SCUOLA NORMALE SUPERIORE
Classe Accademica di Lettere e Filosofia

A cognitive and pragmatic approach to meaning. Behavioral and neural correlates of concept processing

A dissertation submitted for the degree of Doctor of Philosophy in Linguistics
by Marta Virginia Ghio

Committee in charge:
Prof. Pier Marco Bertinetto
Dr. Valentina Bambini
Dr. Lewis Bott
Dr. Marco Tettamanti

Pisa, Italy, 2013

Table of contents

Study A. Fine-grained semantic categorization across the abstract and concrete domains

Study B. Fine-grained semantic categorization across the abstract and concrete domains: an fMRI research

Study C. Differentiating among pragmatic uses of words through timed sensicality judgments: Metaphor, metonymy and approximation

Study D. The time course of figurative language processing: A speed-accuracy tradeoff study on metaphor and metonymy

Introduction

Language is a highly flexible system that allows concepts to be conveyed by symbols. It is thus connected with the semantic or conceptual system. Work on linguistic communication has long recognized two basic facts. First, the conceptual system is flexible and dynamic enough to enable us to construct and use an indefinite number of conceptual representations. Second, the language system, though employing a limited lexicon and a finite set of rules, is productive enough to enable us to convey even conceptual representations that go far beyond anything like the linguistic form itself. How can these two systems work together in producing and comprehending meaning in actual communication?

A fundamental key for understanding the relationship between the concept system and the language system is *context*, which takes different forms and functions in the processing of meaning and in its communication. Context is generally defined as referring to any factors that affect the actual interpretation of utterances, including linguistic, epistemic, physical, and social factors. A shared view is that the meaning of an isolated word may be close to undetermined, and appears to vary even as a function of a minimal linguistic context. For example, the meaning of a verb like 'to use' is closely related to its arguments, with 'use the hammer' and 'use the opportunity' conveying an action-related and an abstract meaning, respectively. The sentential semantic context in which a word like 'grasp' is embedded may convey either a literal (e.g., 'grasp the hammer'), a figurative metaphorical (e.g., 'grasp the idea') or idiomatic (e.g., 'grasp the nettle') meaning. Do we access the same conceptual representation of a word in different contexts? Are the different contexts constitutive of the conceptual representation itself? How do we fill the gap between lexically-encoded concepts and communicated concepts?

The present research investigates the representation and processing of lexical elements across different types of minimal sentential contexts, including non-figurative and figurative sentences. Two different and parallel lines of research were pursued.

The first line of research focuses on work in psychology and cognitive neuroscience on conceptual-semantic knowledge. Scientific advancement in this field, especially in the grounded cognition framework, has emphasized the complexity and the flexibility of the brain mechanisms subserving the processing of linguistic meaning. These mechanisms are thought to be fundamentally experience-dependent and rooted in neural networks extending into experiential brain systems. A still open and highly relevant research question is the representation of abstract meanings, i.e. the meaning of words referring to entities, events, or facts that are not directly experienced in the external world. The experimental studies described in the first part of the present dissertation addressed this question in a series of rating experiments (Study A) and functional magnetic resonance neuroimaging experiments (Study B) by proposing a fine-grained categorization of meanings across the concrete and abstract domains.

The second line of research focuses on work in pragmatics, especially on lexical pragmatics as developed within the Relevance Theory framework. This theoretical approach provides a well-developed explanation of the mechanisms at the interface of words, concepts and communication, investigating the processes by which word meanings are modified in use. All different uses of language are explained in terms of conceptual operations guided by relevance-driven inferential mechanisms. Recent theoretical advancements put forward a more detailed description of different figurative uses, in particular with respect to metaphorical and metonymical meanings. This issue has been experimentally investigated in the second part of the present dissertation. A rating study was carried out to characterize literal and figurative uses at the psycholinguistic level. Standard and speeded sensicality judgments experimental paradigms were employed to investigate the time-course in the processing of figurative, as compared to literal language uses (Study C and Study D).

The present dissertation is thus structured in two parts, I and II. Each part includes experimental papers (either published, submitted, or in preparation). Specifically, Part I includes Study A and Study B, and Part II includes Study C and Study D, briefly summarized below. The experimental orientation provides a common methodological ground.

1.1 Part I

In cognitive neuroscience, the meaning of linguistic units is generally considered to be closely connected with conceptual representations, which are stored in the semantic memory (Vigliocco et al., 2004; Traxler, 2011). Accordingly, cognitive neuroscience research on word meanings mainly deals with the representation of conceptual knowledge in the semantic memory (McRae & Jones, 2013). Cappa (2012) observed that such intrinsic relationship between research on semantics and on conceptual knowledge is paradigmatically illustrated by the fact that Wernicke, in his pioneering paper on the functional neuroanatomy of language, “felt the need to provide a theory of conceptual representations” (see also, Gage & Hickock, 2005).

An impressive body of research on semantic processing suggests that the neural system underlying conceptual knowledge is distributed in the brain, including perisylvian language areas in the frontal and temporo-parietal cortices, and supportive brain networks (for exhaustive reviews, see Binder et al., 2009; Cappa, 2008; 2012; Binder & Desai, 2011). In particular, this supportive brain network seems to consist, at least in part, of the left-hemispheric lateral premotor cortex, sometimes extending more posteriorly into the primary motor area and more anteriorly into the middle frontal gyrus (for a review, see Pulvermueller & Fadiga, 2010). The involvement of sensory and motor systems in conceptual-semantic language processing has led to divergent positions in the field, being interpreted by scholars either in epiphenomenal or in “embodiment” terms (Meteyard et al., 2012). Though several variants of embodiment have been formulated (Meteyard et al., 2012), here I refer to a more general assumption formalized within the theoretical framework of grounded cognition, namely that concepts are represented in multiple, distributed brain networks reflecting the quality of experience that is characteristic for the concepts’ referents.

The work presented in the first part of the present dissertation aimed at contributing to the current debate (extensively presented in the Introduction sections of Study A and Study B), in particular with respect to the hotly-debated issue of grounding the abstract meanings into experiential brain systems. While there is rather consistent evidence in favor of the embodiment of concrete meanings stemming from neuropsychological, behavioral, electrophysiological and neuroimaging studies, evidence about abstract meaning processing, is scarce and quite controversial (for extensive reviews, see Kiefer & Pulvemueller, 2012; Hauk & Tschentscher, 2013).

Recent theoretical advancements within the grounded framework proposed that distributed neural representations of experiential information related to the concepts’ referents might distinguish concepts with a fine-grained specificity also in the abstract domain, by analogy to what has been demonstrated for concrete, action- and object-related knowledge (Barsalou, 2010; Wilson-Mendenhall et al., 2013). For example, the processing of abstract emotion concepts (e.g., happiness) might involve the emotion processing network, whereas the processing of introspective concepts referring to mental states (e.g., memory) might activate the mentalizing neural network. In this sense, grounded cognition may provide a coherent and scientifically testable theoretical framework to account for the processing of both concrete and abstract meanings.

Unlike previous studies, mainly focusing on processing words in isolation, either nouns or action verbs (Hauk & Tschentscher, 2013), in the present research subject-verb-object sentences were used as stimuli. As said above, the meaning of isolated words is mainly underspecified and open to different interpretations. Besides, the meaning of a verb is deeply grounded in syntax, for it involves the information conveyed by its syntactic structure(s) (McRae & Jones, 2013). Placing verbs and nouns within the same sentence structure can thus be considered as a linguistically more constrained access to conceptual knowledge, an access that disambiguates the word meaning in a specific sentential context. In addition, by keeping the syntactic form constant, it is possible to control for the information conveyed by the syntactic structure itself in relation to the target word.

The present research addresses the question about the representation and processing of meaning by distinguishing different kinds of semantic categories, with a high level of specificity within both the concrete and the abstract domain. In particular, for the abstract domain, the following pre-selected categories were considered: mental state-related sentences (e.g., 'She remembers the past'); emotion-related sentences (e.g., 'She shows her disappointment'); mathematics-related sentences (e.g., 'She determines the sum'). For the concrete domain, mouth-related sentences (e.g., 'She clicks her tongue'); hand-related sentences (e.g., 'She embroiders the handkerchief'); leg-related sentences (e.g., 'She kicks the ball').

The two studies in part I (A and B) were conceived of in a cascaded fashion, the first being a necessary precondition to the second.

In particular, ***Study A*** was aimed at providing evidence of such fine-grained distinctions *vis-à-vis* relevant psycholinguistic dimensions. Participants rated sentences with respect either to different semantic domain-related scales, concreteness, familiarity, or context availability. At the broad level, the results consistently reflected the abstract-concrete dichotomy. At a finer level, inferential statistics and correspondence analyses revealed semantic and psycholinguistic traits specific for each category, thus accounting for the semantic variability within either domain. The second aim of the rating research was to develop a carefully controlled linguistic stimulus set to be employed for the subsequent neuroimaging research presented in Study B.

Study B describes a series of functional magnetic resonance imaging experiments, in which healthy participants listened to both concrete and abstract sentences while actively performing an one-back task. Various techniques of fMRI data analysis were applied, providing different views of the functional organization of mental processing. On the one side, univariate general linear model analysis revealed the global engagement in ongoing tasks. On the other side, multivariate pattern analyses revealed a distributed coding of information. In addition, relying on the rating measures established in Study A, the relevant psycholinguistic variables were manipulated in a parametric fashion, either as parameters of interest or as confounds in a parametric analysis.

All the analyses converged in revealing distributed networks of brain regions specific for the processing of both abstract and concrete meanings. This result confirmed that the abstract/concrete distinction is a basic semantic organizational principle of conceptual knowledge. As far as the category specificity is concerned, no specific brain activations were found by applying the univariate analysis. The multivariate pattern analyses, on the contrary, revealed the existence of semantically organized information in the pattern of fMRI-measured brain activation during the semantic processing of different categories. These results contribute to the current debate on the neural representation of conceptual knowledge by suggesting that distributed patterns of brain activity, rather than localized activation in specific brain areas may mirror the processing of different kinds of meanings with high specificity.

In Study A and Study B a detailed discussion of the literature is provided, explaining how this research is in line with recent attempts to provide a more overarching framework for the neural mechanisms underlying the conceptual-semantic processing of concepts, encompassing different types of meaning with a high degree of specificity.

1.2 Part II

Cognitive neuroscience and neuropsychological research provided fundamental evidence about the brain mechanisms subserving the processing of conceptual representations expressed by language in the form of words or sentences. Such research, however, did not provide any information about the mechanisms involved in the interpretation of the meaning conveyed in a given context of use. Pragmatic accounts highlighted the non-isomorphism between ‘what is said’ and ‘what is communicated’ by a linguistic expression; this is particularly evident in the case of figurative language (e.g., 'Mary is an angel'). A shared assumption of pragmatic approaches is that people compute the meaning of a word by enriching the linguistic decoding with contextually driven inferences operating upon the conceptual dimension and integrating information from the context (Bambini & Bara, 2010).

The theoretical background of the work presented in Part II is Relevance Theory, which proposes a cognitively plausible model of communication describing the mechanisms involved in language comprehension (Sperber & Wilson, 1995/2008; Wilson & Carston, 2006; Carston, 2010). Sharing with the Gricean account the claim that the essential feature of human communication is the expression and recognition of intentions, Relevance Theory assumes that the inferential intention-recognition is guided by relevance. Relevance is treated as a property of inputs to cognitive processes, and it is defined in terms of cognitive effects and cognitive efforts: the greater the cognitive effects (respectively, the lower the cognitive efforts), the greater the relevance of the input (Sperber & Wilson, 1995/2008). For example, a linguistic input is relevant in a given context when it produces cognitive effects (such as answering a question or confirming a hypothesis), due to the least effort of parsing and inference. Based on this definition, RT argues that a linguistic input uttered by a speaker in a given context produces expectations of relevance in the listener. The listener is thus “entitled to treat the encoded linguistic meaning as a clue to the speaker’s meaning and to follow a path of least effort in adjusting this encoded meaning to a point where it yields an overall interpretation that satisfies those expectations" (Wilson, 2011: 202).

Recent advancements of Relevance Theory focused on the lexical level (lexical pragmatics) (Carston & Wilson, 2006; Carston, 2010). In this perspective, the comprehension of a word requires a process of pragmatic adjustment of the lexical entry and results in the construction of an *ad hoc concept*. Such concept may be broader or narrower than the encoded one, and contributes to the truth-conditional content of the utterance. Consider, for example, 'Robert is a bulldozer'. The utterance has to do with Robert's persistence and obstinacy and so on, but these properties go clearly beyond the encyclopedic meaning of 'bulldozer' whose features are literally inapplicable to human beings. The lexical entry 'bulldozer' is thus pragmatically adjusted in the context in order to refer to human beings.

In the standard version of Relevance Theory, virtually every word in context requires a pragmatic adjustment that fine-tuned the interpretation of the lexically encoded concept. In this sense, different types of figurative language, such as approximations (e.g., 'Her face is oval'), hyperboles (e.g., 'The garden is a jungle'), and metaphors, are generally categorized together as loose uses of language (Sperber & Wilson, 2008). In recent work, however, some Relevance Theory scholars provided a more detailed articulation of the similarities and differences between such different types of loose uses. For example, it has been argued that, compared to other types of loose uses, metaphors involve both a broadening and a narrowing of the encoded meaning of the metaphor vehicle (Carston & Wearing, 2011). Furthermore, metonymical uses (e.g., 'The saxophone walked out') are not considered as resulting from the standard pragmatic adjustment process, but rather as involving some sort of meaning transfer (Carston, 2010). Similar theoretical distinctions between different types of figurative uses have also been posited by other pragmatic accounts, such as Cognitive Linguistics (Evans & Green, 2006).

Do these differences correspond to variations in language processing? In recent years, an "experimental shift" has taken place in pragmatics: the details of language interpretation can only be defined by experimentally investigating how the comprehension process works (Bambini, 2010). The research programs developed in the so called "Experimental pragmatics" try and find an answer for theoretical questions, conducting experimental studies specifically inspired by pragmatic theories (Noveck & Reboul, 2008). In line with work in Experimental pragmatics, Study C and Study D aimed at comparing the processing of metaphors and metonymies within a single experiment. Initial evidence about approximations was also provided. In comparison to the large quantity of behavioral data on metaphor processing, metonymy processing has been less investigated, and only few studies jointly investigated the two phenomena. To our knowledge, no experimental data are presently available for the notion of approximation.

In ***Study C*** a sensicality judgment paradigm was applied in order to compare the processing of metaphors (e.g., 'Those dancers are butterflies'), and metonymies (e.g., 'That student reads Camilleri'). Furthermore, in order to investigate the distinctiveness of metaphors as compared to other cases of loose uses, approximate uses of language were also considered (e.g., 'Those sunglasses are rectangular'). All stimuli were controlled in a rating study with respect to several psycholinguistic variables. Participants performed sensicality judgments on visually presented sentences. In accordance with theoretical distinctions, differences between the three pragmatic phenomena emerged in terms of costs of interpretation.

Study D aimed at clarifying the results of Study C with respect to the time-course of processing figurative as compared to literal language, focusing on metaphor and metonymy processing. The standard sensicality judgment paradigm provides a good measure of the availability and difficulty of correct interpretation (Klein & Murphy, 2001). Nevertheless, it only allows investigating the late stages of processing, when the sense has already been construed (Frisson, 2009). Moreover, results obtained with the standard reaction time procedure do not allow distinguishing between different interpretations of the time processing differences. For these reasons, in Study D the Multiresponse Speed-Accuracy Trade-Off was applied, a technique that allows to disentangle accuracy from time-course in computing a sensible interpretation. Exponential fits of the time-course functions showed that both metaphorical and metonymical meanings were associated with lower accuracy than their literal counterparts. Metonymy also exhibited slower processing speed than literal expressions, but metaphor did not. The results of Study C and Study D offer new insight for the taxonomy of figurative language, based on the combination of linguistic-pragmatic distinctions and experimental evidence.

Part I

Study A

Fine-grained semantic categorization across the abstract and concrete domains[1]

Abstract

A consolidated approach to the study of the mental representation of word meanings has consisted in contrasting different domains of knowledge, broadly reflecting the abstract-concrete dichotomy. More fine-grained semantic distinctions have emerged in neuropsychological and cognitive neuroscience work, reflecting semantic category specificity, but almost exclusively within the concrete domain. Theoretical advances, particularly within the area of embodied cognition, have more recently put forward the idea that distributed neural representations tied to the kinds of experience maintained with the concepts' referents might distinguish conceptual meanings with a high degree of specificity, including those within the abstract domain. Here we report the results of two psycholinguistic rating studies incorporating such theoretical advances with two main objectives: first, to provide empirical evidence of fine-grained distinctions within both the abstract and the concrete semantic domains with respect to relevant psycholinguistic dimensions; second, to develop a carefully controlled linguistic stimulus set that may be used for auditory as well as visual neuroimaging studies focusing on the parametrization of the semantic space beyond the abstract-concrete dichotomy. Ninety-six participants rated a set of 210 sentences across pre-selected concrete (mouth, hand, or leg action-related) and abstract (mental state-, emotion-, mathematics-related) categories, with respect either to different semantic domain-related scales (rating study 1), or to concreteness, familiarity, and context availability (rating study 2). Inferential statistics and correspondence analyses highlighted distinguishing semantic and psycholinguistic traits for each of the pre-selected categories, indicating that a simple abstract-concrete dichotomy is not sufficient to account for the entire semantic variability within either domains.

---

[1] Ghio, M., Vaghi, M. M. S., & Tettamanti, M. (2013). Fine-Grained Semantic Categorization across the abstract and concrete Domains. *PLoS ONE*, 8(6), e67090.

1. Introduction

Classification in science is crucial. One of the first brilliant examples of it can be found in the work of the Swedish botanist Carl Linnaeus who implemented a naming system for animal and plant organisms that proved to be an elegant solution for the taxonomic literature (Linneaeus, 1735). Maybe the ultimate goal of a good system of classification is to allow the general knowledge of a given phenomenon to go a step further, certainly not classification per se. Even in the research concerning how meaning is represented in the speaker's mind/brain, classification is not a minor detail. A pivotal categorization is the one between concrete (e.g., banana, hand, table, bolt), and abstract (e.g., peace, love, justice, ideal) meanings, respectively defined as referring to something that can either be directly experienced or not through the senses (Hale, 1988). Over the last forty years, the dichotomy between concrete and abstract semantic categories has been suggested by data from: (i) rating studies, describing concrete words as more imageable, easier to think of a specific context for, more familiar, and acquired earlier during infancy than abstract words (Paivio et al., 1968; Schwanenflugel et al., 1988; Barca et al., 2002); (ii) behavioral experiments, demonstrating a concreteness effect, i.e. a cognitive advantage for concrete over abstract meanings in terms of speed and accuracy with which words are processed (Holcomb et al., 1999; Binder et al., 2005); but see (Kousta et al., 2011); (iii) neuropsychological research, reporting double dissociations, i.e. cases of patients more impaired with concrete words, as opposed to other patients more impaired with abstract words (Gainotti, 2004); (iv) neuroimaging studies, suggesting different neural networks supporting abstract and concrete meaning processing (for reviews, see Binder et al., 2009; 2011). At the theoretical level, the differences between concrete and abstract concepts have been explained in terms of greater availability either of both the perceptual and verbal information (Paivio, 1968), or of related contextual information (Schwanenflugel, 1991) for concrete versus abstract concepts. Concrete concepts were also described as being characterized by a higher number of semantic features (Plaut & Shallice, 1991). In contrast to such quantitative accounts, according to which abstract and concrete words differ in terms of the amount of information involved, a recent account rather posited qualitative differences between concrete and abstract words. This kind of alternative theoretical proposal was based on evidence collected in patients (Crutch & Warrington, 2005) and crucially also in healthy subjects (Duñabeitia et al., 2009). Accordingly, it has been suggested that the distinction between concrete and abstract words is embedded in qualitatively different principles of organization for concrete and abstract words, that is, respectively, a categorical versus an associative organization (Duñabeitia et al., 2009).

A limitation of the majority of the aforementioned theoretical accounts on the differences between concrete and abstract meanings is that they do not seem to provide interpretations for subtler sub-categorizations within the concrete and abstract domains. As a matter of fact, beside the more general classification between abstract and concrete meanings, it is also possible to augment the level of categorical resolution both within the concrete and the abstract semantic domains. Within the concrete domain, different categories have been identified. As suggested by Wiemer-Hastings and colleagues (2003), concrete items are characterized by salient dimensions that allow them to be readily classified into categories. For example, given a set of concrete words such as *apple*, *cabbage*, *squirrel*, and *duck*, their sorting into different classes, i.e. vegetables and animals, is straightforward. A potential explanation of this phenomenon is that concrete words belonging to the same category would typically share some features, making them more similar to each other than to other items belonging to distinct categories (Wiemer-Hastings et al., 2003). For example, considering the category of animals, some features such as 'has ears' and 'has a tail' are shared by many members of the same category (Taylor et al., 2007). The distinction of concrete meanings into different sub-categories is also supported by neuropsychological and neuroimaging evidence. Brain damaged patients can show deficits restricted to a single domain (e.g., living things, non-living things), or a category (e.g., animals, fruits, tools, musical instruments, body parts) of knowledge (Tyler et al., 2001). Neuroimaging studies reported sensory modality-specific brain activations for linguistic items referring to entities experienced through senses, such as tactile- (Goldberg et al., 2006), taste- (Simmons et al., 2005), sound- (Kiefer et al., 2008), odor- (González et al., 2006), and visual-related meanings (Martin, 2007). The available literature consistently showed that also action-related concepts identify a category with specific neural substrates (Pulvermueller, 2005), and whose existence can be inferred by means of behavioral experiments (Buccino et al., 2005; Sato et al., 2008). Previous neuroimaging studies (Hauk et al., 2004; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006) also proved that different sub-categories of action-related meanings (such as mouth-, hand-, or leg-related utterances) were somatotopically represented in the left motor and premotor cortex.

The strong overlap between the neural correlates involved in processing semantic knowledge referring to either sensory or motor entities and the neural systems devoted to the sensory-motor experience with those entities, has been formalized particularly over the last fifteen years into the theoretical framework of embodied cognition (Pfeifer & Scheier, 2001). Within this framework, the fine-grain distinction between different categories of concrete concepts naturally follows from the general idea that concepts referring to either sensory or motor entities are stored at least in part in the specific neural systems that mediate the experience with the concepts' referents (Barsalou, 1999; Pulvermueller, 1999).

What about abstract meanings, then? Is it possible to draw fine-grained categorical distinctions within the abstract domain, similarly as for the concrete domain of conceptual knowledge? Embodied cognition accounts have postulated that also in the abstract domain, the storage of conceptual knowledge may reflect the type of experience that is characteristic for the concepts' referents, with for example an involvement of the neural systems processing emotions for affective concepts, and of the mentalizing neural network for introspective concepts referring to mental states (Barsalou, 2008; Simmons et al., 2008).

Evidence compatible with such a generalized embodied account has more recently begun to emerge (e.g., Ghio & Tettamanti, 2010; Moseley et al., 2012), but otherwise the domain of abstract meanings has been scarcely explored and generally regarded as an undifferentiated whole in experimental studies (for a review, see Binder et al., 2009). To start with, the definition of abstract words do not fully characterize abstract concepts, as they are mainly defined by exclusion (Wiemer-Hastings & Xu, 2005), namely as referring to entities that are neither physically nor spatially constrained. It has also been suggested that, in sharp contrast with concrete words in which features are shared within the same category, categories of abstract items have a low inter-category distinctiveness (Wiemer-Hastings et al., 2003). For example, similarity ratings for a pair of items belonging to the same abstract category (e.g., events) were lower than similarity ratings for a pair of items belonging to the same concrete category (e.g., plants) (Wiemer-Hastings et al., 2003). As a consequence, “abstract” has been often used as a wide label including words that do not have physical referents, such as *happiness*, *justice*, and *doubt*, without considering the heterogeneity of this class of meanings (Cappa, 2008).

Only few studies have shed light on whether there exist differences between categories of abstract-related concepts. Setti and Caramelli (2005) investigated three sub-categories of abstract concepts largely related to mental states (nominal kind, state of the self, and cognitive processes), reporting that each semantic domain showed a specific pattern in concreteness/abstractness and imagery ratings, and a specific pattern of information (taxonomic, thematic, and attributive) in a definition production task.

Another semantic category which has generally been confounded among other instances of the generic abstract category is represented by emotion-related concepts. In a rating study, Altarriba and colleagues (1999) showed that, when treated as a separate category, emotion words (e.g., *excited*, *lonely*, *infatuated*, *upset*) were less concrete and lower in context availability, but more imageable than abstract words (e.g., *easy*, *donor*, *travel*, *finish*). In a subsequent memory recall study, the same authors found that emotion words were better remembered than either concrete or abstract words (Altarriba & Bauer, 2004), thus revealing the distinctiveness of emotion meanings in comparison to both concrete and abstract meanings. Kousta and colleagues (2009) showed in a lexical decision task that, irrespective of valence (namely, positive or negative), emotional words were processed more quickly than neutral words. However, evidence is still not clear cut. For example, in terms of reaction times, either a disadvantage (Estes & Adelman, 2008) or an advantage (Nasrallah et al., 2009) was found for negative emotion words. These controversial results could have been due to different task demands that may modulate the effect of emotions, different criteria for item selection, or sampling differences for valence (Kousta et al., 2009; 2011).

As still another potential abstract semantic category, recent studies focused on mathematics-related concepts, considering them as a special case of abstract concepts, with a strong link between numerical representations and the hand fingers used for counting (Ranzini et al., 2011; Previtali et al., 2011; Fischer & Brugger, 2011).

