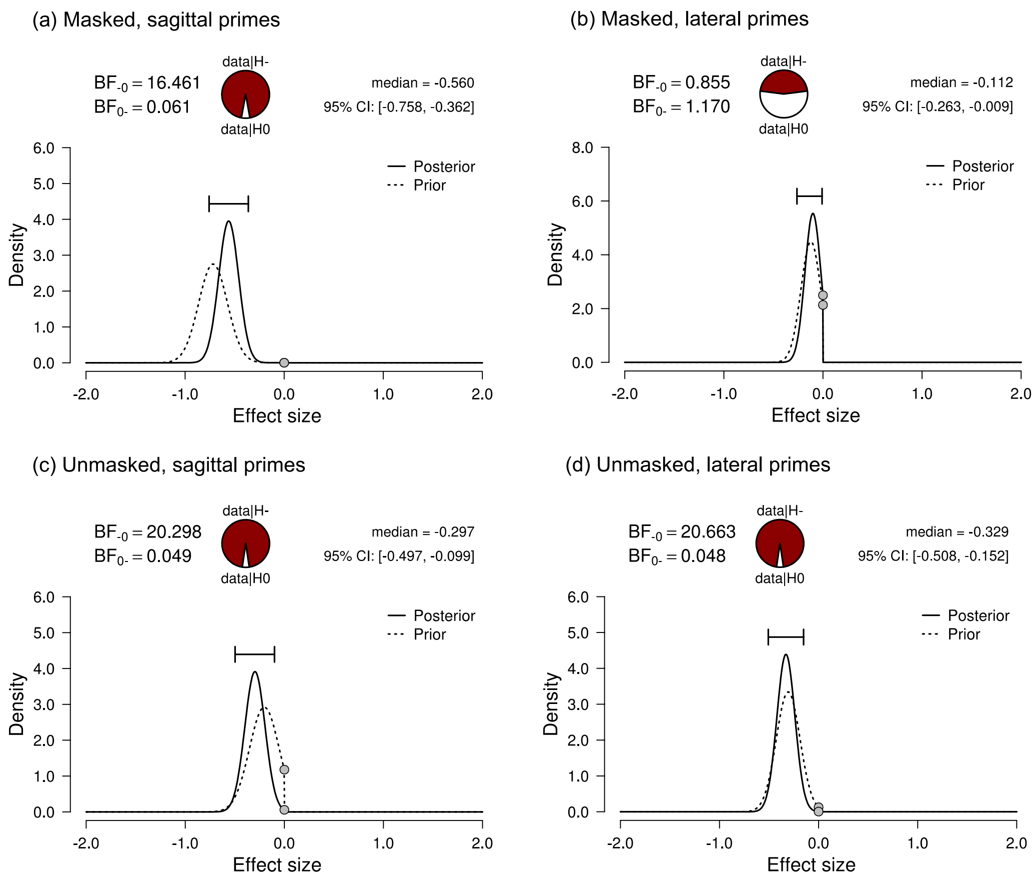

Figure 7. Results from the Bayes factor (BF) replication test for the different conditions.

[4] The values reported here are standardized effect sizes as computed in JASP, and are interpretable similarly to Cohen's d.

General discussion

In this study we investigated what kind of semantic information is extracted when people process words unconsciously. We proposed and tested the hypothesis that sub–liminal processing is limited to language–encoded semantic ties. To this aim, we took advantage of the fact that Westerners scaffold time onto space along a sagittal and a lateral timeline, but only the former is expressed in language (e.g., Monday comes *before*, not *to the lef*t of Tuesday). Consistent with the hypothesis, we found strong and comparable space-time congruity effects along the sagittal (front-back) and lateral (left-right) timelines when primes were visible. By contrast, when participants read the same prime words unconsciously, the sagittal primes produced much stronger effects than the lateral ones, which only yielded very weak facilitation (if any).

These results shed new light on unconscious semantic processing, at least as indexed by masked priming. In most circumstances it is impossible to isolate the role of linguistic experience in the computation of word meaning, because words that are semantically related are typically also related in language use. Our lateral prime-target pairs, by contrast, are related in semantic memory, but not in conventional linguistic expressions (Clark, 1973). Therefore, the finding that lateral spatial primes affected temporal judgments when the prime was read consciously, but much less (if not at all) when it was read unconsciously, supports a reinterpretation of the catalog of results showing unconscious semantic priming (e.g. Kouider & Dehaene, 2007). That is, readers may not access their semantic system to a full extent when exposed to words subliminally. Rather, they may navigate their the lexical–semantic system based on how words are linked to each other in language use (in this case, as related to linguistic metaphors).

An interesting aspect of these experiments, and an improvement as compared to most of the previous literature, is that awareness was manipulated via visual masking, while prime presentation time was kept identical in the sub–liminal and supra–liminal conditions. Thus, we show that prime presentation time is not the main driver of the asymmetry between masked and overt priming—a hypothesis that was compatible with the results from most previous studies where awareness was manipulated via prime presentation time (e.g., Brunellière, Perre, Tran, & Bonnotte, 2017).

These results highlight the role of backward masking, instead. One possible mechanism is that masking prevents words from reaching consciousness by limiting the flow of information within the lexical semantic network (Dehaene et al., 2001). This interpretation is compatible with neuroimaging findings. In fMRI experiments, neural activity related to unconsciously perceived words appears to be limited to occipital–temporal visual areas within the brain word processing network (Price & Devlin, 2011). By contrast, consciously perceived words produce a highly distributed pattern of activations in the cerebral cortex, including not only occipital and temporal areas, but also parietal, motor, and prefrontal areas (Gaillard et al., 2009). These data were taken to support models of consciousness suggesting that stimuli become conscious by activating a "global workspace" (Dehaene & Naccache, 2001), where distant cortical areas can communicate with each other, and a fronto–parietal network can send top-down amplification signals to more posterior and primary sensory areas (Gaillard et al., 2009). The activation of a global workspace network may also facilitate the integration of information coming from different modalities (e.g., visual, auditory) or from brain networks that implement different kinds of mental content (e.g., wordforms, spatial schemas). From this point of view, unconscious processing is likely to be more segregated than conscious processing (Kouider & Dehaene, 2007), and access to the global workspace network with reverberating and sustained activity at the whole-brain scale may be crucial for making the leap from form to meaning in language.

Of course, we did not explore the entire causal chain behind these phenomena. What we observe here is that, when primes are masked, there is no conscious access and semantic priming is bound to linguistic experience. When primes are not masked instead, there is conscious access and semantic priming is not bound to linguistic experience anymore. This is a compatible with a view whereby conscious access is the *main causal factor* behind fully–fledged semantic priming (possibly because it overcomes the limited spread of lexical–semantic information imposed by visual masking). However, the data are also compatible with a milder interpretation where conscious access simply goes together with unbounded semantic priming; and there is a primary cause for both these phenomena, which we did not uncover here. More research is required to clarify this important point.

References

Abrams, R. L., Klinger, M. R., & Greenwald, A. G. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196262

Anaki, D., & Henik, A. (2003). Is there a "strength effect" in automatic semantic priming? *Memory and Cognition*. https://doi.org/10.3758/BF03194385

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., & König, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition*. https://doi.org/10.1016/j.cognition.2010.05.010

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01

Borghesani, V., & Piazza, M. (2017). The neuro-cognitive representations of symbols: the case of concrete words. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2017.06.026

Bottini, R., Bucur, M., & Crepaldi, D. (2016). The nature of semantic priming by subliminal spatial words: Embodied or disembodied? *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000197

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/PBR.15.6.1035

Brunellière, A., Perre, L., Tran, T. M., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2016.1215479

Bueno, S., & Frenck-Mestre, C. (2008). The activation of semantic memory: Effects of prime exposure, prime-target relationship, and task demands. *Memory and Cognition*. https://doi.org/10.3758/MC.36.4.882

Casasanto, D., & Bottini, R. (2014a). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0033297

Casasanto, D., & Bottini, R. (2014b). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews: Cognitive Science*. https://doi.org/10.1002/wcs.1271

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gestures in english speakers. *Cognitive Linguistics*. https://doi.org/10.1515/cog-2012-0020

Clark, H. H. (1973). Space, time, semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York, New York, USA: Academic Press.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*. https://doi.org/10.1177/0956797613504966

Damian, M. F. (2001). Congruity effects evoked by subliminally presented primes: automaticity rather than semantic processing. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(1), 154–165.

De Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000074

Dehaene, S, Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758. https://doi.org/10.1038/89551

Dehaene, Stanislas, & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Devlin, J. T., Jamison, H. L., Matthews, P. M., & Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proceedings of the National Academy of Sciences of the United*

States of America, *101*(41), 14984–14988. https://doi.org/10.1073/pnas.0403766101

Fuhrman, O., & Boroditsky, L. (2010). Cross-Cultural Differences in Mental Representations of Time: Evidence From an Implicit Nonlinguistic Task. *Cognitive Science*. https://doi.org/10.1111/j.1551-6709.2010.01105.x

Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., … Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.1000061

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible (“subliminal”) stimuli: dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology. General*, *124*(1), 22–42. https://doi.org/10.1037/0096-3445.124.1.22

Haspelmath, M. (1997). From Space to Time. In *LINCOM Studies in Theoretical Linguistics* (Vol. 03). Munchen - Newcastle: Lincom Europa.

Hassin, R. R. (2013). Yes It Can: On the Functional Abilities of the Human Unconscious. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612460684

Hassin, R. R., Bargh, J. A., & Zimerman, S. (2009). Automatic and flexible: The case of nonconscious goal pursuit. *Social Cognition*. https://doi.org/10.1521/soco.2009.27.1.20

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*. https://doi.org/10.1016/S0010-0277(00)00125-6

Kiesel, A., Hoffmann, J., & Kunde, W. (2003). Conscious control over the content of unconscious cognition. *Cognition*.

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What’s new in Psychtoolbox-3? *Perception*. https://doi.org/10.1068/v070821

Kline, R. B. (2013). Beyond significance testing: Statistics reform in the behavioral sciences (2nd ed.). In *Beyond significance testing: Statistics reform in the behavioral sciences (2nd ed.)*. https://doi.org/10.1037/14136-000

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2007.2093

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2005.01584.x

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.0.

Louwerse, M., & Connell, L. (2011). A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive Science*. https://doi.org/10.1111/j.1551-6709.2010.01157.x

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience*, *1*(2), 144–149. https://doi.org/10.1038/393

Meteyard, L., & Davies, R. A. I. (2019). Best practice guidance for linear mixed-effects models in psychological science. *PsyArvXiv*. https://doi.org/10.31234/osf.io/h3duq

Miles, L. K., Nind, L. K., & Macrae, C. N. (2010). Moving Thruough Time. *Psychological Science*, *21*(2), 222–223. https://doi.org/10.1177/0956797609359333

Nakamura, K., Makuuchi, M., Oga, T., Mizuochi-Endo, T., Iwabuchi, T., Nakajima, Y., & Dehaene, S. (2018). Neural capacity limits during unconscious semantic processing. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13890

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*. https://doi.org/10.1207/s15516709cog0000_62

Open Science Collobaration. (2015). Estimating the reproducibility of psychological science: Open Science Collobaration. *Science*. https://doi.org/10.1126/science.aac4716

Ortells, J. J., Kiefer, M., Castillo, A., Megías, M., & Morillas, A. (2016). The semantic origin of unconscious priming: Behavioral and event-related potential evidence during category congruency priming from strongly and weakly related masked words. *Cognition*. https://doi.org/10.1016/j.cognition.2015.09.012

Peer, M., Salomon, R., Goldberg, I., Blanke, O., & Arzy, S. (2015). Brain system for mental orientation in space, time, and person. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1504242112

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*. https://doi.org/10.1007/s00426-002-0086-5

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.04.001

Quinn, W. M., & Kinoshita, S. (2008). Congruence effect in semantic categorization with masked primes with narrow and broad categories. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2007.03.004

Rabagliati, H., Robertson, A., & Carmel, D. (2018). The importance of awareness for understanding language. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000348

Radden, G. (2004). The metaphor TIME AS SPACE across languages. In N. Baumgarten, C. Böttger, M. Motz, & J. Probst (Eds.), *Übersetzen, interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung – Das Leben mit Mehreren Sprachen: Festschrift für Juliane House zum 60 Geburtstag*. Bochum, Germany: Aks-verlag.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*. https://doi.org/10.3758/BF03207490

Rinaldi, L., Locati, F., Parolin, L., Bernardi, N. F., & Girelli, L. (2016). Walking on a mental time line: Temporal processing affects step movements along the sagittal space. *Cortex*. https://doi.org/10.1016/j.cortex.2016.02.013

Sackur, J., Naccache, L., Pradat-Diehl, P., Azouvi, P., Mazevet, D., Katz, R., … Dehaene, S. (2008). Semantic processing of neglected numbers. *Cortex*. https://doi.org/10.1016/j.cortex.2007.02.003

Saj, A., Fuhrman, O., Vuilleumier, P., & Boroditsky, L. (2014). Patients With Left Spatial Neglect Also Neglect the “Left Side” of Time. *Psychological Science*. https://doi.org/10.1177/0956797612475222

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*(3), 512–516.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400. https://doi.org/10.1038/nn1549

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1211645109

Strachan, R. W., & Van Dijk, H. K. (2003). Bayesian Model Selection with an Uninformative Prior. *Oxford Bulletin of Economics and Statistics*. https://doi.org/10.1046/j.0305-9049.2003.00095.x

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2012.724694

Ulrich, M., Hoenig, K., Grön, G., & Kiefer, M. (2013). Brain activation during masked and unmasked semantic priming: Commonalities and differences. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_00449

Ulrich, R., Eikmeier, V., de la Vega, I., Fernández, S. R., Alex-Ruf, S., & Maienborn, C. (2012). With the past behind and the future ahead: Back-to-front representation of past and future sentences. *Memory and Cognition*. https://doi.org/10.3758/s13421-011-0162-4

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of Masked Priming: A Meta-Analysis. *Psychological Bulletin*. https://doi.org/10.1037/a0015329

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic*

Bulletin and Review. https://doi.org/10.3758/s13423-017-1323-7

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review, 15*(2), 426–430. https://doi.org/10.3758/PBR.15.2.426

Chapter 2. Word meaning with and without awareness as explored through semantic priming and computational linguistics

Introduction

Over the last decades, several studies have addressed the question of whether readers process subliminal words up to their meaning (Kouider & Dehaene, 2007; Mudrik, Faivre, & Koch, 2014). Masked semantic priming represents one of the most commonly used experimental paradigm to address this issue. In this technique, the recognition of a given word (the target), is facilitated by the quick and masked previous presentation of a related word, (the prime; McNamara, 2005). Specifically, the prime word is presented briefly (50 ms or less), embedded between two sequences of random characters (e.g. "#####", "kxlujwd"). Despite participants would not typically spot its presence, the prime affects the processing of the following target. For example, the word *dog* is processed faster if preceded by the related word *cat* than if preceded by the unrelated word *ship*. Since the pioneering study by Marcel (1983), several experiments have further shown how "invisible" words can prime related targets. Improvements in the experimental procedures have also allowed for a better assessment of the subliminal nature of the masked stimuli, leading to stronger and more reliable results (Greenwald, Klinger, & Schuh, 1995; Reingold & Merikle, 1988).

Despite the existence of subliminal priming is no longer a matter of debate, many studies have tried to shed light onto its cognitive and neural mechanisms. Similarly to overt priming, unconscious semantic priming has been traditionally explained in terms of automatic spread of activation (Collins & Loftus, 1975). Words are represented as nodes within an interconnected network, and links between nodes reflect lexical–semantic ties. When a given word is read, the corresponding node is activated, and activation spreads along the network to related nodes. Crucially, this process has been described as automatic and not liable to strategic control by the reader.

However, some recent discoveries have changed our way to look at masked semantic priming, suggesting that it could be driven by much more dynamic mechanisms (Kiefer, Adams, & Zovko, 2012). In particular, the supposed automaticity of the effect has been challenged as task dependency and top–down influences were found to modulate it. In their meta-analysis of 46 studies, Van den Bussche, Van den Noortgate and Reynvoet (2009) highlighted how the task performed by participants affects priming: different variables moderate priming in semantic categorization and lexical decision, and overall the former provides more reliable results and greater effect sizes than the latter. Similarly, Martens and Kiefer (2009) found that the effect critically depends on the attentional resources currently available, so that a significant reduction was attested if participants, prior to prime presentation, were engaged in an attentional effortful secondary task, as opposed to a less demanding one.

An orthogonal question is what kind of information is grasped subliminally. In fact, words can be semantically related in several different ways (e.g., couch and sofa vs. koala and Australia), and their meaning is extremely multi–faceted (e.g., 'red' refers to visual perception, but is also associated to the meaning 'stop' via our experience with traffic lights, and is metaphorically linked to passion and warmth). Do we capture all these associations and various aspects of words outside of awareness?

According to non-symbolic, embodied accounts of masked semantic priming, the effect would emerge due to the activation of the motor schema associated with lexical meaning. For instance, Ansorge, Kiefer, Khalid, Grassl and König (2010) found that spatially congruent pairs (e.g., up-ABOVE) elicited faster reaction times than incongruent pairs (e.g., down-ABOVE). Critically, this effect interacted with the movement required to provide a response: facilitation was larger when participants had to press an upward button for the target *above,* as compared to when they had to move down to respond to the same target. According to the authors, the effect would be driven by the activation of the motor program associated with the prime word, which would then be grasped subliminally. However, embodied theories of masked priming have been recently challenged. Bottini, Bucur and Crepaldi (2016) tested symbolic and non-symbolic accounts of masked priming in a series of six experiments, showing that no effect was observed once only embodied mechanisms could account for the emergence of priming. On the contrary, priming was attested when embodied explanations were made impossible by the task manipulation, which instead left symbolic ties free to deploy.

Symbolic theories of semantic representations define meaning activation as an a-modal process based on the set of connections that link a given word to others (Louwerse, 2011). Under this perspective, there are at least two main approaches to define the aforementioned set of connections, based either on the conceptual representation of words' referents or on the frequency with which two entities occur together in our experience of the world. This difference has been typically described as an opposition between *semantic similarity* and *association strength* (Mirman, Landrigan, & Britt, 2017).

According to traditional models of semantic similarity (Smith, Shoben, & Rips, 1974; Tversky, 1977), lexical meaning is encoded as a list of descriptive features referring to perceptual, functional and encyclopedic aspects of the words' referent. For example, the words *dog* and *fox* are similar as the two entities share several features—both are mammals, have 4 legs and a tail, are furry, etc. While earlier models did not fully specify how particular features came to be and how they were ranked, more modern implementations of the same idea used data from human raters in property generation tasks to address these issues (McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008). Word meaning can then be represented by a vector keeping track of such features, so that the higher the overlap, the greater the semantic similarity between two words. This approach has been successfully used to explore several issues regarding semantic representation and impairment (Hinton & Shallice, 1991; Randall, Moss, Rodd, Greer, & Tyler, 2004), including semantic priming (McRae & Boisvert, 1998; Vigliocco, Vinson, Lewis, & Garrett, 2004).

The associative approach focuses instead on the link between words whose referents tend to co-occur in the same scenario or event, linguistic or otherwise (De Deyne, Navarro, & Storms, 2013; Nelson, McEvoy, & Schreiber, 2004). For example, the words *dog* and *leash* are associated, at least in Western societies, as every time we encounter a dog, or hear the word dog, it is very likely that we will also encounter a leash, or hear the word leash. Note that dogs and leashes do not really share features, and would thus be considered to be unrelated in feature–based theories, although this is by no means systematic: dogs and cats do share features, and are associated in our experience as well. Association strength is normally estimated through word-generation tasks requiring participants to list one or more words for each target cue. Associative strength gets psycholinguistic validity as significant predictor of various semantic phenomena, such as similarity judgment of word pairs (Deyne, Peirsman, & Storms, 2008) and RTS to the target in a priming context (Anaki & Henik, 2003; de Groot & Nas, 1991).