This brief review of the specialistic literature clearly indicates that evidence on abstract meanings representation and processing is highly fragmentary, and still limited to restricted lexical-semantic domains. In the present study, we propose that in order to improve our understanding of the processing and representation of the abstract conceptual-semantic domain, the time is ripe for developing a more fine-grained classification. As a first step in this direction, considering previous language studies suggesting the existence of different types of abstract meanings, we putatively distinguished between three different categories within the abstract domain: mental state-related meanings, emotion-related meanings, and mathematics-related meanings. Instead of single words as in most previous studies, we used sentences, which, as we will argue, allow for the resolution of many lexical-semantic confounding side-effects.

Mental state-related meanings mainly referred to several cognitive states expressed by mental state verbs (Papafragou et al., 2007) and dealing with abstract entities (e.g., *She contemplates the alternative*).

With respect to emotion-related meanings, differently from most studies aimed at investigating the relationship between language and emotions, we considered only utterances referring to emotions and feelings *per se* (e.g., *She feels disgust*). We in turn excluded highly arousing utterances referring to actions or entities with an emotional connotation (e.g., *She stabs her husband*; see also Moseley et al., 2012 proposing a similar approach).

Mathematics-related concepts, as a special case of abstract knowledge with sensory-motor grounding in hand finger representations, referred to calculations and other mathematical operations (e.g., *She counts the sets*).

We compared mental state-, emotion-, and mathematics-related meanings to three action-related meaning categories within the concrete semantic domain. Based on their relevance for evidence-based sensory-motor embodiment, we distinguished between mouth-related (e.g., *She inflates the balloon*), hand-related (e.g., *She plucks the strings*), and leg-related meanings (e.g., *She bends the knee*), since a fine-grained characterization of effector-specific action-related meanings in psycholinguistic terms is still missing.

The first objective of this study was to provide empirical evidence of fine-grained distinctions within both the abstract and the concrete semantic domains with respect to relevant psycholinguistic dimensions. As we suggested above (see also MacRae & Jones, 2013), the abstract and concrete categories are very heterogeneous, including several different classes of meanings that deserve a thorough psycholinguistic and neuroscientific characterization. In the present study, we start by characterizing meanings with respect to several psycholinguistic dimensions, in order to provide psycholinguistic measures that may guide the selection of stimuli in future studies. In line with this, the second aim of this study was to develop a carefully controlled linguistic stimulus set that may be used for auditory as well as visual neuroimaging studies focusing on the parametrization of the semantic space beyond the abstract-concrete dichotomy.

For these purposes, we created a set of Italian sentences that refer to the six semantic classes described above, and carefully controlled for: (i) psycholinguistic characteristics, such as sentence length, lexical frequency, and syntactic form. The effects of these psycholinguistic variables on behavioral responses and brain processes has been clearly demonstrated for linguistic stimuli presented either in the visual or in the auditory modality (Norris, 2006; Constable et al., 2004); (ii) auditory characteristics, such as prosody, pitch, intensity, and sentence duration, which also influence auditory stimulus processing (Ben-David et al., 2011).

Sentences were characterized at the psycholinguistic level by means of two rating studies. Study 1 was aimed at verifying through a rating procedure whether the literature-based distinction of the abstract and concrete domains into different semantic categories was reflected by speaker's judgments. Participants were asked to evaluate sentences with respect to different semantic domain-related scales, specifically created for measuring if and how sentences were categorized.

In study 2, we measured the concreteness/abstractness of the six semantic categories by means of concreteness ratings. We also characterized the set of stimuli for familiarity and context availability. All these psycholinguistic variables have been used in previous studies to quantify the differences between concrete and abstract meanings at the word level (Coltheart, 1981; Barca et al., 2002; Della Rosa et al., 2010). The current literature does not provide normative data about concreteness, context availability, or familiarity for sentence stimuli, except for studies considering special types of sentences, such as metaphorical sentences (Cardillo et al., 2010). By collecting these ratings, we aimed at providing standard measures to quantify similarities/dissimilarities among different semantic categories within the concrete and abstract domains, also extending previous results at the sentence level.

This set of stimuli may be used in future neuroscientific and behavioral studies on the processing of different semantic categories either through visual or auditory perception. Relying on the provided rating measures, in future research the factors and psycholinguistic variables considered here (i.e. semantic domains, concreteness/abstractness, length, frequency, familiarity, context availability) may be experimentally manipulated in a factorial or a parametric fashion, either as parameters of interest or as confounds.

2. Materials and Methods

2.1 Linguistic stimuli

In a series of normative pre-tests, 150 volunteers (different from the ones mentioned below as participants) evaluated different versions of the sentences with respect to different variables. Pre-tests were paper and pencil questionnaires asking participants to judge all sentences on concreteness, context availability, familiarity, and body-part involvement using 7-point Likert scales. Pre-normative results were statistically evaluated in order to guide the final choice of the sentences to be used in the present study.

The 210 selected Italian sentences all consisted of four words and had the same syntactic structure: third person feminine pronoun, verb in third-person singular, simple present tense, matched to a syntactically and semantically congruent object complement. Thirty-five sentences for each of the three abstract-related semantic domains were created: mental state-related sentences (**Ms**) (e.g., ‘Lei ricorda il passato’, Engl.: *She remembers the past*); emotion-related sentences (**Em**) (e.g., ‘Lei mostra il disappunto’, Engl.: *She shows her disappointment*); mathematics-related sentences (**Ma**) (e.g., ‘Lei calcola la somma’, Engl.: *She determines the sum*). Thirty-five sentences for each action-related semantic domain were also formed: mouth-related sentences (**Mo**) (e.g., ‘Lei schiocca la lingua’, Engl.: *She clicks her tongue*); hand-related sentences (**Ha**) (e.g., ‘Lei ricama il fazzoletto’, Engl.: *She embroiders the handkerchief*); leg-related sentences (**Le**) (e.g., ‘Lei calcia la palla’, Engl.: *She kicks the ball*). For simplicity, example sentences in the remainder parts of the paper are only provided in the form of literal English translations from Italian, omitting in turn the original Italian versions.

Experimental stimuli were controlled for length and frequency of use across the six experimental conditions. The length of sentences was measured by the number of words and letters (important if sentences are to be presented in a visual format), and by the number of syllables (important if sentences are to be presented in a spoken format). The frequency of use was controlled by considering two different measures: (i) a measure of lexical frequency of the content words constituting the sentences (e.g., *kicks* and *ball* are the content words of the sentence *She kicks the ball*) on the basis of the available frequency norm of Italian Corpus and Frequency lexicon of written Italian (ColFIS, Bertinetto et al., 2005); (ii) a subjective measure of the sentence frequency was obtained by means of familiarity rating (for details see section 2.3 *Rating study 1*).

Linguistic stimuli in auditory form

As this study aimed at providing a set of sentences that can be used in future studies not only in a visual format, but also in an auditory format, we created a recorded version of the set of stimuli as well.

Sentences were pronounced by a female, native speaker of Italian in an anechoic room, while registering in stereo modality with a 96.000 Hz sampling rate and a bit-depth of 16 bit. To avoid prosodic effects, and to minimize possible confounding influences of low-level auditory features such as pitch or accent, all sentences were read with a controlled neutral intonation. After recording, a manipulation procedure was applied to all sentences using Praat 5.2.03 software (www.praat.org; Boersma, 2001). Praat scripts, available at the Praat Script Archive (www.sites.google.com/site/praatscripts), were specifically modified for: (i) cutting traces, in order to leave no silence at the beginning and at the end of each sentence; (ii) fixing each audio trace to the same amplitude interval (70 dB); (iii) extracting the values of the following parameters: temporal duration, mean intensity and mean pitch.

The complete set of written and auditory Italian sentences (see Appendix 1) and the modified Praat scripts can be obtained by sending requests to M.T. (tettamanti.marco@hsr.it).

2.2 Participants

Ninety-six undergraduate students from the Vita-Salute San Raffaele University, Milan (63 males, mean age = 20.0 ± 0.7) participated to this study. Half of the participants were randomly assigned to group 1 and performed rating study 1, the other half were assigned to group 2 and performed rating study 2. All subjects were native Italian speakers. Education level was highly matched as all participants were attending the first year Medicine course (years of education mean = 13.5 ± 1.5). They were not paid nor received extra credits for their participation. Participants were unaware of the aim of the study, and they were not experts in linguistics nor in the specialistic psycholinguistic and cognitive neuroscientific literature.

2.3 Rating study 1

Rating study 1 aimed at validating the putative distinction of sentences into six different semantic categories suggested on the basis of the current literature by means of association and body-part ratings.

Association task: for Ms, Em, and Ma sentences, we asked participants to evaluate how much the meaning of each sentence was associated to the meaning of three other sentences (one Ms, one Em, and one Ma) randomly selected from the pool of abstract-related sentences. For example, subjects had to judge how much the meaning of a target sentence like *She feels happy* (Em) was associated to the meaning of the three following sentences: *She memorizes the procedure* (Ms), *She conceals the anger* (Em), and *She calculates the sum* (Ma). For each target sentence, we created a specific triplet in order to use each Ms, Em, and Ma sentence only once; the order of the presentation of the sentences in the triplet was randomized. For each association, a 7-point Likert scale was employed ranging from 1 = “not associated” to 7 = “highly associated”. By way of this association task, we investigated whether different semantic classes could emerge from the rating data, without imposing a priori the semantic categories to which they possibly belonged. More specifically, we expected that Ms, Em, and Ma sentences clustered with their corresponding counterparts.

Body-part task: for Mo, Ha, and Le sentences we asked participants to evaluate how much the action described in each sentence involved the mouth, the hand, and the leg using three body-part Likert scales (mouth scale, hand scale, leg scale) ranging from 1 = “not involved” to 7 = “highly involved” (Hauk & Pulvermueller, 2004; Willems et al., 2010). To better characterize a potential motor dimension of abstract-related sentences, we asked participants to also rate Ms, Em, and Ma sentences.

For both the association and the body-part tasks, two sentence-response examples were provided for reference with the task instructions, using different stimuli than those from the experimental set.

Procedure rating study 1

The pool of 210 sentences was divided into six separate lists. Lists were rotated among the two tasks, i.e. the association task and the body-part task. Five of the lists included 18 target sentences (3 sentences for each of the 6 experimental conditions) for the association rating, and 36 sentences (6 sentences for each of the 6 experimental conditions) for the body-part rating; one list included 15 target sentences for the association rating and 30 sentences for the body-part rating. By means of this procedure, all sentences were scored, avoiding the same subject to rate the same sentence more than once. At the same time, the use of relatively short lists was aimed at preserving a high level of attention throughout the study, and preventing from fatigue. Between lists, the order of the presentation of the tasks was counterbalanced across participants. Within each list, the order of sentences was pseudo-randomized. For each rating, each sentence was rated by 8 participants.

The rating was conducted through a web-based procedure using Survey Monkey (SurveyMonkey.com, LCC, Palo Alto, California, USA, www.surveymonkey.com). Each participant completed the rating study individually on a computer console. Sentences were presented one by one on the screen, and subjects expressed their judgments by clicking on the chosen value of the Likert scales reported under each sentence. This procedure was intended at having a better control over the presentation of items as they were administered in conformity with the sequential order decided by the experimenter. Moreover, participants' rating scores were directly coded on an Excel database file, avoiding mistakes related to the recording of scores. All consent information and instructions for the tasks were provided in Italian, through the same web-based utility. Altogether, the experimental session took no longer than 20 minutes for each subject.

2.4 Rating study 2

To quantify and measure the differences between semantic categories, we designed a second rating study in which sentences were rated on concreteness (CNC), context availability (CA), and familiarity (FAM) by means of 7-point Likert scales. The instructions for the concreteness, the context availability, and the familiarity tasks were largely based on those used by previous investigators for single words (Schwanenflugel et al., 1992; see Della Rosa et al. 2010 for the Italian version of the tasks' instructions), and adapted for use with sentences.

Concreteness task: participants were asked to judge whether the semantic meaning depicted by the sentence either referred to a non-physical situation/state or to a physical action involving objects, materials and/or people (1 = "abstract", 7 = "concrete").

Context availability task: subjects were asked to rate the ease with which they could think of a specific context or circumstances associated with the sentence or in which the sentence could appear (1 = "very difficult", 7 = "very easy").

Familiarity task: participants judged how often they usually listened to or produced each sentence (1 = "unfamiliar", 7 = "very familiar").

A few sentence-response examples were provided for reference with the task instructions, using different stimuli than those from the experimental set.

Procedure rating study 2

Similarly to rating study 1, six lists were created, and rotated among the CNC, CA, and FAM scales so that all sentences were rated on all dimensions but the same subject did not rate the same sentence more than once. An equal number of Ms, Em, Ma, Mo, Ha, and Le sentences were included in each list (3 lists included a total number of 102 sentences, and 3 lists included a total number of 108 sentences). The same procedure of counterbalancing the order of presentation of the rating scales across participants and presenting sentences in a pseudo-randomized order as in rating study 1 was used. Data were collected with the same web-based procedure described for rating study 1.

2.5 Data analysis

Likert scores obtained in rating study 1 and 2 were analyzed using SPSS 13.0 software (IBM, Somers, NY, USA) and R 2.13.0 (R Core Team (2012). Missing responses (0.06%) in the questionnaires were treated as missing data in the analysis.

There is disagreement between scholars about whether Likert data should be analyzed with a parametric statistics ("liberal" approach) or nonparametric statistics ("conservative" approach) (Knapp, 1990; Jamieson, 2004; Carifio & Perla, 2008; Norman, 2010). A recent study comparing type I and II error rates of a parametric t-test vs. nonparametric Mann Whitney-Wilcoxon test for Likert data (De Winter & Dodou, 2010) showed that both tests generally have equivalent power, except for skewed and peaked distributions for which nonparametric test is superior. Nanna and Sawilowsky (1998) found that the Mann-Whitney-Wilcoxon test was superior in all investigated cases of seven-point Likert data which allows for longer tails and more skewness than five-point data. Leys and Schumann (2010) also showed that nonparametric tests are more powerful when assumptions underlying the use of parametric tests are violated. For each rating, we analyzed the distribution of Likert data showing that the assumption of normality of data distribution was never verified, and some distributions (e.g., concreteness and leg scales) were skewed. Consequently, for each rating, Likert data were analyzed by applying the following procedure: (i) as far as descriptive statistics is concerned, we used median as a measure of central tendency and inter-quartile range as a measure of dispersion. However, given that the largest majority of literature articles report means and standard deviations for descriptive purposes, we also reported these values to facilitate comparisons with previous studies; (ii) we applied the nonparametric Kruskall-Wallis test on raw data to assess differences in mean ranks across the six experimental conditions; (iii) we used post-hoc Mann–Whitney U tests with Bonferroni correction for multiple comparisons. To further control the results obtained following this procedure, for each rating scale we also conducted parametric analyses, both by items and by subjects, by applying the Univariate General Linear Model. In all cases, the results confirmed those obtained with the non-parametric procedure described above, and are not reported in the Results section (3).

In addition, in rating study 2, in order to find the latent patterns underlying our stimuli, CNC, CA, and FAM ratings were explored in R statistical software using the “languageR” package (Murtagh, 2005; Baayen, 2011) by means of correspondence analysis, an exploratory data technique used to analyze categorical data (Benzécri, 1973). The correspondence analysis provides an informative and concise means of visualizing data and it is capable of uncovering relationships both among and between variables. In statistical terms, it tests the association between two variables tallied in the form of a contingency table; graphically, it enables a low dimensional configuration of the associations between the rows and the columns of the contingency table. The goals of the correspondence analysis are to reduce the dimension original space, and to find an optimal subspace that is closest to the cloud of points in the chi square-metric. The loss of information associated with this dimension reduction is quantified in terms of the proportion of the so-called inertia that is explained by the axes displayed. To decide how many dimensions (hereafter named as “factors” according to Baayen, 2011) are needed to explain the variation in the data we used the screeplot, in which the factors' eigenvalues are plotted in order of magnitude from largest to smallest. An “elbow” in the plot, that is a change in slope in the diagram, corresponds to the point where there is a marked drop in the amount of variation explained. Factors with inertia contribution higher than this elbow were selected for interpretation, whereas the factors forming the elbow or lower than the elbow were not further considered. The coordinates of both row and column points of the chi-square contingency table were projected onto the selected low-dimensional subspace: in this representation, row and column points that are close together are more alike than points that are far apart. Finally, in order to describe the distribution of points with respect to the six semantic categories, for each factor we plotted the mean coordinates of the points of each category by means of barplots. These mean coordinates were also statistically compared with respect to the six semantic categories.

Non-parametric Spearman's rank-order correlations (r_s) were calculated in order to assess the relations among: (i) CNC, CA, FAM ratings with respect to all sentence categories; (ii) CNC and body-part ratings with respect to abstract-related categories.

3. Results

3.1 Linguistic and auditory characteristics

Linguistic and auditory characteristics are shown in Table 1. Nouns and verbs frequency were balanced across the six semantic categories (nouns: F(5,204) = 1.861; p = 0.103; verbs: F(5,204) = 1.723; p = 0.131; noun-verb combinations: F(5,204) = 1.824; p = 0.110). The length of the stimuli was also controlled: all sentences had four words and the number of letters was balanced across categories (F(5,204) = 1.250; p = 0.287). However, when considering the number of syllables, we found a trend toward a main effect of the semantic category ($\chi^2(25) = 36.371$; p = 0.066). Statistical analysis of auditory features revealed that mean intensity (F(5,204) = 1.465; p = 0.203), and mean pitch (F(5,204) = 1.433; p = 0.214) of sentences were balanced across the six semantic categories. We found that the difference of sentence duration across categories reached the threshold of significance (F(5,204) = 2.259; p = 0.050).

3.2 Rating Study 1

Association rating

Table 2 presents descriptive statistics (median, inter-quartile range, mean, standard deviation) showing how Ms, Em, and Ma sentences were associated to the meaning of sentences belonging, respectively, to the mental-state, emotion, and mathematics-related semantic domain.

We found a significant effect of the semantic domain for each group of abstract-related sentences (Figure 1). Specifically, Ms sentences received higher scores for the mental-state association scale than for the two other scales ($\chi^2(2)= 148.484$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001); Em sentences received higher scores for the emotion association scale than for the two other scales ($\chi^2(2) = 360.371$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001); Ma sentences received higher scores for the mathematics association scale than for the two other scales ($\chi^2(2) = 381.572$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001). To exclude similarities across different semantic domains, for each association scale we compared the median association scores obtained by the sentences belonging to the three different semantic domains (Figure 1). We found that Ms sentences were significantly more associated with Ms sentences than were Em and Ma sentences ($\chi^2(2) = 151.455$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001); Em sentences were significantly more associated with Em sentences than were Ms and Ma sentences ($\chi^2(2) = 342.740$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001); Ma sentences were significantly more associated with Ma sentences than were Ms and Em sentences ($\chi^2(2) = 381.909$; p < 0.001; Mann Whitney pairwise comparisons, all p < 0.001).

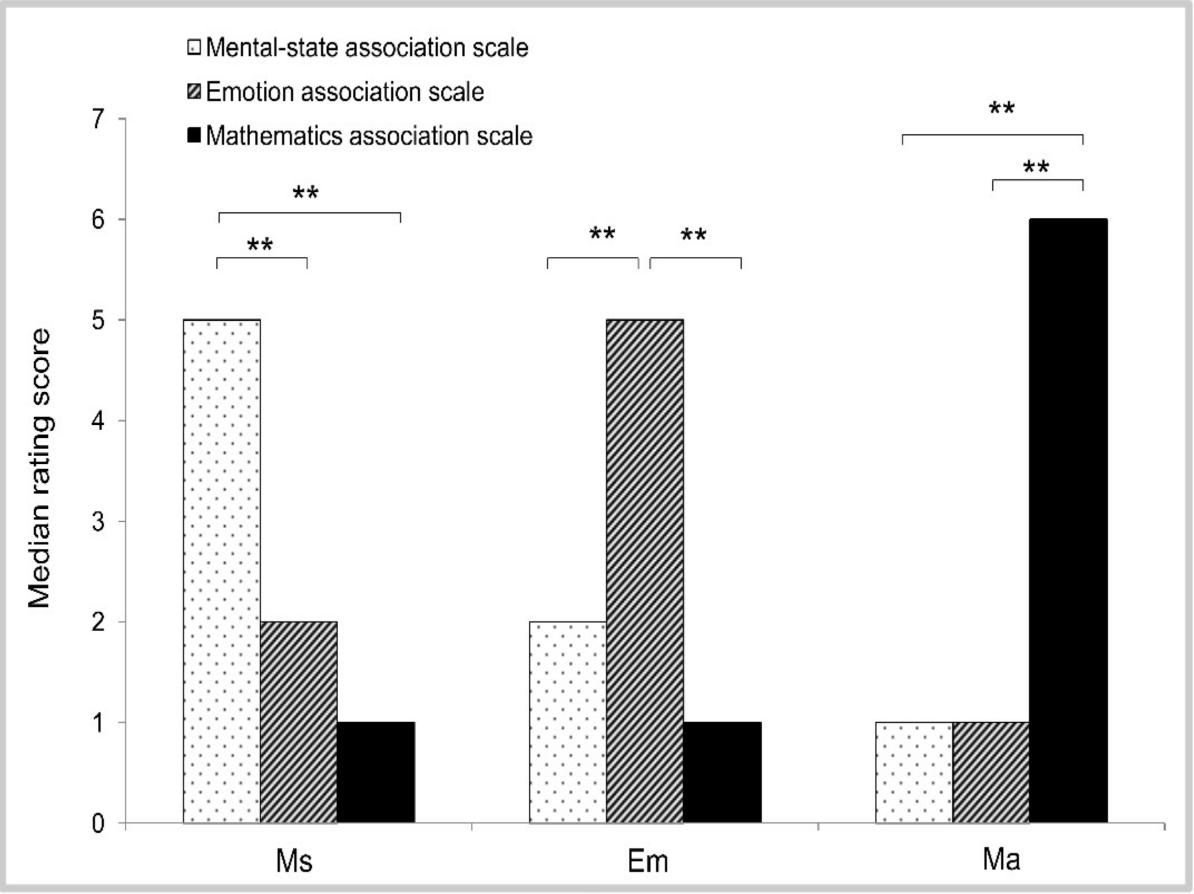


Figure 1. Association rating scores. Bar plot showing median association rating scores for (Ms) mental state-, (Em) emotion-, and (Ma) mathematics-related sentences (* $p < 0.05$, ** $p < 0.01$).

Body-part rating

Table 3 presents descriptive statistics (median, inter-quartile range, mean, standard deviation) describing how each group of sentences was judged for the three action-related scales.

For action-related sentences, we found that the three groups of sentences were different from each other, and also significantly different from abstract-related sentences (Figure 2). Specifically, actions described by Mo sentences were judged as involving the mouth significantly more than the hands or the legs ($\chi^2(2) = 665.939$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$); actions described by Ha sentences were judged as involving hands significantly more than the mouth or the legs ($\chi^2(2) = 608.299$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$); actions described by Le sentences were judged as involving the legs significantly more than the mouth or the hands ($\chi^2(2) = 568.916$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$).

For each body-part scale, we also verified the hypothesis of an association between each group of action-related sentences and the specific effector involved (Figure 2A).

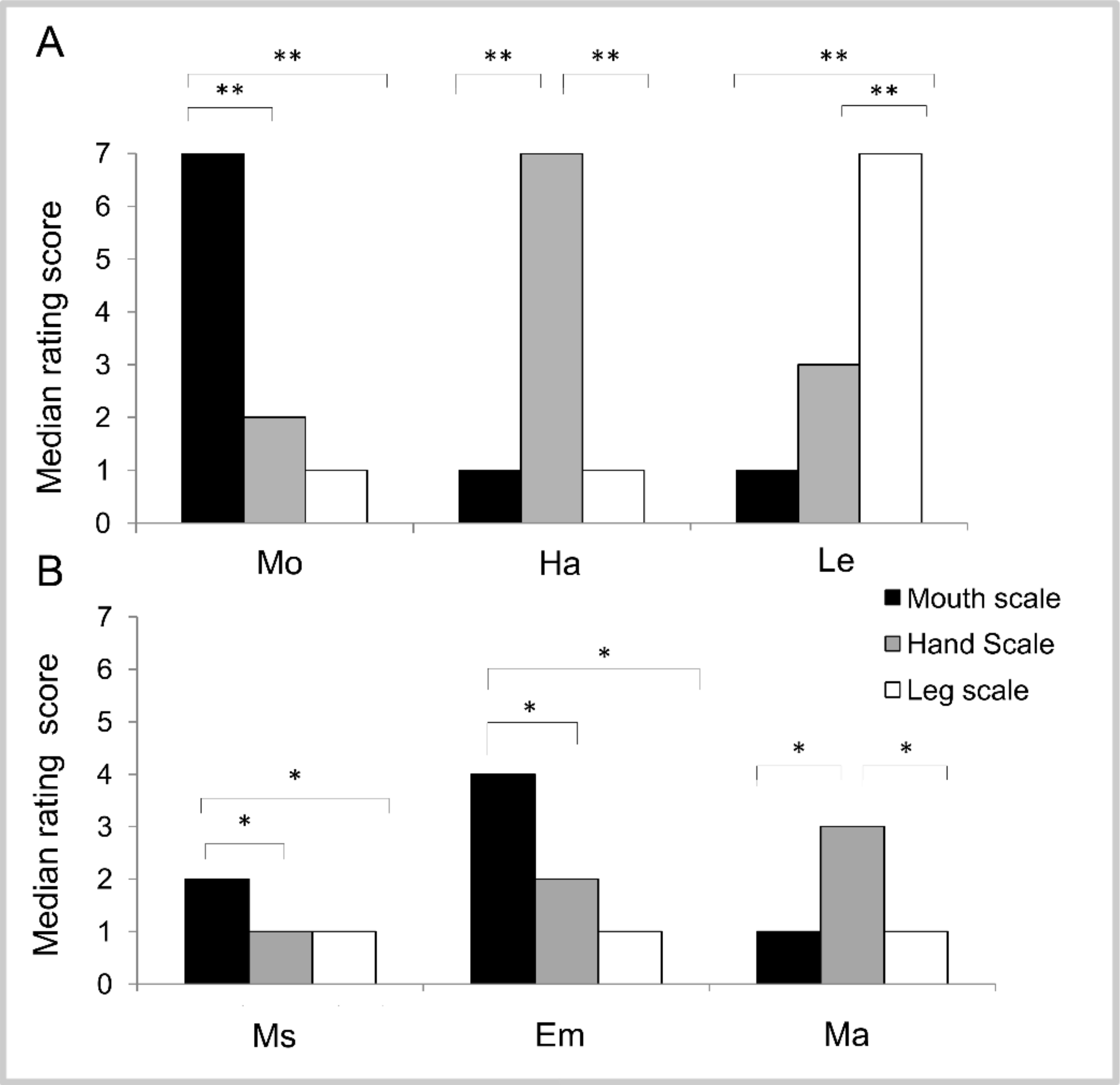


Figure 2. Body-part rating scores. Bar plots showing median body-part rating scores for: A) (Mo) mouth-, (Ha) hand-, and (Le) leg-related sentences, and B) (Ms) mental state-, (Em) emotion-, and (Ma) mathematics-related sentences (* $p < 0.05$, ** $p < 0.01$).