It is not very clear how these different aspects of word meaning characterize semantic access outside of awareness. Despite semantic similarity and associative strength have been proposed and contrasted as the mechanisms underlying the emergence of priming (Ferrand & New, 2003), results are mixed, leaving the question still highly debated (Hutchison, 2003; Lucas, 2000). Indeed, it is not easy to tear the two apart, as highly associated items tend to share semantic features as well (Brainerd, Yang, Reyna, Howe, & Mills, 2008; Guida & Lenci, 2007). Part of the problem may also stem from the definition of the two types of relatedness, which was often rather loose. In a broad sense, semantic similarity may reflect any kind of relations that link two words based on their meaning. For instance, prime-target pairs were considered semantically similar if the two words were synonyms (e.g. boat-SHIP; Bueno and Frenck-Mestre, 2002), or if they share perceptual (e.g., pizza-COIN) or functional similarity in the way they are used (e.g. house key–SCREWDRIVER; Lam, Dijkstra and Rueschemeyer, 2015). Category membership has also been proposed as a proxy of semantic relatedness; yet, results are mixed. Some studies provided evidence for subliminal semantic priming based on category membership (Dell'Acqua & Grainger, 1999; Van Den Bussche & Reynvoet, 2007) others found that the effect was attested only for stimuli belonging to small categories (e.g., farm animals, mule-SHEEP; Abrams, 2008). Quinn and Kinoshita (2008) demonstrated how category membership cannot be considered as the main engine of masked priming: in their first experiment, each target was paired with a highly similar category coordinate (hawk-EAGLE), with a category coordinate that did not share many features (mole-EAGLE) and a category incongruent prime (knee-EAGLE). Only the former condition elicited significant priming, the other two not being different. Crucially, the authors also showed that a significant effect was observed for prime-target pairs like *moon* and *earth*, that are highly similar in terms of feature overlap despite not belonging to the same category (planets). Interestingly, the authors suggested that their pattern of results could also be explained by associations in language use (e.g., the words *moon* and *earth* occur relatively often within the same sentence).

Another way to explore the issue of what aspects of word meaning are captured outside of awareness is to compare masked and overt semantic priming. While most of the above-mentioned experiments focused either on the masked or on the unmasked condition, few studies have directly contrasted the two. Again, results are intermixed: some spoke in favor of a qualitative difference, others instead suggested rather a quantitative distinction. For example, Gomez, Perea and Ratcliff (2013) provided behavioral and computational evidence that masked and unmasked primes are processed in a qualitative different manner. More precisely, they developed a drift diffusion model fed with behavioral data collected from participants engaged in a lexical decision with primes presented either consciously or unconsciously. In the former condition, priming was clearly observed, while in the latter semantic facilitation, if any, was weak. Model parameters were differently affected by visible and masked primes, leading the authors to conclude that the effect elicited by attended stimuli is qualitative different from the effect elicited by unattended stimuli. However, as already mentioned, masked priming is known to be task dependent, and the study from Gomez and colleagues considered lexical decision only.

On the other hand, De Wit and Kinoshita (2015) compared subliminal and supraliminal priming across different tasks. Crucially, they observed that masking the primes affects priming only in the lexical decision, while in the semantic categorization the effect was attested independently of prime visibility. Thus, priming is not tied to the relation between prime and target, but it hinges upon the nature of the experimental task. Rather than merely identifying words, readers collect information that is relevant to address the task they are required to perform. In the case of lexical decision, the optimal strategy would exploit relatedness between the prime and the target as a cue of target lexical status (retrospective semantic matching). This strategy critically depends on prime visibility, as masking makes the comparison with the target impossible.

Yet, in the case of semantic decision, priming is a byproduct of processes of evidence accumulation and source confusion. Information to optimally performed the task – that the authors described in terms of shared semantic features – is extracted from the stimuli. Under masking condition, the prime and the target are presented so close in time that readers cannot distinguish between the two sources of information. As a consequence, when the task requires them to address a semantic question (e.g., does this word refer to something you can eat?), readers will unconsciously process the prime meaning and gather question–relevant information (the prime *lasagna* provides information toward a YES response), which is not distinguished from the information later obtained from the target, so that when the word *pasta* comes up, they will become convinced of a YES response more quickly.

Overall, we have learned a great deal from the studies described above, but we are still far from having a clear picture on what aspects of word meaning are captured subliminally. One issue that surely contributes to cloud this picture is a less than rigorous definition of the various facets of word meaning—category membership, feature overlap, and associative strength were often confounded, or used to explore different types of semantic relationships across different studies, or again, operationalized in different ways, and sometimes sub–optimally (e.g., only based on the authors' intuition).

Luckily, useful tools to characterize meaning-based similarity in a very precise, quantitative manner were recently developed in the field of computational linguistics. Distributional semantics assumes that lexical meaning can be described on the basis of statistical analysis of the way words are used in large text corpora (Baroni & Lenci, 2010; Sahlgren, 2008). The main idea under this view is that words that tend to share the same linguistic contexts will be similar in meaning; words themselves act as semantic features and the corresponding occurrence frequencies define the strength of the semantic link in a quantifiable and objective manner. By making no assumption about the organizational principles contributing to the observed similarity, it is then possible to avoid the theoretical weakness of postulating a-priori a given set of semantic features. Moreover, similarity estimates can be obtained for most of the words attested in a text corpus (normally in the range of hundred thousands), while feature-lists and associated words are available only for a limited set of stimuli.

Distributional Semantic Models (DSMs) represent lexical meaning via vectors that populate a high-dimensional space where similar words tend to cluster together. Early models (LSA, Landauer and Dumais, 1997; HAL, Lund and Burgess, 1996) built word vectors from co-occurrence matrices that keep track of how words are used in relation to each other in a given corpus. Meaning relatedness between two words is computed by applying geometrical techniques to these vectors; for example, one can approximate relatedness as the cosine proximity (henceforth COS) between the two word vectors:

$$\mathrm{COS}\theta = \frac{a \cdot b}{||a|| \cdot ||b||}$$

DSMs have been proposed as a psychologically plausible models of semantic memory, with particular emphasis on how meaning representations are achieved and structured. In particular, the state-of-art model (word2vec; Mikolov, Chen, Corrado and Dean, 2013) represents a simple neural network consisting of an input, an output and a hidden layer, and is based on a predictive mechanism that allows to infer a target given a cue. Word2vec provides similarity estimates that cover a wide range of classic lexical-semantic relationships, like synonymy[5] (e.g., car-automobile, 0.45), antonymy (e.g., young-old, 0.51), meronymy (e.g., cherry-fruit, .49). Although word2vec is not specifically designed to capture associative relationships, these can be grasped as well (e.g., carrot-stick, .41). Finally, featural similarity can be accounted for beyond category membership; to get back to Quinn's and Kinoshita's (2008) study described above, word2vec clearly teases apart similar members of the same category (e.g., lion-tiger, .54) from dissimilar members (e.g., lion-mole, .17).

Experimental evidence has shown that word2vec (and DSMs in general) explains human behavior well in a variety of tasks, such as synonym detection, concept categorization and synonym detection (Baroni, Dinu, & Kruszewski, 2014; Marelli, 2017). Interestingly, DSMs were also used to account for supraliminal, overt semantic priming. Mandera, Keuleers and Brysbaert (2017) tested word2vec performance on a large dataset of behavioral data comprising reaction times to word targets in primed lexical decision and naming tasks. Model estimates nicely fit the data, better than (or as good as) those based on association norms or feature lists. Whether these data and theoretical insights would also hold for masked semantic priming, thus characterizing the computation of word meaning outside of awareness, it is currently unexplored.

A simpler and more immediate way to model meaning based on the linguistic context is to look at surface co-occurrence, i.e., how much two words are used together within a given window of text (Spence & Owens, 1990). Borrowing from information theory, computational linguistics has adopted

[5] All model estimates taken from Mandera, Keuleers and Brysbaert (2017)

Pointwise Mutual Information (henceforth PMI) to express association between two words in this terms, according to the formula:

$$\mathrm{PMI}(w_1,w_2) = \log_2 \frac{p(w_1,w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ corresponds to the probability of occurrence of the word pair, while $p(w_1)$ and $p(w_2)$ refer to the individual probabilities of the two components (Church & Hanks, 1989). In essence, what we are capturing here is how likely two words will occur together, given their individual probability of occurrence.

PMI expresses how a given word can be used as a proxy for expecting another word, thus can be rightfully considered as an index of associative relationship. Another important property of this metric is that, despite the window of text in which co–occurrence is counted can vary, they are typically quite small, which makes PMI a strong index of local, short range relationships.

The metrics has been used to model a wide range of psycholinguistics phenomena, such as similarity judgements (Recchia & Jones, 2009), reading speed (Ellis, Simpson-Vlach, & Maynard, 2008), and free association and syntactic parsing (Pitler, Louis, & Nenkova, 2010). Moreover, PMI has also been shown to successfully generalize to non-linguistic fields (e.g., reasoning; Paperno, Marelli, Tentori and Baroni, 2014).

Despite they are both based on word co–occurrence counts, cosine proximity and PMI capture rather different information about word meaning. The former is more geared towards higher order relationships: two words may never occur together, but will come up as related as long as they occur similarly with all the other words in the vocabulary. The words *car* and *automobile* are not likely to appear close to each other in a given text; still they represent the same referent, and therefore will be used in similar contexts. PMI is instead more geared towards local, shallower relationships, and rely only on the effective co-presence of two words within the same window of text. For example, the words *glove* and *oven* do not really entertain any obvious semantic relationship (e.g., they are not synonyms, do not belong to the same category, do not share many features), but are likely to be used together in language due to the fact that every time you need to take out a baking pan from the oven, you need a glove to avoid getting burnt.

If cosine proximity and PMI can be disentangled, several items can be found where the two metrics diverge, thus allowing to address their contribution separately. In addition, these metrics provide a more precise and consistent definition, and therefore a neat quantification, of the dynamics that govern meaning construction outside of awareness (at least as far as masked priming can tell). This is the goal of this paper—we will use these metrics to create a set of items that tease apart local ties vs. higher–level relationships, therefore allowing us to further our knowledge on what kind of semantic information we can process outside of awareness. Hopefully, the more rigorous approach that is brought about by computational semantics will clarify some of the inconsistent results that we have highlighted above.

The present study features several other novelties as compared to the existing literature. Because computational linguistics brings us a precise quantification of the strength of words' relationships, we do not need to dichotomize these relationships. Accordingly, we don't have related and unrelated primes in this experiment; rather, prime–target pairs vary continuously for the strength of their relationship, either according to PMI or cosine proximity, and priming is captured by regressing response times on these computational indexes. This approach has several advantages. It reflects more naturally the nature of words' semantic ties, which are genuinely continuous—words are never totally related or unrelated, but rather vary from very weak to very strong associations with no obvious discrete steps. With this design, we also avoid the baseline problem: in classic studies it is not easy to understand whether priming comes from quicker response on related trials, or slower responses on unrelated trials, or, quite likely, a mixture of the two.

A second important feature of the study is that the trial timeline was identical in the supraliminal and subliminal conditions, which differed only for the presence\absence of visual masks. This implies that primes were presented for the same amount of time, thus ensuring that any difference would only depend on awareness, not on prime presentation time (Kanwisher, 2001).

Next, to make sure that our masking technique was effective and to consider individual variability appropriately, we asked participants to perform a prime visibility task after they concluded the masked priming experiment. Based on their performance in this task, we computed a d–prime score (d') for each participant, a signal detection theory metric that, in this context, provides a quantitative measure of prime visibility (Reingold & Merikle, 1988).

Finally, we made use of the exact same set of stimuli in the masked and unmasked priming conditions, so as to be able to compare subliminal and overt priming directly. In fact, the comparison between masked and overt priming that we have described above is mostly based on data from different studies, where target and prime words obviously changed in several different ways.

Experiment 1

In the first experiment, we explored masked semantic priming via a set of 300 prime–target pairs with varying degree of PMI and cosine proximity—participants performed a semantic decision on the target words after having seen a more or less related prime. Critically, the correlation between PMI and cosine proximity was kept as low as possible, so as to be able to disentangle their contribution to priming. Also importantly: (i) participants underwent a prime detection task after the main task was carried out, so that prime visibility was kept under appropriate control; and (ii) a perfectly symmetrical supraliminal version of the experiment was also carried out, allowing us to contrast semantic priming within and outside of awareness.

Methods

Participants. 102 healthy volunteers (68 females and 34 males; mean age= 24 years) were recruited into the experiment. Ten participants were left-handed. All participants were native Italian speakers, with normal or corrected–to–normal vision and no history of neurological diseases. They all provided their informed consent to take part into the experiment, and were compensated for their time with 8 Euros.

Stimuli. 100 Italian words were used as target stimuli, 50 of which referred to animals (e.g., *aquila*, eagle) and 50 to tools (e.g., *forbice*, scissor). Each target was paired with three words from the same category (animals were paired with animals, and tools with tools), resulting in 300 unique prime–target pairs.

For each of these pairs, we computed two indexes of semantic relatedness, Pointwise Mutual Information (PMI, henceforth) and Cosine Proximity between the corresponding word vectors (COS). For PMI, cooccurrence data were gathered by means of a 5–words window sliding across the Itwac corpus, a lemmatized and part–of–speech annotated database of nearly 2 billion Italian words built by web crawling (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). All characters were set to lowercase, and special characters were removed together with a list of stop–words. The raw counts were subsequently transformed into PMI scores according to the following equation:

$$\mathrm{PMI}(w_1, w_2) = \log_2 \frac{p(w_1, w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ represents the probability of encountering the two words within the same 5–word window, and $p(w_1)$ and $p(w_2)$ represents the overall probability of encountering w_1 and w_2.

Cosine proximity between word vectors was obtained training a word2vec model (Mikolov et al., 2013) on the same corpus. Model's parameters were set according to the WEISS model (Marelli, 2017). All words attested at least 100 times were included in the model, which was trained using the continuous-bag-of-word (CBOW) architecture, based again on a 5-word window and on 200 dimensions. The parameter k for negative sampling was set to 10, and the subsampling parameter to 10^{-5}. Among the two different architectures implemented in word2vec, CBOW has been proved to gain better results than Skip-Gram in semantic priming simulations (Baroni et al., 2014). Negative sampling reduces the computational load of the model by selecting a restricted set of items in the output layer for each learning phase, when the probabilities are estimated. Subsampling allows the model to reduce the influence of very high–frequency words, which are known to provide little information for distributional analysis.

Prime–target pairs were selected to obtain nice PMI and COS distributions (see Figure 1), and to avoid excessive correlation between the two indexes (r= .541), so that it is possible to disentangle their specific contribution to semantic priming.

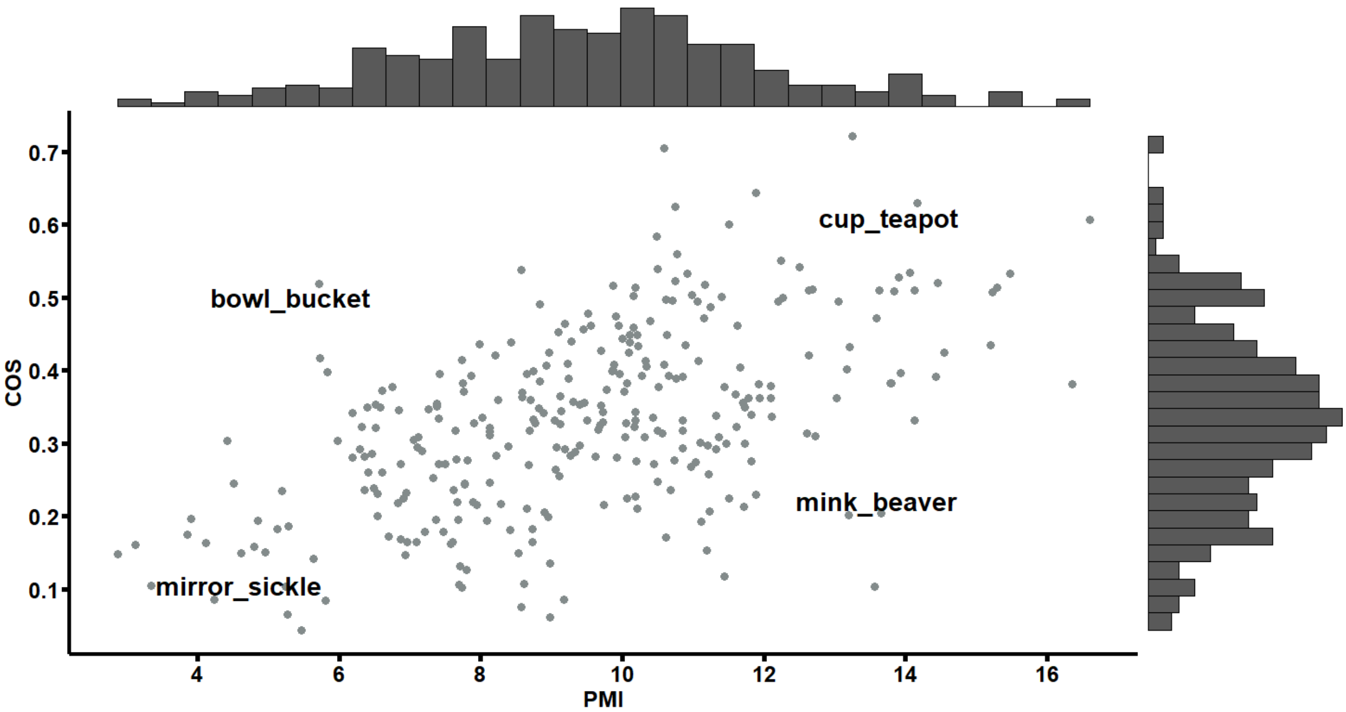

Figure 1. Scatterplot showing of the prime-target pairs used in the study.

Prime and target features are reported in Table 1.

	Prime	Target
Zipf Frequency	3.83 (0.49)	3.22 (0.47)
Length	6.24 (1.39)	6.56 (1.19)
Old20	1.88 (0.58)	2.11 (0.47)

Table 1. Prime and target lexical features - mean (sd).

We also selected an additional sample of 100 filler prime–target pairs, which worked as NO–response trials. These items were not included in the analysis. We used abstract words as target stimuli, roughly comparable in frequency (m= 3.40, sd= 0.57), length (m= 6.51, sd= 1.25) and orthographic neighborhood size (m= 2.15, sd= 0.53) to the target words in the experimental trials. These filler targets were paired with animal and tool word primes, different from those presented in the experimental set, but, again, similar to them in frequency (m= 4.03, sd= 0.49), length (m= 6.34, sd= 1.39) and orthographic neighborhood size (m= 1.93, sd= 0.56). This way, we ensured that the response to the target was not predictable on the basis of the prime.

Procedure. Each trial began with a 750 ms fixation-cross (+). The prime word was then shown for 50 ms, either embedded between two visual masks (i.e. sequences of random uppercase letters as long as the prime word), for the masked condition; or embedded between two blank screens, for the unmasked condition. The visual masks/blank screens lasted 200 ms (before the prime) and 50 ms (after the prime). Finally, the target word was presented for 1500 ms, or until a response was provided (see Figure 2). In the masked condition, participants were not informed of the presence of a prime word.

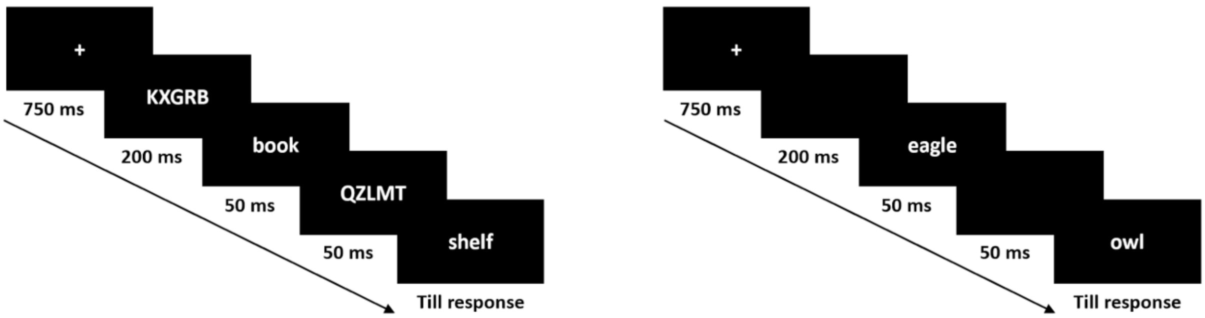

Figure 2. Exemplar trials in the masked (left) and visible (right) conditions.

All stimuli were presented in Arial (font size=32), in white against a black background. We used Matlab Psychtoolbox (Brainard, 1997) to control the presentation of the stimuli and gather participants' response times, which were collected via a Cedrus button box. Stimuli were presented on a 22'' monitor with a refresh rate of 120 Hz.

Participants were engaged in a classic YES/NO task, requiring them to classify target words as members of either the animal or tool category, according to the instructions. YES responses were always provided with the dominant hand. Primes were rotated over target words in a classic Latin Square design, so that each participant was exposed to each target word only once. Because each target was associated with three different prime words, this procedure generated three experimental lists. Each list was composed of 200 trials, which were divided into two blocks. In one block, subjects were asked to press the YES-button if the target word referred to an animal, while in the other block they were asked to press the yes-button if the target word referred to a tool. The proportion of YES responses was .50 in both blocks. The order of the two blocks was counterbalanced across subjects. Ten practice and two warm-up trials were presented before each block. Participants were allowed to take a short break halfway through each block.

Each participant took up both the masked and the overt priming conditions, in two separate sessions that were held between 2 and 5 days far from each other. The condition order was also counterbalanced across participants.

Prime visibility task. Once participants had completed the masked version of the experiment, they were informed about the presence of the prime. Because there is variability in the participants' ability to perceive masked primes, and we wanted to control for this variability, they were then engaged into a prime visibility task requiring them to spot the presence of the letter "n" within the masked word. The trial timeline and presentation parameters remained exactly the same as in the main task; essentially, the trials were just played back to the participants. In order to ensure that participants understood the prime's position within the trial, two examples were presented before the proper task where prime duration was increased to 150ms, in order to make it visible despite the visual masks. Then, 10 practice and 80 experimental trials were displayed. The 80 experimental trials were taken from the main task and were selected randomly, but in such a way that the proportion of YES-response was .50 again.

Data analysis. Analyses were conducted on accurate YES responses only. Individual subjects and items were excluded if they departed substantially from the group distribution, based on visual inspection. Response Times (RTs) were inverse transformed to approximate a normal distribution and used as a dependent variable in linear mixed-effects regression models using the package lme4 (Bates, Maechler, Bolker, & Walker, 2015) of the statistical software R (Chambers, 2008). Outliers were controlled for by fitting a random-effect-only model and excluding those individual data points with standardized residuals exceeding 2.5 standard deviations. This technique allows to discard outliers "a-priori" and to avoid any bias toward the effects of interests.

This analysis allows us to control for all the covariates that may have affected the performance, such as trial position in the randomized list, rotation, RT and accuracy on the preceding trial, the response required in the preceding trial, frequency and length of the target. All these variables were modeled as fixed effects, with participant and item as random intercepts, in a baseline model. Only those covariates that significantly contributed to the goodness of fit were retained into the model. The variables of interest, PMI and COS, were then added to the baseline model, and we checked both whether they provided additional goodness of fit (via a Chi–Square test) and whether their parameters in the model were significantly different from zero (via a t test). In order to compare the specific contribution of PMI and COS, we used the same statistical approach and inserted (i) PMI in the baseline model augmented with COS, and (ii) COS in the baseline model augmented with PMI. PMI and COS were both scaled before entering the model. Finally, p-values were computed using the Satterthwaite approximation to degrees of freedom (Luke, 2017) provided by the *jtools* package (Long, 2018).

Data from the prime visibility task were analyzed in terms of sensitivity index (d'), which computes, for each participant, the ratio between correct hits and false alarms, according to the formula:

$$d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$$

where $Z(p)$, $p \in [0,1]$, is the inverse of the cumulative distribution function of the Gaussian distribution. Prime visibility can thus be indexed by each participant's d', so that the higher its value, the better s\he is able to detect the masked stimulus. Unawareness of the primes is assumed when d' does not differ significantly from 0, despite values below .5 are interpreted as flagging scarce ability to detect the target (in their review of 58 papers, Van den Buscche and colleagues reported d' values ranging from -0.06 to 0.66).

Open practices statement. All data and analysis code are available at https://osf.io/zcdba/, and can be accessed independently from the authors.

Results

Masked primes and prime visibility task. The overall accuracy in this condition was 97%. The mean RTs on accurate trials was 727 ms. No individual participant was taken out because of a particularly anomalous performance. Inaccurate trials (~2.6%) and outliers (~1.6%) were identified and removed, leaving an overall set of 9750 available data points for the analysis.

The d' distribution is shown in Figure 3; the average value was 0.54 [95% CI= 0.41 – 0.67], comparable to previous studies assessing prime awareness (e.g., Bottini et al., 2016; Kouider and Dupoux, 2005).

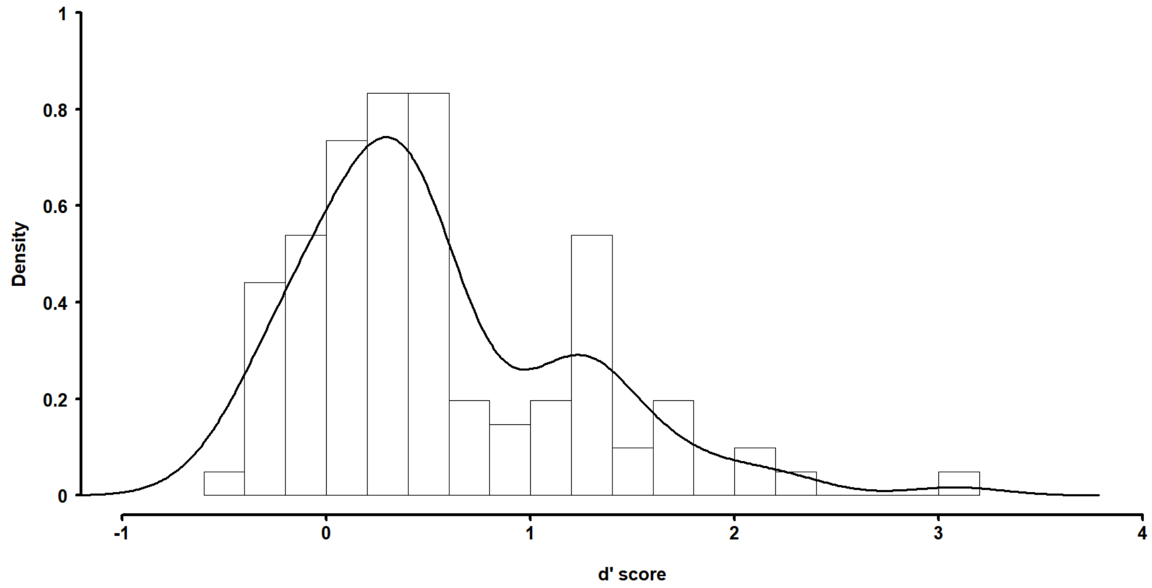

Figure 3. Density plot of the distribution of the d'

RT analysis showed no main effect of semantic similarity—neither PMI nor COS led to a significant increase in goodness of fit ($\chi^2_{(1)}$= 0.58, *p*< .001 and $\chi^2_{(1)}$= 0.29, *p*= .591 respectively), nor their parameters in the model were significantly different from zero (PMI: *β*= -0.002, *t*(9582)= -0.76, *p*=.449; COS: *β*= -0.002, *t*(9361)= -0.54, *p*= .591)

Interestingly though, model fit increased when semantic indexes were tested in interaction with prime visibility as tracked by participants d' in the letter detection task, $\chi^2_{(1)}$= 12.56, *p*= .446 and $\chi^2_{(1)}$= 10.11, *p*= .001, for PMI and COS respectively. As illustrated in Figure 6, the higher the d', the more response times shrink as PMI (β= -0.012, *t*(9547)= -3.54, p< .001) and COS (β= -0.010, t(9546)= -3.18, p= .001) grow. That is, the higher the d', the larger the semantic priming. Also, Figure

4 illustrates quite clearly that semantic priming is likely null when d–prime is low (see the red line, which refers to a d–prime value of 0).

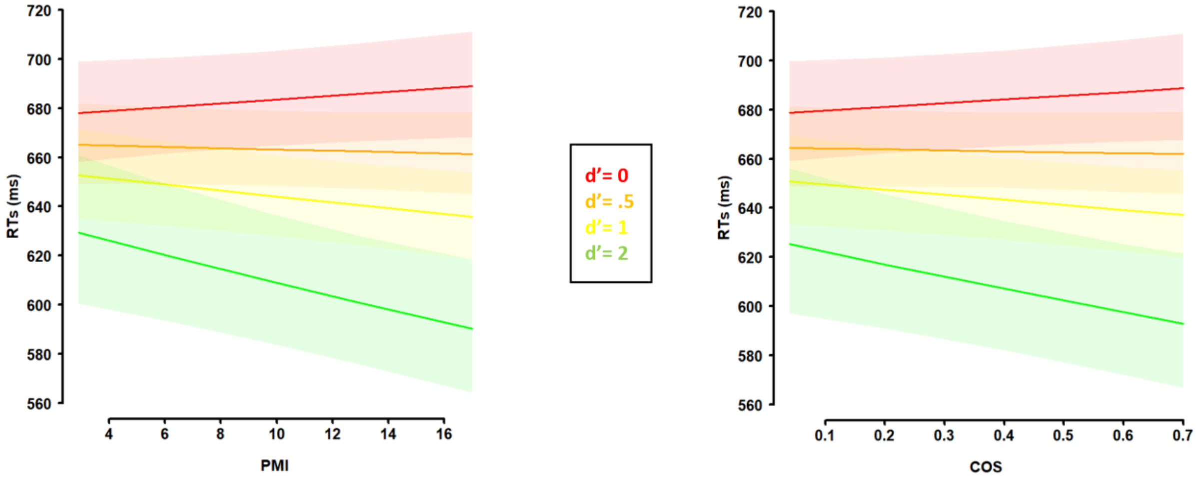

Figure 4. Interaction between d' and prime–target association. Both PMI (left) and COS (right) effects become stronger as prime visibility (d') increases. Shaded areas refer to 95% C.I.

Visible primes. The overall accuracy in this condition was 97% and the mean response time on accurate trials was 720 ms. No individual participant was excluded because of a particularly anomalous performance. Removal of incorrect trials (~2.5%) and outliers (~ 1.75%) led to a total 9770 datapoints for modelling.

Relative to the baseline model with non semantic variables only, we observed a better goodness of fit resulting from the inclusion of either PMI ($\chi^2_{(1)}$= 10.13, *p*= .001) or COS ($\chi^2_{(1)}$= 6.50, *p*= .011). This is in line with the model parameters, which are significantly different from zero for both PMI (*β*= -0.010, *t*(9400)= -3.18, *p*= .001) and COS (*β*= -0.008, *t*(8870)= -2.55, *p*= .011).

When we compared the two metrics, we found out that adding PMI to the COS model improved the overall fit to the behavioral data ($\chi^2_{(1)}$= 4.16, *p*= .041), but not vice-versa ($\chi^2_{(1)}$= 0.52, *p*= .469). Correspondingly, the parameter analysis in the model with both PMI and COS reveals that while the former is significantly different from zero (*β*= -0.008, *t*(8623)= -2.03, *p*= .042), the latter is not (*β*= -0.003, *t*(7817)= -0.73, *p*= .465). The pattern of results is shown in Figure 5.

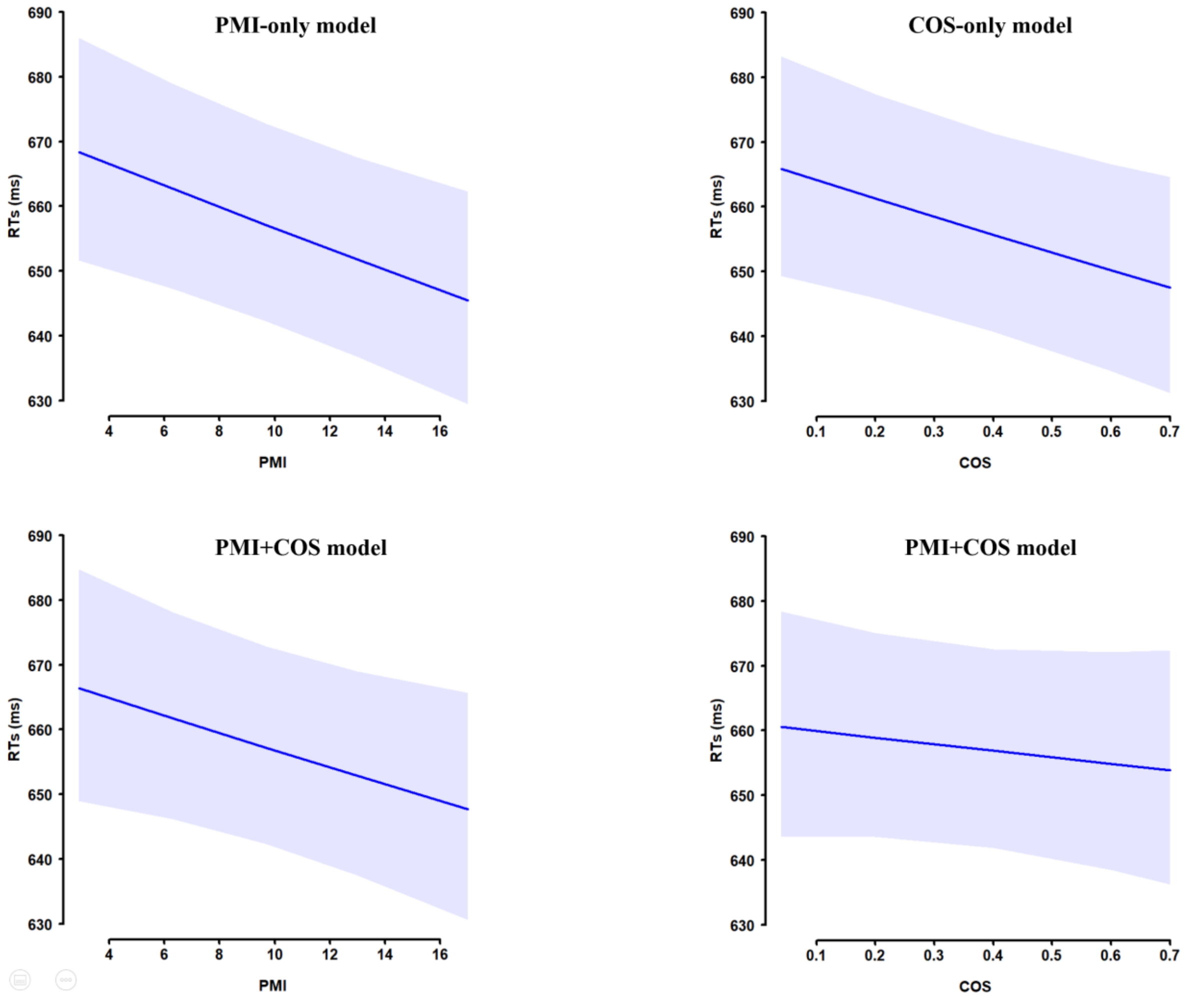

Figure 5. Significant effects of PMI (upper-left) and COS (upper-right) in isolated models. When the two predictors are contrasted, PMI (lower-left) outperformed COS (lower-right). Shaded areas refer to 95% C.I.

Discussion

Based on these data, genuine masked semantic priming seems dubious, no matter what semantic index is taken into consideration. Neither PMI nor COS were, by themselves, significant predictors of the emergence of priming in the masked condition; and both interacted with prime visibility, in a way that facilitation increases with participants' ability to detect the prime. Thus, some degree of prime visibility may be required for processing words up to the semantic level.

These results are at odds with several previous studies supporting the existence of masked semantic priming. Those studies, however, used the classic, dichotomous design contrasting related and unrelated primes. Perhaps, when one explores the effect along the entire relatedness continuum, subliminal semantic effects may actually turn out to be weaker than previously thought. Also, it is hard to tear apart local associations (i.e., PMI) from more distributed, high–level relatedness (i.e., DSM) at the extremities of the semantic continuum, where words tend to be associated (or not associated) on both indexes. So, perhaps, masked semantic priming in previous studies benefitted from multiple levels of relatedness, which we explicitly tried to separate here.

Semantic facilitation, instead was clearly attested when primes were fully visible. In the overt condition, both PMI and COS successfully predicted the emergence of priming—the higher the strength of the link between the prime and the target, the shorter the response time. Yet, when both the indexes were entered in the same model, PMI outperformed DSM in the fit to the behavioral data. These results seem to suggest that overt semantic priming is primarily driven by local association ties as tracked by word co-occurrence, rather than by higher–level semantic relationship as tracked by state–of–the–art DSMs.

A comparison between masked and overt priming —which is possible here for the first time on the same subjects, items and prime presentation time— clearly reveals a strong asymmetry: while priming does not seem to emerge subliminally, at least for those participants who really had no awareness of the primes, facilitation is solid supraliminally.

Of course, some of the conclusions we draw here need further testing. For example, we are aware that 50 ms is quite atypical for prime presentation time in studies on *conscious* semantic priming, and several experiments have shown how different prime durations may affect facilitation depending on the particular kind of semantic link being processed (Lam et al., 2015). This calls for longer prime durations, which we tested in Experiment 3.

Before that, however, we turned our attention to masked priming, and tested one prediction of the interpretation offered above for this phenomenon. The interaction between the semantic indexes and prime visibility, and the d–prime distribution itself, shows that some participants were still able to somehow detect the presence of the masked primes. So, in Experiment 2 we reduced prime duration to 33ms, thus enforcing lower prime visibility. If semantic facilitation does indeed need some awareness of the primes to emerge, then it should completely disappear under such conditions. In other words, in Experiment 2 we expect (i) lower, possibly around zero d–primes; and, consequently, (ii) no sign of priming, nor interaction between priming and d–prime.

Experiment 2

Methods

Participants. 75 healthy volunteers (56 females and 19 males; mean age= 23 years) were recruited into the experiment. They all provided their informed consent, and were compensated for their time with 8 Euros. None of the subjects took part in the previous experiments.

Stimuli and *Procedure* were kept the same as in the masked priming condition of Experiment 1, with the only difference that primes remained on the screen for 33 ms now. We adapted the duration of the backward mask consequently (67 ms), so as to keep the overall prime-target stimulus onset asynchrony (SOA) fixed at 100ms.

As for Experiment 1, once the participants had completed the main task, they were informed about the presence of the prime and underwent the prime visibility task.

Data analysis were conducted exactly as in Experiment 1.

Results

The overall accuracy in this experiment was 97%. The mean RTs on accurate trials was 675 ms. No individual participant was excluded because of a particularly anomalous performance. After inaccurate trials (~2.4%) and outliers (~2%) were removed, 7196 available data points were considered for the analysis.

From each subject's performance in the prime visibility task, we computed the corresponding d' score. Mean d' was 0.03 [95% CI= -0.03 – 0.10]; the overall distribution is illustrated in Figure 6. Participants' ability to spot the prime was, as expected, lower than in the previous experiment, as confirmed via Welch Two Sample t-test between the two d' distribution, $t(146)= -6.77$, $p< 00.1$. Moreover, all participants except 4 (95%) had a d' below .5, thus being effectively unaware of the primes.

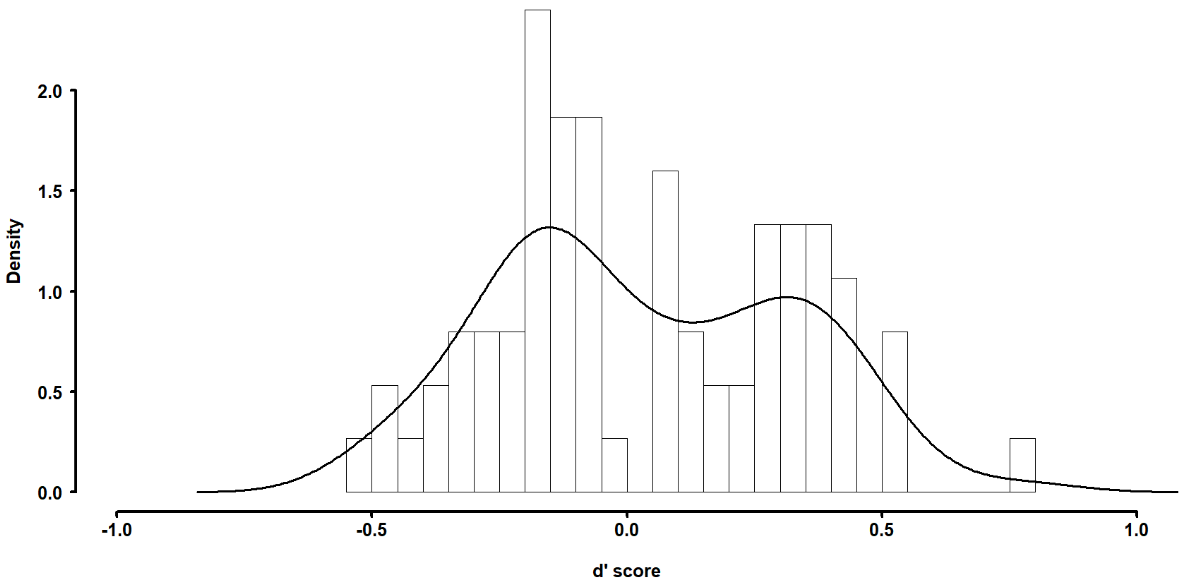

Figure 6. Density plot of the distribution of the d'.

Consistently with these d' data and the results of Experiment 1, the RT analysis revealed no effect of semantic similarity—goodness of fit of the baseline model did not benefit from adding PMI ($\chi^2_{(1)}$= 0.47, p= .492) or COS ($\chi^2_{(1)}$= 0.38, p= .538) as predictors. Model parameter further confirmed that the two indexes had no effect on the dependent variable (β= -0.002, $t(7024)$= -0.69, p= .492 and β= 0.002, $t(6820)$= 0.62, p= .538 for PMI and COS respectively), nor yielded an interaction with d' scores (β= 0.007, $t(7021)$= 0.74, p= .457 and β= 0.005, $t(7021)$= 0.57, p= .569 for PMI and COS respectively).