Ratings for the mouth scale revealed that Mo sentences were significantly more associated with the mouth than were Ha, Le, Ms, Em, and Ma sentences ($\chi^2(5) = 848.326$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$). Considering the hand scale, Ha sentences were significantly more associated with the hands than were Mo, Le, Ms, Em and Ma sentences ($\chi^2(5) = 607.613$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$). Consistently, Le sentences were judged as significantly more associated with the legs than were Mo, Ha, Ms, Em, and Ma sentences ($\chi^2(5) = 1013.41$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$).

For abstract-related sentences, results showed that, when explicitly required, subjects judged the content described by Ms, Em, and Ma sentences as significantly involving different effectors (Figure 2B). Specifically, the semantic content of Ms sentences was more associated with mouth actions than with hand or leg actions ($\chi^2(2) = 146.577$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$). The semantic content of Em sentences was more associated with mouth actions than with hand or leg actions ($\chi^2(2) = 88.742$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$). Finally, the semantic content of Ma sentences was more associated with hand actions than with mouth or leg actions ($\chi^2(2) = 227.500$; $p < 0.001$; Mann Whitney pairwise comparisons, all $p < 0.001$). Considering each scale, Mann Whitney pairwise comparisons showed significant differences between Ms, Em, and Ma sentences. Ratings for the mouth scale indicated that Em sentences were significantly more associated to mouth actions than were either Ms and Ma sentences ($p = 0.001$); moreover Ms sentences received higher median score than Ma sentences ($p < 0.001$). Ratings for the hand scale revealed that Ma sentences and Em sentences were significantly more associated to hand actions than were Ms sentences (all $p < 0.001$). Considering the leg scale, Em sentences were significantly more associated with leg actions than were Ma and Ms sentences (all $p < 0.001$).

3.3 Rating Study 2

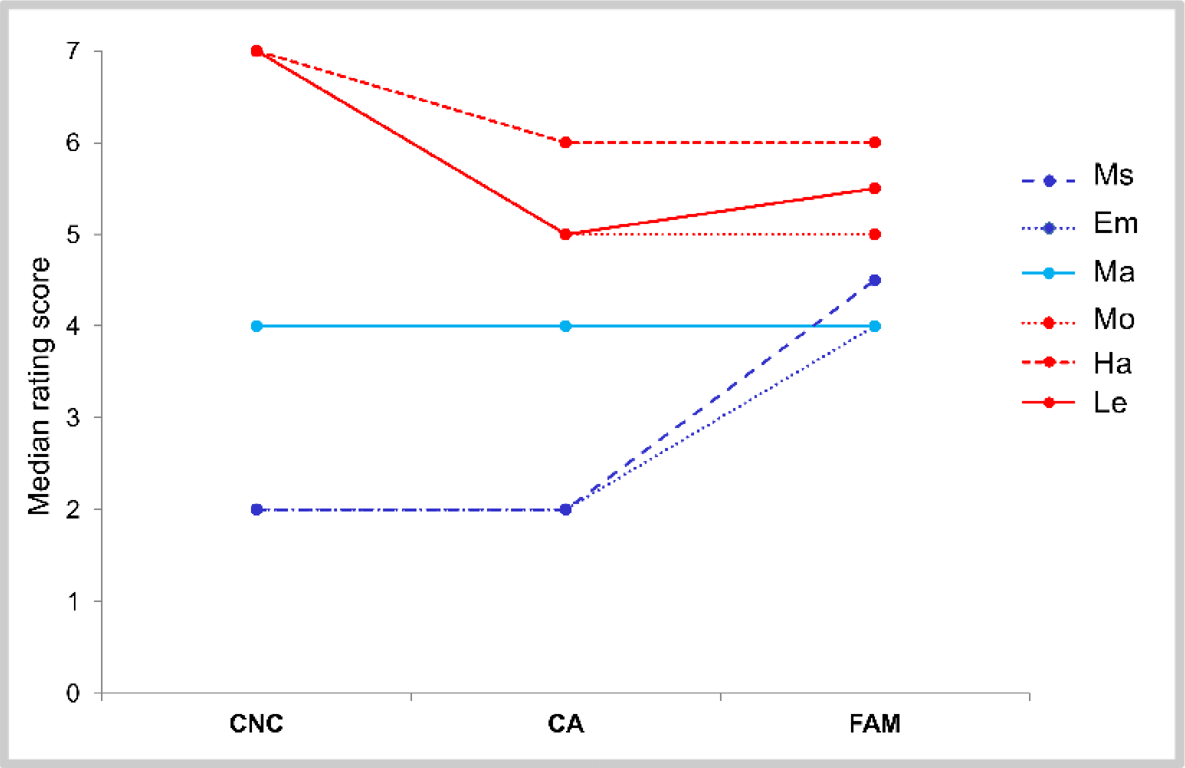


Figure 3. Concreteness, context availability, and familiarity rating scores. Line graph showing median (CNC) concreteness, (CA) context availability, and (FAM) familiarity rating scores for the six categories of sentences. (Ms) mental state-, (Em) emotion-, (Ma) mathematics-, (Mo) mouth-, (Ha) hand-, and (Le) leg-related sentences.

Concreteness rating

We found a significant effect of the semantic domain ($\chi^2(5) = 1117.396$; $p < 0.001$). Based on Mann Whitney pairwise comparisons, four significantly different groups were identified: (i) Ms and Em sentences (Ms vs. Em, $p = 0.297$; all other comparisons: $p < 0.001$); (ii) Ma sentences (all $p < 0.001$); (iii) Mo sentences (all $p < 0.001$); (iv) Ha and Le sentences (Ha vs. Le, $p = 0.211$; all other comparisons: $p < 0.001$) (Table 4, Figure 3). A correspondence analysis was performed with the 210 sentences as one variable (35 Ms, 35 Em, 35 Ma, 35 Mo, 35 Ha, 35 Le) and Likert scores as the other variable. The Chi-square test was significant ($\chi^2(1254) = 2624.613$; $p < 0.001$), indicating an association between variables. The resulting scree plot revealed a marked decrease in the proportion of inertia explained by the third and subsequent eigenvalues, thus suggesting that a two-factor solution comprising only the first and second factors provided a parsimonious decomposition of the original data. The first and the second factors accounted for 48.5% and 19.1% of the total inertia, respectively.

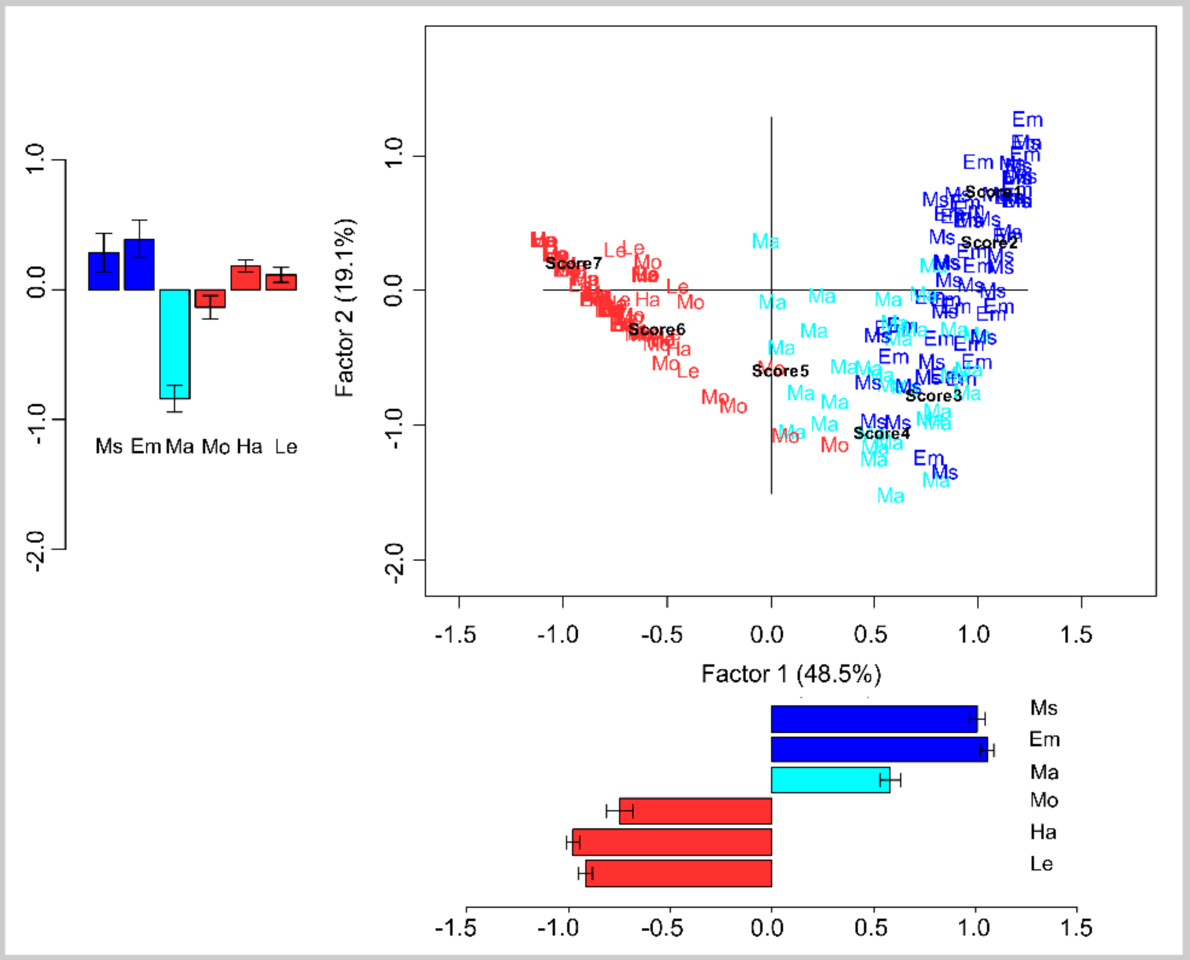


Figure 4. Correspondence analysis for concreteness rating scores. The 210 sentences belonging to the six categories and the 7 Likert points are plotted at their corresponding coordinates. The first and the second factor accounted for the 48.5% and the 19.1% of the total inertia, respectively. Barplots indicate mean coordinates for each factor and category of sentences; error bars indicate standard error means. Action-related (Ha, Mo, Le) sentences are shown in red. Abstract-related sentences are displayed in blue (Ms,Em) and cyan (Ma).

As shown in Figure 4, the first factor roughly separated Mo, Ha, and Le from Ms, Em, and Ma sentences, and may be interpreted to reflect the abstract-concrete dichotomy. By statistically comparing the coordinates along the first factor with respect to sentence categories, a significant difference was found between action-related and abstract-related sentences (t(208) = -41.405; $p < 0.001$). Considering the first factor with respect to the Likert scores, we observed that it was organized according to the exact order of the Likert scale values, with 7 as the leftmost score on the plot and subsequent scores in decreasing order taking a more and more rightward position (Figure 4). As for the second factor, we observed a separation between Ma sentences on the one side and Ms and Em sentences on the other side, thus highlighting a dissociation within the abstract domain. By statistically comparing the coordinates along the second factor, a significant difference was found between Ms and Em vs. Ma (t(88.097) = 8.057; $p < 0.001$). Moreover, the coordinates of Ma were significantly different from those of action-related sentences (t(44.853) = 7.854; $p < 0.001$).

Context availability rating

A significant effect of semantic domain was found ($\chi^2(5) = 345.279$; $p < 0.001$). Mann Whitney pairwise comparisons revealed significant differences between the following subgroups: (i) Ms and Em sentences (Ms vs. Em, p = 0.327; all other comparisons: $p < 0.001$); (ii) Ma sentences (all $p < 0.001$); (iii) Mo and Le sentences (Mo vs. Le, p = 0.120; all other comparisons: $p < 0.001$); (iv) Le and Ha sentences (Ha vs. Le, p = 0.057; all other comparisons: $p < 0.001$) (Table 4, Figure 3).

The correspondence analysis revealed an association between the sentences belonging to the six semantic categories and CA Likert scores ($\chi^2(1254) = 1576.656$; $p < 0.001$). The scree plot indicated a marked decrease in the proportion of inertia explained by the second and subsequent eigenvalues; the second and the following factors were therefore not further considered (for additional confidence, we analyzed the second factor coordinates and did not find any significant effects). The first factor, accounting for 37.9% of the total inertia, roughly separated action-related sentences from Ms and Em sentences, with Ma sentences showing a more dispersed distribution (Figure 5). Factor 1 thus seems to reflect the abstract-concrete dichotomy, but with Ma sentences forming a separate category. By statistically comparing the coordinates along the first factor with respect to sentence categories, we observed a significant difference between: action-related and abstract-related sentences (t(136.562) = -16.962; $p < 0.001$), Ma and abstract-related sentences (t(59.756)= 5.523; $p < 0.001$), and Ma and action-related sentences (t(48.140) = -5.766; $p < 0.001$). As for the Likert scores, the first factor was organized according to the exact order of the Likert scale values, with 7 as the leftmost score on the plot and subsequent scores in decreasing order taking a more and more rightward position (Figure 5).

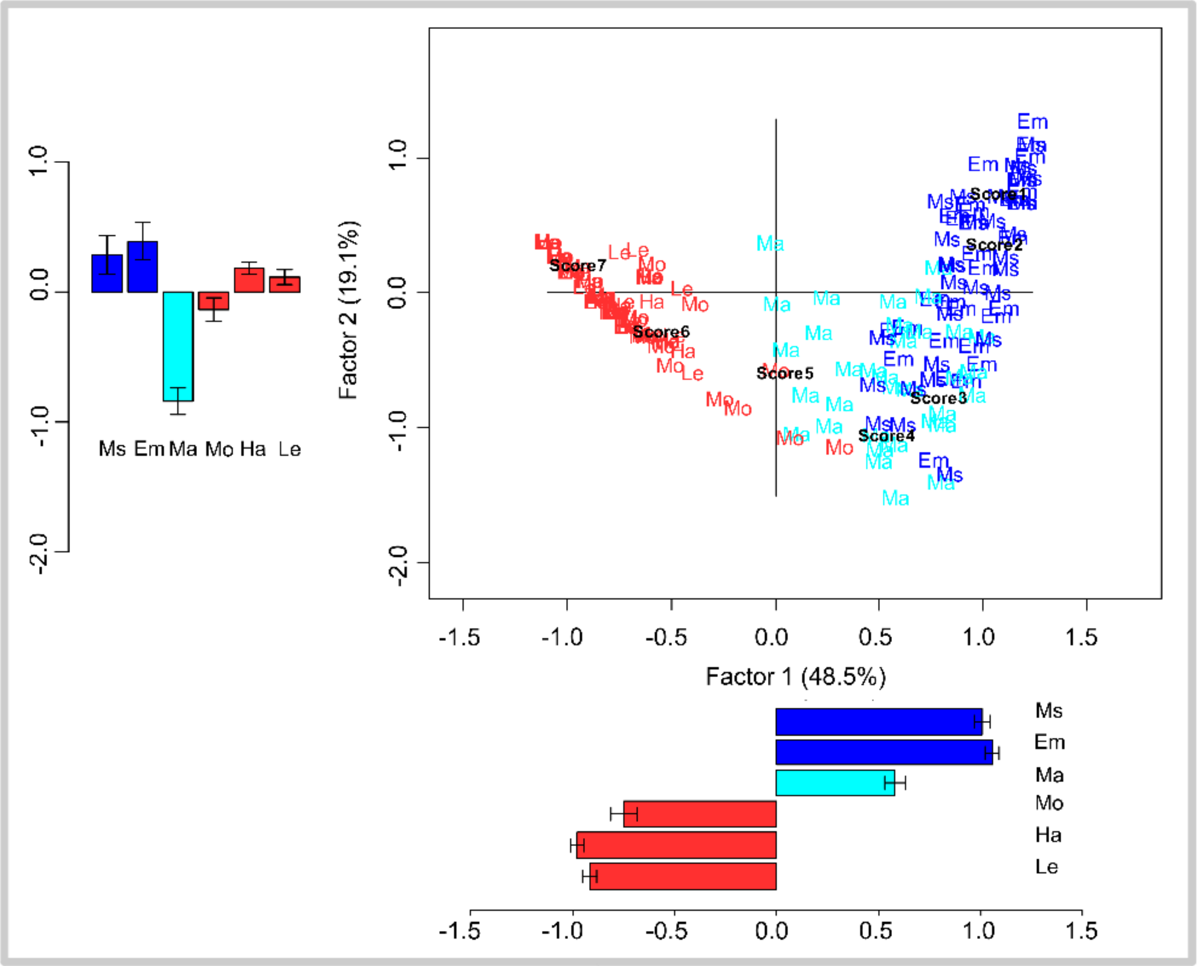


Figure 5. Correspondence analysis for context availability rating scores. The 210 sentences belonging to the six categories and the 7 Likert points are plotted at their corresponding coordinates. The first and the second factor accounted for the 37.9% and the 14.8% of the total inertia, respectively. Barplots indicate mean coordinates for each factor and category of sentences; error bars indicate standard error means. Action-related (Ha, Mo, Le) sentences are shown in red. Abstract-related sentences are displayed in blue (Em, Ms) and cyan (Ma).

Familiarity rating

We found a significant effect of semantic domain ($\chi^2(5) = 109.383$; $p < 0.001$), with Ms, Em, and Ma sentences judged as significantly less familiar than action-related sentences (Mann Whitney comparisons, all $p < 0.001$). No differences were found neither between abstract-related sentences (all $p > 0.05$; alpha level corrected for multiple comparisons = 0.003) nor action-related sentences (all $p > 0.04$; corrected alpha level = 0.003) (Table 4, Figure 3).

Also for familiarity, an association between the sentences belonging to the six semantic categories and Likert scores was revealed by the correspondence analysis ($\chi^2(1254) = 1776.257$; $p < 0.001$) (Figure 6).

The scree plot indicated a marked decrease in the proportion of inertia explained by the second and subsequent eigenvalues; the second and the following factors were therefore not further considered (for additional confidence, we analyzed the second factor coordinates and did not find any significant effects). The first factor, accounting for 35.3% of the total inertia, roughly separated action-related sentences from abstract-related sentences (Figure 6), thus again most likely reflecting the abstract-concrete dichotomy. The distinction of action-related vs. abstract-related sentences into two clusters was confirmed by the analysis of the coordinates along the first factor (t(203.871) = -6.496; p < 0.001). To exclude a possible alternative interpretation in terms of lexical frequency instead of familiarity, we compared the coordinates of high vs. low frequency sentences, and no differences were found (t(208) = 1.244; p = 0.215). As for the Likert scores, the first factor was organized according to the exact order of the Likert scale values, from 7 as the leftmost score to 1 as the rightmost score on the plot (Figure 6).

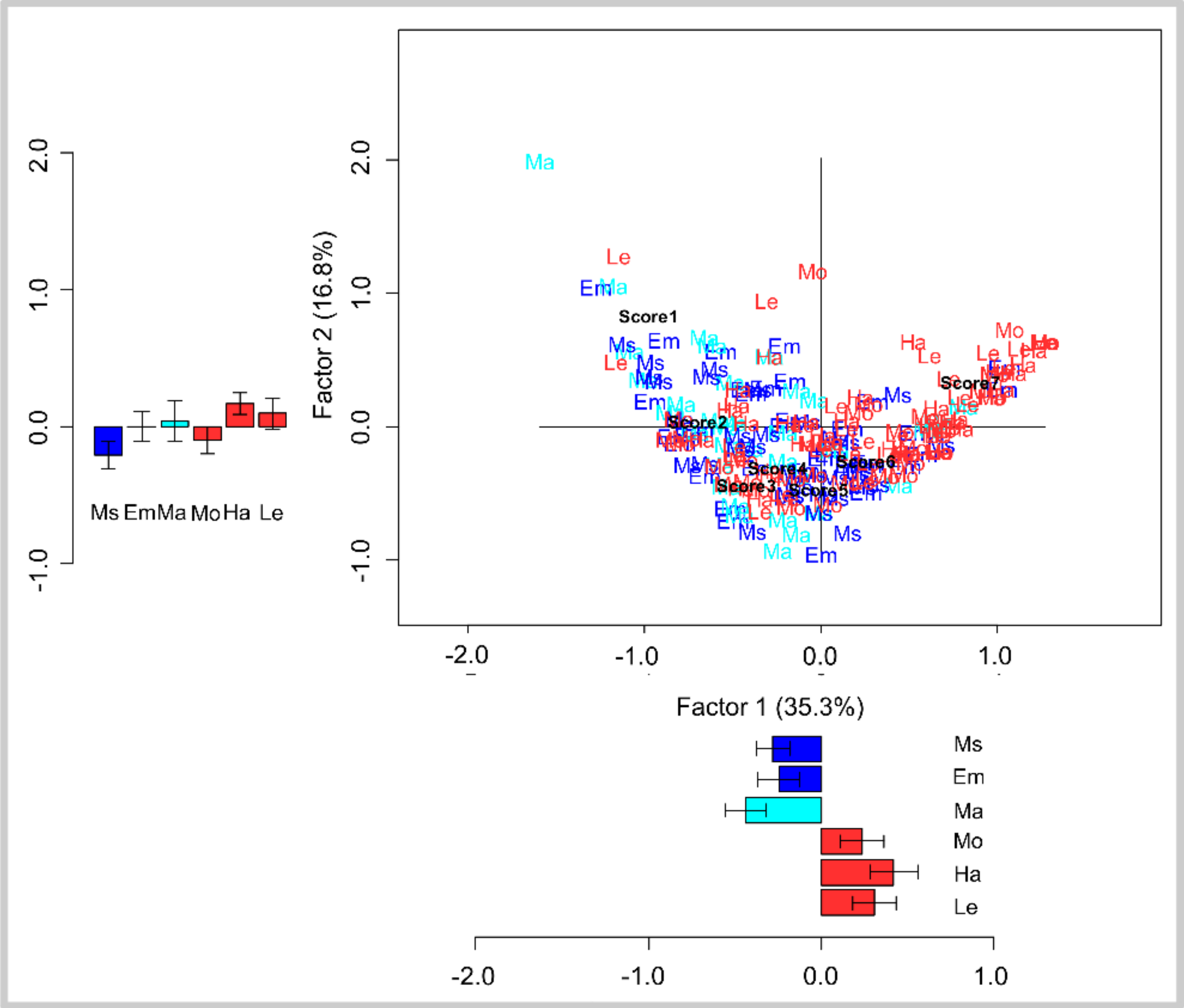


Figure 6. Correspondence analysis for familiarity rating scores. The 210 sentences belonging to the six categories and the 7 Likert points are plotted at their corresponding coordinates. The first and the second factor accounted for the 35.3% and the 16.8% of the total inertia, respectively. Barplots indicate mean coordinates for each factor and category of sentences; error bars indicate standard error means. Action-related (Ha, Mo, Le) sentences are shown in red. Abstract-related sentences are displayed in blue (Ms,Em) and cyan (Ma).

3.4 Correlation analysis

We calculated the correlations for CNC, CA, FAM variables across semantic categories. Consistently with the extant literature (Cardillo et al., 2010), all the variables significantly correlated with each other: CNC and CA ($r_s = 0.745$; $p < 0.01$); CNC and FAM ($r_s = 0.440$; $p < 0.01$); CA and FAM ($r_s = 0.521$; $p < 0.01$).

Following Altarriba et al. (1999), we also calculated the relation among variables within each semantic group. We found that CNC and CA did not correlate (Spearman's correlation on median scores: all $p > 0.05$). CA and FAM were significantly correlated for Ms ($r_s = 0.568$; $p < 0.001$), Em ($r_s = 0.381$; $p < 0.05$), Ma ($r_s = 0.635$; $p < 0.001$), and Mo ($r_s = 0.449$; $p < 0.001$) sentences, but did not correlate for Ha and Le sentences ($p > 0.05$). FAM and CNC were correlated for Em sentences only ($r_s = 0.456$; $p < 0.001$).

Finally, in order to characterize the possible relationship between the rated involvement of the three body parts and the perceived concreteness of each abstract-related category, we also calculated abstract category-specific correlations between CNC and body-part ratings (Table 5). Significant correlations were found between CNC and all three body-part scores for Em sentences, and between CNC and mouth-related scores for Ms sentences.

3.5 Cross-study validation

For cross validation purposes, we conducted correlations between our data at sentence-level and relevant word-level normative data publicly available. As our stimuli are in Italian, we referred to data of a norming study on Italian words by Della Rosa et al. (2010), which is the yet widest normative study in Italian providing concreteness, context availability, and familiarity scores. In this study (Della Rosa et al., 2010), nouns were taken from the MRC Psycholinguistic Database (Coltheart, 1981), translated from English to Italian, and rated for the variables of interest. No verbs were included in Della Rosa et al. (2010), so our correlations are limited to the noun grammatical category.