Discussion

The critical manipulation in this experiment, that is, prime presentation time brought down to 33 ms, worked as expected—prime visibility decreased dramatically from Experiment 1, and is now effectively null, as indexed by d–primes in a letter detection task on the primes themselves. As predicted on the basis of the results in Experiment 1, this prevented semantic priming—we did not observe any evidence for a main effect of PMI or COS, similarly to Experiment 1, and also, more importantly, we did not observe any interaction with d–primes either. Essentially, priming does not emerge consistently across the d–prime spectrum that we captured in this experiment.

Putting together the results from Experiment 1 and 2, it seems that priming would only start to emerge for d–prime values around 1, which does indicate some prime visibility. Thus, no semantic priming seems to be attested when primes are strictly kept outside of awareness.

How does this go together with the several reports of masked semantic priming that populate the literature? The most apparent difference between this study and the previous one is in the design—while classic masked priming experiments are based on taking the difference between response times in a related (e.g., cat–DOG) and unrelated condition (e.g., tip–DOG), here we modeled the strength of the prime–target relatedness continuously. Essentially, instead of tapping only onto the extremes of the relatedness distribution, we explored its effect all along its continuum. If this is the reason why we do not find evidence for subliminal semantic priming, then we should be able to see this priming emerge if we just apply the more classic, dichotomic approach to these very same data. We illustrate this analysis in the next section.

A dichotomic re–analysis of the masked priming data

Of the 300 prime–target pairs that we employed in Experiment 1 and Experiment 2, we selected as related pairs those that were concurrently above the upper quartile of the distribution of both the metrics considered (11.04 for PMI, 0.41 for COS); and those that were below the lower quartile of the distribution (7.69 for PMI, 0.24 for COS), as unrelated pairs. Unfortunately, we could not ensure the within-target comparison between related and unrelated primes, as normally done in priming experiments, because not all the targets in the related condition appeared in the unrelated condition as well. Yet, possible confounding from unbalanced design could be controlled for in the analysis. Finally, selecting pairs only from the extremes did not allow us to disentangle between PMI and COS as specific sources of priming because of the high correlation (r= .9) between the two metrics.

We then took all the response times we gathered on these pairs in Experiment 1 (prime duration=50ms) and Experiment 2 (prime duration=33ms), which generated a sample of 4193 datapoints. We submitted these data to mixed–effect modelling, with semantic relatedness and experiment/prime duration, as well as their interaction, as fixed effects, and participant and item as random intercepts. All other details about the modelling of the data were the same as in Experiment 1 and Experiment 2. We also collected the d–prime values for all the participants involved in this re–analysis (n= 177), and regressed them against each participant's priming effect. This method does not only allow us to assess the correlation between facilitation and prime visibility, but also to estimate facilitation when d–prime is zero, that is, when prime visibility is null (Greenwald et al., 1995).

Results are illustrated in Figure 7. We observed a significant interaction between relatedness and experiment/prime duration ($\chi^2_{(1)}$= 5.39, *p*= .020); with a 33ms presentation time for the primes priming does not seem to emerge (*β*= 0.002, *t*(1048)= 0.13, *p*= .900), while facilitation is more strongly attested for primes lasting 50 ms (*β*= 0.036, *t*(3963)= 2.32, *p*= .020). In this latter condition, the correlation between prime visibility and facilitation at the subject level was .19 (95% CI: -.004 – .371; p= .55. See Figure 7b), which suggests, similarly to the original analysis of the Experiment 1 data, that masked priming partially depends on prime visibility. However, the 95% CI at the intercept lies entirely above the origin (5ms – 37ms; point estimate=21ms), indicating that priming is indeed estimated to be higher than zero even when primes are completely outside of awareness.

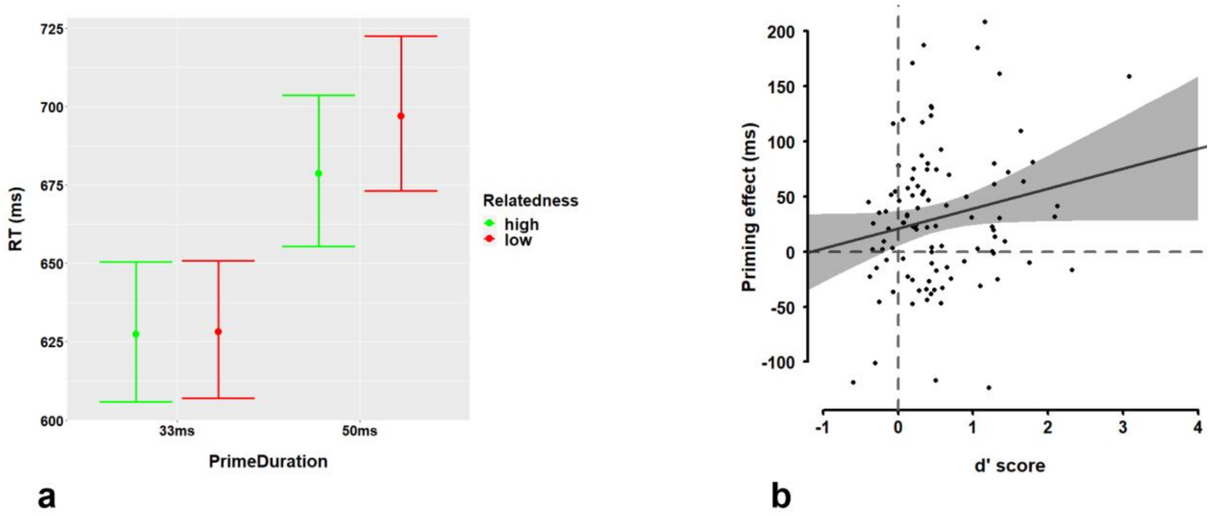

Figure 7 ***(a)*** *Priming effect across different prime exposures. The congruent condition is plotted in green, and the incongruent condition in red. Error bars refer to the 95% confidence intervals.* ***(b)*** *Relationship between priming and prime visibility. Points represent individual participants, and the shaded area indicates the 95% confidence interval of the regression line. Note that priming is measured by subtracting mean RTs on congruent trials from mean RTs on incongruent trials, that is, positive values indicate facilitation.*

This re–analysis of the masked priming data with the more classic, dichotomic approach reveals that, at least when the prime duration was 50ms (which is a very typical value in the masked priming literature; Van den Bussche et al., 2009), facilitation does seem to emerge outside of awareness. Or at least, this would be the interpretation of the pattern of results that we observe here: related trials yield quicker response times than unrelated trials, and the regression analysis shows that priming would be significantly higher than zero when the d–prime is zero.

So, at 33ms of prime presentation time the effect is virtually null, and therefore a continuous rather than a dichotomous modelling does not really affect the outcome. However, when the prime is available for 50ms, we are only able to see it when the extremes of the semantic continuum are considered. Thus, subliminal priming effects may be the result of an “all-or-nothing” phenomenon (or illusion?), which requires a strong difference in relatedness to emerge clearly in the data. Should we “believe” more in the dichotomic analysis, and therefore claim genuine subliminal semantics? Or rather, we should trust the continuous analysis, and therefore deny masked semantic priming? We will take up this issue in the General Discussion. We were not able to disentangle the different sources of information contributing to meaning similarity due to the high overlap between the measures considered. More research, possibly adopting a mega study approach with thousands of datapoints taken into consideration, is necessary to further explore the dynamics of subliminal semantic processing.

Experiment 3

In this experiment, we assess whether the results observed in Experiment 1 on overt priming, that is, that PMI accounts for the phenomenon better than COS, are confirmed when we adopt a prime duration that is more comparable with previous studies. In particular, we tested 150ms and 1150ms.

Methods

Participants. 85 healthy volunteers (59 females and 26 males, mean age= 24 years) were recruited into the experiment, which involved two different sessions with 2 to 5 days in between. They all provided their informed consent , and were compensated for their time with 10 Euros. None of the subjects took part in the previous experiments.

Stimuli and *Procedure* were kept identical to the overt priming condition in Experiment 1, with the only difference that primes were now presented for 150 ms and 1150 ms, in two separate sessions. Participants always underwent the shorter prime duration session first.

Data analysis. Data were analyzed exactly as in Experiment 1, with the exception that there was an additional variable of interest here, prime presentation time (150ms vs. 1150ms), which we modeled as a further fixed effect.

Results

The overall accuracy in this condition was 97%. The mean RTs on accurate trials was 674 ms. No individual participant was taken out because of a particularly anomalous performance. Inaccurate trials (~2.4%) and outliers (~1.8%) were removed, leaving a total of 16261 overall observations for the analysis.

Entering either PMI ($\chi^2_{(1)}$= 21.65, p< .001) or COS ($\chi^2_{(1)}$= 10.98, p< .001) in the model with non-semantic covariates improved the fit to the data. According to model estimates, both PMI (β= -0.012, t(15291)= -4.65, p< .001) and COS (β= -0.009, t= -3.31, p= .001) significantly predict priming, so that the higher the semantic similarity, the shorter the RT to the target. Remarkably, we found no evidence of an interaction between priming and prime duration ($\chi^2_{(1)}$= 0.04, β= 0.001, t(16059)= 0.19, p= .848 and $\chi^2_{(1)}$= 0.99, β= 0.004, t(16059)= 0.99, p= .320, for PMI and COS respectively).

Next, we contrasted the two measures one against the other. Adding PMI improved the overall fit to the data relative to the model testing for COS in isolation ($\chi^2_{(1)}$= 10.96, p= .001), but not vice-versa ($\chi^2_{(1)}$= 0.29, p= .591). LMM analysis confirmed the strong facilitation determined by PMI (β= -0.011, t(13502)= -3.31, p= .001), while the COS effect drastically dropped off (β= -0.002, t(11574)= -0.54, p= .591). Again, there was no interaction between the observed PMI-led priming and different prime timing\SOA ($\chi^2_{(1)}$= 0.17, β= -0.002, t(16057)= -0.42, p= .677). Results are shown in Figure 8.

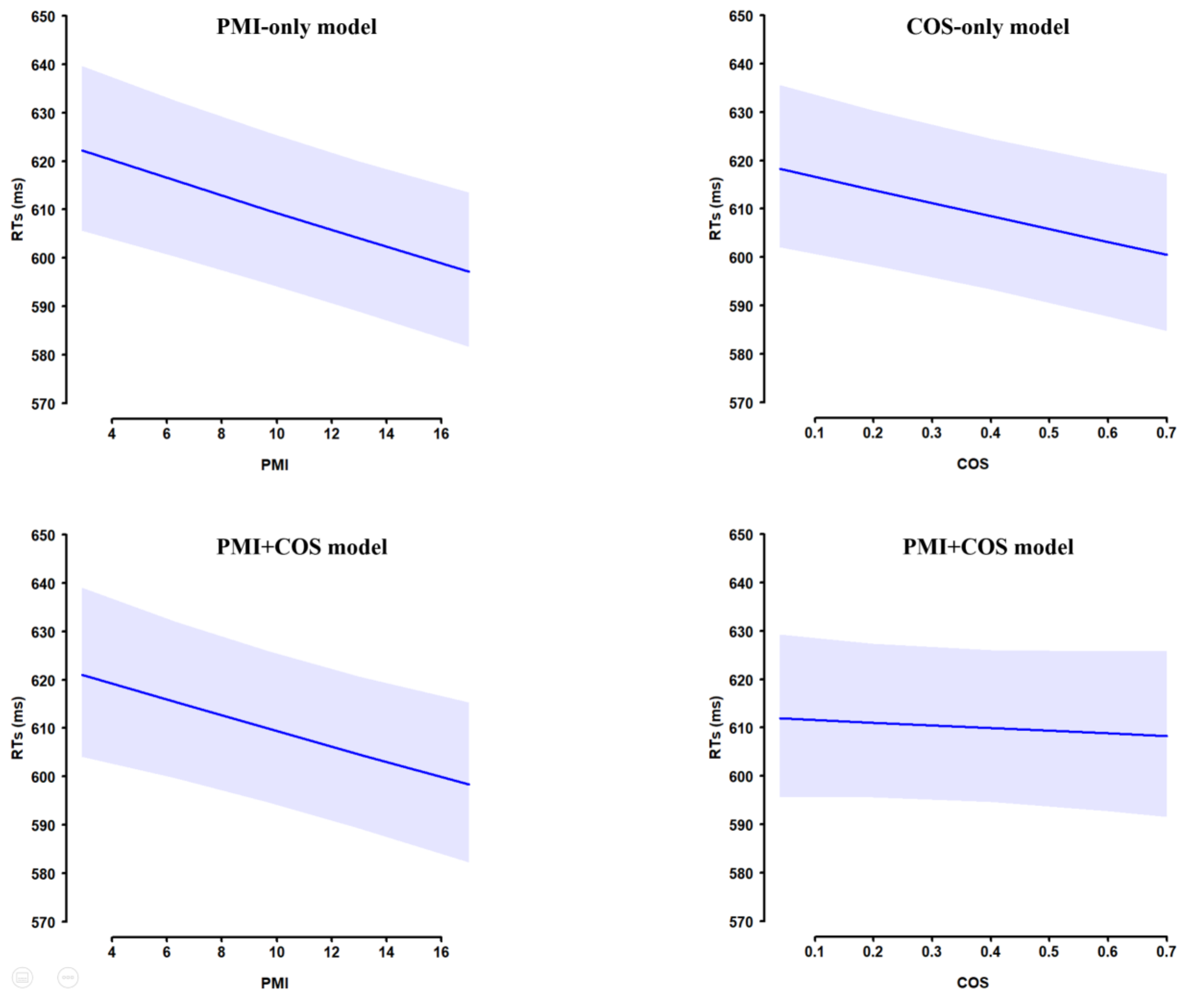

Figure 8. Significant effects of PMI (upper-left) and cosine proximity (upper-right) in isolated models. When the two predictors are contrasted, PMI (lower-feft) outperformed cosine proximity (lower-right). Shaded areas refer to 95% C.I.

Discussion

We perfectly replicated the results observed in the supraliminal condition of Experiment 1. Semantic facilitation was successfully accounted for by both distributional metrics, in line with the previous literature addressing semantic priming with computational linguistics tools (Günther, Dudschig, & Kaup, 2016; Mandera et al., 2017). Yet, when we contrasted PMI and COS, the former clearly outperformed the latter. This seems so be true irrespective of the time available to process the prime word.

General Discussion

This study represents a large-scale attempt at gauging semantic priming while modeling quantitatively and in a principled way different types of semantic relationships. To this aim, we considered a state–of–the–art Distributed Semantic Model (DSM), namely wordToVec (Mikolov et al., 2013), which track various types of high–level, long–distance semantic relationships (e.g., sofa-hammock, worm-caterpillar), and Pointwise Mutual Information (PMI), which specifically captures associative, more local ties (e.g., tank–paint, scissors–razor). In a series of experiments manipulating prime visibility and prime duration, we obtained the following core results:

(i) When we gauge semantic priming along the whole relatedness continuum, we do not observe a reliable effect; only when primes are at least partially visible facilitation starts to emerge.

(ii) When semantic relatedness is modeled dichotomically instead, thus contrasting strongly related prime–target pairs with unrelated ones, subliminal priming does seem to arise.

(iii) Overt priming is nicely accounted for by both DSM and PMI similarity, when these indexes are assessed in isolation; however, when the two are contrasted, PMI seems to provide a far better account of semantic facilitation.

(iv) This pattern of results is unaffected by prime duration; as long as the prime is visible, PMI dominates DSM.

It is not obvious what to make of (i) and (ii). On the one hand, they may just offer a methodological warning: dichotomizing naturally continuous variables may create effects that are not confirmed (or, at the very least, are much weaker) when the entire continuum is considered. We believe, however, that these results also carry an important theoretical message. Previous studies typically used words from small/closed classes (e.g., spatial words, planet names; e.g., Bottini et al., 2016; Quinn and Kinoshita, 2008), thus allowing explanations of the effect based on target predictability, or at least potentially limiting the scope of their conclusions. Conversely, here we drew stimuli from across the lexicon, and sampled form very large categories such as animals and tools. Together with the regression design, which considers all levels of semantic relatedness, these features make this study the widest–scope investigation to date of masked semantic priming. The fact that this approach does not result in solid subliminal priming casts doubts on a wide, across–the–lexicon processing of semantic information outside of awareness.

These results are in line with previous behavioral data suggesting a primary role for local linguistic ties in supraliminal semantic priming. Günther et al. (2016) showed that similarity estimates derived from a semantic space based on local context information (based on word-by-word matrix) predict priming better than those derived from a semantic space based on global context information (based on a word-by-document matrix). Similarly, Brunellière, Perre, Tran and Bonnotte (2017) probed that, while keeping semantic similarity constant, the magnitude of priming was greater as prime-target pairs co–occur more frequently.

These data are difficult to reconcile with theoretical accounts of priming based on automatic activation spreading within a semantic network coding for high–level, relatively complex relationships (Collins & Loftus, 1975; Neely & H., 1991). Taking PMI at face value, these results may suggest that priming is based on expectancy generation—the prime is taken as a cue for the coming target, and expectation is computed based on local, relatively simple association links. Interestingly, this makes connection with models of sentence processing, where it is very well established that upcoming words are predicted based on the current and previously encountered ones (Kuperberg & Jaeger, 2016). Perhaps, a similar mechanism is in action with isolated word priming; given that syntax and discourse level information is just not available, the reader is left with mere word–level prediction, for which PMI offers a nice metric. The lack of the same kind of results with masked priming would further suggests that this strategy requires awareness.

The operationalization of associative strength in terms of information conveyed by the prime-target pairs based on their weighted surface co-occurrence (what PMI codes for, essentially) may inform us about the nature of priming. The effect seems to be better explained by associative mechanisms that link lexical items in our mental lexicon, rather than by the activation of conceptual information in semantic memory. A similar perspective has been proposed by Recchia and Jones (2009), who showed that PMI-based similarity estimates collected from very large amounts of data more closely matched with human semantic similarity ratings than do several more complex models. Our results support these findings and provide further psychological validation of this modeling via semantic priming. What has been traditionally thought of as semantic processing could be largely an epiphenomenon of such processes. This would be in line with previous literature suggesting that the behavior of the human cognitive system may be effectively described by Information Theory principles aimed at transforming perception into information (Crupi, Nelson, Meder, Cevolani, & Tentori, 2018; Sayood, 2018).

Our study could speak in favor of a semantic match account of priming (Jones, 2010), according to which the effect would be due to a retrospective strategy applied by subjects who may check for a relationship between the two stimuli after target presentation. Unfortunately our best predictor, PMI, is by definition a symmetric measure, and therefore we cannot assess whether the prospective expectancy generation or the retrospective semantic match could better account for the current results.

Should we merely take these computational indexes as useful metrics that, for some reason, happen to reflect well human behavior? Or should we rather consider them as realistic models of how we come to acquire this information? The methodological advantages provided by distributional techniques are undeniable; not only they outperform (or match) similarity estimates from feature lists or association norms in accounting for a variety of language-related behaviors, but they are also much easier to collect and share. More importantly, all the measures developed within the distributional framework are based on an inferential mechanism that exploits the effective presence or absence of a given stimulus to predict the presence or absence of another stimulus. This learning procedure, that has a long tradition in cognitive psychology and neuroscience that traces back to Rescorla and Wagner (1972), can be observed in several biological and psychological systems. Therefore, it is not specific to language modelling but rather may offer a general mechanism of learning that humans exploit to pick up statistical regularities in the environment and construct complex conceptual representations (Günther, Rinaldi, & Marelli, 2019).

As a final remark, we would like to acknowledge that contrasting PMI and DSM is a rather gross oversimplification of the complexity of the human semantic system. We followed on several recent attempts (e.g., Mandera et al., 2017; Paperno et al., 2014) and tried to use the nice quantitative tools developed in the field of computational semantics to shed light on a psychological phenomenon, whose investigation, we believe, had suffered the lack of such tools, and the precision in defining constructs that they bring about. We think that this gave us important insight already—we saw here that subliminal semantic priming is not as clear as it might seem, and that overt priming is better accounted for by local associations rather than by general, higher–level semantic models. These latter, however, and particularly the metric that we specifically investigated here, capture a number of very different semantic relationships, which may well deploy their effect on priming (and, potentially, on several other meaning–based human behaviors) very differently from one another. Future work will try to dig deeper in this respect, and tease apart more precisely the mechanics that govern the human lexical–semantic system.

References

Abrams, R. L. (2008). Influence of category size and target set size on unconscious priming by novel words. *Experimental Psychology*. https://doi.org/10.1027/1618-3169.55.3.189

Anaki, D., & Henik, A. (2003). Is there a "strength effect" in automatic semantic priming? *Memory and Cognition*. https://doi.org/10.3758/BF03194385

Ansorge, U., Kiefer, M., Khalid, S., Grassl, S., & König, P. (2010). Testing the theory of embodied cognition with subliminal words. *Cognition*. https://doi.org/10.1016/j.cognition.2010.05.010

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The waCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-009-9081-4

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*.

Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*. https://doi.org/10.1162/coli_a_00016

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01

Bottini, R., Bucur, M., & Crepaldi, D. (2016). The nature of semantic priming by subliminal spatial words: Embodied or disembodied? *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000197

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/PBR.15.6.1035

Brunellière, A., Perre, L., Tran, T. M., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2016.1215479

Bueno, S., & Frenck-Mestre, C. (2002). Rapid activation of the lexicon: A further investigation with behavioral and computational results. *Brain and Language*. https://doi.org/10.1006/brln.2001.2511

Chambers, J. (2008). Software for Data Analysis. In *Springer*. https://doi.org/10.1007/978-0-387-75936-4

Church, K. W., & Hanks, P. (1989). *Word association norms, mutual information, and lexicography*. https://doi.org/10.3115/981623.981633

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*. https://doi.org/10.1037/0033-295X.82.6.407

Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2018). Generalized Information Theory Meets Human Cognition: Introducing a Unified Framework to Model Uncertainty and Information Search. *Cognitive Science*. https://doi.org/10.1111/cogs.12613

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*. https://doi.org/10.3758/s13428-012-0260-7

de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*. https://doi.org/10.1016/0749-596X(91)90012-9

De Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000074

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*. https://doi.org/10.1016/S0010-0277(99)00049-9

Deyne, S. De, Peirsman, Y., & Storms, G. (2008). Sources of Semantic Similarity. *Animals*.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Ferrand, L., & New, B. (2003). Semantic and Associative Priming in the Mental Lexicon. In *The Mental Lexicon: "Some Words to Talk About Words."* https://doi.org/10.1.1.3.705

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/a0032333

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology. General*, *124*(1), 22–42. https://doi.org/10.1037/0096-3445.124.1.22

Guida, A., & Lenci, A. (2007). Semantic properties of word associations to Italian verbs. *Italian Journal of Linguistics*.

Günther, F., Dudschig, C., & Kaup, B. (2016). Predicting lexical priming effects from distributional semantic similarities: A replication with extension. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01646

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619861372

Hinton, G., & Shallice, T. (1991). Hinton&Shallice(1991)PsychologicalReview_Lesioning an Attractor Network- Investigations of Acquired Dyslexia.pdf. *Psychological Review*.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196544

Jones, L. L. (2010). Pure Mediated Priming: A Retrospective Semantic Matching Model. *Journal of*

Experimental Psychology: Learning Memory and Cognition. https://doi.org/10.1037/a0017517

Kiefer, M., Adams, S. C., & Zovko, M. (2012). Attentional sensitization of unconscious visual processing: Top-down influences on masked priming. *Advances in Cognitive Psychology*. https://doi.org/10.2478/v10053-008-0102-4

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2007.2093

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2015.1102299

Lam, K. J. Y., Dijkstra, T., & Rueschemeyer, S.-A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00659

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*. https://doi.org/10.1037/0033-295X.104.2.211

Long, J. A. (2018). jtools: Analysis and Presentation of Social ScientificData. *R Package*.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03212999

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0809-y

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03204766

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2016.04.001

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*. https://doi.org/10.1016/0010-0285(83)90009-9

Marelli, M. (2017). Word-embeddings Italian Semantic spaces: A semantic model for psycholinguistic research. *Psihologija*. https://doi.org/10.2298/PSI161208011M

Martens, U., & Kiefer, M. (2009). Specifying attentional top-down influences on subsequent unconscious semantic processing. *Advances in Cognitive Psychology*. https://doi.org/10.2478/v10053-008-0067-3

McNamara, T. P. (2005). Semantic priming: Perspectives from memory and word recognition. In *Semantic Priming: Perspectives from Memory and Word Recognition*. https://doi.org/10.4324/9780203338001

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.24.3.558

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*. https://doi.org/10.3758/BF03192726

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013. *ArXiv Preprint ArXiv:1310.4546*.

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*. https://doi.org/10.1037/bul0000092

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496. https://doi.org/10.1016/j.tics.2014.04.009

Neely, & H., J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading: Visual word recognition*.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03195588

Paperno, D., Marelli, M., Tentori, K., & Baroni, M. (2014). Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2014.07.001

Pitler, E., Louis, A., & Nenkova, A. (2010). Automatic evaluation of linguistic quality in multi-document summarization. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.

Quinn, W. M., & Kinoshita, S. (2008). Congruence effect in semantic categorization with masked primes with narrow and broad categories. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2007.03.004

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and Correlation in Conceptual Structure: Behavioral and Computational Studies. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.30.2.393

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*. https://doi.org/10.3758/BRM.41.3.647

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*. https://doi.org/10.3758/BF03207490

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning. *Classical Conditioning II Current Research and Theory*. https://doi.org/10.1101/gr.110528.110

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*.

Sayood, K. (2018). Information theory and cognition: A review. *Entropy*. https://doi.org/10.3390/e20090706

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*. https://doi.org/10.1037/h0036351

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/BF01074363

Tversky, A. (1977). Features of similarity. *Psychological Review*. https://doi.org/10.1037/0033-295X.84.4.327

Van Den Bussche, E., & Reynvoet, B. (2007). Masked priming effects in semantic categorization are independent of category size. *Experimental Psychology*. https://doi.org/10.1027/1618-3169.54.3.225

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of Masked Priming: A Meta-Analysis. *Psychological Bulletin*. https://doi.org/10.1037/a0015329

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*. https://doi.org/10.1016/j.cogpsych.2003.09.001

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*. https://doi.org/10.3758/BRM.40.1.183

Chapter 3. Electrophysiological correlates of semantic processing as revealed by priming and distributional semantics

Introduction

Semantic knowledge represents a fundamental feature of human cognition: it allows us to assign meaning to different entities in the world, and consequently to make inferences about how to interact with these entities, as well as how they may interact between each other. Such knowledge is clearly reflected in language, as it allows speakers to extract meaning from the words stored in the mental lexicon and to link them via meaning-based similarity relationships.

Pivotal insights into the internal organization of the mental lexicon have been provided via semantic priming experiments (McNamara, 2005). This paradigm is based on faster recognition times when a *target* word (e.g., dog) is preceded by a semantically related *prime* word (e.g., cat) vs. a semantically unrelated one (e.g., cap). The word "semantic" in semantic priming implies that the observed facilitation is due to overlap in meaning between the two words. The effect is very robust, as it can be observed in a variety of tasks, such as lexical decision, semantic categorization or naming.

Originally, the observed facilitation was accounted for via spreading activation mechanisms: words are represented as nodes within an interconnected network, and links between nodes reflect lexical-semantic ties (Collins & Loftus, 1975). When a given word is read, the corresponding node is activated, and activation spreads to related nodes, proportionally to their association strength.

This latter has been traditionally computed by presenting subjects with a given seed word and asking them to produce one or more words that the seed brought to their mind (Nelson, McEvoy, & Schreiber, 2004). Association norms–the documents where these responses are collected–have been used in psycholinguistic research as significant predictor for the emergence of priming (Anaki & Henik, 2003). Despite their successful application, it is not clear what those norms represent as a psychological construct. Their definition is rather loose (participants can walk the semantic space from the seed word in any different way), and therefore they end up capturing several different types of relationships, like category membership (hare-rabbit)[6], collocation (keg-beer), synonymy (stone-rock), meronymy (cheddar-cheese), antonymy (north-south), scripts (cinema-movie), function (lock-key), even proper names of notorious entity (flipper-dolphin).

A different approach to model semantic association is based on featural similarity (McRae, De Sa, & Seidenberg, 1997). Under this view, lexical meaning is represented by means of features describing perceptual, functional and encyclopedic aspects of the corresponding referent. The more features two words share, the higher their semantic similarity. For example, the words 'dog' and 'wolf' are similar as the two entities they refer to share much of the same characteristics (have a fur, four legs, a tail, both yowl, etc.). Operationally, this approach relies on human participants performing feature-production tasks. Words are then encoded as vectors keeping track of the presence\absence of such features, and semantic similarity is numerically defined as the cosine of the angle between vectors (McRae, Cree, Seidenberg, & McNorgan, 2005).

While feature-based approaches performed quite well in modelling a wide range of language related behaviors (McRae & Boisvert, 1998; Vigliocco, Vinson, Lewis, & Garrett, 2004), they are not immune to criticism. For example, they are not perfectly suited to represent the semantic content of abstract entities, whose features can be quite difficult to define.

Computational Semantics now offer another approach to define and quantify semantic relationship, based on how words are used together in language. The main theoretical assumption behind this approach is that humans process words in relation to a context, i.e., words get their meaning due to the linguistic context they appear in (Lenci, 2008; Sahlgren, 2008). The idea is not new (Firth, 1957; Harris, 1954), but it has only recently become a critical aspect of contemporary research in Computer Science and Cognitive Neuroscience. Over the past two decades, great

[6] All the example are taken from the University of South Florida Free Association Norms (Nelson et al., 2004)

advancements have been made in the mathematical manipulation on word co–occurrence data and in the development of ever more precise estimates of word distributions in the language, mainly thanks to the development of larger linguistic corpora. In this approach, words themselves represent the organizational principles of the semantic system, making it possible to avoid the theoretical weakness of postulating a-priori a given set of semantic features. Moreover, similarity estimates can be obtained for most of the words attested in a text corpus, including abstract words of course; feature-lists and association norms, instead, are available only for a relatively limited set of stimuli.

More precisely, in distributional semantic models (DSMs; Günther, Rinaldi, & Marelli, 2019), lexical items are represented as vectors that populate a high-dimensional space where semantic relatedness is reflected in spatial proximity. Words with similar meaning tend to cluster together, and such similarity can be quantified by applying geometrical techniques to these vectors. For example, one can approximate relatedness as the cosine similarity (henceforth COS) formed by two word-vectors:

$$\mathrm{COS}\theta = \frac{a \cdot b}{||a|| \cdot ||b||}$$

DSMs have been proposed as a psychologically plausible models of semantic memory, with particular emphasis on how meaning representations are achieved and structured. In particular, state-of-art models (e.g., word2vec; Mikolov, Yih, & Zweig, 2013) represent a simple neural network consisting of an input, an output and a hidden layer, and is based on a predictive mechanism that allows to infer a target given a set of cue words. Thus, words are similar if they are similarly predicted in similar linguistic contexts. For example, in a sentence about pets, it's likely to encounter the word 'dog', as well as the word 'cat'. Word2vec provides similarity estimates that cover a wide range of classic lexical-semantic relationships, like synonymy (e.g. student-pupil, 0.54), antonymy (e.g. rich-poor, 0.57), meronymy (e.g. hound-dog, .53). Associative relations can be grasped as well (dog-leash, .50). Finally, it can account for featural similarity beyond category membership (e.g. eagle-hawk, .45 vs penguin-hawk, .19). Word2vec has been shown to perform better than (or as well as) other DSMs in a variety of task, such as synonym detection, concept categorization, semantic priming (Baroni, Dinu, & Kruszewski, 2014; Mandera, Keuleers, & Brysbaert, 2017).

Cosine similarity in vector models is not the only computational linguistic metric that one can use to measure semantic proximity/association. A more immediate way to model linguistic context is by looking at surface co-occurrence, i.e., simply counting how many times two are used close together. This approach can be psychologically interpreted as how strong of a cue word A is for word B (Spence & Owens, 1990). A useful mathematical tool to operationalize this assumption is Pointwise Mutual Information (henceforth PMI):

$$\mathrm{PMI}(w_1,w_2) = \log_2\frac{p(w_1,w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ corresponds to the probability of occurrence of the word pair in a given window of test (e.g., five consecutive words), while $p(w_1)$ and $p(w_2)$ are the individual probabilities of occurrence of the two words in isolation (Church & Hanks, 1989). PMI has been used to model a wide range of psycholinguistics phenomena, such as similarity judgements (Recchia & Jones, 2009), reading speed (Ellis & Simpson-Vlach, 2009), and free association and syntactic parsing (Pitler, Louis, & Nenkova, 2010). Moreover, PMI has also been shown to successfully generalize to non-linguistic fields, such as reasoning and induction (e.g., Paperno, Marelli, Tentori, & Baroni, 2014).

So, cosine similarity and PMI allow us to investigate semantic processing with tools that provide a precise and consistent definition, and therefore a neat quantification, of the word relationships that govern meaning construction. Importantly, they also seem to roughly map onto different psychological constructs that were heavily investigated in the past: while PMI seems to specifically track local associations, COS more generally captures a variety of higher–level relationships (e.g., category membership, feature similarity, synonymy, antonymy) that most often do not result into direct co–occurrence in language use. The paper builds onto these considerations, and addresses the processing of psychologically relevant aspects of word meaning via rigorously defined mathematical tools.

One important aspect in which the different facets of word meaning tracked by PMI and COS may differ is timing. If PMI truly tracks local, relatively shallow associations, one might imagine that its effect will deploy quickly after word presentation; while perhaps the complex, higher–level relationships captured by COS may take more time to emerge. To keep track of the time-course of the processes underlying priming, we recorded participant's EEG signal. Several event related potentials (ERPs) have been associated with language related phenomena; in particular, the N400, has been acknowledged as an index of lexical and semantic processing (Lau, Phillips, & Poeppel, 2008). In the context of priming, N400 reflects a more pronounced negativity for unrelated primes compared to related ones, typically emerging in a time window between 300 ms to 500 ms after word onset.

There is no unique interpretation of what kind of processes are reflected by N400. At least two major components seem to be at stake: accessing long-term representations of words' meaning and integrating such representations into a more complex mental structure. Early explanations defined the effect in terms of semantic match between a target word and the preceding context, in sentences like "He spread the warm bread with *butter/socks*" (Kutas & Hillyard, 1980). Later results challenged this interpretation; while controlling for semantic congruency with the preceding context, N400 seemed to track the likelihood with which a given target was expected, like in 'Don't touch the wet *paint/dog*' (Kutas & Hillyard, 1984).

The debate on the N400 interpretation is still open today, although the focus has moved somewhat on whether the N400 modulation reflects information processing at the semantic or at the lexical level. According to the integration theory (Federmeier & Kutas, 1999; Kutas & Federmeier, 2011), the semantic features associated with the upcoming target are preactivated, making the integration with the preceding context less effortful. Conversely, the prediction theory posits that N400 truly reflects pre-activation of the critical word itself, resulting in an easier lexical access

(Bornkessel-Schlesewsky & Schlesewsky, 2019; Lau, Namyst, Fogel, & Delgado, 2016; Szewczyk & Schriefers, 2018).

Interestingly, PMI and COS are particularly fit to attack this debate. While cosine proximity should mostly represent relatively high–level semantic aspects of word representation, PMI may more genuinely reflect association/prediction at a pure lexical level.

Of course, this is not the first attempt at looking at the electrophysiological correlates of semantic similarity from a distributional perspective. In an MEG study, Parviz, Johnson, Johnson, & Brock (2012) tested several variables as possible predictors for the emergence of N400m, the neuromagnetic analog of N400. They define the strength of the link between a given sentence and the corresponding final word in terms of surprisal and semantic congruency. The former was operationalized as the likelihood with which the ending word was expected given the preceding context, based on co-occurrence patterns emerging from a large text corpus. The latter was implemented with Latent Semantic Analysis representations derived from word-by-documents matrices, that is, matrices keeping track of how words distribute across the several different documents (e.g., books, newspaper articles) that were considered in this model (Landauer & Dumais, 1997). Crucially, both the metrics could successfully account for the modulation of the MEG signal in the N400 time window. Similarly, Frank and Willems (2017) showed how semantic similarity—i.e. word2vec similarity estimates—and word expectancy—i.e. probability estimate based on the preceding words—elicit distinct patterns of brain activity as revealed by fMRI data, despite such difference was not attested at the ERP level.

Yet, differently from these studies that analyzed the N400 from a computational linguistic perspective in a sentence context (see also Ettinger, Feldman, Resnik, & Phillips, 2016), the current work attempts to dip further into this issue with isolated word processing. In addition to setting a bridge with the vast behavioral literature that is dominated by individual word experiments, this adds a further element of interest—we check whether and how metrics that are based on how words go together in language deploy their effects when words are presented in isolation, without any broader contextual information. More specifically, we designed a priming experiment where we contrast related and unrelated prime–target pairs in three conditions: (i) association is quantified via PMI, while COS is controlled for; (ii) association is quantified via COS, while PMI is controlled for; (iii) association is quantified via both PMI and COS, so that related and unrelated primes are such on both metrics. With this design, we hope to identify the separate contribution and timing of relatively shallow, associative ties (PMI) vs. higher–level, more abstract semantic relationships (COS), as well as their eventual interaction (through the PMI+COS condition).

The experiment

Method

Participants. 30 students at the University of Trieste were recruited into the experiment (12 males, 18 females; mean age=25y, age range=20y-32y). All subjects were right-handed, native Italian speakers, and had normal or corrected-to-normal vision and no history of neurological disorders. Subjects gave written informed consent for participation, and received 15 Euros in exchange for their time.

Design. The experiment was based on a 2-by-3 design comparing congruent and incongruent prime-target pairs across 3 categories that differed with regard to the type of semantic similarity linking the two words. Target words, that were not the same across categories, were paired with one congruent and one incongruent prime. Participants saw all the prime-target pairs once in the experiment.

Material. Ninety Italian words were selected to be used as target stimuli and were equally divided (N=30) across three categories, PMI , COS and PMI+COS. Each target was paired with one related and one unrelated prime (e.g., PMI: cheese\monument-MOUSE; COS: lamp\missile-TORCH; PMI+COS: prawn\veal-CRAB)

PMI was computed by first collecting cooccurrence data by means of a 2–words window sliding along the Itwac corpus, a lemmatized and part–of–speech annotated database for Italian of nearly 2 billion words (Baroni et al, 2009). All characters were set to lowercase, and special characters were removed together with a list of stop-words. The raw counts were subsequently transformed into PMI scores according to the following equation:

$$\mathrm{PMI}(w_1,w_2) = \log_2 \frac{p(w_1,w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ represents the probability of encountering the two words within the same 2–word window, and $p(w_1)$ and $p(w_2)$ represents the overall probability of encountering w_1 and w_2.

Cosine proximity between word vectors was obtained training a *word2vec* model (Mikolov, Chen, Corrado, & Dean, 2013) on the same corpus. Model's parameters were set according to (Marelli, 2017). All words attested at least 100 times were included in the model, which was trained using the continuous-bag-of-word (CBOW) architecture, based again on a 5-word window and on 200 dimensions. The parameter k for negative sampling was set to 10, and the subsampling parameter to 10^{-5}. Among the two different architectures implemented in word2vec, CBOW has been proven to gain better results than Skip-Gram in semantic priming simulations (Baroni et al., 2014). Negative sampling reduces the computational load of the model by selecting a restricted set of items in the output layer for each learning phase, when the probabilities are estimated. Subsampling allows the model to reduce the influence of very high–frequency words, which are known to provide little information for distributional analysis.

In order to test for the specific contribution to the emergence of priming provided by semantic similarity as indexed by PMI and COS, we constructed the categories so that the two indexes could be kept as separated as possible. That is, we ensured that when testing for one variable (e.g. PMI), the other (e.g. COS) was as matched as possible across the comparison. Thus, in the PMI category, average PMI for related and unrelated conditions was 7.77 (sd 1.17) and 0.13 (sd 0.69) respectively, while average COS was 0.17 (sd 0.04) and 0.13 (sd 0.04) respectively. Viceversa, in the COS category, average PMI for related and unrelated conditions was 1.80 (sd 1.76) and 0 (sd 0) respectively, while average COS was 0.43 (sd 0.04) and 0.13 (sd 0.05) respectively. Finally, in the PMI+COS category, related pairs had an average value of 8.79 (sd 1.97) and 0.45 (sd 0.12) for PMI and COS respectively, while unrelated pairs had an average value of 0.21 (sd 0.83) and 0.12 (sd 0.04) for PMI and COS respectively. Figure 1 shows the distribution of the two metrics across the three categories.

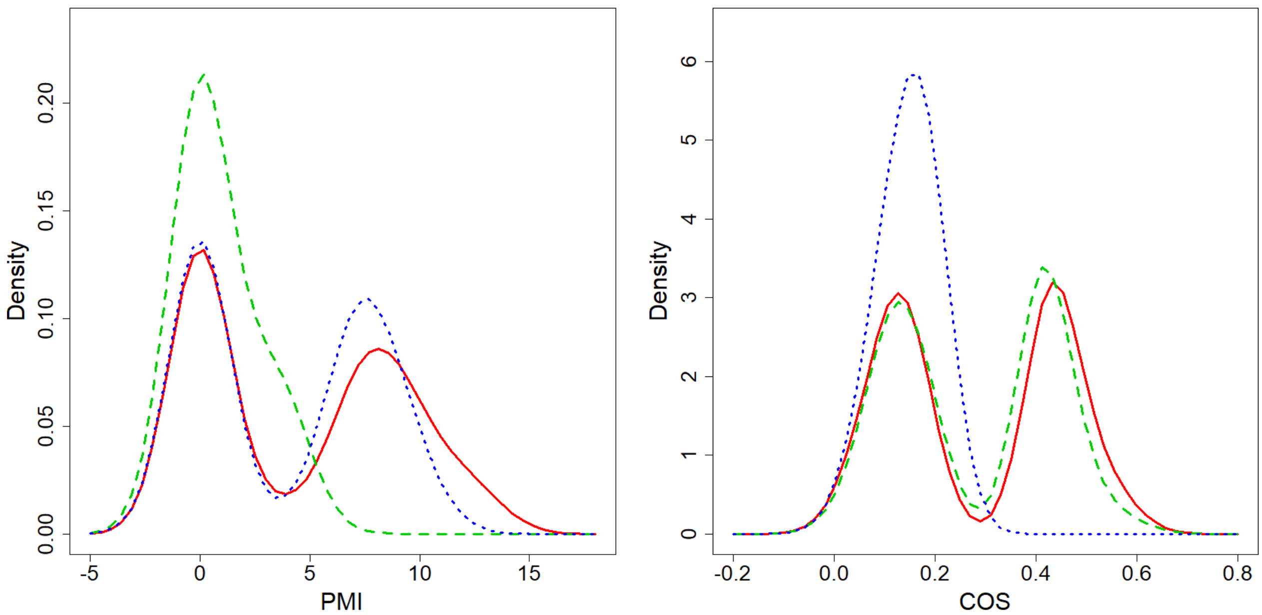

Figure 1. Distribution of the semantic indexes considered (PMI, left; COS, right)

Primes and targets in the three categories were matched in frequency, length, and orthographic neighbourhood, as shown in Table 1.

	Prime Frequency	Prime Length	Target Frequency	Target Length	Prime OLD20	Target OLD20
PMI	4.18 (0.35)	6 (1)	3.68 (0.33)	6 (2)	1.73 (0.52)	1.95 (0.59)
COS	4.17 (0.41)	6 (1)	3.62 (0.36)	7 (1)	1.77 (058)	2.05 (0.49)
PMI+COS	4.20 (0.40)	6 (1)	3.67 (0.32)	7 (1)	1.76 (0.53)	2.12 (0.37)

Table 1. Prime and target lexical features - mean (sd).

Finally, 90 non-word targets were constructed by shuffling the letters from the target words and recombing them without violating phonotactic rules (e.g., *tabio* < *abito*). Each non-word target was paired with two word primes, different from those used in the word-trials. Thus, the word/non-word target ratio was equal to .5.

Procedure. Participants performed a lexical decision task, requiring them to assess whether the target stimulus was an existing Italian word. Stimuli presentation was done using using MatLab Psychtoolbox (Brainard, 1997). All words were shown in Arial font, 32 in size, in white against a black background, displayed on a 22’’ monitor with a refresh rate of 120 Hz. Responses were collected by keyboard press. The experiment comprised 4 blocks of 90 trials. Each trial started with a fixation point (+) displayed for 500ms. Then, the prime was shown for 200ms, followed by a 100ms blank screen, and then by the target, which stayed on screen for 1000ms. Finally, a question mark (?) was presented, triggering the participants to respond (see Figure 2).

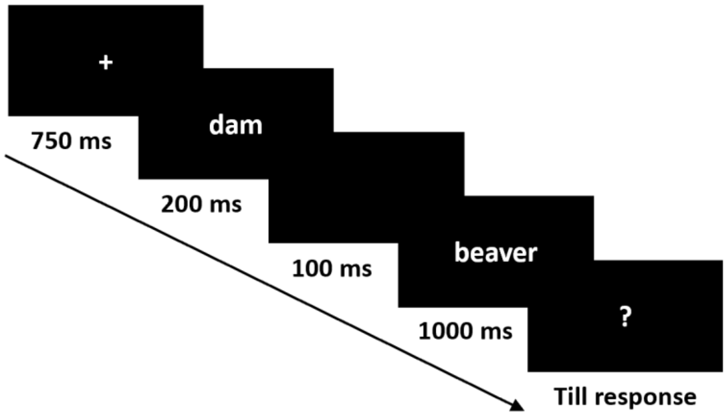

Figure 2. Exemplar trial of the experiment.

The delayed–response design prevented us from performing an analysis of the response times (RT), but crucially avoided motor interference in the target-related EEG signal. Each participant was provided with a few practice trials before the actual experiment, and s\he was invited to have a one-minute break between blocks.

EEG Recording. Data acquisition was conducted via a Biosemi ActiveTwo system. Throughout the experiment, EEG signal continuously recorded from a 128-electrode cap at a sampling rate of 1024 Hz. All electrodes were referenced to a common mode sense (CMS) electrode and grounded to a driven right leg (DRL) passive electrode.

EEG Preprocessing. Preprocessing was implemented using EEGLab (Delorme & Makeig, 2004). Out of the 30 participants who took part into the experiment, two were excluded for technical problem in the recording (prime triggers were missing), and three for a noisy signal. Data were first filtered with 0.1 Hz high-pass and 40 Hz low-pass filters, and resampled at 256 Hz. The continuous recording was segmented into 1500ms epochs, from 500ms before the onset of the target until its offset. Noisy channels (~9 per subject) were removed and ICA was run to detect blinks and ocular movements; automatic artifact correction was performed via ADJUST (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). Data were then re-referenced to the average activity at all electrodes, and baseline corrected. Automatic epoch rejection was conducted by removing epochs during which the signal exceeded the limit of ±100mV in any of the channels (7.9% of the data). Finally, missing channels were interpolated from neighboring electrodes, and grand-averages per subject per condition were computed.

Statistical analysis. Comparisons between the conditions of interest at the group level were conducted on the preprocessed EEG data via non-parametric cluster based permutation test (CBPT) as implemented in the FieldTrip toolbox (CBPT; Maris & Oostenveld, 2007). This analysis allows to tackle the multiple comparison problem in a straightforward manner. Due to the spatio-temporal structure of EEG data, a reliable effect should be attested across different electrodes and time bins. Rather than checking for differences between conditions point-wise, which would result in a huge number of comparisons, CBPT groups together observations that are close in both space and time. More precisely, for each condition, single channel-by-time observations are statistically compared via a t-test. The t values of adjacent spatio-temporal points with p values < 0.05 are grouped together and a cluster-level statistic is computed; in our case, we used cluster-mass, which is the sum of the t-values within the cluster. The next step is to compute the distribution of the cluster size under the null hypothesis of no difference between conditions. This is achieved via non-parametric permutation test: conditions are shuffled, and cluster-level statistics are computed again. This step is repeated several times (e.g. 2500) and on each iteration the highest cluster-mass is retained. Finally, cluster level p values are calculated as the proportion of cluster-mass resulting from the null hypothesis that are higher than the observed one.

In order to assess the reliability of the group level results, we additionally performed a test at the subject level. For each subject, we extracted the activity averaged over space and time as determined by the group-level cluster. Conditions of interest were compared via t-test, and corresponding t values were then set to 1 if they matched the difference observed at the cluster level, or to 0 if they did not. Finally, these transformed t values underwent a one-tailed binomial test. With this analysis, we could assess the strength of an effect observed at the group level by looking at how many participants show a difference between conditions in the same direction.

Open practices statement. All data and analysis code are available at https://osf.io/qs4fr/, and can be accessed independently from the authors.

Results

The cluster-based permutation tests were run across all electrodes in the N400 time window (300-500 ms) for each category separately. The analyses revealed a significant main effect of Relatedness in the PMI category (p= .034, g= 0.42483 [0.10924 - 0.75869], significant time window= 379 ms–426 ms). The topography corresponding to this effect (Figure 3-A) is broadly suggestive of an N400, being particularly pronounced over centro-frontal electrodes. No such difference was observed when comparing related and unrelated conditions in the COS category (p= .680). Conversely, a significant difference between related and unrelated conditions emerged in the PMI+COS condition (p= .032, g=0.45232 [0.1569 - 0.7672]; significant time window=309 ms–383 ms). As shown in Figure 3-C, the negativity was particularly prominent over central electrodes—again, roughly consistent with a classic N400 effect.

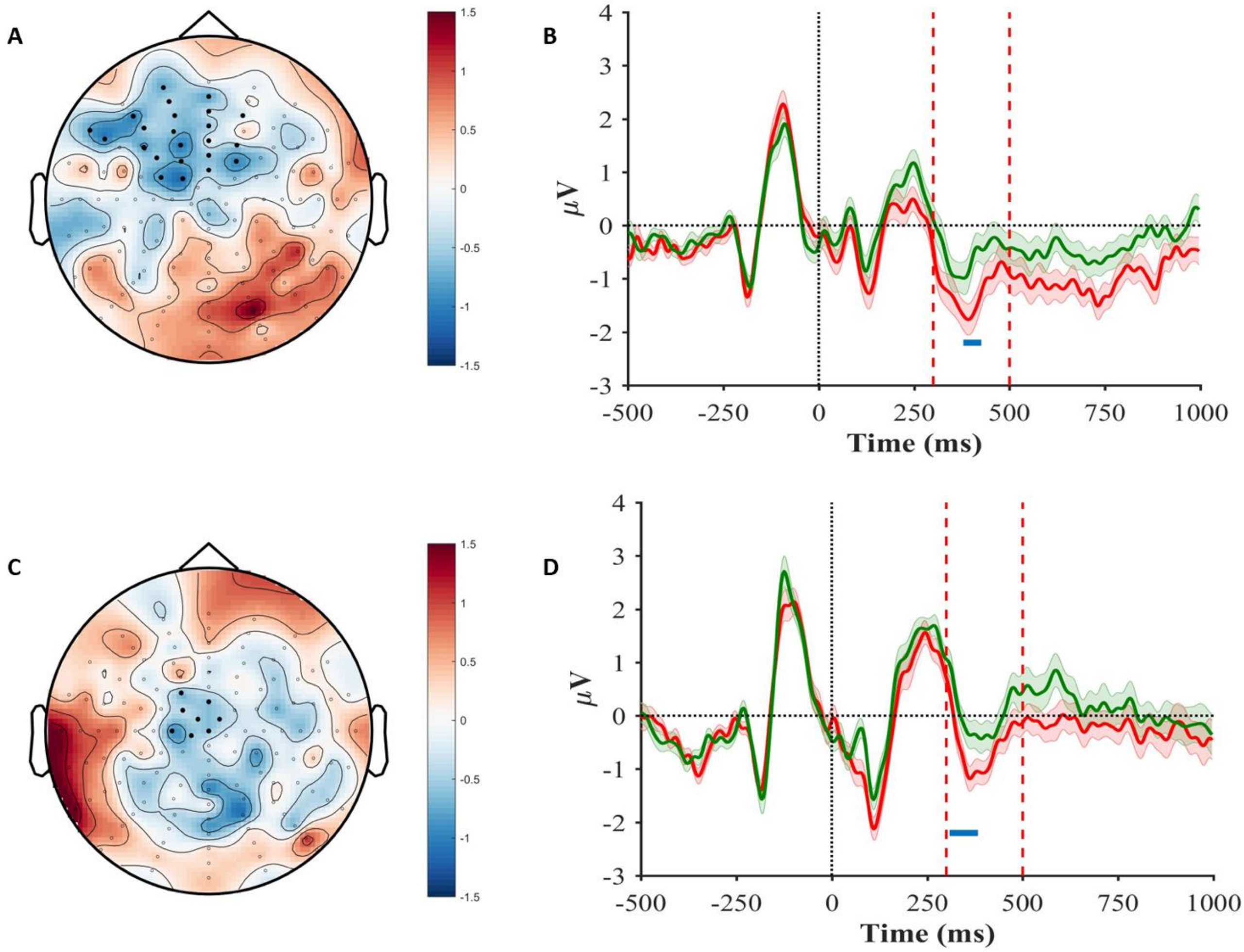

Figure 3. A N400 response was observed for word pairs related according to the PMI (A B) and for word pairs related on both cosine proximity and PMI (C D). On the right, grand averages over significant electrodes in the CBPT. Time zero indicates the onset of the target word. Shaded areas denote 95% CI. Vertical dashed red line delimits time window of analysis. Horizontal solid blue line indicates p < 0.05 (cluster corrected).

Results from individual participants mirrored the results observed at the group level. This analysis showed that the majority of the participants displayed a difference between conditions in the direction congruent with the tested hypothesis in both PMI category (22/25, 88%, *p*< .001) and PMI+COS category (18/25, 85.71%, *p*= .021).

Discussion

In this paper we investigated the electrophysiological correlates of semantic priming taking advantage of computational linguistics metrics that allow for a neat definition of the specific relationship linking primes and targets. Relatedness was defined as either local association between words, as captured by Pointwise Mutual Information (PMI); or neighborhood in a multi-dimension semantic space, as tracked by cosine similarity (COS) in a word2vec model for Italian built on the same corpus. We also considered a third condition where prime–target relatedness was based on both metrics, so as to assess their eventual interaction and\or additive effect. We recorded participants' EEG signal while they were performing a primed lexical decision task, and analyzed the data at the ERP level, focusing on the N400 component. While a robust effect emerged for locally associated words (PMI), with incongruent trials eliciting a higher negativity over fronto-central electrodes, the effect for semantic neighbors (COS) was quite weaker, and did not reach significance. Yet, when items were both strong associates and close in the semantic space (PMI+COS), N400 was observed again, with a slightly different topography, though, more posterior than in PMI alone and mostly driven by central electrodes. A slight difference in time also emerged, with a slightly earlier effect for PMI+COS pairs as compared to the PMI only condition.

Overall, these results suggest that semantic priming in the brain is primarily driven by local association. In a review of 26 papers addressing semantic and associative priming, Lucas (2000) demonstrated that purely semantic relationships tend to elicit smaller effect sizes than associative ones, and put forward the idea of an "associative boost"—priming would be stronger when an associative relationship top up a semantic tie. Here we show that associative priming is stronger than semantic priming even when the two are tested independently. More recently, Brunellière, Perre, Tran, & Bonnotte (2017) showed that semantic priming was boosted when the primes and the targets co-occurred frequently. Other studies modeling semantic similarity as a continuous variable corroborated these results. Günther, Dudschig, & Kaup (2016) showed that similarity estimates derived from a semantic space based on local context information predict priming better than those derived from a semantic space based on global context information. Our own work brought behavioral evidence in support of these claims; in the previous chapter, we tested how PMI and cosine proximity perform in accounting for response times in a set of sematic priming experiments, and the former systematically outperformed the latter, independently of prime visibility and duration.

These considerations would suggest that the cognitive and neural mechanics behind semantic priming are not primarily driven by spreading activation, or feature overlap, or, more generally, by the way the semantic network is arranged in the brain. Rather, the prime is taken by the system as a cue to the target, and the information that this cue activates is primarily associative in nature—more than predicting semantically similar words, or category associates, or synonym (which it may surely activate, to some extent), the prime predicts words with which it often co–occur. This interpretation of semantic priming, at an even more general level is in line with previous literature suggesting that the behavior of the human cognitive system may be effectively described by Information Theory principles, aimed at transforming perception into information (Crupi, Nelson, Meder, Cevolani, & Tentori, 2018; Paperno et al., 2014).

Another interesting insight coming from these data is that the effects of association/PMI and semantic relatedness/COS do not simply sum up; it is not simply the case that the brain reacts more strongly to prime–target pairs that are related both on PMI and COS. Rather, the brain pattern seems to change *qualitatively*—priming in the PMI+COS condition emerged earlier and was captured by more posterior electrodes as compared to priming in the PMI–only condition. Although different time–space distributions cannot be directly mapped onto different cognitive processes, this observation does suggest that local association and higher–level semantic relatedness interact in a complex way. Perhaps, the presence of a semantic tie potentiates dynamics in the semantic network, thus reducing the dominance of the more shallow predictive process suggested in the previous paragraph.

Our results also shed light onto the nature of the information processing behind the N400 component. They do not seem to sit well with theoretical accounts according to which the modulation of this ERP is primarily due to semantic integration. The lack of a significant difference between related and unrelated condition in the COS condition, where congruent prime-target pairs were close in semantic space but not predictively related, rather suggests that N400 is first and foremost an index of lexical access, and particularly of word prediction. Several studies reported larger N400 responses for semantically incongruent words relative to semantically congruent ones. However, they might have mixed up congruity and predictability, making the congruent condition also highly predictable given the preceding the context—indeed, the two correlate quite strongly. However, when predictability and semantic relatedness are disentangled, like in the present study, the former is clearly a stronger modulator of N400.

Furthermore, the different topographies in the N400 window for the PMI and PMI+COS condition lend support to suggestions that the N400 is hardly a unitary component. Lau et al. (2016), for example, demonstrated that predictability highly affected the amplitude of N400, while semantic congruity resulted in a smaller effect, and with a quite different distribution. More precisely, the effect of predictability could be observed at electrode Fz, where instead semantically congruent and incongruent conditions could not be distinguished. Szewczyk and Schriefers (2018) showed that an already predicted target word that was semantically incongruent with the preceding text, still did not elicit N400.

As a final remark, we want to stress that the current results were obtained using a lexical decisions task, and semantic priming is known to be highly task dependent (De Wit & Kinoshita, 2015). On the one hand, this makes these data even more interesting and convincing: lexical decision, in fact, typically yields weaker semantic effects (than semantic decision tasks, for example); and yet, we find solid brain signatures for semantic priming here. On the other hand though, we cannot exclude that using a task tapping more explicitly on word meaning may facilitate the activation of semantic features proper, eliciting a stronger effect for COS as well. Similarly, varying the stimulus onset asynchrony between the stimuli, and thus giving participants more time to process the prime, can affect the observed results. For example, Lam, Dijkstra and Rueschemeyer (2015) found that action similarity (i.e., similarity in how objects are manipulated; e.g. piano-typewriter) elicited priming already at a SOA of 100 ms, while facilitation from visual similarity (e.g. pizza-coin) emerged only at a SOA of 1000 ms. Again, it is possible that allowing for a longer processing of the prime may elicit an effect in the COS category. More research is clearly required to address these issues.

References

Anaki, D., & Henik, A. (2003). Is there a “strength effect” in automatic semantic priming? *Memory and Cognition*. https://doi.org/10.3758/BF03194385

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.00298

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brunellière, A., Perre, L., Tran, T. M., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2016.1215479

Church, K. W., & Hanks, P. (1989). *Word association norms, mutual information, and lexicography*. https://doi.org/10.3115/981623.981633

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*. https://doi.org/10.1037/0033-295X.82.6.407

Crupi, V., Nelson, J. D., Meder, B., Cevolani, G., & Tentori, K. (2018). Generalized Information Theory Meets Human Cognition: Introducing a Unified Framework to Model Uncertainty and Information Search. *Cognitive Science*. https://doi.org/10.1111/cogs.12613

De Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent:

Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000074

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. https://doi.org/10.1016/j.jneumeth.2003.10.009

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/CLLT.2009.003

Ettinger, A., Feldman, N. H., Resnik, P., & Phillips, C. (2016). Modeling N400 amplitude using vector space models of word representation. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society (CogSci 2016)*.

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1999.2660

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In *Studies in Linguistic Analysis*.

Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1323109

Günther, F., Dudschig, C., & Kaup, B. (2016). Predicting lexical priming effects from distributional semantic similarities: A replication with extension. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01646

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation

From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619861372

Harris, Z. S. (1954). Distributional Structure. WORD. https://doi.org/10.1080/00437956.1954.11659520

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*. https://doi.org/10.1016/0301-0511(80)90046-0

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*. https://doi.org/10.1038/307161a0

Lam, K. J. Y., Dijkstra, T., & Rueschemeyer, S.-A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00659

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*. https://doi.org/10.1037/0033-295X.104.2.211

Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A Direct Comparison of N400 Effects of Predictability and Incongruity in Adjective-Noun Combination. *Collabra*. https://doi.org/10.1525/collabra.40

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2532

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

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03212999

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2016.04.001

Marelli, M. (2017). Word-embeddings Italian Semantic spaces: A semantic model for psycholinguistic research. *Psihologija*. https://doi.org/10.2298/PSI161208011M

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. https://doi.org/10.1016/j.jneumeth.2007.03.024

McNamara, T. P. (2005). Semantic priming: Perspectives from memory and word recognition. In *Semantic Priming: Perspectives from Memory and Word Recognition*. https://doi.org/10.4324/9780203338001

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*. https://doi.org/10.3758/BF03192726

McRae, K., De Sa, V. R., & Seidenberg, M. S. (1997). On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.126.2.99

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013. *ArXiv Preprint ArXiv:1310.4546*.

Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous spaceword representations. *NAACL HLT 2013 - 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Main Conference*.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03195588

Paperno, D., Marelli, M., Tentori, K., & Baroni, M. (2014). Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2014.07.001

Parviz, M., Johnson, M., Johnson, B., & Brock, J. (2012). Using language models and latent semantic analysis to characterise the N400 m neural response. *Clinical EEG and Neuroscience*. https://doi.org/10.1177/1550059412444821

Pitler, E., Louis, A., & Nenkova, A. (2010). Automatic evaluation of linguistic quality in multi-document summarization. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*. https://doi.org/10.3758/BRM.41.3.647

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/BF01074363

Szewczyk, J. M., & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1401101

Chapter 4. Semantic priming in neglect patients

Introduction

Nowadays, the idea that words presented below the threshold for conscious perception can activate cognitive representations is uncontroversial. In particular, lexical meaning is generally held to be accessed outside of awareness (Kouider & Dehaene, 2007)—evidence from priming experiments suggest that words can be processed up to the semantic level even when the speaker/reader did not perceive them consciously. This paradigm shows how words are recognized faster if preceded by a semantically related prime (cat-DOG) rather than a semantically unrelated prime (cap-DOG).

In order to test unconscious processing, this paradigm is often used with prime words presented very briefly (for at most 50ms), sandwiched between visual masks (e.g., a string of hashmarks, "#########", or a random string of letters, e.g., "aljfkhs", or the target word itself). This paradigm is specifically called "masked priming". Despite participants are generally unaware of the presence of the primes, these can still make semantic judgments on the subsequent target words faster. For example, Perea and Rosa (2002) observed that category coordinates (table-CHAIR; dog-CAT) elicited similar priming both in visible and in masked conditions. Similarly, Bueno and Frenck-Mestre (2008)reported faster response times (hereafter, RT) to targets that were preceded by prime words with a high overlap in semantic features (yacht-SHIP; eagle-HAWK), again independently of their visibility. Thus, it may seem that semantic representations are accessed similarly with and without awareness.

Conversely, other studies described different patterns of semantic facilitation depending on whether the prime was available to conscious report. Gomez, Perea and Ratcliff (2013) provided behavioral and computational evidence that masked and unmasked priming involve different cognitive processes. Some studies reported weak (if any) priming in the masked condition, while facilitation clearly showed up if primes were visible (Brunellière, Perre, Tran, & Bonnotte, 2017;

Montefinese, Buchanan, & Vinson, 2018). The effect was also shown to be task dependent, as its emergence is most often attested when participants are engaged in a semantic task rather than in lexical decision or naming (De Wit & Kinoshita, 2015)

As outlined above, masked and overt semantic priming data are mixed, and it is not clear whether semantic relationships are processed in the same way with or without awareness. Furthermore, it is not clear what kind of semantic information can be extracted subliminally, nor the depth of processing up to which it may undergo. For example, *cat* may prime *dog* due to feature overlap (they are both furry, have four legs, are kept as pets by humans; Quinn and Kinoshita, 2008), or due to category membership (animals; Abrams, Klinger & Greenwald, 2002), or due to associative strength (which is also reflected in their high co–occurrence in language use; Anaki & Henik, 2003). While these different aspects of lexical meaning are accessible when words are conveyed above the threshold for conscious perception, the unconscious reader may only grasp part of them.

Indeed, visual masking is only one of several techniques to make stimuli "invisible", each with its own relative strengths and weaknesses (see Kim & Blake (2005) for an exhaustive review). Awareness may be also disrupted by visual crowding (Whitney & Levi, 2011), or by bistable perception, as in binocular rivalry (Tong, Nakayama, Vaughan, & Kanwisher, 1998). Similarly, overloading participants' attentional resources may fail them to report the presence of a given stimulus, as in the attentional blink paradigm. The choice of a specific method may affect the overall results. For example, in a study comparing unconscious processing under continuous flash suppression (CFS) and meta-contrast masking, while keeping stimuli and tasks the same, Peremen and Lamy (2014) found that unconscious processing was substantial with meta-contrast masking, but absent with CFS.

Crucially, all the aforementioned techniques represent psychophysical "tricks" that induce unawareness experimentally. However, unawareness also emerges spontaneously in several real–life situations, and, in some cases, it is even un unfortunate stable trait of individuals who have suffered a psychological and/or neurological trauma. For example, brain–damaged patients, particularly when the neurological insult has affected the right parietal lobe, may present a complex syndrome whose fundamental feature is the failure to report consciously events that happened in the contralateral (most often, left) visual hemifield (Corbetta & Shulman, 2011). Of course, Spatial Neglect is a much more complex syndrome than the characterization we offered above. The deficit can hit the visual domain only, or multiple senses (Beschin, Cazzani, Cubelli, Della Sala, & Spinazzola, 1996). It can also affects imagination, in addition to perception (Bisiach & Luzzatti, 1978). Moreover, patients may or may not have motor symptoms (Punt & Riddoch, 2006). The core feature of the syndrome, however, remains the inability to report events in the hemifield contralateral to the lesion; and this is the feature of interest in this study.

It is well-known that neglected stimuli are not simply ignored, but they activate cognitive representations that seem to exert an influence upon high-level cognitive processes. Marshall and Halligan (1988) reported the case of a patient who was shown simultaneously with two pictures of a house, one of which had its left side on fire. While she did not report any difference between the two, when asked to choose which house she would prefer to live in, she consistently manifested preference for the one spared by the flames.

Other studies directly tested if a stimulus, and particularly a word, presented in the left hemifield of a neglect patient can be processed up to semantic level. In a single case study, Làdavas, Paladini and Cubelli (1993) found that centrally presented target words were primed by related words that were presented in the neglected hemifield (silver-GOLD). Similar results were provided a few years later by McGlinchey-Berroth et al. (1996) in a group study involving seven patients. More recently, Sackur et al. (2008) tested a group of four patients in a magnitude judgement task, where each target number was preceded by a number prime that was presented either in the neglected or in the intact hemifield. Priming emerged independently of prime position, both at the group and at the single subject level.

Thus, there seems to be evidence supporting semantic processing of neglected words. Yet, all the aforementioned studies are not exempt of problems. In McGlinchey-Berroth et al. (1996) and Sackur et al. (2008), for example, primes were presented only 1.5 or 2 degrees of visual angle, respectively, to the left of the central targets. It is not obvious, then, that participants were entirely unaware of them—the separation between the visible and the invisible hemifield is never abrupt, of course, and this close distance from the center of the visual field may have left some partial conscious access available.

This was not an issue in the study by Làdavas et al. (1993), where primes were presented 5.5 degrees of visual angle away from the center of the visual field. However, these authors used different semantic relationships in their stimulus set, which included noun-adjective collocates (*blood-red*) together with highly related category co-ordinates (*dog-cat*). It is perhaps clear, then, that their patients were accessing word meaning, at least to some extent, but it is not all clear *which* specific semantic information they were processing—it may well be, in fact, that only some of the several facets of word meaning remain available outside of awareness.

In the current experiment, we fix the issues highlighted above by implementing a strictly controlled priming experiment that tests conscious and unconscious semantic processing in neglect patients. To make sure that prime words were truly neglected, before the main experiment patients performed a visibility task requiring them to assess whether a square box appeared either on the left, on the right, or on both sides of a centrally presented fixation point; this way, we guarantee that primes were truly presented in parts of the visual field where patients had no conscious access.

Also, we carefully define different types of meaning-based similarity, taking advantage of distributional semantics techniques. These procedures stem from the theoretical assumption that words with similar meaning will tend to be used in similar linguistic context. Words themselves act as semantic features and their distribution observed over large text database define the strength of the semantic link in a quantifiable and objective manner. In particular, we compare word embedding and local cooccurrence.

Word embedding represents a computational technique to create distributional semantic models (DSMs), where words are mapped to numerical vectors derived from word-by-word contingency tables. Words with similar meaning tend to cluster together, and such similarity can be quantified by applying geometrical techniques to these vectors. For example, one can approximate relatedness as the cosine of the angle formed by two word-vectors:

$$\cos\theta = \frac{a \cdot b}{||a|| \cdot ||b||}$$

DSMs have been proposed as a psychologically plausible models of semantic memory, with particular emphasis on how meaning representations are achieved and structured. In particular, the model we employed (word2vec; Mikolov, Yih, & Zweig, 2013) represents a simple neural network consisting of an input, an output and a hidden layer, and is based on a predictive mechanism that allows to infer a target given a cue. Thus, words are similar if their presence is expected in roughly the same linguistic context; for example, in a sentence about domestic pets, it's likely to encounter the word *dog*, as well as the word *cat*. Word2vec provides similarity estimates that cover a wide range of classic lexical-semantic relationships, like synonymy (e.g., car-automobile, 0.45), antonymy (e.g., young-old, 0.51), meronymy (e.g., cherry-fruit, .49). Associative relations as well can be grasped (carrot-stick, .41). Finally, it can account for featural similarity beyond category membership (e.g. lion-tiger, .54 vs lion-mole, .17). Word2vec has been shown to perform better than (or as well as) other DSMs in a variety of task, such as synonym detection, concept categorization, semantic priming (Baroni, Dinu, & Kruszewski, 2014; Mandera, Keuleers, & Brysbaert, 2017)

Local co–occurrence was instead captured by simply counting how many times two words are used close to one another. As behavioral and computational studies have shown, words that are likely to be used together, tend to be associated in meaning. This type of local relationship is also reflected in the likelihood with which a given word recalls a second one (Spence & Owens, 1990). A useful tool to test for this assumption is to compute Pointwise Mutual Information (henceforth PMI) between two words, according to the formula:

$$\mathrm{PMI}(w_1,w_2) = \log_2\frac{p(w_1,w_2)}{p(w_1)p(w_2)}$$

where $p(w_1,w_2)$ corresponds to the probability that word w_1 and word w_2 co–occur in a window of test of a given size, while $p(w_1)$ and $p(w_2)$ are the individual probabilities of occurrence of word *w1* and word *w2* in isolation (Church & Hanks, 1989). PMI has been used to model a wide range of psycholinguistics phenomena, as similarity judgements (Recchia & Jones, 2009), reading speed (Ellis & Simpson-Vlach, 2009), free association and syntactic parsing (Pitler, Louis, & Nenkova, 2010). Moreover, PMI has also been shown to successfully generalize to non-linguistic fields as epistemology and psychology of reasoning (Paperno, Marelli, Tentori, & Baroni, 2014). Most critically for the purpose of the present experiment, this metric is specifically suited to capture local associations (e.g., leash–dog, kangaroo–australia, white–flag), and is known to fail on several higher–level semantic relationships, such as synonymy. This kind of relationships require methods, such as DSM, that consider wider contexts and "abstract away" from mere local co–occurrence.

By contrasting DMS– and PMI–associated prime–target pairs, and showing them to Neglect patients in either their visible or affected hemifield, we investigate whether word meaning is available outside of awareness and, most importantly, which aspects of word meaning are captured in subliminal word perception.

Experiment

Method

Patients. Seven right-handed patients (2 males, 5 females; 62 to 87 years old) were recruited into the experiment, who suffered from left unilateral neglect secondary to right hemisphere strokes. I saw the patients between 3 and 9 days after stroke; thus, all were in sub-acute conditions (see Table 1).

Testing was performed in the hospital, in a dedicated and quiet room when patients could manage a sitting position; or at the patients' bed otherwise. For each patient, neglect was assessed by non standardized pen-and-pencil neuropsychological testing, which included line bisection, star cancellation, the bell test and clock drawing.

	Age	**Gender**	**Education**	**Stroke Day**	**Test Day**	**Δ stroke-test**	**Site of the lesion**
1	69	M	8	31/07/2018	03/08/2018	3	F-T
2	70	F	13	09/09/2018	13/09/2018	4	F-T-P + basal ganglia
3	65	F	13	05/10/2018	10/10/2018	5	F-T-P + internal capsule
4	83	F	8	14/10/2018	18/10/2018	4	F-T
5	87	F	8	11/12/2018	15/12/2018	4	F-Insula
6	83	F	8	16/01/2019	25/01/2019	9	T-F
7	62	M	11	29/01/2019	01/02/2019	3	T-P-Insula

Table 2. Clinical details of the patients involved in the current experiment

Design. The independent variables were prime–target relatedness (related vs. unrelated), type of similarity (local association/PMI only, higher–level semantics/COS only, or both PMI and COS), and prime awareness (aware, that is, presented in the spared hemifield vs. unaware, that is, presented in the neglected hemifield). These variables were fully crossed, thus generating a 2-by-2-by-3 full design.

Material. The materials were the same as in Chapter 3. Ninety Italian word per category were equally divided across three category and used as target stimuli. Each of them was paired with a congruent and an incongruent prime, according to the semantic category it was assigned to (e.g., PMI: cheese\monument-MOUSE; COS: lamp\missile-TORCH; PMI+COS: prawn\veal-CRAB). Target across categories were matched on length and frequency. Next, ninety pronounceable non-word targets were added, and each of them was couple with two prime words, different from those used in the word-trials.

Procedure. Patients performed a lexical decision task, which required them to assess whether the target stimulus was a real Italian word (e.g., *tavolo*, table) or not (e.g., *tevolo*, lit. teble). All stimuli were shown in Arial font 32, in white against a black background, and were displayed on a 17'' monitor with a refresh rate of 60 Hz, using MatLab Psychtoolbox (Brainard, 1997). Responses were collected by mouse press.

The experiment was comprised of 720 trials. Each prime-target pair was shown twice, one with the prime word displayed on the left, and the other with the prime word displayed on the right side of the screen. Each trial started with a fixation point (+) displayed for 750ms. Then, the prime was shown for 200ms, at 5 degree of visual angle to the left or to the right of the fixation point; contralaterally to the prime word, a visual foil (#####) of the same length was presented. Finally, the target word appeared and remained on the screen until a response was provided (Figure 1).

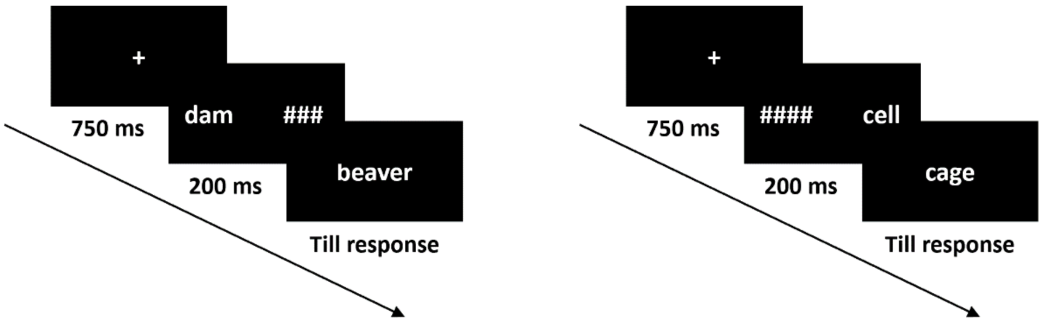

Figure 1. Exemplar trials used in the current experiment. "Subliminal" primes were presented in the left side of the screen, and visible primes on the right side of the screen.

Data analysis. Accurate, YES–response trials were retained for the analyses, which were carried out via mixed–effects linear regression using the package lme4 of the statistical software R (Chambers, 2008). Reaction times (RTs) were logarithmically transformed to approximate a normal distribution, and were employed as dependent variable. The factors constituting our main experimental manipulations – semantic category (PMI only, COS only and PMI+COS), congruency (congruent vs incongruent) and prime presentation (left vs right) – were tested as main effects, as well as their interactions. We additionally added random intercepts for each individual patient and target word. P-values were computed using the Satterthwaite approximation to degrees of freedom (Luke, 2017) provided by the lmerTest package. Model–based estimated of RTs in each design cell were eventually obtained via the R package *emmeans* (Lenth, 2018). We construct the model for the analysis in order to explore the parameters that are most relevant to our experimental questions, that is, (i) weather priming differs according to the type of semantic similarity linking the prime to the target; (ii) weather masking the prime changed the results relative to the visible condition.

Open practices statement. All data and analysis code are available at *https://osf.io/bdwp4/*, and can be accessed independently from the authors.

Results

Patients mean accuracy in the priming task was 84% (sd 37%). Mean RT on accurate word trials was 1.97 second (sd 0.87 second). RT distributions, at the group and individual level, are shown in Figure 2. Based on visual inspection, datapoints with RT higher than 6 seconds were removed (2 in totals), leaving a total of 2149 observations for the analysis.

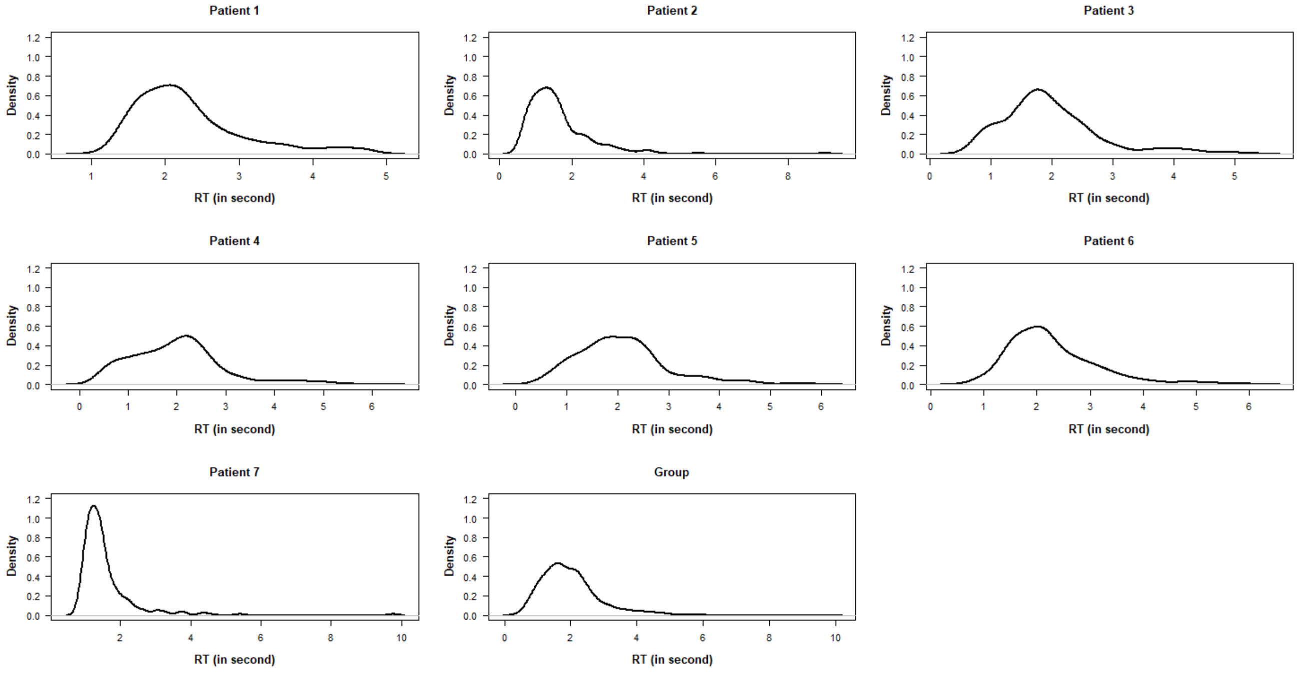

Figure 2. Density distribution of RT in corrected trials for each individual patient and at the group level.

Data were firstly analyzed by means of a full model testing the main effect of semantic category, congruency and prime presentation, as well as their interactions. Yet, this model faced high collinearity between predictors, so that the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. The variance inflation factor for the 3-way interactions was 30, while it should not be higher than 10 (VIF; Fox & Monette, 1992). Thus, we fitted two individual model, one for each prime position level, testing the emergence of priming across the 3 different semantic categories.

When primes were presented on the right hemifield – thus, they were clearly visible – we observed main effects of congruency ($F(1,998)= 7.72$, $p= .006$) and category ($F(2,88)= 4.90$, $p= .010$), while their interaction was not significant ($F(2,998)= 0.44$, $p= .644$). Although the congruency by category interaction was not significant, the model parameters revealed that priming was attested for the PMI ($t(987)= -2.02$, $p= .043$) and BOTH ($t(996)= -2.07$, $p= .039$) categories, but was much weaker (actually, absent) for the COS category ($t(1006)= -0.79$, $p= .430$).

When primes were presented on the left side of the screen – thus, they did not reach awareness – we did not find any effect of Congruity ($F(1,982)= 0.10$, $p= .757$) nor of Category ($F(2,87)= 0.75$, $p= .476$); their interaction was not significant as well ($F(2,982)= 2.00$, $p=.136$). Model estimates of the RTs per condition are represented in Figure 3.