Thirty five out of the total of 210 nouns in our stimulus set were available in Della Rosa et al.'s dataset. Correlations were done on this small subset of stimuli on CNC ($r_s = 0.815$; $p < 0.001$), CA ($r_s = 0.670$; $p < 0.001$), and FAM ($r_s = 0.195$; $p = 0.262$) scales. As expected, a high level of coherence was found for CNC and CA between the word-level noun ratings of Della Rosa et al. (2010) and the corresponding sentence-level ratings collected in our study for sentences containing the same nouns. In turn, we did not find any significant correlations for the FAM scale. This may not be much surprising, since the noun *horse*, for example, may be rated as highly familiar as an isolated word, whereas a sentence including *horse*, such as *She rams the horse*, might have been encountered/used relatively infrequently and thus obtain a low familiarity score.

We believe that, in spite of the limited sample (35 out of 210 cases), the fact that the between-sets correlations for CNC and CA were highly significant allows us to conclude with sufficient confidence that the ratings we have collected at the sentence level do not provide a biased picture with respect to available data at the word level. For further confidence in the generalizability of our between-sets correlation results, we also performed nonparametric bootstrapping simulations (Efron, 1979) in order to estimate the extent to which the results obtained with such a limited sample may still hold in the probabilistic scenario of much larger samplings. We let the R statistical software randomly resample the 35 rating pairs (i.e. the pairs constituted by our ratings and those of Della Rosa et al., 2010) 10'000 times with replacements, and we then calculated the ensuing distribution of the Spearman correlation scores for each simulated sample together with the 95% percentile Confidence Intervals (CI) (Davison & Hinkley, 1997). The results of the bootstrapping simulations confirmed the high correlation between the word-level noun ratings of Della Rosa et al. (2010) and the corresponding sentence level ratings collected in our study for CNC (r_s = 0.815; 95% CI = 0.629–0.917) and CA (r_s = 0.670; 95% CI = 0.409–0.806), but not for FAM (r_s = 0.195; 95% CI = -0.139–0.477).

4. Discussion

Until now, psycholinguistic studies investigated semantic knowledge by showing a dichotomy between abstract and concrete meanings (Wiemer-Hastings et al., 2001). However, there is increasing evidence from neuroimaging studies that the neural networks involved in the representation of meanings are flexible and extended throughout the cerebral cortex (Martin, 2007; Pulvermueller, 2005), thus suggesting that the simplistic classical dichotomy between abstract and concrete meanings has little explanatory power. Such evidence brings into question theoretical accounts explaining the differences between concrete and abstract concepts, both in terms of quantitative (Paivio, 1986; Schwanenflugel, 1991) or qualitative (Crutch & Warrington, 2005; Duñabeitia et al., 2009) differences, without considering within-domain distinctions. Furthermore, experimental data are pivotal to grounded theories of semantics, according to which the conceptual representation of a semantic category can be viewed as a collection of the multimodal information that has been experienced and processed for instances of that category (Wilson-Mendenhall et al., 2011; Kiefer & Barsalou, 2012). In general, concrete meanings are thought to mainly rely on modalities and systems that process perception and action, while abstract meanings have been suggested to bear on internal states (Barsalou, 1999; Pulvermueller, 1999; Barsalou, 2008). Assuming a more specific categorization of the concrete domain, it has been shown that the conceptual-semantic language processing of, for example, utterances whose semantic content is related to a particular sensory modality relies on distributed neural networks including the sensory-motor system (Kiefer et al., 2008; González et al., 2006; Martin et al., 1996; Barrós-Loscertales et al., 2012). Conversely, evidence about the semantic networks supporting the processing of different types of abstract meanings is sparse. One reason may be the under-specification of abstract-related meanings so far. A much finer distinction of subordinate referential domains in the abstract domain is nevertheless possible and should by now be taken into consideration. For instance, the above mentioned "internal states", considered relevant for abstract-related meanings, include: interoception (e.g., affective valence, arousal, hunger, pain, visceral activity, muscle tension), mentalizing (e.g., self-related thoughts, evaluations, representing the thoughts of others, representing how one is perceived by others), attention, reward, affects, executive processing, memory, and reasoning (Wilson-Mendenhall et al., 2011). All these different internal states could be systematically operationalized at the experimental level in future studies, as done at least in part here.

In this study we have offered a psycholinguistic characterization of different conceptual-semantic categories, with a special focus on abstract-related meanings. These data may be quite helpful for future studies aimed at unraveling the grounding of semantic language processing, mainly for two reasons: i) a more accurate description of the psycholinguistic characteristics of categories within the concrete and abstract domains may provide further hints on the type of information included/aggregated to form a conceptual representation; ii) data about psycholinguistic variables such as length, frequency, concreteness, context availability, familiarity, and body-part involvement can be better controlled, as we will suggest, within a parametric experimental approach. Notably, this stimulus set may be suitable for behavioral and neuroimaging research aimed at investigating semantic processing by means of experimental paradigms employing either visually or auditorily presented linguistic stimuli. Relevant linguistic features, i.e. sentence length and lexical frequency have been controlled for all the sentence categories. Familiarity ratings, considered as a subjective measure of frequency (Gernsbacher, 1984), revealed that action-related categories were significantly more familiar than abstract-related categories. In order to extend the range of utilization of these stimuli and to make auditory presentation feasible as well, the digitally recorded sentences were matched for mean intensity, mean pitch, and temporal duration, minimizing the possible influences of low-level auditory features. Indeed, a measurable impact of these linguistic characteristics on language processing has been demonstrated both at the behavioral and neural level not only for words, but also when more complex linguistic structures are used (Cardillo et al., 2010). As a further feature of this stimulus set, syntactic complexity was comparable across sentences, with all sentences having the same phrasal structure (i.e., subject + verb + object). While most of the previous studies investigated concrete/abstract differences at the single words level, here we used sentences, thus contributing to the depiction of domain-specific meanings at the sentence level. The use of single words in the research on conceptual processing could have suffered from some confounding side-effects. It has been shown that processing a single verb requires not only to determine its meaning and its syntactic category, but also to establish what arguments it may or must take and what general types of meanings these arguments must have (Liversedge et al., 2003). For example, Ferretti and colleagues (2001) found that verbs immediately prime typical agents and patients, suggesting that readers immediately compute typical entities fitting thematic roles associated with verbs on the basis of their schematic knowledge representations. It has also been observed that many nouns, without an available context, contain elements of vagueness or indeterminacy of their meaning (e.g., ambiguous or polysemous nouns) (Cacciari, 2001). These observations suggest that single words, especially verbs (e.g., *to grasp*, *to kick*), if presented in isolation, could trigger different interpretations, ranging from a concrete one (e.g., *to grasp the pen*, *to kick the ball*) to an abstract one (e.g., *to grasp the concept*, up to the idiomatic expression like *to kick the bucket*), thus potentially yielding to an inconsistent classification of experimental stimuli. Providing verbs and nouns within a sentence structure, we linguistically contextualized the meanings thus avoiding also this potential drawback.

With our cross-study correlations and bootstrapping simulations, comparing the word-level noun ratings of Della Rosa et al. (2010) and the corresponding sentence-level ratings collected in our study for sentences containing the same nouns, we nevertheless controlled that, from a psycholinguistic point of view, the data we have collected at the sentence level do not provide a biased picture with respect to available data at the word level.

In particular, we considered three categories of concrete, action-related meanings, namely mouth-, hand-, and leg-related sentences, and three categories of abstract meanings, namely emotion-, mathematics-, and mental state-related sentences. By this, we aimed to validate by means of psycholinguistic rating methods, a set of semantic domains – particularly the abstract Ms, Em, Ma semantic categories – for which some evidence on their category status was already available in the extant literature. This is obviously not meant to exclude that a number of other relevant categories may be identified in either the concrete and abstract domains, such as, just to mention one, the category of "social concepts" (Cappa, 2008).

At a broad level, our results consistently reflected the classical dichotomy between concrete and abstract meanings: action-related sentences resulted as more concrete, easier to think a context for, and more familiar than Ms, Em, and Ma sentences. This is in agreement with the vast literature on concrete and abstract single words (Paivio et al., 1968; Schwanenflugel et al., 1988; Barca et al., 2002; Cacciari, 2001) , but, importantly, it extends the validity of these findings from single word to sentence processing.

At a finer level, in rating study 1 we showed that abstract sentences were clustered into three groups, demonstrating that different types of abstract-related meanings were identified by language users, even if they were not asked to explicitly distinguish between different categories. Alternatively, the results of rating study 1 may be interpreted as an evidence of sentence clustering based not solely on semantic relatedness, but possibly also on the association strength between lexical items. However, we believe that this does not jeopardize an interpretation of our findings in terms of semantic relatedness, given that associative and semantic relations seem to be intrinsically intertwined. The distinction between association based on lexical co-occurrence and semantic relatedness has been questioned in a number of research studies (McRae & Jones, 2013; McNamara, 2005). Indeed, it seems empirically difficult to consider the net effect of one type of relation after excluding the other one: for instance, McNamara (2005) directly challenged anyone to find two highly associated words that are not semantically related in some plausible way. The observation that associatively related words are almost unavoidably semantically related has been empirically corroborated by Brainerd et al. (2008), showing a correlation between a number of semantic variables and word association strength. It has been shown that lexical co-occurrence is correlated with associative strength (Spence & Owens, 1990) and lexical co-occurrence has been proposed as a less costly and more reliable source of association norms (Church & Hanks, 1990). The dividing line between associative and semantic relatedness is then completely blurred in models of semantic representations based on word co-occurrence over text corpora, such as Latent Semantic Analysis (Landauer & Dumais, 1997) and Hyperspace Analogue to Language (Lund & Burgess, 1996), in which semantic spaces are derived from co-occurrence statistics. In this sense, the association strength between lexical items of sentences belonging to the same semantic category (e.g., *anger* and *happiness* in Em sentences) may be higher than for lexical items of different semantic categories (e.g., *procedure* in Ms sentences, and *sum* in Ma sentences), as lexical co-occurrence is intrinsically related to meaning aspects.

Moreover, the correspondence analysis of rating study 1 then revealed that the dichotomy between abstract and action-related meanings was not sufficient to account for the total data variability. The category-specific correlation patterns provided further indication for differences between the six semantic categories. We also complemented this evidence with data of body-part ratings for both action- and abstract-related sentences. Exploiting the classic method of identifying a category of entities by means of the combination of different traits, we provide a tentative synthetic table summarizing the main results of the present study (Table 6). Based on this table, we suggest the possibility of describing a particular pattern of characteristics for each category of sentence, which will be the main focus of the remaining part of our discussion.

4.1 Action-related sentences

With respect to action-related meanings, we found a specific involvement of the mouth, the hands or the legs in the actions referred to, respectively, by mouth-, hand-, and leg-related sentences. Indeed, the distinctiveness of these action-related sentences has been observed in previous behavioral (Buccino et al., 2005), and neuroimaging studies (Hauk et al., 2004; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006), and it is in general agreement with embodied cognition accounts (Barsalou, 1999; 2008; Kiefer & Barsalou, 2012; Chatterjee, 2010; Meteyard et al., 2012) highlighting the relevance of specific motor information for the semantic representation of action-related sentences. Here we completed the characterization of action-related sentences by ratings on concreteness, context availability and familiarity. In particular, we showed that mouth-related sentences were similar as far as familiarity is concerned, but were otherwise considered as being less concrete than hand- and leg-related sentences and less easily connected to a specific context than hand-related sentences, while still receiving higher concreteness and context availability scores than abstract-related meanings. Sentences with the lowest concreteness median scores (< 6) were: *She mimes a face*; *She twists her lips*; *She tastes the wine*; *She savors the food*; *She relishes the champagne*.

The two sentences *She mimes a face* and *She twists her lips* can be considered as referring to non-verbal oro-facial communicative actions (verbal communicative actions were intentionally excluded from the present stimulus set), and thus considered of a more symbolic (i.e., “abstract”) kind than the remainder group of mouth-related sentences, in which an oro-facial motor involvement was generally coupled to a physical object to be ingested (e.g., *She bites the sandwich*; *She crunches the fruit*; *She swallows the pill*). In turn, the three sentences, *She tastes the wine*, *She savors the food*, and *She relishes the champagne*, albeit also referring to ingestive actions, were arguably associated with a somewhat peculiar function of "pleasure", rather than strictly of "nourishment". This more hedonistic function may be associated to increased sensory rather than solely motor attributes, thus maybe explaining the relatively lower concreteness scores. These data may suggest that the function of an action might be a component of its conceptual-semantic representation. Indeed functional knowledge is considered part of the information constituting the representation of object concepts, including knowledge about objects' function and more abstract propositional properties (Gernsbacher, 1984). Neuropsychological and neuroimaging studies provided data showing how object concepts are represented in the brain as distributed networks including areas preferentially involved in the processing of sensory or functional knowledge (Canessa et al., 2008; Rueschemeyer et al., 2010). The hypothesis might be tested and further extended to the other domains of action-related meanings in future research, by operationalizing the type of information available in processing action concepts. In any case, differences on concreteness and context availability between mouth- vs. hand- and foot-related sentences reveal that, even within the well-defined domain of concrete, action-related meanings, subtle differences between different categories can be identified that might be more deeply investigated in future studies.

4.2 Mathematics-related sentences

Mathematics-related sentences were judged as significantly engaging the hands more than the mouth and the legs. From a linguistic perspective, it is worth noting that there exist some Amazonian languages (such as Mundurukú) that lack words for numbers beyond 5 and use a broad variety of expressions such as "more than one hand", "two hands", "some toes", "all the fingers of the hands" for referring to quantities greater than 5 (Pica et al., 2004). Several lines of evidence indeed posit in favor of a possible relationship between finger counting and number processing, with number considered as a special kind of abstract concept (Ranzini et al., 2011). Finger counting is a basic numerical learning strategy that develops spontaneously in infancy (Butterworth, 1999), supporting and preceding the acquisition of more advanced mathematical achievements (Bryant, 1988). Recent findings suggest that even in adults, finger counting patterns modulate arithmetic performance (Klein et al., 2011). An increase in amplitude of motor-evoked potentials was found for the right hand muscles of subjects performing a visual parity judgment task on Arabic numerals (Sato et al., 2007), and on numbers and letters (Andres et al., 2008). Recently, in a functional magnetic resonance imaging experiment, a signal increase was observed in the hemisphere contralateral to the hand used for counting when low numerosity numbers were presented, despite the absence of overt hand movement (Tschentscher et al., 2012). Our results extend such evidence in showing that hand-related semantic features can be identified at the semantic level in mathematics-related sentences. These results can be interpreted in the light of embodiment accounts, with the hand-related motor information as one of the possible modalities relevant for mathematics-related meaning.

Moreover, mathematics-related sentences appeared to be more concrete and more easily associated to a specific context than emotion- and mental state-related meanings, but lower in concreteness and context availability than action-related meanings. Interestingly, Dehaene and colleagues (1999) proposed that internal representations of language-specific number words have a special role in mathematical thought: the use of number words (e.g., ‘ninety-eight’) is connected to the appreciation that each such number word names a distinct quantity (98-ness). Complementing the more basic biological capacities of individuating small quantities (such as, ‘1-ness’, ‘2-ness’, ‘3-ness’ and ‘more-than-that-ness’) and approximating magnitudes (for example, discriminating arrays of 8 dots from arrays of 16, but not more closely matched arrays) with the ability to use number words, humans can benefit of a simple and flexible method to think about an unlimited set of exact quantities. Speakers of Amazonian languages which do not have words for representing exact quantities rely on analogue magnitude estimation for estimating large quantities (Gordon, 2004). This may also occur in numerical-savvy English speakers when they are prevented from using linguistic resources by means of verbal interference tasks (Frank et al., 2008; 2012). Although we didn’t use number words, but sentences describing mathematical operations, we might interpret the degree of concreteness and context availability as reflecting the fact that processing mathematics-related meanings may lead to the construction of quantities, which can easily be associated to contextualized concrete entities.

In sum, a strict classification of mathematics-related concepts as either concrete or abstract doesn’t seem to be appropriate. In this sense, mathematics-related concepts may constitute a case study of hybrid embodiment across the abstract and concrete domains, with a grounding in both abstract, reasoning mental processes and concrete, sensory-motor finger representations.

4.3 Mental state- and emotion-related sentences

Even if emotion and mental-state meanings resulted similar with respect to concreteness, context availability and familiarity, they exhibited dissimilarities in the involvement of body parts, with emotion sentences more associated with mouth, hand and leg movements than mental-state and mathematics-related sentences. Recently, by means of event-related functional magnetic resonance imaging it has been shown that, in addition to a range of brain regions previously found to be active in emotion word processing, sensorimotor areas were also activated during the silent reading of abstract emotion words (Moseley et al., 2012). Specifically, signal increase was observed in the same areas entailed during the processing of face- and arm-related words, possibly suggesting that emotion words are associated to the involvement of specific districts of the body that are pivotal for displaying typical behaviors related to emotion. Importantly the emotional stimuli used in the experiment were words whose semantic meaning was either related to concrete or sensorimotor emotional actions (e.g., *frown*, *gnash*, *retch*) or not (e.g., *ail*, *rile*, *gloat*). Results were obtained for emotional words of both types, and further confirmed when only emotion stimuli not related to sensorimotor features were considered. By employing abstract emotion-related sentences (e.g., *She reveals the embarrassment*; *She mocks the disappointment*; *She experiences the excitement*) our results provide further evidence of an involvement of body-part representations (not limited to the mouth and the hands, but also including the legs) related to the semantics of emotion-related linguistic utterances.

It's worth noting that emotions and actions are supposed to be inter-related at anatomical and functional levels as follows (LeDoux, 1996): i) the projections from the amygdala, which mediates emotional responses, to the brain stem may have influences on the generation of relatively simple, stereotypical motor responses and facial expressions; ii) the projections from the amygdala to the prefrontal cortex and the cingulate cortex may have influences on working memory and executive functions, which are crucial to higher-level planning and control of voluntary movements; iii) the emotional responses involve the autonomic and endocrine systems and provoke changes in the bodily states that may have some effects on action execution and control. It seems likely that emotion-related linguistic utterances evoke action-related features. According to embodied theories, emotion perception is linked to action simulation, since covert emotional states are often associated with overt motor behavior. Thus, observers can simulate and understand the observable emotional state of others by embodying their observable motor behavior (Bastiaansen et al., 2009; Gallese & Sinigaglia, 2011). In this view, emotion perception and action simulation are closely bounded together. Another line of research has suggested that emotional processing can trigger the motor system to prepare a motor act (Lang, 1993; Frijda, 2009; Tettamanti et al., 2012). Defensive and approaching movements are triggered by unpleasant and pleasant cues, respectively (Chen & Bargh, 1999; Rotteveel et al., 2004). Accordingly, we may speculate that high rating scores for the involvement of the legs in emotion related sentences may be due to defensive movement preparations elicited by emotion-related sentences. Still another possibility, however, is that motor components are tied to emotion-related linguistic utterances due to arousing semantic content, rather than as intrinsic embodied features.

In turn, mental-state meanings were specifically associated only to mouth movements. The mental-state related sentences that obtained the highest scores on the mouth scale (≥ 5) were: She memorizes the procedure; *She determines the fate*; *She discerns the opinion*; *She influences the choice*; *She pretends an interest*; *She assesses the views*. Within an embodied cognition framework, it is plausible that the meaning of these sentences integrates motor information about typical oro-facial activities that might be performed during a cognitive process, such as subvocal repetition during memorization processes or talking in order to take position or express personal opinions or views.

Although emotion and mental-state sentences seem to involve motor representations, they received very low concreteness and context availability scores. Abstract concepts are relational structures resulting from the integration of many different concepts in a situated conceptualization. For example, the concept of *to convince* integrates an agent, other people, an idea, communicative acts, possible changes in belief, talking with another, etc. (Wilson-Mendenhall et al., 2011). The low context availability of emotion- and mental state-related sentences might reflect the difficulty in retrieving all such elements for the representation of the entire situated conceptualization. The body-part involvement can be considered as one of the dimensions of a relational structure that can dynamically become more or less relevant depending on the context.

5. Conclusions

Altogether, the present study provided a fine-grained characterization of abstract meanings at the psycholinguistic level. We discussed the characterization of abstract-related categories especially in the light of recent proposals in the embodied cognition literature, suggesting that other theoretical accounts do not seem to explain within-domain meaning differences. These results are consistent with previous studies showing the distinctiveness of emotion-related concepts in terms of rating measures and neural underpinnings, and add important clues toward the possibility of identifying mathematics-related sentences as characterized by specific features within a hybrid abstract-concrete domain. Further research is necessary in order to investigate other important features related to abstract meanings. For example, in line with the traditional approach used by Russell (1980) concerning emotion, investigating valence and arousal of linguistic utterances may reveal that these dimensions could differently mark emotion-related meanings.

In conclusion, these data inform future studies aimed at investigating the nature of different categories of concepts, indicating, for example, that also in the representation of abstract meanings sensory-motor maps may be significantly involved. Specifically, the ratings collected allow for a quantification of different profile of characteristics for action and abstract concepts, thus enabling the parametric manipulation of these characteristics in future research.

7. Tables

Table 1.
Descriptive statistics of linguistic and auditory characteristics for (Ms) mental state-,(Em) emotion-, (Ma) mathematics-, (Mo) mouth-, (Ha) hand-, and (Le) leg-related sentences.

	No. of words	**No. of syllables**	**No. of letters**	**Frequency verb**	**Frequency noun**	**Frequency verb+noun**	**Intensity (dB)**	**Pitch (Hz)**	**Duration (sec)**
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Ms	4.00	8.14	19.66	45.96	50.56	96.52	70.09	232.85	1.47
	(.00)	(.88)	(1.86)	(59.82)	(47.61)	(67.87)	(0.02)	(4.26)	(0.11)
Em	4.00	7.63	19.03	61.43	30.01	91.44	70.07	230.41	1.44
	(.00)	(1.11)	(2.67)	(82.57)	(46.34)	(88.69)	(0.04)	(6.03)	(0.17)
Ma	4.00	8.00	19.29	57.07	56.25	113.32	70.07	232.09	1.44
	(.00)	(.69)	(1.43)	(93.17)	(90.27)	(139.53)	(0.02)	(5.12)	(0.12)
Mo	4.00	7.37	18.91	12.05	32.15	44.20	70.08	230.59	1.39
	(.00)	(1.09)	(2.85)	(34.84)	(52.22)	(65.77)	(0.03)	(4.52)	(0.15)
Ha	4.00	7.49	18.43	46.70	26.50	73.20	70.07	231.75	1.38
	(.00)	(.82)	(2.23)	(131.54)	(38.35)	(136.25)	(0.02)	(4.40)	(0.13)
Le	4.00	7.43	18.71	66.37	27.29	93.66	70.07	230.81	1.40
	(.00)	(.88)	(2.35)	(93.20)	(44.78)	(104.40)	(0.03)	(4.03)	(0.12)

Table 2.
Descriptive statistics of association ratings for (Ms) mental state-, (Em) emotion-, (Ma) mathematics-related sentences.

	Mental-state association scale		**Emotion association scale**		**Mathematics association scale**	
	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)
Ms (n=35)	5 (3-6)	4.31 (2.31)	2 (1-3)	2.38 (1.77)	1 (1-4)	2.4 (1.97)
Em (n=35)	2 (1-3)	2.34 (1.73)	5 (3-5)	4.75 (1.98)	1 (1-1)	1.46 (1.19)
Ma (n=35)	1 (1-4)	2.41 (1.82)	1 (1-1)	1.53 (1.27)	6 (4-7)	5.23 (1.86)

Table 3.
Descriptive statistics of body-part ratings for (Mo) mouth-, (Ha) hand-, (Le) leg-, (Ms) mental state-, (Em) emotion-, (Ma) mathematics-related sentences.

	Mouth scale		**Hand scale**		**Leg scale**	
	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)
Mo (n= 35)	7 (7-7)	6.81 (0.57)	2 (1-4)	2.88 (1.87)	1 (1-1)	1.13 (0.49)
Ha (n= 35)	1 (1-2)	1.49 (1.04)	7 (7-7)	6.61 (0.87)	1 (1-2)	1.50 (1.13)
Le (n= 35)	1 (1-1)	1.41 (0.98)	3 (1-4)	2.93 (1.92)	7 (7-7)	6.59 (1.05)
Ms (n=35)	2 (1-5)	3.15 (2.23)	1 (1-3)	2.21 (1.74)	1 (1-1)	1.30 (0.93)
Em (n=35)	4 (1-6)	3.84 (2.34)	2 (1-5)	2.93 (2.17)	1 (1-3)	2.06 (1.74)
Ma (n=35)	1 (1-3)	2.24 (1.70)	3 (1-5)	3.06 (1.90)	1 (1-1)	1.11 (0.59)

Table 4.
Descriptive statistics of CNC (concreteness), CA (context availability), and FAM (familiarity) ratings for (Ms) mental state-, (Em) emotion-, (Ma) mathematics-, (Ha) hand-, (Le) leg-, and (Mo) mouth-related sentences.

	CNC		**CA**		**FAM**	
	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)	Mdn (IQR)	Mean (SD)
Ms (n=35)	2 (1-3)	2.41 (1.52)	2 (1-4)	2.9 (1.9)	4.5 (2-6)	4.19 (1.97)
Em (n=35)	2 (1-3)	2.27 (1.43)	2 (1-4)	2.75 (1.85)	4 (2-6)	4.19 (2.09)
Ma (n=35)	4 (2-5)	3.61 (1.72)	4 (2-6)	3.86 (2.06)	4 (2-6)	3.89 (2.06)
Mo (n=35)	7 (6-7)	6.24 (1.19)	5 (3-6.75)	4.68 (1.97)	5 (4-7)	4.98 (1.89)
Ha (n=35)	7 (7-7)	6.64 (0.8)	6 (4-7)	5.25 (1.79)	6 (4-7)	5.23 (1.97)
Le (n=35)	7 (6-7)	6.52 (1)	5 (3-7)	4.93 (1.90)	5.5 (4-7)	5.04 (1.98)

Table 5.
Correlations between (CNC) concreteness and Mouth, Hand, and Leg ratings calculated for (Ms) mental state-, (Em) emotion-, (Ma) mathematics-related sentences (Spearman's rank-order coefficients (r_s) on median value; * $p < 0.05$; ** $p < 0.01$).