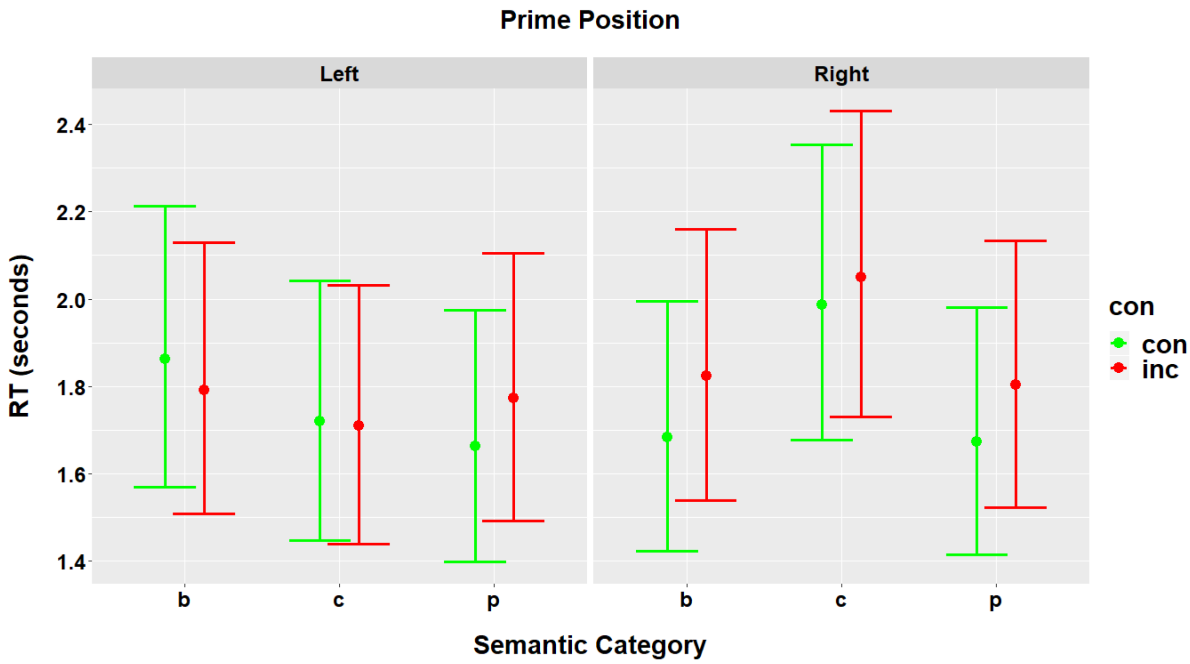

Figure 3. Model estimates of the RT for each category (p= PMI, c=COS, b=PMI+COS). On the left, results observed with "subliminal" primes, on the right with visible primes. Congruent condition is shown in green, incongruent condition in red. Error bars refer to the 95% confidence intervals.

Discussion

In the current study we explored the mechanisms underlying semantic processing via a primed lexical decision experiment. To explore weather meaning is accessed similarly when words are processed above or below the threshold of conscious perception, we recruited patients suffering from spatial neglect—by delivering the prime either on the left (neglected) or on the right (spared) hemifield, we were able to compare overt and masked priming without the need to manipulate the way the prime was presented.

The semantic relationship between prime and target was also defined in a quantitatively and principled manner, taking advantage of distributional semantics technique to model meaning similarity based on word usage. In particular, prime-target pairs could be related according to local association as tracked by Pointwise Mutual Information (PMI); higher–level semantic similarity, as tracked by spatial proximity in a multidimensional semantic space (cosine similarity, COS); or both.

When primes were presented subliminally, in the neglected hemifield, priming did not emerge, regardless of the semantic relationship being considered. Conversely, when primes were presented on the right side of the screen, and thus they were processed consciously by the patients, solid facilitation emerged in the PMI category, where congruent primes resulted in shorter RT to the target than incongruent ones. Similarly, the effect was also attested when primes and targets were related according to both PMI and COS. Yet, only weak – if any – priming was elicited in the COS only category.

The lack of subliminal priming is not surprising, and it echoes previous reports showing that masking the prime makes the effect unstable and difficult to reproduce (Brunellière et al., 2017; Montefinese et al., 2018). Furthermore, priming has been shown to be dependent on the specific task being performed: whereas related masked and visible primes prompt faster response to the target in a semantic categorization, presenting the prime out of conscious perception deletes the effect in a lexical decision (De Wit & Kinoshita, 2015). We implemented a lexical decision due to comparability with previous studies; clearly, more research adopting a semantic task is required to further explore semantic priming with neglect patients.

The lack of semantic priming outside of awareness would be in line with other data reported in this thesis. In Chapter 2, we report that, even in a condition that would be considered masked priming by most, facilitation only emerges when at least some residual prime visibility is attested in a detection task performed on the prime itself. When prime presentation time is short enough to entirely prevent its visibility, the effect disappears. However, in that same paper, we also showed that subliminal priming re–emerges when data are analyzed dichotomously, by only taking items at the extremes of the relatedness distribution, thus drawing a comparison between related and unrelated primes. This is exactly the approach we adopted here; so, those data would have predicted that we should obtain facilitation here too.

These data are also inconsistent with Làdavas et al. (1993). Although they did not differentiate types of semantic relationship in their paper, they do report overall semantic priming in the neglected hemifield. There are various reasons that can explain the discrepancy between their results and ours. The main difference between the two studies is at the participant level: while our patients were in sub-acute condition and were still hospitalized, the single patient involved in the study by Làdavas and colleagues was tested two years after the stroke, and was monitored for six months before the doing the experiment. Even more importantly, his education level (18 years) was much higher than the one of our patients (8-13 years). It is well known that linguistic competence is a factor modulating lexical processing, as more educated speakers are likely to have been exposed to more varied language during their school/university years (Dabrowska, 2015; Yap, Hutchison, & Tan, 2016).

Data from the visible condition suggest instead that semantic facilitation is particularly strong for word pairs linked by local association; if this latter is prevented, cosine similarity alone is not enough for words to fully prime each other. These results matched those of the previous chapters, showing that simple measures based on local, surface information are more effective in predicting priming than the more complex ones based on word embedding. Without reiterating what we described extensively before, this might indicate that the processes underlying the emergence of priming are better described in terms of associative mechanisms that link lexical items in our mental lexicon, rather than by the activation of conceptual information in semantic memory.

Finally, a word of caution on these data. Finding sub–acute stroke patients who are amenable to testing, semantically intact, and also show neat symptoms of Spatial Neglect is not easy; the numbers illustrated in the Participants section attest to this. Despite the effort, then, the final sample of participants, albeit larger than in most of the previous investigations of this issue (Làdavas et al., 1993; McGlinchey-Berroth et al., 1996; Sackur et al., 2008), is still rather limited. In addition, RTs in brain–damaged patients are typically very noisy, and do make it difficult for neat effects to come up. So, clearly, the current data must be taken with caution, and, although we surely believe that they provide useful insights into the dynamics of lexical-semantic processing, they should be replicated in a sample of chronic patients, whose neglect is more stable (and therefore stimuli presentation can be tight up more precisely to the unattended part of their visual field) and whose general condition would also be likely better, thus providing a better signal–to–noise ratio with their response times.

References

Abrams, R. L., Klinger, M. R., & Greenwald, A. G. (2002). Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196262

Anaki, D., & Henik, A. (2003). Is there a "strength effect" in automatic semantic priming? *Memory and Cognition*. https://doi.org/10.3758/BF03194385

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*.

Beschin, N., Cazzani, M., Cubelli, R., Della Sala, S., & Spinazzola, L. (1996). Ignoring left and far: An investigation of tactile neglect. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(95)00063-1

Bisiach, E., & Luzzatti, C. (1978). Unilateral Neglect of Representational Space. *Cortex*. https://doi.org/10.1016/S0010-9452(78)80016-1

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brunellière, A., Perre, L., Tran, T. M., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2016.1215479

Bueno, S., & Frenck-Mestre, C. (2008). The activation of semantic memory: Effects of prime exposure, prime-target relationship, and task demands. *Memory and Cognition*. https://doi.org/10.3758/MC.36.4.882

Chambers, J. (2008). Software for Data Analysis. In *Springer*. https://doi.org/10.1007/978-0-387-75936-4

Church, K. W., & Hanks, P. (1989). *Word association norms, mutual information, and lexicography*. https://doi.org/10.3115/981623.981633

Corbetta, M., & Shulman, G. L. (2011). Spatial Neglect and Attention Networks. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev-neuro-061010-113731

Dabrowska, E. (2015). Individual differences in grammatical knowledge. In *Handbook of Cognitive Linguistics*. https://doi.org/10.1515/9783110292022-033

De Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000074

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*. https://doi.org/10.1515/CLLT.2009.003

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1992.10475190

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/a0032333

Kim, C. Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible “invisible.” *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2005.06.012

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2007.2093

Làdavas, E., Paladini, R., & Cubelli, R. (1993). Implicit associative priming in a patient with left visual neglect. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(93)90100-E

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.0.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0809-y

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2016.04.001

Marshall, J. C., & Halligan, P. W. (1988). Blindsight and insight in visuo-spatial neglect. *Nature*. https://doi.org/10.1038/336766a0

McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., Grande, L., D'Esposito, M., & Alexander, M. (1996). Semantic processing and orthographic specificity in hemispatial neglect. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.1996.8.3.291

Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous spaceword representations. *NAACL HLT 2013 - 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Main Conference*.

Montefinese, M., Buchanan, E., & Vinson, D. (2018). How well do similarity measures predict priming in abstract and concrete concepts? *PsyArvXiv*. https://doi.org/10.31234/osf.io/ypvgw

Paperno, D., Marelli, M., Tentori, K., & Baroni, M. (2014). Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2014.07.001

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*. https://doi.org/10.1007/s00426-002-0086-5

Peremen, Z., & Lamy, D. (2014). Comparing unconscious processing during continuous flash suppression and meta-contrast masking just under the limen of consciousness. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00969

Pitler, E., Louis, A., & Nenkova, A. (2010). Automatic evaluation of linguistic quality in multi-document summarization. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.

Punt, T. D., & Riddoch, M. J. (2006). Motor neglect: Implications for movement and rehabilitation following stroke. *Disability and Rehabilitation*. https://doi.org/10.1080/09638280500535025

Quinn, W. M., & Kinoshita, S. (2008). Congruence effect in semantic categorization with masked primes with narrow and broad categories. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2007.03.004

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*. https://doi.org/10.3758/BRM.41.3.647

Sackur, J., Naccache, L., Pradat-Diehl, P., Azouvi, P., Mazevet, D., Katz, R., … Dehaene, S. (2008). Semantic processing of neglected numbers. *Cortex*.

https://doi.org/10.1016/j.cortex.2007.02.003

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/BF01074363

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*. https://doi.org/10.1016/S0896-6273(00)80592-9

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.02.005

Yap, M. J., Hutchison, K. A., & Tan, L. C. (2016). Individual differences in semantic priming performance: Insights from the semantic priming project. In *Big Data in Cognitive Science*. https://doi.org/10.4324/9781315413570

Conclusions

In the current thesis, I have investigated the cognitive and brain processes underlying access to lexical meaning, and weather semantic processing is held similarly when words are presented below or above the threshold for conscious perception.

In the first experimental contribution presented in this thesis, I advanced and tested the hypothesis that subliminal processing is limited to language–encoded semantic ties. A perfect test bed for this hypothesis was offered by the metaphorical relationship linking time to space. At least in Western societies, time is spatially arranged along a sagittal and a lateral mental timeline, but only the former is linguistically encoded. That is, while people normally speak of the future as located in front of us and the past at our back (look *ahead* to the weekend; think *back* to the childhood), no languages is known to rely on the lateral mapping. Thus, I developed a priming experiment where temporal targets (e.g. yesterday, tomorrow) were paired with spatial primes (e.g. left, back) that were presented either consciously or unconsciously.

Coherently with the hypothesis tested, we found evident and comparable space-time congruity effects along the sagittal and lateral timelines when primes were visible. By contrast, in the masked condition, sagittal words strongly primed related targets, while the lateral words led only to a weak (if any) facilitation. According to these results, readers may not be able to activate fully fledged semantic representations when exposed to subliminal words. Rather, they may navigate their the lexical–semantic system based on how words are linked to each other in language use (in this case, as related to linguistic metaphors).

In the second experimental contribution, we followed up these results by further exploring how meaning-based similarity is encoded in language. To this aim, we took advantage of distributional semantics methods that allow to define lexical meaning by looking at words distribution over large text corpora. Words themselves represent semantic features in these models, and by looking at how they are used in relation to each other, it is possible to define the strength of the semantic link in a quantifiable manner. Clearly, there are several ways to do so; here, I considered cosine similarity (COS) derived from the state–of–the–art Distributed Semantic Model, namely wordToVec (Mikolov, Chen, Corrado, & Dean, 2013), which tracks various types of high–level, long–distance semantic relationships (e.g., sofa-hammock, worm-caterpillar), and Pointwise Mutual Information (PMI), which specifically captures associative, more local ties (e.g., tank–paint, scissors–razor). Thanks to these metrics, I was able to explore the entire relatedness continuum, rather than selecting only the extreme values like in most published studies, which adopted a dichotomous design.

In a series of experiments manipulating prime visibility and prime duration, we observed that genuine semantic priming seems not to emerge in the masked condition. Neither PMI nor COS led to a significant facilitation in the processing of the target stimuli when prime visibility was strictly controlled for. When, instead, some room for prime detection was allowed, priming started to emerge; the interaction between prime visibility and both PMI and COS clearly showed how the effect increases with participants' ability to spot the presence of the primes. Yet, when we restricted our stimulus set by selecting only word pairs that were either strongly related or strongly unrelated on both the metrics, subliminal priming showed up.

Conversely, when primes were fully visible, a clear modulation of the semantic index on the response times to the target was observed. Even with the same presentation time - but, most likely, a better information uptake by the participants - semantic facilitation was fully observed. Both PMI and COS successfully predict the emergence of priming, replicating effects already shown in the literature. Yet, when the two metrics were pitted one against the other, PMI clearly outperformed COS in the fit to the behavioral data, independently from how much time is given to process the prime. Overall, semantic priming seems to be primarily driven by local word associations that can be extracted from surface co-occurrence patterns emerging from natural language documents.

In the third experimental contribution, I investigated the electrophysiological correlates of semantic priming. As the previous experiments showed more solid and reliable results when the primes were visible, I choose to focus on this condition only. Similarity between word pairs was again defined via either PMI or COS, disentangling as much as possible the specific contribution provided by each metrics. Furthermore, we included a third category where congruent prime-target pairs were related according to both PMI and COS. In order to explore the event related potentials associated with the processing of the semantic information reflected by these metrics, we contrasted dichotomously related and unrelated conditions. Thus, I recorded participants' EEG signal while they were performing a primed lexical decision task, and analyzed the data focusing on the N400 component.

A strong effect emerged for word pairs in the PMI category, with incongruent trials resulting in higher negativity than congruent trials, mostly over fronto-central electrodes. Conversely, the effect of cosine similarity was much weaker, and did not reach significance. Yet, N400 was observed again for items that were both strong associates and close in the semantic space (PMI+COS); the topography of the effect was slightly different though, more posterior than in PMI alone and mostly attested over central electrodes.

Finally, in the fourth experimental contribution, I tested subliminal and supraliminal semantic processing in patients suffering from Spatial Neglect. This syndrome is characterized by a deficit in attending and responding to stimuli presented on one side of the visual field, which is often contralateral to the hemisphere of the brain where a damage had been sustained. By delivering the prime either on the left (neglected) or on the right (spared) hemifield, we were able to compare overt and masked priming without the need for psychophysical "tricks" that induce unawareness experimentally, such as visual masking.

Mirroring results from the previous experiments, no facilitation showed up with subliminal primes, those presented in the neglected hemifield, independently of the semantic relationship being considered. Conversely, when primes were delivered on the right side of the screen, and thus they were processed consciously by the patients, priming emerged in the PMI category, where incongruent primes resulted in longer RT to the target than congruent ones. Similarly, the effect was also attested when primes and targets were related according to both PMI and COS. Yet, only weak – if any – priming was elicited in the COS category.

To sum up, during my PhD I have conducted a series of priming experiments aimed at better understanding how lexical meaning is computed with and without awareness. Subliminally, when we estimated priming taking the entire relatedness continuum into consideration, we observed only a weak effect which strongly depended on prime visibility. However, unconscious semantic facilitation showed up only when related and unrelated prime-target pairs laid at the extreme tails of the semantic continuum (Chapters 1 & 2). The lack of subliminal priming in the experiment presented in Chapter 4 instead does not match our experimental hypothesis, which conversely predicted its presence, especially in the category with relatedness defined by both PMI and COS. Yet, such results may be accounted for by the task adopted (but see Làdavas, Paladini and Cubelli, 1993), or they may be due to the patients we managed to test, whose clinical situation and education level were not optimal for experimental testing.

Thus, the current thesis not only offers the methodological warning that forcing into categorical terms naturally continuous variables may create effects that are not attested (or, at the very least, are much weaker) when the entire distribution is considered. More importantly, these data cast some doubts on a wide, across–the–lexicon processing of semantic information outside of awareness.

On the other side, semantic processing was clearly attested when primes were visible. More interestingly, the effect was better explained by local association measures (PMI) than by more complex metrics that take into account long–distance, higher–level semantic relationships more generally (COS). This pattern of results held both behaviorally and at the ERP level, suggesting the strength and reliability of the current findings.

These data clearly contradict theoretical accounts of masked priming whereby the effect would origin from automatic spread of activation within a semantic network (Collins & Loftus, 1975; Neely & H., 1991). Rather, priming may mostly arise due to expectancy generation—the prime is taken as a cue for the coming target, and expectation is computed based on local, relatively simple association links (Jones, 2010).

This is in line with previous behavioral data suggesting a primary role for local linguistic ties in structuring our lexical-semantic system. Günther, Dudschig and Kaup (2016) showed that similarity estimates derived from a semantic space based on local context information predict priming better than those derived from a semantic space based on global context information. Similarly, Brunellière, Perre, Tran and Bonnotte (2017) probed that, while keeping semantic similarity constant, the magnitude of priming was greater as prime-target pairs co–occur more frequently.

Next, these results may be informative of the kind of information processing reflected in the N400. The lack of a significant difference between related and unrelated condition in the COS category suggested that N400 is first and foremost an index of lexical access, which is more strongly modulated by predictability than incongruity (Bornkessel-Schlesewsky & Schlesewsky, 2019; Lau, Namyst, Fogel, & Delgado, 2016). Thus, this experiment goes against theoretical accounts according to which modulation of the ERP reflects the effort of integrating lexical meanings in a semantically coherent way (Federmeier & Kutas, 1999; Kutas & Federmeier, 2011).

Clearly, there are several issues left open in the current thesis. First of all, our best predictor, PMI, is by definition a symmetric measure, and therefore we cannot assess weather expectations proceed prospectively or retrospectively. This is a crucial point, as association can be directional; for example, the words *surgeon* and *hospital* are clearly related to each other, but *surgeon* is a much stronger cue to predict *hospital* than the other way around. Asymmetric association can exert an influence on human behavior. With regard to priming, evidence is intermixed. In his review, Hutchison (2003) reported that the size of the backward priming was statistically equivalent to the size of the forward priming effect. However, Zeelenberg, Shiffrin, & Raaijmakers (1999) found that backward association was mandatory for priming to be attested. Similarly, false memory formation seems to depends more on backward than on forward associative strength (Roediger, Watson, Mcdermott, & Gallo, 2001).

Moreover, it would be interesting to follow up the present EEG study with a more naturalistic experimental setting, that is, making participants read sentences rather than words in isolation. These methodological changes may allow for a better understanding of how semantic congruency and lexical predictability interact during on-line language comprehension. The same reasoning holds for eye-tracking methods.

Thus, more research is clearly required to address these issues, and I hope that this thesis may represent the starting point of an amazing journey exploring the fascinating dynamics of human language.

References

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.00298

Brunellière, A., Perre, L., Tran, T. M., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2016.1215479

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*. https://doi.org/10.1037/0033-295X.82.6.407

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1999.2660

Günther, F., Dudschig, C., & Kaup, B. (2016). Predicting lexical priming effects from distributional semantic similarities: A replication with extension. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01646

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196544

Jones, L. L. (2010). Pure Mediated Priming: A Retrospective Semantic Matching Model. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/a0017517

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.093008.131123

Làdavas, E., Paladini, R., & Cubelli, R. (1993). Implicit associative priming in a patient with left visual neglect. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(93)90100-E

Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A Direct Comparison of N400 Effects of Predictability and Incongruity in Adjective-Noun Combination. *Collabra*. https://doi.org/10.1525/collabra.40

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013. *ArXiv Preprint ArXiv:1310.4546*.

Neely, & H., J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading: Visual word recognition*.

Roediger, H. L., Watson, J. M., Mcdermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196177

Zeelenberg, R., Shiffrin, R. M., & Raaijmakers, J. G. W. (1999). Priming in a free association task as a function of association directionality. *Memory and Cognition*. https://doi.org/10.3758/BF03201226