		Mouth scale	Hand scale	Leg scale
Ms	**CNC**	0.385*	0.272	0.211
Em	**CNC**	0.426*	0.453**	0.463**
Ma	**CNC**	-0.038	0.308	-

Table 6.
Synthetic summary of the main results of the present study.

	Mouth scale	Hand scale	Foot scale	CNC	CA	FAM
Ms	+	-	-	- -	- -	-
Em	+	+	+	- -	- -	-
Ma	-	+	-	+/-	+/-	-
Mo	++	-	-	+	+	+
Ha	-	++	-	++	++	+
Le	-	-	++	++	++	+

Study B

Fine-grained semantic categorization across the abstract and concrete domains: an fMRI research[2]

Abstract

In the past decade, neuroscientific research based on the framework of embodied cognition shed light on a fine-grained categorization of concrete conceptual knowledge emphasizing the role of sensory and motor information in the representation and processing of different concrete-related meanings. Only recently has abstract conceptual knowledge been addressed from an embodied perspective. The aim of the present work was to complement the available neuroimaging evidence suggesting modality-specific neural representations for different types of concrete concepts by distinguishing semantic categories with a high degree of specificity both within the abstract and the concrete domain. We report evidence from a series of functional Magnetic Resonance Imaging (fMRI) experiments, in which healthy participants listened to sentences belonging to different abstract-related categories (mental-state-, emotion- and mathematics-related meanings) and concrete categories (mouth-, hand-, and leg-related meanings). In order to take into account the multidimensional nature of the data space, various techniques of fMRI data analysis were applied, including standard univariate analysis, multivariate pattern analysis, and parametric analysis. Evidence of distributed networks of brain regions specific for the processing of abstract versus concrete meanings confirmed that the abstract/concrete distinction is, at a broad level, a basic semantic organizational principle of conceptual knowledge. At a finer level, univariate statistical analysis mainly failed to reveal specific brain activations. In turn, the application of multivariate techniques revealed the existence of semantically organized structure in the pattern of fMRI-measured brain activation during the semantic processing of different categories both within the abstract and the concrete domain. These results extend the discussion on the neural representation of conceptual knowledge by suggesting that fine-grained categories might be represented by specific distributed patterns of brain activity mainly reflected by small relative changes in activation across populations of voxels, rather than by global activation in specific brain areas.

---

[2] Ghio M., Vaghi, M. M. S., Tettamanti, M., Fine-grained semantic categorization across the abstract and concrete domains: an fMRI research, (*in preparation*).

1. Introduction: Grounding the abstract knowledge

Conceptual knowledge (i.e., all we know about the world) virtually intervenes in every cognitive activity, including perception and action, memory, thought, and language. The recognition and the use of different objects (e.g., a laptop vs. a cup) is an example of everyday cognitive activities that depend on conceptual knowledge. Language allows retrieving, manipulating and communicating knowledge even of an entity that is not, when talking about it, also the object of concomitant perception (e.g., concrete entities), or that is indeed not perceivable through senses (e.g., abstract entities). In language, auditory/written information is mapped into conceptual knowledge about the referent of an utterance. The mapping between linguistic units and the conceptual knowledge they signify is a constitutive part of the “semantics” of a language.

In cognitive neuroscience, two main classes of theories about conceptual/semantic knowledge have been proposed, both aiming at serving as integrated frameworks for both concrete and abstract meanings. On the one hand, amodal theories claim that conceptual representations are abstract, symbolic and represented in specific conceptual systems (Caramazza & Shelton, 1998; McClelland & Rogers, 2003; Caramazza & Mahon, 2003). Often discussed as examples of amodal accounts are the computational models, claiming that meanings result from word co-occurrence statistics (for recent reviews, see Bullinaria & Levy, 2007; McRae & Jones, 2013).

On the other hand, modal theories maintain that concepts are represented in multiple, distributed brain networks reflecting the type of experience that is characteristic for the concepts’ referents. This view has been formalized within the theoretical framework of grounded cognition. As suggested by Barsalou (2010), the label “grounded cognition” better captures the broad scope of grounded mechanisms, including not only the bodily states (embodied cognition), but also the physical environment, the social environment, affective and internal states (for a recent review, see Pezzulo et al., 2013). Also, some feature-based accounts assume that distributed semantic features building a concept are represented according to the modality by which we acquire and experience events and things (Vigliocco et al., 2004; Meteyard et al., 2007; Martin, 2007). It is worth noting that the notion of modality-specific representations does not rule out the possibility of higher-order integrative memory systems/hubs that contain systematically organized units for binding cross-modal feature correlations (Simmons & Barsalou, 2003; Patterson et al., 2007; Hoenig et al., 2008; Kemmerer et al., 2008; Kiefer & Pulvermueller, 2012).

Up to now, most of the empirical research on conceptual/semantic processing has focused on concrete-related knowledge. Recently, excellent reviews on neuropsychological, behavioral, and neuroimaging studies illustrated the role of sensory and motor information in the representation and processing of different types of concrete-related knowledge, including: (i) object-related knowledge, mainly organized by taxonomic categories (e.g., tools, animals, vegetables) or feature-based categories (e.g., color, sound, odour, functional features, etc.) (Martin, 2007; Cappa, 2008); (ii) action-related knowledge, organized by categories according to the body-parts involved in the action described by the linguistic item (e.g., mouth-related actions, hand-related actions, leg-related actions), or the type of action (e.g., grasping vs. reaching) (Pulvermueller et al., 2010; Cappa & Pulvermueller, 2012). In particular, there is a large amount of neuropsychological and neuroimaging research demonstrating that processing concrete concepts expressed by words or sentences activate distributed neural networks including sensory and motor brain regions, connected with linguistic areas in the perisylvian cortex, and supramodal temporal and/or parietal structures (Patterson et al., 2007; Ghio & Tettamanti, 2010; Binder & Desai, 2011; Kiefer & Pulvermueller, 2012; Meteyard et al., 2012).

Based on this growing body of research, the focus of the current debate has been shifted to some specific issues mainly concerning the grounded framework. Among others, the probably most crucial issue is the representation and the processing of abstract-related knowledge (Barsalou, 2010; Kiefer & Pulvermueller, 2012; Meteyard et al., 2012). Indeed, what is still controversial is whether the processing of abstract-related conceptual knowledge (e.g., thought, happiness, rumination) is supported by distributed modality-specific representations, by analogy to what has been demonstrated for the processing of concrete-related knowledge. The present research specifically addressed this question in a series of functional Magnetic Resonance Imaging (fMRI) studies using as stimuli different types of concrete/action-related and abstract-related sentences.

In the remainder, we briefly review current modal accounts of semantic representation that explicitly consider the question of abstract meanings (section 1.1). The core theoretical assumption shared by these accounts is that, like concrete meanings, abstract meanings are represented in modality-specific brain regions reflecting the type of experience that is characteristic for the concepts' referents. In section 1.2, we argue that most of the available neuroimaging evidence comes from studies investigating the abstract/concrete dichotomy, regardless of the characteristics of different types of abstract meanings. In this sense, most previous research does not allow discussing specific predictions of modal theories. In section 1.3, we present some methodological issues in investigating the abstract domain. First, a specific characterization of the types of abstract knowledge. As outlined above, organizing the concrete-related knowledge by categories has proved to be a useful means for testing specific modal hypotheses. In a previous study, we proposed a fine-grained categorization also with respect to the abstract-related knowledge (Ghio et al., 2013; Study A in the present dissertation). Other methodological issues concern: the experimental stimuli (words vs. sentences); the experimental task (e.g., passive listening vs. active response). Finally, in section 1.4 we describe the rationale of the fMRI experimental studies here reported.

1.1 Theoretical accounts

The seminal theory on abstract and concrete knowledge is the Dual Coding theory (Paivio, 1971). This account endorsed a form of modality-specific representation, but only for concrete meanings. Basically, processing concrete knowledge relies on both verbal and non-verbal, imagery-based representations, whereas processing abstract knowledge relies on verbal representations only. In this sense, the Dual Coding theory, often contrasted with the Context Availability theory (Schwanenflugel et al., 1988) – another quite influential theory – shares with the latter the assumption that abstract concepts are verbally represented in the left hemisphere.

In turn, accounts based on the grounded framework advocated modality-specific representations for both concrete and abstract meanings. Among these accounts, different nuances exist with respect to the type of experience/information considered characteristics for abstract concepts. According to a (strong) version of embodiment, abstract concepts are grounded in sensory and motor experience via conceptual metaphors (Boroditsky & Ramscar, 2002; Gallese & Lakoff, 2005; Gibbs, 2006; Glenberg et al., 2008). For example, the mental representation of an abstract concept such as “time” may be metaphorically built out on sensory and motor representations that result from the physical experience of “space” (Casasanto & Boroditsky, 2008). Criticisms have been raised about the foundational role of the metaphorical mechanisms and its generalizability to all abstract concepts (Barsalou, 1999; Vigliocco et al., 2009; Tettamanti & Moro, 2012), and the empirical evidence is controversial (Aziz-Zadeh et al., 2006; Rueschemeyer et al., 2007; Boulenger et al., 2009).

One other prominent version of embodiment proposed that, in addition to sensory and motor information, abstract meanings rely on emotional and introspective information about internal states (e.g., interoception, mentalizing, beliefs, affects, self-thoughts, intention recognition) (Barsalou, 2008; Ghio & Tettamanti, 2010; Kiefer & Pulvermueller, 2012; Wilson-Mendenhall et al., 2011; 2013). Abstract concepts may be experientially more complex than concrete concepts, being tied both to situations in which people experienced them and to internally generated cognitive and emotionally states (Barsalou & Wiemer-Hastings, 2005; McRae & Jones, 2013). For example, the representation for “convince” may include several sets, characters, actions of self and others, as well as intentions, beliefs, internal states, affects, etc. (Wilson-Mendenhall et al., 2013). Also linguistic information (in the form of word associations resulting from co-occurrence patterns and syntactic information) has been considered to be relevant for abstract meaning (Simmons et al., 2008; Vigliocco et al., 2009). In general, experiential information (motor, sensory, affective, and introspective) and linguistic information contribute to the representation of both concrete and abstract meanings, though in different proportion. Concrete meanings may more strongly depend on sensory-motor information, while abstract meanings on affective and linguistic information (Barsalou, 2008; Vigliocco et al., 2009). The relative contribution of different types of information may vary depending on the context and the task (Wilson-Mendenhall et al., 2011; Kiefer & Pulvermueller, 2012).

Grounded accounts seem to have a great potential also in explaining the existing variety of abstract meanings, encompassing various entities and processes such as social relationships or facts, events, and introspective states (Barsalou & Wiemer-Hastings, 2005). Depending on the particular concept, a particular profile of systems that process perception, action, language, emotion, and internal states may be more or less relevant. This mechanism has been clearly implicated to explain differences across concrete meanings. For example, visual information is thought to be more relevant for animals than for tools, whereas action and functional information is thought to be more relevant for tools than for animals, also depending on the context and on the task (Martin, 2007; Cappa, 2008; Hoenig et al., 2008). Similarly, some authors suggested, that within the abstract domain, introspective information about intentions and beliefs may be more relevant for social concepts (e.g., convince), while affective information may be more relevant for emotion concepts (e.g., fear) (Wilson-Mendenhall et al., 2011; 2013). At present, however, little research has been conducted to identify different categories of abstract entities with respect to their relevant properties, as we illustrate in the following section.

In sum, although grounded theories of abstract knowledge are still in their infancy, it is likely that the future development of the research on abstract meanings will be anchored on the following points: (i) Modality-specific representations support the processing of both concrete and abstract meanings, beyond the classic abstract/concrete dichotomy. This does not mean denying the existence of differences between abstract and concrete meanings. Such differences are clearly suggested by behavioral evidence of a concreteness effect – i.e. a typically superior performance for concrete meanings relative to abstract ones in behavioral tasks – and neuropsychological evidence of a double dissociation – i.e., patients showing a concreteness effect and patients showing an abstractness effect (for a review, see Macoir et al., 2009). (ii) Differences across and within different semantic domains bear on the relative contribution of sensory-motor, emotional, introspective, and linguistic representations. Depending on the particular experience-based information relevant for each concept, a multimodal representation becomes stored in distributed neural networks. Based on this assumption, (iii) specific hypotheses about the neural underpinnings of different types of conceptual representations can be formulated.

The last point marks an important difference between proposals in the grounded framework as compared to other accounts with respect to referring to clear neuroanatomical hypotheses. For example, an alternative account on semantic knowledge suggests a qualitative distinction between concrete and abstract meanings in terms of different principles of organization, i.e. a categorical organization for concrete meaning vs. an associative organization for the abstract ones (Crutch & Warrington, 2005). Although this proposal was based on neuropsychological observations, the neuroanatomical correlates of such organizational principles are difficult to hypothesize (see also Vigliocco et al., 2009). A promising alternative approach for the study of abstract meanings is represented by corpus-based distributional models, according to which the meaning of a word (either concrete or abstract) is a function of a pattern of values of different vector elements (HAL model, Lund & Burgess, 1996; Landauer & Dumais, 1997; for a review of recent accounts, see Bullinaria & Levy, 2007). But again, computational models *per se* do not allow formulating neuroanatomical predictions. Intriguingly, however, they may be combined with grounded approaches to form hybrid models (Vigliocco et al., 2009; Kriegeskorte, 2011; MacRae & Jones, 2013; Pezzulo et al., 2013).

1.2 Neuroimaging evidence

Until now, most of the neuroimaging research on abstract knowledge was restricted to testing the hypothesis that processing abstract meanings should be associated with left-hemispheric activations in linguistic brain regions only (Dual Coding theory), possibly to a lesser extent than concrete meanings (Context Availability theory). In partial agreement with this prediction, a coordinate-based meta-analysis combining data across 19 neuroimaging (fMRI and PET) studies revealed that abstract meanings elicited greater activity in the left inferior frontal gyrus and in the left middle temporal gyrus compared to concrete meanings (Wang et al., 2010; see also the meta-analysis of Binder et al., 2009).

However, for the “abstract > concrete” contrast reported in the meta-analysis it has been observed that: (i) the activation in the left inferior frontal gyrus was found only in 10 out of 19 studies; (ii) activations in extra-linguistic brain regions that were reported by several studies were not identified by the meta-analysis results, including right-hemispheric activations in the superior frontal gyrus and the precuneus (D’Esposito et al., 1997), in the parieto-occipital junction, anterior cingulate cortex, and in amygdala (Perani et al., 1999), occipital gyrus (Jessen et al., 2000). The processing of abstract concepts has also been associated with several foci of activation in a bilateral fronto-temporal-parietal network (Pexman et al., 2007), and in the retrosplenial cingulate cortex (Tettamanti et al., 2008). These findings remain difficult to explain based on the Dual Coding theory and the Context Availability theory. Nevertheless, interpreting previous results with respect to the modality-specific hypotheses formulated within the grounded framework seem to be difficult as well, primarily due to the type of stimuli employed in the studies (but see Pexman et al., 2007; Tettamanti et al., 2008; Ghio & Tettamanti, 2010). In particular, previous studies considered the abstract domain as a whole, regardless of the variety of abstract meanings (Wang et al., 2010; Binder et al., 2009). Consequently, whether the representational differences between abstract and concrete concepts were content specific could not be inferred. Moreover, psycholinguistic dimensions characterizing abstract linguistic stimuli with respect to the type of information they are related with (e.g., valence, arousal, action relatedness) were usually not controlled.

Only recently have modal accounts been tested in neuroimaging studies, by using experimental paradigms suitable to address specific hypotheses about abstract knowledge. Two different, though complementary, experimental approaches have been applied. A first feature-based approach consists in testing the hypothesis that abstract meanings depend on experiential information, such as affective or motor information, by controlling or manipulating these dimensions. For example, in a recent fMRI study (Vigliocco et al., *in press*), participants were asked to perform a visual lexical decision task on abstract and concrete words highly matched on a wide range of sub-lexical and lexical variables. The results revealed that processing abstract words selectively activated a rostral portion of the anterior cingulate cortex, a brain region plausibly involved in emotion processing. In order to verify the general role of affective information in word processing, the authors performed a series of regression analyses using valence and arousal ratings. Results showed that more highly valenced words (both concrete and abstract) engaged the visual processing system extending over occipital, temporal, and subcortical regions, but that the activation of this system extended to the rostral anterior cingulate cortex only in the case of more highly valenced abstract words. These findings highlighted the relevance of affective information relying on an emotion-specific brain area for processing abstract words. Similarly, Moseley et al. (2012) conducted an fMRI study in order to test the specific role of sensory-motor information in processing abstract affective-emotional meanings. It has been shown that sensory-motor and affective information are often associated, as demonstrated by the fact that high-arousal words (e.g., 'explosion') partially activate the same areas involved in sensory-motor processing (Kiefer et al., 2008). In order to dissociate sensory-motor and affective information in the representation of abstract words, Moseley et al. (2012) used tightly controlled stimuli: (i) only emotion words with low emotional ratings (i.e., low arousal, low valence) were employed; (ii) based on ratings on imageability, concreteness, and action-relatedness, the category of emotion words was split into two subgroups: emotion words related to emotional actions (e.g., *frown*, *gnash, retch*) vs. emotion words unrelated to emotional actions (e.g., *ail, rile, gloat*). The results revealed that silent-reading of emotion words activated not only a range of brain regions previously found to be active in emotion word processing, but also the same sensorimotor areas activated by arm and face-related verbs. Importantly, such result has been obtained both for the emotional action-related and emotional action-unrelated words.

An alternative categorical approach consisted in organizing the abstract domain by categories, and testing specific hypotheses about the involvement of modality-specific brain networks reflecting the more relevant type of experience related to the concept's referents (Cappa, 2008; Wilson-Mendenhall et al., 2011; 2013). Consider, for example, the category of abstract social concepts, defined as words referring to social behavior or properties of human beings (e.g., psychological characteristics such as *loyal, ambitious, assertive*). For this category, activations in brain systems relevant to social perception/interaction can be predicted. This abstract category has been investigated in several fMRI studies using either single words (Mitchell et al., 2002; Zahn et al., 2007) or sentences (Simmons et al., 2009; Contreras et al., 2011), and applying different experimental paradigms (meaning-relatedness judgments: Zahn et al., 2007; semantic judgments: Mitchell et al., 2002; Contreras et al., 2010; learning tasks: Simmons et al., 2009). Across these studies, comparable activations were found in brain regions typically implicated in social cognition, such as the temporal poles, the medial prefrontal cortex, the posterior superior temporal sulcus, the precuneus/posterior cingulate bilaterally, and the fusiform gyrus. In particular, the superior anterior temporal lobe plays a key role in representing abstract conceptual knowledge of social behaviors (Zahn et al., 2007; Simmons et al., 2009). Interestingly, Zahn et al. (2007) observed that activations in the same portion of the temporal lobe were reported in previous studies on abstract concepts (Noppeney & Price, 2004; Sabsevitz et al., 2005). The authors suggested a "category-specific re-interpretation" of the results reported by previous studies as due to the social relevance of used words, including indeed several social concepts such as *courage*, *glory, esteem* (Zahn et al., 2007) .

Altogether, these findings suggested that distinguishing between abstract and concrete concepts, broadly defined, may not be sufficient to understand how abstract-related knowledge is represented and processed. The variety of abstract meanings should be examined in the same way as the variety of concrete concepts has been examined, i.e. by taking into account the kind of information that may be more relevant for different categories of meanings. This can be achieved by controlling/manipulating the type of information, by organizing the abstract meanings by categories, or both.

1.3 Some methodological issues

At present, only few abstract-related categories have been selectively investigated, as illustrated above. Although the classification of abstract meanings is less straightforward than the one of concrete meanings (Hampton, 1981; Wiemer-Hastings & Xu, 2005), it seems to be both linguistically (e.g., WordNet, Miller et al., 1990) and psychologically plausible (Setti & Caramelli, 2005; Altarriba, et al., 1999; 2004; Kousta et al. 2011). Based on the previous literature we recently carried out a rating study with the purpose of providing psycholinguistic evidence of fine-grained distinctions within both the abstract and the concrete semantic domains with respect to relevant psycholinguistic dimensions. As for the abstract domain, we considered the following categories: mental state-related (e.g., *She contemplates the alternative*), emotion-related (e.g., *She feels disgust*), and mathematics-related (e.g., *She counts the sets*) meanings (Ghio et al., 2013). Such categories were selected so that clear hypotheses about their neural representation and processing could be formulated. The review of the specialistic literature revealed that some evidence on their category status was indeed already available, although neuroimaging evidence on their processing when accessed through words/sentences is still scarce and controversial.

Indeed, mental state-related meanings have often been used as abstract stimuli in previous fMRI studies, though mixed with other types of abstract meanings. For example, in a PET study, Perani et al. (1999) compared concrete nouns (e.g., *hammer*, *scissors*) and verbs (e.g., *to cut*, *to comb*) to abstract nouns (e.g., *justice, hope*) and abstract verbs related to psychological states. The latter class of experimental stimuli included verbs such as *to think*, *to believe* – referring to mental states – and verb such as *to hope, to desire* – referring to emotional states. Hence, the results cannot be interpreted with respect to the specific abstract-related category of the mental states. Intriguingly, a growing number of studies suggested that mental/internal states (such as thoughts, beliefs, self-reflection, a.k.a. mentalizing) are primarily associated with the so-called default mode network (Spreng et al., 2009; Lombardo et al., 2010).

As for emotion-related meanings, most fMRI studies focused on *emotional* words/sentences, i.e. words characterized by positive/negative valence and/or high/low arousal (e.g., *baby*, *killer, earthquake, accident*). Processing linguistic items with high arousal and extreme valence activated brain regions known to be involved in emotion processing, such as the orbitofrontal cortex, the insula, the anterior and posterior cingulate cortex (for a review, see Citron, 2012). Although these studies provided important evidence about the representation of emotion knowledge in general (for a review, see Niendenthal, 2007), they are not decisive for the case of abstract emotion meanings since concreteness has generally not been controlled/manipulated. Differently from most of previous research, in the present study we considered only utterances referring to emotions and feelings *per se* (e.g., *She feels disgust*). In turn, highly arousing utterances referring to actions or entities with an emotional connotation (e.g., *She stabs her husband*) were excluded. A similar approach has been proposed by Moseley et al. (2012), providing a first evidence of modality-specific neural network associated with the processing of abstract emotion verbs.

As far as mathematical meanings are concerned, from a behavioral point of view some evidence in favor of the grounding of numerical cognition has been provided (Pezzulo et al., 2013): (i) the universal association of smaller numbers with lower space and of larger numbers with upper space (Ito & Hatta, 2004; Schwarz &

Keus, 2004); (ii) the SNARC effect (spatial-numerical association of response codes), i.e., the fact that in behavioral tasks small and large numbers are responded to faster with the left and the right hand, respectively. This effect is weaker in people who start counting on the fingers of their right hand (Fischer, 2008; Lindemann et al., 2011), presumably because right-starters have initially learned to associate small numbers with their right side. Recently, Wilson-Mendenhall et al. (2013) carried out an fMRI study in which participants were repeatedly presented with the word 'Arithmetic' in a concept-scene match task. The results revealed that brain regions underlying numerical cognition (e.g., bilateral intraparietal sulcus) were also active during the presentation of the word 'Arithmetic'. Accordingly, it should be plausible to investigate the mathematics-related knowledge also when it is accessed through sentences.

Importantly, in our rating study we also considered three action-related meaning categories, including mouth-related (e.g., *She inflates the balloon*), hand-related (e.g., *She plucks the strings*), and leg-related meanings (e.g., *She bends the knee*). Consistent evidence suggests that processing action-related words activated a neural network including not only language areas but also motor areas in a somatotopic fashion (for a review, see Pulvermueller & Fadiga, 2010). Moreover, it has been shown that, when the action described by the linguistic items is directed toward an object, the action-related brain network extended over the posterior parietal lobe (Tettamanti et al., 2005; Tettamanti et al., 2008).

For each concrete- and abstract-related category, we created a set of sentences, and characterized them with respect to several psycholinguistics dimensions, including concreteness, context availability, familiarity, action relatedness, and abstract-meaning relatedness. By means of these measures, we highlighted not only psycholinguistic, but also semantic traits characteristics for each category. For example, we provided rating data about the motor information associated to each action-related and abstract-related category. These highly controlled sentences were meant to overcome the possible drawbacks of previous researches and were used as stimuli in the present neuroimaging research, as illustrated in the next paragraph.

Before concluding, we briefly discuss two other methodological issues, concerning the use of sentences vs. words and the selection of the appropriate task for investigating semantic processing.

Experimental stimuli. Most previous fMRI studies on processing abstract meanings used single words (both noun and verbs), underestimating the lexical-semantic ambiguities that presenting words without a linguistic context may introduce. For example, the same word (e.g., *to grasp*) can be interpreted as more concrete in some contexts (e.g., *I grasp the pen*), or as more abstract in other contexts (e.g., *I grasp the idea*). In this sense, specifying the linguistic context by using sentences instead of single words can be considered as a means for better controlling the concreteness/abstractness of the linguistic items. Moreover, by employing sentences with the same syntactic form, it is also possible to control, at least in part, the information conveyed by the syntactic structure of single words. Some studies demonstrated, for example, that readers immediately compute typical entities fitting thematic roles associated with verbs on the basis of their schematic knowledge representations (Ferretti et al., 2001; McRae & Jones, 2013). Sentences have successfully been used in previous experiments mainly focusing on concrete, action-related meanings (Tettamanti et al., 2005; 2008; Buccino et al., 2005; Sato et al., 2008; Beilock et al., 2008).

Experimental task. There is growing evidence that the type of experimental task performed by subjects may have an impact on semantic processing. Several points have been discussed in the literature. First, the use of a passive task (in which subjects are given no tasks) or a minimally demanding task (e.g., fixating a point in the visual field) has been considered quite problematic since, during such states, people experience "task-unrelated" thoughts, such as vivid thoughts and mental images (Binder et al., 2009; McKiernan et al., 2006). Binder et al. (2009) argued that such experiences are essentially semantic, primarily involving the manipulation of the encyclopedic/world knowledge. Second, among active-linguistic tasks it is possible distinguishing between: (i) tasks that elicit a shallow representation of the meaning since they can be performed by analyzing the form (orthographical or phonological) of words, for example, lexical decision task in the context of non-words that violate rules of phonology; (ii) tasks that elicit a deep processing of the meaning (e.g., semantic similarity judgments). It has been observed that the first type of task may involve only linguistic representation, and not modality-specific representations (Simmons et al., 2008; Binder & Desai, 2011). Third, Hoenig et al. (2008) demonstrated that the activation of the distributed semantic representations varies in a highly flexible manner depending on the type of concept retrieval (e.g., visual properties vs. action properties) that is required by the given task (visual task vs. action task).

1.4 The present research

The main aim of the present research was to test, in a series of fMRI studies (Study B1 and Study B2), the hypothesis of modal theories that semantic processing of both concrete and abstract meanings is mediated by a cross-talk between language parsing networks and modality-specific brain regions . To this purpose, we set out to further specify this hypothesis with respect to a fine-grained categorization within both the abstract and the concrete domains in order to investigate category-specific modal hypotheses. Based on our previous study (Ghio et al., 2013), the following categories were considered for the abstract domain: mental state-related, emotion-related, and mathematics-related categories. Given their relevance for evidence-based sensory-motor embodiment, we also investigated three action-related categories: mouth-related, hand-related, and leg-related action categories. By including these action-related categories we were also able to examine the abstract (mental state-related + emotion-related + mathematics-related meanings) versus concrete (mouth-related + hand-related + leg-related meanings) contrast, similar to previous research studies. We used as stimuli the set of sentences carefully characterized for several psycholinguistic variables in our previous rating study (Ghio et al., 2013).

In order to avoid the shortcomings associated with the use of passive tasks, an active task was employed. Participants were asked to carefully listen to sentences. Each sentence was followed by either a fixation cross or a question mark. When the fixation cross appeared on the screen, subjects were not asked to perform any additional task. In turn, when the question mark appeared on the screen, it was followed by a written sentence, and participants were asked to perform a 1-back cross-modal task, i.e. to judge whether the written sentence matched the auditorily sentence presented immediately before. In order to elicit a deep semantic processing, mismatches between auditory and written sentence were possibly due to (i) a mismatch in the verb (e.g., auditory sentence: *She learns the doctrine*; written sentence: *She explains the doctrine*), (ii) a mismatch in the object complement (e.g., auditory sentence: *She instils the joy*; written sentence: *She instils the warmth*), or (iii) a mismatch in both the verb and the object complement (e.g., auditory sentence: *She resets the calculator*; written sentence: *She reads the book*). Mismatches did not include any semantic or syntactic violations. In order not to interfere with the semantic processing of mouth-, hand-, and leg-related sentences, participants were asked to respond yes/no by blinking their eyelids once or twice respectively, thus avoiding the involvement of the mouth, the hand or the foot in the response.

As for abstract-related categories the following specific hypotheses were tested: (i) mental state-related meanings would activate the default mode network, a set of brain regions usually involved in self-reflection, mind wandering, and introspective states; (ii) emotion-related meanings would activate a network of areas including regions involved in processing emotions; (ii) mathematics-related meanings would activate a network of areas including regions involved in mathematical calculation and in the representation of numbers and quantities. As for action-related categories, consistent with previous results (for a review, Fadiga & Pulvermueller, 2010), we tested the involvement of an action-specific fronto-parietal network in a somatotopic fashion.

In order to provide what is considered a more rigorous test of these predictions, during the final scans of the first experiment (Study B1) participants also performed six functional localizer tasks. Each localizer task was meant to be specifically related to one semantic category. For the mental state category, the rationale was to pinpoint a neural network involved in self-reflection, mind wandering, and introspective states. Recent evidence suggested that such internal states are primarily associated with the involvement of the regions forming the default mode network (Spreng et al., 2009; Lombardo et al., 2010). It has been observed that such brain regions are more active at rest than during task performance (for a review, see Fox & Raichle, 2007). Hence, a resting state localizer was performed in which subject laid still in the scanner, without performing any task. To localize brain areas involved in emotion processing, subjects were visually presented with pictures of human faces with emotional facial expressions of joy, anger, and neutral expressions (for a review, see Fusar-Poli et al., 2009). To localize regions involved in number and mathematical processing, participants covertly performed mathematical calculations (for reviews and meta-analyses, see Dehaene et al., 2003; Pinel et al., 2007; Nieder & Dehaene, 2009). Localizers for actions performed with the mouth, the hand, and the foot were analogously used to establish brain areas predicted to be relevant for the action-related concepts (for a review, see Pulvermueller & Fadiga, 2010). To this aim participants were asked to move their hand, leg, or tongue. We used these independent localizer tasks to identify unique Regions Of Interest (ROIs) in which to compare the activations for the semantic processing of sentences belonging to the six semantic categories under investigation.

Stages of the research

This research has been developed in two stages, each corresponding to one fMRI study, differing with respect to some methodological points. In the first stage of the research (Study B1), we applied a sparse image acquisition technique that allows the presentation of the stimuli in the silent intervals between brain volumes acquisitions, thus ensuring that participants could hear the auditory stimuli under optimal listening conditions devoid of contaminations with scanner noise (Hall et al., 1999; Belin et al., 1999). This procedure is particularly suited to experiments in the domain of auditory language processing to prevent fMRI noise to interfere with stimulus processing, although a negative side effect is the partial reduction of the statistical power (Zaehle et al., 2007).

In this preliminary study, in addition to the experimental factor concerning the semantic content of the abstract and action-related categories, we also explored the impact of a linguistic operator, namely sentential negation, on the semantic processing. Psycholinguistic and neuroimaging findings point to a “disembodiment effect” of sentential negation in terms of (i) reduced access to mental representations of negated conceptual information (Kaup, 2001; Kaup & Zwaan, 2003); (ii) reduced activation of concept-specific embodied neural systems in processing negative vs. affirmative sentences, leaving neural resources more free for concurrent motor performance (Tettamanti et al., 2008; Tomasino, 2010; Liuzza et al., 2011; Bartoli et al., 2012; Foroni & Semin, 2013). We aimed at confirming these findings by testing the specificity of the negation effect with respect to the six categories of meanings presented either in the affirmative or in the negative form (6 x 2 experimental design).

Results obtained in this first research stage provided interesting data according to some of the a priori hypotheses, nevertheless the statistical power was low (see StudyB1, section 2.2). The failure of statistical significant results can plausibly be ascribed to the modality of data acquisition, namely the sparse acquisition technique.

Accordingly, in the second stage of the research (Study B2), we collected data in an independent sample of experimental subjects using standard continuous sampling acquisition, which is known to yield higher fMRI statistical power. Moreover, in order to further increase the statistical power, we reduced the number of experimental conditions by considering only the main factor related to the semantic domain (6 levels corresponding to the three abstract-related and the three action-related categories, 6 x 1 experimental design) and avoiding the presentation of sentences in the negative form, given that this latter aspect of research was deemed as subjacent to first finding positive evidence for the semantic domain factor.

In study B2, in addition to the standard, univariate general linear model analysis, we applied two further statistical approaches:

(i) a multi-voxel pattern analysis (MVPA), in which a pattern-classification algorithm is applied to multi-voxel patterns of activity in order to decode the information that is represented in that pattern of activity (for methodological reviews see Haynes & Rees, 2006; Norman et al., 2006; O'Toole et al., 2007; Pereira et al., 2009). The idea is that if decoding identifies categories better than chance, then there must be information about the stimuli in the response patterns. The main benefit of MVPA is that it allows identifying patterns of activity that are specifically diagnostic of a particular stimulus type, thus providing much greater specificity to neuroimaging results (Norman et al., 2006; Poldrack, 2008).

This approach has been successfully applied in several studies investigating how the semantic information is represented in the cerebral cortex, including: (i) studies on visual object categories accessed through pictures (Haxby et al., 2001; Carlson et al., 2003; Cox & Savoy, 2003; Henson et al., 2004; Hanson & Halschenko, 2007; Kay et al., 2208; Reddy et al., 2009); (ii) studies on sound-related object categories accessed through sounds (Staeren et al., 2009); (iii) studies on concrete meanings accessed through concrete nouns (Just et al., 2010; Shinkareva et al., 2011; Quadflieg et al., 2011). Up to now, there is only a preliminary study applying MVPA approach to the study of abstract words (Anderson et al., 2012).

In this study we applied the MVPA in order to verify the possibility of distinguishing abstract and concrete meanings from the brain activity. In particular, this approach was particularly appropriate to explore the distributed nature of semantic representations accessed through the six categories of meanings described above. MVPA was also exploited to perform the analysis across subjects to test whether information is represented in the same way across individuals (Shinkareva et al., 2011; Quadflieg et al., 2011; Kaplan & Meyer, 2012).

(ii) Parametric analysis, in which we investigated the specific correlates of the abstract- and action-related domains, by directly taking into account the data about psycholinguistic variables collected in the rating study, including length/duration, frequency, concreteness, context availability, familiarity, body-part ratings, and category-specific association ratings (Ghio et al., 2013). These measures allowed for a quantification of different profiles of action and abstract meaning characteristics, but without any a priori categorical distinctions. The parametric experimental approach has the potential to be particularly useful in controlling the role of these psycholinguistic dimensions in sentence processing. This approach has been successfully applied by Vigliocco et al., (*in press*) in order to investigate the correlation of valence and arousal ratings with the brain activations.

2. Study B1

2.1 Materials and Methods

2.1.1 Participants

Fifty volunteer subjects (25M/25F; mean age 23.02 years, s.d. 4.88) of comparable education level participated in the experiment. All subjects were right-handed (mean score 0.94, s.d. 0.05) according to the Edinburgh Inventory (Oldfield, 1971). They were all native monolingual speakers of Italian, with no history of neurological or psychiatric disorders. Participants gave written consent to participate in the study after receiving a careful explanation of the procedures. The study was approved by the Ethics Committee of the San Raffaele Scientific Institute, Milano, Italy. Two subjects (1M/1F) were excluded due to the incidental findings of structural anomalies.

2.1.2 Experimental design

The experiment consisted of a 6x2 factorial design, the two factors being the Semantic domain (six levels, mental state-, emotion-, mathematics-, mouth-, hand-, leg-related), and Polarity (two levels: affirmative, negative). The experimental stimuli consisted of 210 base form sentences (35 for each level of the first factor). Each sentence was repeated twice, with minimal variations reflecting the two levels of the second factor. Namely, for the affirmative condition, sentences consisted of a third person feminine pronoun (‘Lei’, Engl.: *She*), a verb in third-person singular, simple present tense, matched to a syntactically and semantically congruent object complement. Corresponding negative sentences were created by replacing the third person feminine pronoun with sentential negation. As sentential negation in Italian is expressed by a single word (‘Non’, Engl.: *Not*), both affirmative and negative sentences consisted of four words. The presence of the optional subject pronoun served to equate the number of words between affirmative and negative sentences, and it was chosen because it does not add extra semantic information, since the features of number and person that the pronoun carries are already included in Italian verbal inflections. Indeed, in Italian, as opposed to other languages like English, the subject is not obligatorily expressed. The complete set of experimental stimuli thus included 420 sentences.

The twelve experimental conditions, corresponding to the 6x2 factorial design, were: **(AMs)** affirmative mental state-related sentences (e.g., ‘Lei ricorda il passato’, Engl.: *She remembers the past*); **(NMs)** negative mental state-related sentences (e.g., ‘Non ricorda il passato’, Engl.: (*She does) Not remember the past*); (**AEm)** affirmative emotion-related sentences (e.g., ‘Lei mostra il disappunto’, Engl.: *She shows her* *disappointment*); **(NEm)** negative emotion-related sentences (e.g., ‘Non mostra il disappunto’, Engl.: (*She does) Not show her disappointment*); **(AMa)** affirmative mathematics-related sentences (e.g., ‘Lei calcola la somma’, Engl.: *She determines the sum*); **(NMa)** negative mathematics-related sentences (e.g., ‘Non calcola la somma’, Engl.: (*She does) Not determine the sum*); **(AMo)** affirmative mouth-related sentences (e.g., ‘Lei schiocca la lingua’, Engl.: *She clicks her tongue*); **(NMo)** negative mouth-related sentences (e.g., ‘Non schiocca la lingua’, Engl.: (*She does) Not click her tongue*); **(AHa)** affirmative hand-related sentences (e.g., ‘Lei ricama il fazzoletto’, Engl.: *She embroiders the handkerchief*); **(NHa)** negative hand-related sentences (e.g., ‘Non ricama il fazzoletto’, Engl.: (*She does) Not embroider the handkerchief*); (**ALe)** affirmative leg-related sentences (e.g., ‘Lei calcia la palla’, Engl.: *She kicks the ball*); **(NLe)** negative leg-related sentences (e.g., ‘Non calcia la palla’, Engl.: (*She does) Not kick the ball*).

2.1.3 Linguistic stimuli

In order to generate the set of stimuli, we used the same base form sentences employed in a previous psycholinguistic rating study (Ghio et al., 2013, Study A in the present dissertation). Comprehensive details on how stimuli were characterized for several psycholinguistic dimensions, such as length, frequency, concreteness, context availability, familiarity, body-part involvement, and specific-category association were provided in the rating study. Here we reported additional information relevant to the present experiment, in which stimuli are presented auditorily both in the affirmative and negative form.

Sentences were digitally recorded by a female native Italian speaker in a sound-proof room and edited using Praat (Version 5.2.35, www.praat.org). All sentences were normalized and intensity was set at 70 dB. Pitch was balanced across the six semantic categories both in the affirmative (F (5,204) = 1.43, p = 0.21) and the negative (F (5,204) = 0.65, p = 0.66) condition. The average playback duration of the recorded sentences was 1.43 sec, s.d. = 0.13 sec (AMs: 1.47 sec, s.d. = 0.11 sec; NMs: 1.49 sec, s.d. = 0.10 sec; AEm: 1.44 sec, s.d. = 0.17 sec; NEm: 1.44 sec, s.d. = 0.18 sec; AMa: 1.44 sec, s.d. = 0.12 sec; NMa: 1.45 sec, s.d. = 0.12 sec; AMo: 1.39 sec, s.d. = 0.15 sec; NMo: 1.41 sec, s.d. = 0.13 sec; AHa: 1.38 sec, s.d. = 0.12 sec; NHa: 1.40 sec, s.d. = 0.11 sec; ALe: 1.40 sec, s.d. = 0.12 sec; NLe: 1.42 sec, s.d. = 0.13 sec). In the affirmative condition, we found a trend toward a main effect of the semantic category (F (5, 204) = 2.23, p = 0.052, Tukey’s post hoc comparisons, all p > 0.05). In the negative condition, duration was not balanced across the six semantic categories (F (5, 204) = 2.68, p = 0.022). Tukey’s post hoc comparisons revealed a significant difference between NMs vs. NHa (difference between conditions = 0.09 sec) and NMs vs. NMo (mean difference between conditions = 0.08 sec)[3].

---

[3] It is very unlikely that these duration differences may be perceived at the sentence level, since the just noticeable difference

We also considered the polarity factor separately for each semantic domain. The following independent t-tests were carried out: AMs vs. NMs, AEm vs. NEm, AMa vs. NMa, AMo vs. NMo, AHa vs. NHa, ALe vs. NLe. For each semantic domain, no significant differences between affirmative and negative sentences were found with respect to the mean pitch and the mean duration (all $p > 0.05$).

2.1.4 Experimental procedures

Participants underwent different tasks during fMRI scanning: 1) linguistic task; 2) six localizers tasks (Figure 1). The software package Presentation 14.9 (Neurobehavioral Systems, Albany, CA, USA) was used for auditory and visual stimuli presentation. Visual stimuli were projected on a back-projection screen located in front of the scanner and viewed through a mirror placed on the head coil.

2.1.4.1. Linguistic task

Subjects were instructed to carefully listen to all sentences administered via MRI-compatible headphones integrated in the Philips MRI scanner, connected to a personal computer. 2000 ms after the end of each auditory sentence presentation, a fixation cross or a question mark were visually presented for 500 ms. When a fixation cross appeared on the screen, subjects were not asked to perform any tasks (experimental trial). When a question mark appeared on the screen it was followed by a written sentence (duration: 1000 ms), and subjects were asked to perform a cross-modal 1-back task, i.e. to judge whether the written sentence matched the auditory sentence presented immediately before (catch trial) (Figure 1.1). Participants were instructed to either slowly blink their eyelids once if the written sentence was identical to the auditory sentence presented immediately before (match trial), or to blink their eyelids twice if the written sentence did not match to the previous auditory sentence (non match trial). Mismatches between auditory and written sentence were possibly due to a mismatch in the verb, a mismatch in the object complement, or a mismatch in both the verb and the object complement (for the examples, see the Introduction, section 1.4). Eyelid-blink responses were video recorded and separately scored by two independent judges. The items used in catch trials were 4-word sentences of the same form of the experimental stimuli (e.g., 'Lei apre la porta', Engl.: *She opens the door*).

Before fMRI scanning and outside the magnet room, participants completed a brief training to familiarize with the experimental linguistic task. For this purpose, subjects performed the task on 11 sentences (including 2 catch trials, 1 match, 1 mismatch). Training sentences were related to social, medical, and law semantic domains to avoid biasing the subjects' attention toward the semantic domains of the experimental stimuli. The training sentences were not included in the actual experiment.

---

between two phonetic sounds is 0.02-0.03 sec (Bertinetto, 1981: p. 133, p. 142).

In order to reduce the duration of the experiment and to avoid the exposure of each individual subject to the same sentence in both the affirmative and negative conditions, the pool of 420 sentences was split in two counterbalanced lists of 210 experimental stimuli. Each list included only one version of each base form sentence, either with affirmative or with negative polarity. Each list included all the twelve experimental conditions, 17 items for each experimental condition. The lists were alternated between two groups of subjects (Group A and Group B). 24 subjects were semi-randomly assigned to Group A (**GA**) and 24 to Group B (**GB**). The sex and the age of the participants were balanced across the two groups (GA: 12M/12F, mean age = 23.71, s.d. = 6.14, Oldfield mean score = 0.94, s.d. = 0.06. GB: 12M/12F, mean age = 22.42, s.d. = 3.61, Oldfield mean score = 0.95, s.d. = 0.05).

The linguistic stimuli were presented in an event-related experimental design with sparse sampling, divided in four separate fMRI runs (12 minutes and 40 seconds long each). Each linguistic run was constituted by 67 randomized trials: 51 experimental trials, 9 catch trials, and 8 null trials. The first and the third runs included six out of the 12 experimental conditions, one condition for each level of the Semantic domain factor, either in the affirmative or negative polarity form (GA: NMs, AEm, NMa, AMo, NHa, ALe; GB: AMs, NEm, AMa, NMo, AHa, NLe); the second and the fourth runs included the remaining six experimental conditions, i.e. those with reversed polarity (GA: AMs, NEm, AMa, NMo, AHa, NLe; GB: NMs, AEm, NMa, AMo, NHa, ALe). A minimum of 8 and a maximum of 9 experimental trials for each conditions were included in each run. Null trials were trials of no events and were included to increase statistical contrast efficiency. Stimulus Onset Asynchrony was jittered in order to maximize the detection of the hemodynamic response elicited by the auditory sentence. SOAs were calculated from the end of the sentence to the onset of the TR coming immediately after, and ranged between 2750 and 4750 ms (250 ms jitters; average SOA: 3847 ms).

2.1.4.2. Localizer tasks

Following the linguistic task, participants completed six different localizer tasks. The purpose was to localize functional regions involved in emotion, mathematics and mental-states related processing and in actions performed with the mouth, the hand and the foot. For the emotion, mathematics and action localizers we used a standard block-design localizer task in which three task-related blocks alternated with three rest blocks. For the mental state-related meanings, a resting state localizer was performed in order to identify the default mode network. A detailed description of each localizer is provided in the following.

Emotion localizer task (Figure 1.2A): Rest blocks during which the word '"Rest" appeared on the screen (duration 24000 ms) were alternated with task-related blocks during which subjects were visually presented with pictures of human faces with emotional facial expressions of joy, anger, and neutral expressions in semirandomized order. Stimuli were selected from the NimStim Emotional Face Database (Tottenham et al., 2009). Each block (duration 24000 ms) included 8 pictures, each lasting 3000 ms. Participants were required to carefully observe each picture without any additional tasks.

Mathematics localizer task (Figure 1.2B): Rest blocks during which the word “Rest” appeared on the screen (duration 24000 ms) were alternated with task-related blocks during which participants covertly performed mathematical calculations following visually presented calculations. Each block started with the word “Calculate!” (2500 ms) followed by a series of calculations, like “47” (2500 ms), “+ 13” (5500 ms), “x 2” (5500 ms). Finally, an equal mark and a question mark “= ?” appeared on the screen (5500 ms), and subjects were required to mentally calculate the final results. They were instructed not to move their lips and mouth, nor to articulate the verbal calculations. In order to verify that the participants actually performed the calculation, at the end of the localizer task subjects were asked to recall the results obtained. In order to avoid any memory effects, participants were not told in advance that they would be asked to recall the results.

Action localizer tasks (Figure 1.2C): To identify motor cortex activation corresponding to mouth, hand and foot movements, three different motor localizer tasks were administered. In each localizer tasks, the sentences 'Move your tongue /right hand /right foot' appeared on a computer screen for 24000 ms each, and participants had to execute the corresponding movement during the whole period that the sentence remained on the screen, alternated by 24000 ms of “Rest”. In the tongue movement localizer task, subjects were required to move their tongue with a continuous rotatory movement. In the hand movement localizer task, participants were required to move their right hand (open and close the fingers, once every 1-2 seconds). In the leg movement localizer task, participants were required to move their right foot (backward and forward along the MRI bed horizontal axis, once every 1-2 seconds).

Resting state localizer (Figure 1.2D). Subjects laid still in the scanner with their eyes closed for three minutes.

In order not to bias the participants’ attention toward emotion-, mathematics, mental state-, or action-related conceptual aspects of the linguistic stimuli, the participants were kept unaware of the instructions of the localizer tasks until the end of the linguistic task.

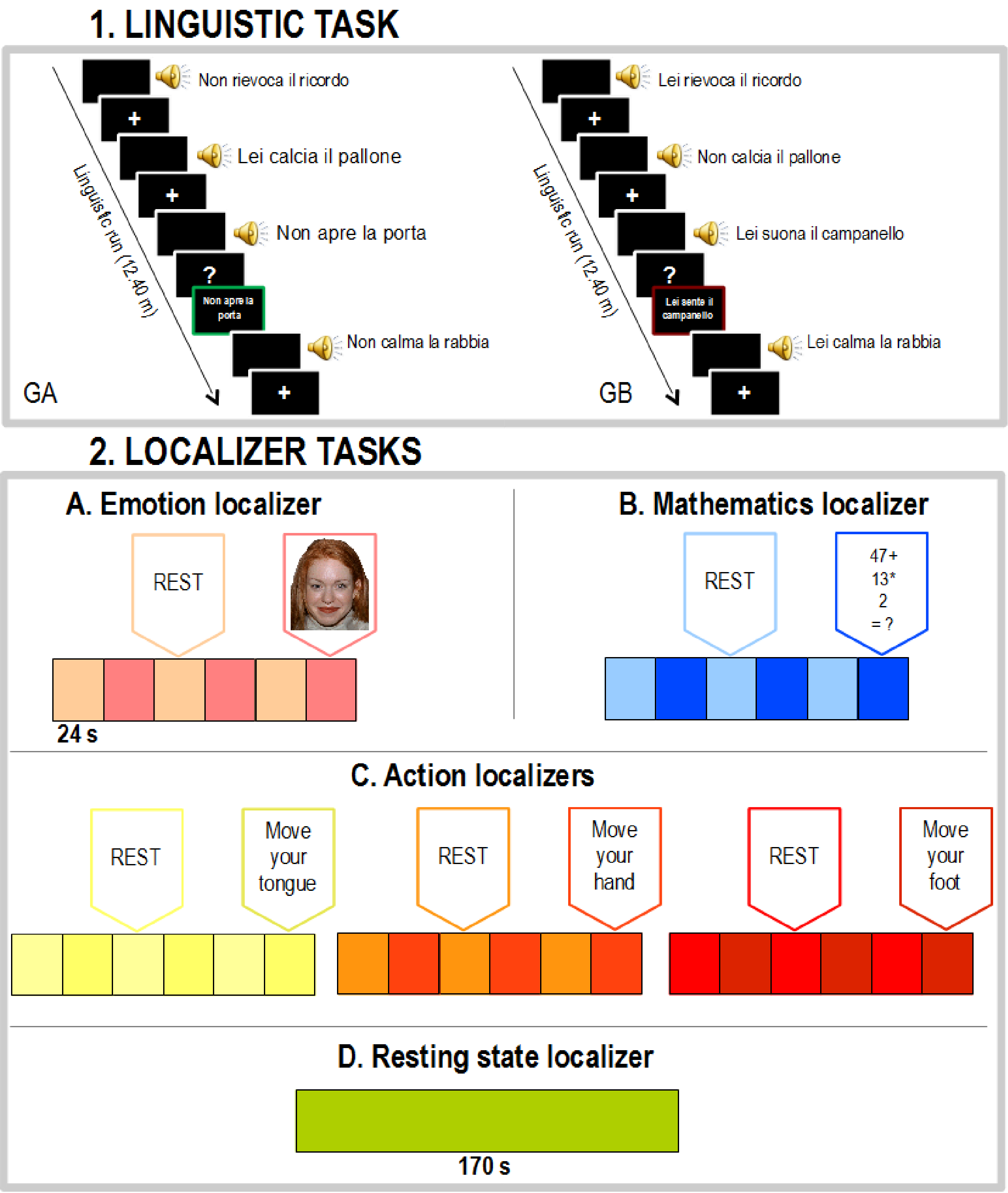


Figure 1. Schematic representation of the experimental procedure (Study B1). (1) Linguistic task; Green = match trials; red = non match trials. GA = Group A; GB = Group B. **(2)** Localizer tasks: (1.2A) Emotion localizer; (1.2B) Mathematics localizer; (1.2C) Action localizers, including: the mouth localizer, the hand localizer, the leg localizer; (1.2D) Resting state localizer.

2.1.5 Data acquisition

MRI scans were acquired on a 3T Intera Philips body scanner (Philips Medical Systems, Best, NL) using an 8 channels-sense head coil (sense reduction factor = 2). Whole-brain functional images were obtained with a T2*-weighted gradient-echo, echo-planar sequence, using blood-oxygenation-level-dependent contrast. Each functional image comprised 35 contiguous axial slices (3.2 mm thick, 0.8 mm gap), acquired sequentially from bottom to top (field of view = 240 mm x 240 mm, matrix size = 128 x 128).

To ensure that participants could hear the auditory stimuli under optimal listening conditions devoid of contaminations with scanner noise, a sparse image acquisition technique was used during the linguistic task. By applying this technique, a silent period was interleaved between each volume acquisition (repetition time: 10500 ms, acquisition time: 2915 ms, echo time: 30 ms). Each participant underwent four functional scanning sessions. Each scanning session comprised 71 scans, preceded by 2 dummy scans that were discarded prior to data analysis.

A continuous acquisition technique was in turn used during the localizer tasks since they did not include any auditory stimulation (repetition time: 3000 ms, acquisition time: 2915 ms, echo time: 30 ms; all the other parameters were the same as for the sparse acquisition). Each participant underwent six localizer scanning sessions. Each localizer session comprised 56 scans, preceded by 5 dummy scans, which did not enter in the analysis.

A high resolution T1-weighted anatomical scan (three-dimensional (3D), spoiled-gradient-recalled sequence, 200 slices, TR = 600 ms, TE = 20 ms, slice thickness = 1 mm, in-plane resolution 1x1x1 mm) was acquired for each subject.

2.1.6 Data analysis

Imaging data were processed using Statistical Parametric Mapping (SPM8, Wellcome Department of Imaging Neuroscience, London, UK).

The New Segment procedure was first performed to produce a bias corrected version of the structural images to improve the segmentation. The New Segment procedure was then performed a second time to segment the bias corrected structural MRI image of each subject into six partitions, including grey matter, white matter, cerebrospinal fluid, skull, and non-brain regions of the image. The segmentation made use of a custom template based on a sample of 317 images of healthy subjects acquired with the same 3T Intera Philips body scanner used in the present experiment. The custom template was previously normalized to approximate the Montreal Neurological Institute (MNI) standard space. The images underwent a very light bias regularization (bias FWHM 60 mm cut-off) and were spatially normalized using an affine spatial normalization. The resulting segmentation structural images were resampled with a spatial resolution of 2×2×4 mm and smoothed with a 2 mm full width at half maximum isotropic Gaussian kernel.

Separately for linguistic and localizer scans, functional images were corrected for slice timing, and realigned to the first scan of the first session acquired in each subject. The same New Segment procedure was applied separately to functional linguistic images and the localizer images. The tissue probability maps used in this procedure were the images obtained through the New Segmentation of the structural image described above. Images were normalized into the approximate MNI space and spatially smoothed with a 6-mm FWHM Gaussian kernel.

General Linear Model statistical analysis was used (Friston et al., 2002). We adopted a two-stage random-effects approach to ensure generalizability of the results at the population level (Penny & Holmes, 2004). The linguistic task and the localizer tasks were analysed separately.

2.1.6.1 Linguistic task

First-level General Linear Models. The time series of each participant were high-pass filtered at 128 sec. The autoregressive model AR(1) was not applied due to sparse sampling acquisition. No global normalization was performed. For each participant, we modelled a 6x2 factorial design with 4 separate sessions. For subjects belonging to GA, the first and the third sessions included AMo, NHa, ALe, AEm, NMa, NMs; the second and the fourth sessions included NMo, AHa, NLe, NEm, AMa, AMs. Vice versa, for subjects belonging to GB, the first and the third sessions included NMo, AHa, NLe, NEm, AMa, AMs; the second and the fourth sessions included the AMo, NHa, ALe, AEm, NMa, NMs conditions. Separate regressors modelled catch trials and task instructions. To correct for motion artefacts, subject-specific realignment parameters were modelled as covariates of no interest.

The following Student's t-test contrasts were defined at the first level: (i) a t-contrast including a weight of -1 for the regressors modelling the experimental conditions in each session, and a weight of 6 for the regressor modelling the catch trials; (ii) a set of t-contrasts specified for the purpose of the random effects group analysis, each including a weight of one for a particular regressor modelling one experimental condition and a weight of zero for all the other regressors, thus resulting in one contrast per experimental condition (12 contrasts) for each participant.

Second-level General Linear Model. (i) In order to control for statistically significant effects associated with performing the linguistic task versus the cross-modal 1-back task, the contrast images, one for each subject, obtained at the single-subject level were entered into a one-sample t-test: “Linguistic task > 1-back task” ([|Ms| + |Em| + |Ma| + |Mo| + |Ha| + |Le|] – Catch), and “1-back task > Linguistic task” (Catch – [|Ms| + |Em| + |Ma| + |Mo| + |Ha| + |Le|]).

(ii) The twelve contrast images of each subject obtained at the single-subject level were entered into a second-level random effects model using the flexible factorial design. The model included the subject factor (n

= 48 participants), the Semantic domain factor (six levels: Ms, Em, Ma, Mo, Ha, Le), and the Polarity factor (two levels: affirmative, negative). In this random-effects model, we modelled independence across images from the same subject and equal variances between conditions and subjects as implemented in SPM8. The following one-sample t-contrasts were defined: i) Main effect of concreteness: “abstract-related > action-related” [(AMs + NMs + AEm + NEm +AMa + NMa) - (AMo + NMo+ AHa + NHa + ALe + NLe)], and “action-related > abstract-related” [(AMo + NMo+ AHa + NHa + ALe + NLe) – (AMs + NMs + AEm + NEm +AMa + NMa)]; ii) Main effect of polarity: “affirmative > negative” [(AMo + AHa + ALe + AMs + AMa + AEm) – (NMo – NHa – NLe – NMs – NMa – NEm)], and “negative > affirmative” [(NMo – NHa – NLe – NMs – NMa – Nem) – (AMo + AHa + ALe + AMs + AMa + AEm)]. All reported effects relate to voxel-level statistics $p < 0.05$, Family Wise Error (FWE) type correction for multiple comparisons.

2.1.6.2 Localizer tasks

Emotion, Mathematics, and Action localizer tasks

First-level General Linear Models. The time series of each participant were high-pass filtered at 128 sec. The autoregressive model AR(1) was applied. No global normalization was performed. As explained above, for the emotion, mathematics, and action localizers a block design was used. fMRI responses were modelled with a canonical hemodynamic response function aligned to the onset of each block. First-level t-Student contrasts were specified, each representing the activation evoked by the localizer task relative to the rest condition.

Second-level General Linear Model. The contrast images obtained for each localizer task at the single subject level entered in the second-level analysis. The following one-sample t-contrasts were calculated: i) Emotion task vs. rest; ii) Calculation task vs. rest; iii) Mouth movement vs. rest; iv) Hand movement vs. rest; v) Leg movement vs. rest.

Resting state localizer

Analysis of the resting state localizer scan was performed using the GIFT toolbox2 (www.icatb.sourceforge.net/groupica.htm) according to the procedure indicated by Allen and colleagues (2011). Data were decomposed into functional networks using spatial Independent Components Analysis (ICA). This approach, when applied to fMRI data, identifies temporally coherent networks by estimating maximally independent spatial source. A relatively high-order ICA model (number of components, C = 40) was used. Subject-specific data reduction principal components analysis (PCA) retained T1 = 45 principal components (PCs) using a standard economy-size decomposition. Group data reduction retained C = 40 PCs using the expectation-maximization (EM) algorithm, included in GIFT. The Infomax ICA algorithm was repeated

20 times in Icasso3 and resulting components were clustered to estimate the reliability of the decomposition. Subject-specific spatial maps and time courses were estimated using GICA3 back-reconstruction method. By inspecting power spectra, components considered as plausible candidates for constituting the resting state network (as opposed to physiological artifacts) were selected. The component with the highest power spectrum was retained for the univariate analysis (t-test on beta images, FWE correction at 0.05).

2.2 Results

2.2.1 Behavioral results

The videos with the participants' eyelid blinking responses during the 1-back task in the linguistic fMRI runs were analysed by two independent judges. Qualitatively, all participants were judged as alert throughout the entire duration of the experiment. Quantitative analysis was done by measuring the accuracy for each run for each participant. GA mean accuracy was 98.93% (run 1: 99.14%; run 2: 98.29%; run 3: 99.14%; run 4: 99.14%). The mean accuracy did not differ across runs (F (3, 100) = 0.35, p = 0.78). GB mean accuracy was 98.49% (run 1: 99.07%; run 2: 97.22%; run 3: 98.61%; run 4: 99.07%). The mean accuracy did not differ across runs (F (3, 92) = 1.07, p = 0.36). No differences between the mean accuracy of the two groups were found (F (1, 48) = 0.92, p = 0.34).

2.2.2 Neuroimaging results

2.2.2.1 Linguistic task

First, we explored the pattern of activation for the linguistic task compared to the cross-modal 1-back task (Table 1A). The results revealed most prominent clusters of activation in the Heschl's gyrus and in the superior temporal gyrus, bilaterally. Additional activation foci were located in bilateral angular gyrus, in the right inferior frontal gyrus (*pars orbitalis*), in the left superior orbital frontal gyrus, in the left hippocampus, and in the cerebellum (at the level of lobules VIIa, VIIb, and VI). Activation patterns specific to cross-modal 1-back task compared to linguistic task are reported in Table 1B. Clusters of activation were observed in subcortical regions (parahippocampus, bilateral pallidum, bilateral thalamus, right putamen) and in the insula lobe, bilaterally. Bilateral occipital gyrus, right calcarine gyrus and bilateral lingual gyrus were also found to be significantly activated.

Subsequently, we analysed the specific effects of the linguistic task. The main effect of Semantic domain was investigated irrespective of Polarity in order to test which brain systems are engaged by the processing of abstract-related versus action-related sentences (Table 2). For abstract-related sentences, significant activations were found in the left temporal pole, the left superior temporal gyrus, the left middle temporal gyrus, the right superior temporal gyrus, and the right temporal pole (Figure 2.1). Action-related sentence processing was not associated with any specific activation. No evidence of significant activations was found even based on an a priori hypothesis for the left hemispheric fronto-parieto-temporal network.

We also assessed the main effect of Polarity independently of the level of Semantic domain. No significant results were found for the main effect of Polarity (Table 3). Therefore, polarity was not investigated any further with respect to the interactions with the Semantic domain.

2.2.2.2 Localizer tasks

For the emotion localizer task, comparison of processing pictures emotional facial expressions against rest showed significant activations in the following regions: bilateral fusiform gyrus (x = 38; y = -46; z = -20 and x = -40, y = -78, z = -16); right and left amygdala (x = 22, y = -8, z = -16 and x = -22, y = -10, z = -16); basal ganglia (left caudate nucleus x = -14, y = 4, z = 12, right putamen x = 30, y = 8, z = -4); right rectal frontal gyrus (x = 2, y = 52, z = -24) (Figure 2.2A).

Calculation task as compared to rest elicited activations in the bilateral inferior parietal lobule (x = -48, y = -36, z = 44; x = 46, y = -38, z = 44), in the left middle frontal gyrus (x = -40, y = 32, z = 32), in the right inferior frontal gyrus (x = 52, y = 10, z = 16; x = 44, y = 8, z = 28), and in the right angular gyrus (x = 30, y = -62, z = 48) (Figure 2.2B).

Hand, mouth, and leg localizers yielded activations in somatotopically organized primary motor areas involved in mouth (x= -50, y = -10, z = 32), hand (x= -38, y = -24, z = 52), and leg (x= -4, y = -30, z = 68) movements, respectively. In addition, premotor and supplementary motor areas, somatosensory areas and cerebellum were activated (Figure 2.2C).

The resting state localizer analysed by means of ICA revealed a network of brain regions including the precuneus bilaterally (x = 2, y =-52, z = 20), the superior medial frontal gyrus (x = 0, y = 62, z = -12; x = 2, y = 56, z = 28), the left inferior frontal gyrus (x = -36, y = 20, z = 24), the angular gyrus bilaterally (x = -46, y =-64, z = 28; x = 58, y = -62, z = 28), the right inferior frontal gyrus (x = 42, y =18, z = 32), and the left middle frontal gyrus (x = -34, y =18, z = 44) (Figure 2.2D).

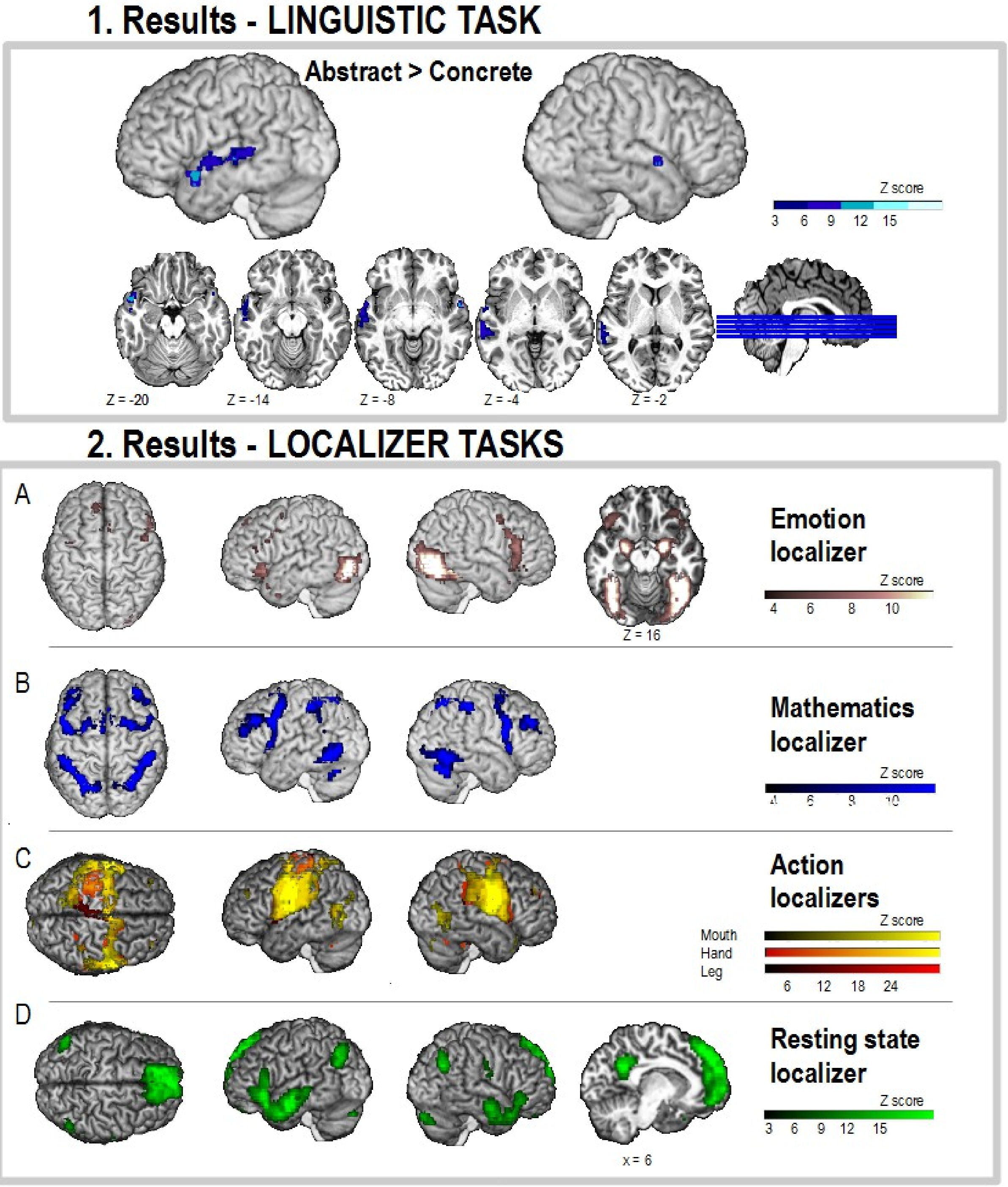


Figure 2. Functional localization of the effects for the linguistic task and the localizer tasks. Significant activations ($P < 0.05$, FWE corrected for multiple comparisons) are displayed on cortical renderings and on sagittal (x coordinate level in mm) and axial (z coordinate levels in mm) slices of the anatomical image of one of the participants (warped to the MNI coordinate space). **(1)** Activations specific for the Abstract > Concrete contrast. **(2)** Results of the localizers tasks. (2.2A) Activations specific to the Emotion localizer; (2.2B) Activations specific to the Mathematics localizer; (2.2C) Activations specific to the Action localizers, including: the mouth localizer (yellow); the hand localizer (orange); the leg localizer (red); (2.2D) Activations specific to the Resting state localizer.

2.3 Discussion

The results of the present study need to be discussed taking into account that a sparse image acquisition technique was used. On the one side, the main advantage of this approach is the possibility of presenting the stimuli during the silent intervals between brain volume acquisitions, thus allowing participants to listen to the sentences in a more ecological condition as compared to continuous acquisition (Hall et al., 1999; Belin et al., 1999). Moreover, it has been shown that the fMRI noise has an effect on brain function, activating the superior and middle temporal gyri and the primary auditory cortex (MacSweeney et al., 2000). This effect can be problematic for fMRI studies on language processing employing auditory stimuli that partially activate the same brain regions (Elliot et al., 1999). The event-related design employed in the present experiment avoided this potential confound by interleaving sentence administration with the sparse acquisition of images. On the other side, the main disadvantage of the sparse acquisition procedure is the partial reduction of the statistical power (Zaehle et al., 2007). In order to limit this side effect, data were collected from a large sample of subjects, and a high number of stimuli for each condition was employed.

Despite that, we generally observed a low statistical power. Significant results were only found for the control contrast (linguistic task vs. 1-back task), and for the abstract-related > action-related contrast. On the contrary, none of the activations found in the action-related > abstract-related, affirmative > negative, and negative > affirmative contrasts survived a stringent statistical threshold ($p < 0.05$), after correcting for multiple comparisons (Family-Wise Error correction). It must be noted that we also performed exploratory analyses of the specific effects related to the processing of individual semantic categories within both the abstract and concrete domains, but again with no significant results. Nevertheless, we believe that the lack of significant results can plausibly be ascribed to the low statistical power associated with the sparse acquisition technique. Indeed, by applying less stringent statistical thresholds, we found interesting data according to the *a priori* hypotheses with respect to the condition-specific semantic attributes. In particular, exploratory analyses at a lower significance threshold revealed some evidence of activations consistent with the predictions of modality-specific networks underlying the processing of different abstract-related and concrete-related meanings (not reported here, in compliance to stringent statistical thresholding criteria). However, even at lower significance thresholds, no specific activations were found for the polarity factor.

For these reasons, we decided to further investigate the semantic domain factor only, by carrying out an fMRI study using a continuous sampling acquisition technique (Study B2). This acquisition technique is the standard in the fMRI research as it yields higher statistical power. Moreover, by excluding the polarity factor we further increased the statistical power.

In the light of the second study, we briefly discuss the significant results of the present experiment in the remainder of the discussion. First, we illustrate how the findings of the control contrast (linguistic task vs. 1-back task) were particularly relevant with respect to some methodological issues concerning the choice of the experimental task (see also the Introduction). Second, we present the results obtained for the processing of abstract-related meanings compared to concrete-related meanings. Third, we discuss the results of the localizer tasks since they were subsequently used in the Study B2.

2.3.1 Discussion of the experimental task and behavioral results

The experimental design employed in the present experiment allowed distinguishing between the linguistic task – i.e. listening to sentences without performing any additional tasks – from the cross-modal 1-back task – i.e. judging whether the written sentence matched the auditory sentence presented immediately before.

Behavioral results revealed that participants performed the 1-back task with high accuracy. These results assured that the subjects were able to perform the task easily. At the same time, high accuracy rates indicate that participants were alert throughout the entire duration of the experiment, thus ruling out the possibility that the observed brain activity might reflect a passive/resting state. Indeed, the use of the active task in combination with listening to the sentences was aimed at avoiding task-unrelated semantic processes, and at preserving a high level of attention on sentence processing throughout the study (Binder et al., 2009).

The neuroimaging results revealed that the 1-back task elicited a pattern of activity that reflects: (i) working memory processes for the active maintenance and processing of the linguistic inputs. These processes were associated with activations in the frontal cortex and in the parahippocampal cortex, also extending to subcortical regions (Jansma et al., 2000). These regions are known to be involved also in recognition and in familiarity memory tasks (Martin et al., 2013; Staresina et al., 2011). It must be however acknowledged that these are at most tentative interpretations of the findings, as the specific working memory network is not clearly revealed by the data in its entirety, given for instance the lack of middle frontal and parietal activations (but see Study B2) – possibly due to low statistical power; (ii) supportive processes, such as visual perception of written sentences – associated with activations in visual areas, including the right calcarine gyrus, the bilateral lingual gyrus, the bilateral inferior occipital gyrus, and the right fusiform gyrus – and motor execution of the response – associated with activations in the subcortical regions. Importantly, no activations in modality-specific brain networks relevant for the investigated semantic categories were found. In particular, no activation in mouth, hand, or leg motor cortices was found.

In turn, listening to sentences was associated with bilateral activations in the Heschl's gyrus due to the auditory processing of the stimuli. Language-specific activations were found in the left superior temporal gyrus, the left angular gyrus, and in their contralateral counterparts.

The finding of two distinct networks underlying the linguistic task and the n-back task suggests that two different processes were elicited. This observation was particularly relevant since the 1-back task was intended merely as a means for eliciting a deep semantic processing of sentences during the linguistic task, without interfering with specific semantic aspects of meaning processing.

In conclusion, based on these behavioral and neuroimaging results, in Study B2 we employed the same cross-modal, 1-back task in combination with the linguistic task.

2.3.2 Discussion of the semantic domain factor

The first objective of the fMRI data analysis was to identify the brain regions subserving the processing of abstract- and concrete-related meanings. We found that processing abstract-related sentences compared to the processing of concrete sentences correlated with activations in the bilateral superior temporal gyrus, the left middle temporal gyrus, and the bilateral temporal pole. These findings are largely consistent with previous evidence on the processing of abstract-related meanings (for meta-analyses see Binder et al., 2009; Wang et al., 2010).

First, although the involvement of the left superior temporal gyrus has been primarily related to speech perception and phonological processing, several fMRI studies suggested that this region plays a role in the semantic processing of abstract concepts (Wang et al., 2010). In the present study, abstract and action-related sentences were carefully matched for mean intensity, mean pitch, and temporal duration, minimizing the possible influences of low-level auditory features. For this reason, the involvement of the left superior temporal gyrus may be ascribed to the semantic processing of abstract-related meanings, according to previous evidence. In turn, the involvement of the right superior temporal gyrus in the semantic processing is more controversial. According to some studies, this brain region seems to be more activated when the emotional prosody is incongruous with semantic content (Wittfoth et al., 2010). Interestingly, such an interpretation may be tenable in the present context as all sentences were read with controlled neutral intonation. Since abstract-related meanings, in general, receive higher ratings for affective/emotional association (both valence and arousal) than concrete meanings (Vigliocco et al., 2009; *in press*), the neutral intonation may have been considered by subjects less congruous for the abstract sentences than for the concrete ones. However, further studies are needed in order to validate this interpretation.

Second, the role of the left middle temporal gyrus as an heteromodal cortex involved in supramodal integration and concept retrieval has been widely demonstrated in neuroimaging studies (for a review, see Price, 2010). Neuropsychological data also demonstrated that lesions in the middle temporal gyrus are associated with semantic deficits (Dronkers et al., 2004). In addition, the meta-analysis of Wang et al. (2010)

suggested the specific involvement of this region for the abstract-related meaning compared to the concrete ones.

Perhaps the most interesting result was the bilateral involvement of the anterior temporal lobe (ATL). The ATL is thought to be critical for semantic knowledge, although there is disagreement over its precise role in semantic processing (Bonner & Price, 2013). On the one hand, based on neuropsychological and neuroimaging data, the ATL is generally considered as serving as an amodal hub that integrates the various information associated with a concept (Patterson et al., 2007). On the other hand, it has been recently observed that the ATL includes several anatomically and functionally distinct regions, suggesting that the semantic processing in the ATL may not be fully amodal (Bonner & Price, 2013). In particular, a growing amount of evidence suggests that the temporal pole receives strong affective input from the ventral frontal lobe and amygdala and it is better characterized as a modal region for processing emotions and social concepts (Olson et al., 2007; Zahn et al., 2007). The finding of the present study is consistent with the idea that abstract meanings mainly rely on affective information, as proposed by some versions of embodiment (Barsalou, 2008; 2010; Vigliocco et al., 2009).

On the contrary, the analysis of fMRI data failed to reveal activations for the processing of action-related sentences compared to abstract-related ones. In particular, no evidence of significant activations was found, based on an a priori hypothesis, in the left hemispheric fronto-parieto-temporal network (Tettamanti et al., 2005; 2008; Pulvermueller & Fadiga, 2010). The absence of significant activations for concrete meanings is consistent with several studies demonstrating a set of brain regions more activated by abstract words, and no brain regions showing the opposite tendency (Kiehl et al., 1999; Perani et al., 1999; Grossman et al., 2002; Noppeney & Price, 2004; Pexman et al., 2007; Vigliocco et al., *in press*). However, there are also several studies demonstrating, on the contrary, a significant involvement of specific brain areas in the processing of concrete relative to abstract meanings (for reviews, see Binder et al., 2009; Wang et al., 2010). We believe that the absence of significant results in the present study was rather due to the low statistical power associated with the sparse acquisition technique. As anticipated previously, by applying less stringent statistical thresholds, we found data consistent with the a priori hypothesis. These exploratory analyses encouraged us to conduct further investigations (i.e., Study B2).

2.3.3 Discussion of the localizer tasks' results

For abstract-related categories, localizer tasks were performed in order to identify brain regions activated when participants (i) actually processed emotional facial expressions (emotion localizer); (ii) performed mathematical calculations (mathematics localizer); (iii) were involved in mentalizing processes (resting state localizer). Within localizer regions more active during each localizer task, in Study B2 we tested the predictions that neural activity would be greater for processing emotion-, mathematics-, and mental state-related sentences, respectively. In good agreement with previous evidence on emotional facial expression processing, the emotion localizer lead to identify a brain circuit including the bilateral amygdala, the basal ganglia, and the bilateral fusiform gyrus (for a review, see Fusar-Poli et al., 2009). The mathematics localizer task elicited parietal and frontal activations in specific brain regions classically associated to the processing of numerical quantities and calculations, such as the intraparietal sulcus and the middle frontal gyrus (for a review, see Pinel et al., 2007). Finally, by employing the resting state localizer, we identified the well-known default mode network, including the posterior cingulate cortex, medial superior frontal gyrus, bilateral angular gyrus, left middle frontal gyrus (for a review, see Fox & Raichle 2007).

For the action-related categories, action localizer tasks were used to identify brain regions activated when participants moved their tongue, their left hand, and their left foot. The group mean coordinates of the peak activations within the motor and the premotor cortex for mouth, hand, and leg movements were in good accordance with previous data (Tomasino et al., 2010; Carota et al., 2010; Moseley et al., 2012). Within mouth, hand, and leg localizer regions, in Study B2 we tested the prediction that neural activity would be greater for processing mouth-, hand-, and leg-related sentences, respectively.

3. Study B2

3.1 Materials and Methods

3.1.1 Participants

Thirty-six volunteer subjects (18M/18F; mean age 21.16 years, s.d. 3.87) of comparable education level participated in the experiment. All subjects were right-handed (mean score 0.96, s.d. 0.04) according to the Edinburgh Inventory (Oldfield, 1971). They were all native monolingual speakers of Italian, with no history of neurological or psychiatric disorders. Participants gave written consent to participate in the study after receiving a careful explanation of the procedures. The study was approved by the Ethics Committee of the San Raffaele Scientific Institute, Milano, Italy.

3.1.2 Stimuli and Experimental design

The experimental stimuli consisted of the same 210 affirmative sentences described in Study B1 (section 2.1.3), belonging to different semantic domains, i.e. three abstract-related semantic domains (mental state-, emotion-, mathematics-related sentences), and three action-related semantic domains (mouth-, hand, leg-related sentences).

The experiment consisted in a single-factor design, the factor being the Semantic domain (six levels). The six experimental conditions were: **(Ms)** mental state-related sentences (e.g., 'Lei ricorda il passato', Engl.: *She remembers the past*); (**Em)** emotion-related sentences (e.g., 'Lei mostra il disappunto', Engl.: *She shows her disappointment*); **(Ma)** mathematics-related affirmative sentences (e.g., 'Lei calcola la somma', Engl.: *She determines the sum*); **(Mo)** mouth-related affirmative sentences (e.g., 'Lei schiocca la lingua', Engl.: *She clicks her tongue*); **(Ha)** hand-related affirmative sentences (e.g., 'Lei ricama il fazzoletto', Engl.: *She embroiders the handkerchief*); (**Le)** leg-related affirmative sentences (Le) (e.g., 'Lei calcia la palla', Engl.: *She kicks the ball*).

3.1.3 Experimental Procedure

Participants performed the same linguistic task described in Study B1 (section 2.1.4). Experimental trials are illustrated in Figure 3. Before the experimental session, subjects completed a brief training to familiarize with the Linguistic task. Unlike Study B1, participants performed the training while in the scanner in order to set the audio volume level to be optimal for each subject. This procedure was adopted since this study was performed with a continuous fMRI acquisition technique and, thus, sentences were presented with background scanner noise. In order to ensure optimal listening conditions, we employed high-quality MRI-compatible headphones controlled by an optical audio-control unit (both MR Confon GmbH, Magdeburg, Germany). The software package Presentation 14.9 (Neurobehavioral Systems, Albany, CA, USA) was used for stimuli presentation.

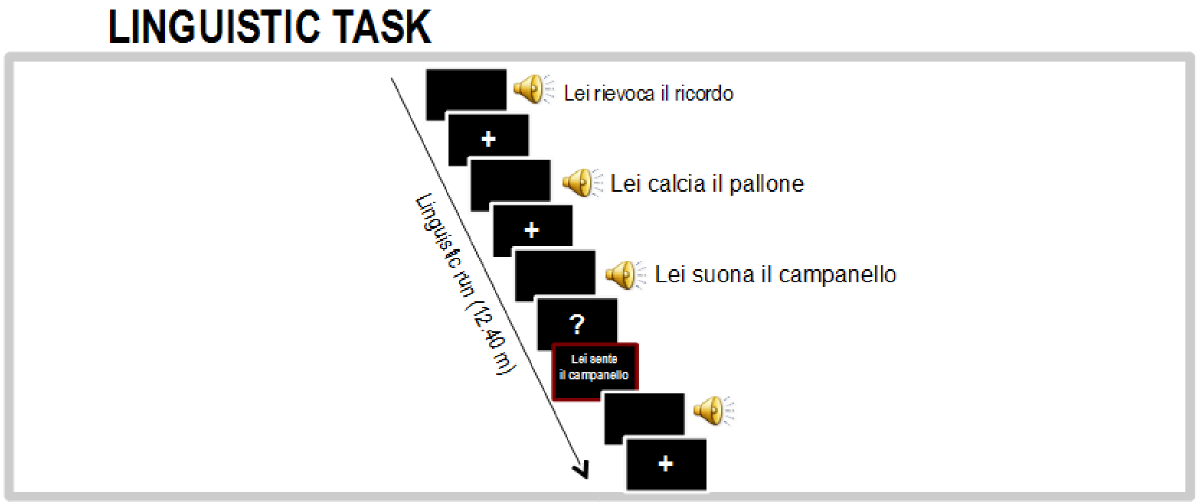


Figure 3. Schematic representation of the experimental procedure during the linguistic task (Study B2).

All subjects were presented with all the 210 sentences. The linguistic stimuli were presented in a jittered event-related experimental design, divided in three separate fMRI runs (11 minutes and 47 seconds long each), each including a minimum of 11 and a maximum of 12 sentences for each experimental condition. Each run was constituted by 89 randomized trials: 68 experimental trials (auditorily presented sentence – fixation cross); 12 catch trials (auditorily presented sentence – question mark – written sentence); 9 null trials (for a detailed description of the experimental and the catch trials see 2.1.4.1). The order and the inter-trial intervals were determined by OPTseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/) to maximize the hemodynamic signal sensitivity of the event-related design. Three inter-stimulus intervals were used, corresponding to 3000 ms, 5000 ms, and 7000 ms (proportions: 4:2:1). Also the intervals within each trials (i.e., the intervals between the auditory presented sentence and the fixation cross or the question mark, and the interval between the question mark and the written sentence) were determined by OPTseq2, ranging between 400 and 600 ms (400, 500, 600; proportions: 1:1:1).

In order to reduce the overall scanning duration for each participant, we did not include the localizer tasks described in Study B1.

3.1.4 Data acquisition

MRI scans were acquired on a 3T Intera Philips body scanner (Philips Medical Systems, Best, NL) using an 8 channels-sense head coil (sense reduction factor = 2). Whole-brain functional images were obtained with a T2*-weighted gradient-echo, echo-planar sequence, using blood-oxygenation-level-dependent contrast. Each functional image comprised 31 contiguous axial slices (3.4 mm thick, 0.6 mm gap), acquired sequentially from bottom to top (field of view = 240 mm x 240 mm, matrix size = 96 x 96). A continuous acquisition technique was used (repetition time: 2000 ms, echo time: 30 ms). Each participant underwent three functional scanning sessions. Each scanning session comprised 240 scans, preceded by 5 dummy scans that were discarded prior to data analysis. A high resolution T1-weighted anatomical scan (3D, spoiled-gradient-recalled sequence, 200 slices, TR = 600 msec, TE = 20 ms, slice thickness = 1 mm, in-plane resolution 1x1x1 mm) was acquired for each subject.

3.1.5 Data Analysis

Statistical parametric mapping (SPM8, Wellcome Department of Imaging Neuroscience, London, UK) was used for the New Segment procedure with structural images, and for processing functional images, including: slice timing, image realignment and unwarping, normalization, smoothing by 6 mm FWHM Gaussian isotropic kernel (for more details see Study1, Section 2.1.6).

General Linear Model statistical analysis was used. We adopted a two-stage random-effects approach to ensure generalizability of the results at the population level.

3.1.5.1 Whole-brain analysis

First-level General Linear Models. The time series of each participant were high-pass filtered at 128 sec and pre-whitened by means of autoregressive model AR(1). No global scaling was performed. Hemodynamic evoked response for all experimental conditions were modelled as canonical hemodynamic response functions. For each participant, we modelled 3 separate sessions with the semantic domain factor (six levels: Ms, Em, Ma, Mo, Ha, Le). Separate regressors modelled catch trials, task instructions, and movement parameters.

The following Student's t-test contrasts were defined: (i) a t-contrast including a weight of -1 for the six regressors modelling the experimental conditions and a weight of 6 for the regressor modelling the catch trials; (ii) a set of t-contrasts specified for the purpose of the random effects group analysis, each including a weight of one for a particular regressor of interest and a weight of zero for all the other regressors, thus resulting in one contrast per experimental condition (6 contrasts) for each participant.

Second-level General Linear Model. (i) In order to control for statistically significant effects associated with performing the linguistic task versus the cross-modal 1-back task, the contrast images, one for each subject, obtained at the single-subject level was entered into a one-sample t-test: "Linguistic task > 1-back task" ([Ms + Em + Ma + Mo + Ha + Le] – Catch), and "1-back task > Linguistic task" (Catch – [|Ms| + |Em| + |Ma| + |Mo| + |Ha| + |Le|]).

(ii) The six contrast images of each subject obtained at the single-subject level were entered into a second-level random effects model using the flexible factorial design. The model included a subject factor (n = 36 participants) and the Semantic domain factor (six levels: Ms, Em, Ma, Mo, Ha, Le). We modelled independence across images from the same subject and equal variances between conditions and subjects as implemented in SPM8. In order to exclude with a greater level of confidence the activations associated with low-level auditory processing of all stimuli, an exclusive explicit mask (Linguistic task – 1-back task) was applied in the analysis. We first identified the group average response for the main effect of concreteness (one sample t-contrast): abstract-related > action-related [(Ms + Em + Ma) – (Mo + Ha + Le)], and action-related > abstract-related [(Mo + Ha + Le) – (Ms + Em + Ma)]. Subsequently, we investigated the effects for each semantic category within the abstract and the action-related domains by specifying the following T contrasts: (i) for abstract-related categories: (A) Ms – (Em + Ma); (B) Em – (Ms + Ma); (C) Ma – (Em + Ms); ii) for action-related semantic categories: (D) Mo – (Ha + Le); (E) Ha – (Mo + Le); (F) Le – (Mo + Ha). Finally, we investigated the specific effects for each semantic category. For abstract-related categories, we exclusively masked the contrast of a given semantic category vs. action-related categories with the corresponding contrasts for all the other abstract-related conditions, i.e.: (A) [Ms – (Mo + Ha + Le)] exclusively masked by [Em – (Mo + Ha + Le)] and [Ma – (Mo + Ha + Le)]; (B) [Em – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Ma – (Mo + Ha + Le)]; (C) [Ma – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Em – (Mo + Ha + Le)]. For action-related categories, we exclusively masked the contrast of a given semantic category vs. abstract-related conditions with the corresponding contrasts for all the other action-related conditions, i.e.: (D) [Mo – (Ms + Em + Ms)] exclusively masked by [Ha – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]; (E) [Ha – (Ms + Em + Ms)] exclusively masked by [Mo – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]; (F) [Le – (Ms + Em + Ms)] exclusively masked by [Mo – (Ms + Em + Ms)], [Ha – (Ms + Em + Ms)].

All reported effects are related to cluster-level statistics $p < 0.05$, FWE type correction for multiple comparisons. We used a Small Volume Correction (SVC) (peak-level statistics $p < 0.05$, FWE corrected) to test for the activation of category-specific brain regions which did not survive a whole-brain correction. Small-volume corrections were performed using literature-based and/or anatomically-constrained functional regions of interest described in the following section (see also Table 4).

3.1.5.2 Region of Interest (ROI) analysis

In addition to whole-brain voxel-by-voxel analysis, a set of ROIs were scrutinized statistically for semantic category effects. ROI definition and analysis were performed by using Marsbar (http://marsbar.sourceforge.net/).

For abstract-related categories, we considered regions of interest predicted by the extant literature. In particular, the ROI analysis was based on peak coordinates taken from the coordinate-based meta-analysis of 19 neuroimaging studies proposed by Wang et al. (2010). Coordinates of foci consistently activated across all studies for abstract versus concrete concepts were used to create 6-mm spherical ROIs (Table 4A). In addition, a ROI analysis of emotion-specific, mathematics-specific and mental state-specific brain regions was performed based on the literature and on the results of the localizer tasks of Study B1, including the emotion localizer task, the mathematics localizer task, and the resting state localizer task. For the emotion-related category, we created two anatomically-constrained functional regions by (i) considering those voxels which were significantly active during the emotion localizer task (threshold of $p < 0.05$, FWE corrected, $k = 5$); (ii) considering only those voxels of the functional ROIs that were located within the cytoarchitectonically defined maximum probability maps of the left amygdala and the left fusiform gyrus. Literature-based regions of interest were created by taking the peak coordinates reported by Moseley et al. (2012) (Table 4B). For the mathematics-related category, we defined a functional ROI in the left middle frontal gyrus and a functional ROI in the left inferior parietal lobule based on the results of the localizer task (Table 4C). For the mental state-related category, we defined functional ROIs in the posterior cingulate cortex, medial superior frontal gyrus, bilateral angular gyrus, left middle frontal gyrus. ROIs definition was based on the results of the resting state localizer analysed by applying the ICA analysis.

Guided by results from the motor localizer scans in Study B1 for tongue, hand and foot movements, a ROI analysis of the primary motor cortex and the premotor cortex was carried out for action-related categories. We applied a combined anatomical and functional approach, since the (anatomical) cytoarchitectonically defined probability maps do not specify the (functional) mouth, hand or leg representations within the motor areas (Tomasino et al., 2010). According to this approach, each ROI of premotor and primary motor cortex was defined in two steps: (i) we considered only those voxels which were significantly more active for performing tongue, hand, or leg movements vs. rest during either the tongue, the hand, or the leg localizer task, respectively (threshold of $p < 0.05$, FWE corrected, $k = 5$). (ii) Subsequently, we considered only those voxels of the functional ROIs that were located within the cytoarchitectonically defined maximum probability maps of the primary motor cortex (Brodmann area 4a and 4p) and the premotor cortex (Brodmann area 6) provided by the SPM Anatomy toolbox. Only for mouth functional ROIs, we also considered the probability map of area 44.

In addition, for mouth-related category, we created literature-based regions of interest by taking the peak coordinates reported by Simmons et al. (2005). ROIs created for each action-related category are listed in Table 4(E-G).

The beta values associated with each experimental condition (i.e., Ms, Em, Ma, Mo, Ha, Le) were extracted from all voxels within the defined anatomically-constrained functional ROIs by using Marsbar. Then, these beta values were entered into repeated measures ANOVA with semantic categories as within-subjects factors. For all the ANOVA results, Huynh–Feldt correction was applied to correct for sphericity violations.

3.1.5.3 MultiVoxel Pattern Analyses (MVPA)

First, we ran a univariate GLM analysis in order to account for the HFR by convolving the stimulus time-series with a canonical hemodynamic function. Schrouff et al. (2013) have suggested this approach for analysing event-related design experiments in which the effect of the HFR can be felt over multiple scans. SPM8 was used for slice timing, image realignment and unwarping, and normalization following the same procedure described in Study1 (section 2.1.6). For the purpose of the multivoxel pattern analyses no spatial smoothing was applied. For each participant, we modelled three separate sessions with the semantic domain factor (six experimental conditions: Em, Ms, Ma, Mo, Ha, Le). Separate regressors modelled catch trials, task instructions, and movement parameters. The resulting beta images of the 6 experimental conditions for each of the 3 runs were merged into a single 4D image containing the 18 beta images of each participant's design matrix, by using the FMRIB's Software Library (FSL). Then, the 4D images obtained for each subject were merged into a single 4D image containing all the 18*36 (subjects) beta images of all participants.

Subsequently, the beta images were entered in the Multivoxel Pattern analyses as implemented in PyMVPA 2.2 software (http://www.pymvpa.org/; Hanke et al., 2009) running under Python 2.7.5 (http://www.python.org/). The 4D image was loaded and masked by the mask image of the 2nd-level group random effects analysis (see above, section 3.1.5.1), to ensure that the voxels included in MVPA contained BOLD signal across all participants. Z score normalization of each voxel for each run was performed in order to control for global variations of hemodynamic response across runs and subjects. No linear detrending was applied since the beta images are not correlated.

The LinearCSVMC as implemented in PyMVPA (where the default C parameter automatically scales C according to the Euclidean norm of the data (Kaplan & Meyer, 2012)) was used as the algorithm for the classification. Two multivariate classification analyses were carried out:

Abstract/action-related classification analysis. We investigated whether the fMRI data contained sufficient information to predict the processing of abstract and action-related sentences. For the purpose of this two-way classification, the beta images for each subject were (i) averaged across the three abstract-related experimental conditions (Ms, Em, Ma) and across the three runs; (ii) averaged across the three action-related experimental conditions (Mo, Ha, Le) and across the three runs. This procedure was adopted in order to reduce the intrasubject variability and improve the signal to noise ratio of the data (Quadflieg et al., 2011; Pereira et al., 2009).

Between-subjects analysis was performed by using a leave-one-subject-out cross validation procedure in which the LinearCSVMC was trained on data from 35 subjects and then tested on data from the 36th subject. The procedure was repeated 36 times, leaving each subject out once. The classification accuracy reported was the simple mean of the 36 classification results (1 for correct; 0 for incorrect) (Akama et al., 2012; Kaplan & Meyer, 2012). This cross-individual analysis was performed in order to verify whether information is represented in the same way across individuals. This procedure is analogous to standard analyses that treat subject as random factor in order to generalize to human population.

The following analyses were performed: (i) a classification using all brain mask voxels; (ii) a sensitivity analysis by applying a recursive feature elimination algorithm (Hanson & Halchenko, 2008). The recursive feature elimination was performed strictly within the training partitions, by iteratively eliminating the less sensitive 50% of the brain mask voxels, and then selecting the reduced brain voxels partition having the greatest sensitivity.

The statistical significance of the results (i.e., whether the classification accuracy was significantly above the chance-level of 50%) was assessed by using Monte Carlo permutation testing. We randomly re-sampled the target attributes within each fold (/subject) 1000 times with replacements, and we then calculated the ensuing distribution of the classification accuracy values for each simulated sample. We compared the empirical result (i.e., the one computed from the original training dataset) to the resulting distribution. The probability of the empirical result under the no signal condition is the fraction of results from the permutation runs that is smaller than the empirical.

Semantic category-specific classification analysis. We investigated whether fMRI data contained sufficient information to predict the processing of each and every semantic category (Ms, Em, Ma, Mo, Ha, Le). For the purpose of this six-way classification, the beta images of each participant were averaged across the three runs according to the experimental conditions (Quadflieg et al., 2011; Pereira et al., 2009). The multi-way classification was performed as implemented in PyMVPA which provides a framework to create meta-classifiers, among others for the LinearCSVMC classifier used here (Hanke et al., 2009).

For the between-subject analysis, we applied a leave-one-subject-out cross validation procedure in which the LinearCSVMC was trained on data from 35 subjects and then tested on data from the 36th subject. The procedure was repeated 36 times, leaving each subject out once. The classification accuracy reported was the simple mean of the 36 classification results (1 for correct; 0 for incorrect) (Akama et al., 2012; Kaplan & Meyer, 2012).

The following analyses were performed: (i) a classification using all brain mask voxels; (ii) a sensitivity analysis by applying a recursive feature elimination algorithm (Hanson & Halchenko, 2009). The recursive feature elimination was performed strictly within the training partitions, by iteratively eliminating the less sensitive 50% of the brain mask voxels, and then selecting the reduced brain voxels partition having the greatest sensitivity.

For the 6-way classification, we verified whether the classification accuracy was significantly above the chance-level of 16.6% (1/6). In the case of the n-way classification (n > 2), the Monte Carlo permutation testing is not appropriate since it does not allow to estimate how likely it is that the employed classifier is capable of discriminating all stimulus categories from each other, or just between subsets of categories. For this reason, we applied the procedure suggested by Olivetti et al. (2012), which developed a Bayesian hypothesis testing in order to evaluate the posterior probability of each possible partitioning of distinguishable subsets of test classes. For example, taking three classes, potential subsets of test classes are [1][2][3]; [1,2][3]; [1,3][2]; [1][2,3]; [1,2,3] (Anderson et al., 2012). As explained in Anderson et al. (2012), each of these subsets would be assigned a posterior probability where a probability in excess of 1/K (where K is the number of hypotheses) would be seen as informative evidence. We performed the Bayesian analysis as implemented in PyMVPA (the algorithm is available in the BayesConfusionHypothesis node). The output of this analysis is the most likely hypothesis to explain this confusion matrix.

3.1.5.4 Parametric analysis

In order to investigate the specific correlates of the abstract- and action-related domains, we also directly took into account the data about psycholinguistic variables collected in the rating study (Ghio et al., 2013). We focused on the following variables: concreteness, context availability, body-part ratings, and category-specific association ratings. In particular, the following regression analyses were performed:

(i) In order to assess whether the degree of concreteness predicted the signal change, a regression analysis including all stimuli (abstract-related and action-related sentences) was carried out. The regressor of interest was the linear component of concreteness (mean values). Duration, frequency, and familiarity were included as confound regressors.

(ii) In order to assess whether the degree of context availability predicted the signal change, a regression analysis including all stimuli (abstract-related and action-related sentences) was carried out. The regressor of interest was the linear component of context availability (mean values). Duration, frequency, and familiarity were included as confound regressors.

(iii) For abstract-related categories, we considered the association ratings indicating how much each abstract sentence was semantically associated with the emotion-, mathematics-, and mental state-related group of sentences (for details see Ghio et al., 2013). Three separate regression analyses were carried out including, respectively, (A) the emotion-, (B) the mathematics-, and (C) the mental state-association ratings as the regressor of interest. In each regression analysis, the association ratings of no interest were model led as confound regressors in order to exclude their effects. For example, in the regression analysis (A), in which the regressor of interest was the linear component of the emotion-association rating, the mathematics- and the mental state-association ratings were modelled as confound regressors. Abstract-related and action-related sentences were modelled separately. Abstract-related sentences were entered as regressors of interest, whereas the regressor for action-related sentences was entered in the analysis as a confound regressors along with the regressors for duration, frequency, and familiarity.

(iv) Body-part ratings indicating how much the action described in each sentence involved the mouth, the hand, and the leg were also considered. Three separate regression analyses were carried out including, respectively, (A) the mouth-, (B) the hand-, and (C) the leg-related rating scores as the regressor of interest. In each regression analysis, the body-part ratings of no interest were modelled as confound regressors in order to exclude their effects. For example, in the regression analysis (A), in which the regressor of interest was the linear component of the mouth rating, the hand and the leg ratings were modelled as confound regressors. All abstract-related and action-related sentences were included in the analysis as regressors of interest. Duration, frequency, and familiarity were entered as confound regressors.

For each regression analysis described above, a two-level analysis was performed by using SPM8. At the first level, we examined the responses of individual subjects by modelling the presentation times of sentences along with the parametric regressor of interest and the confound regressors for each sentence to compute individual SPM maps. Parameter estimates reflecting the height of the HRF for each of these regressors were calculated at each voxel. The resulting first-level images were used to calculate second-level group contrasts using one-sample t-tests identifying the positive and the negative main effects of interest. All reported effects are related to peak-level statistics $p < 0.05$, Family Wise Error (FWE) type correction for multiple comparisons. We used Small Volume Correction ($p < 0.05$, FWE corrected) to test for the activation of category-specific brain regions (see Table 4) which did not survive a whole brain correction.

3.2 Results

3.2.1 Behavioral results

The videos with the participants' eyelid blinking responses during the 1-back task in the linguistic fMRI runs were analysed by two independent judges. Qualitatively, the participants were judged as alert throughout the study. Quantitative analysis was done by measuring the accuracy for each run for each participant. Mean accuracy was 96.14% (run 1: 97.91%; run 2: 93.98%; run 3: 96.52%). The mean accuracy did not differ across runs ($F(2, 105) = 0.85$, $p = 0.43$).

3.2.2 fMRI results

First, we explored the pattern of activation for the linguistic task compared to the cross-modal 1-back task (Table 5A). The results revealed most prominent clusters of activation in superior temporal gyrus, bilaterally, and in the left middle temporal gyrus. Additional activation foci were located in the middle temporal gyrus, in the right inferior frontal gyrus (*pars triangularis*), in the right paracentral lobule, and in the middle cingulate cortex. Activation patterns specific to cross-modal 1-back task against linguistic task are reported in table 5B. Several cortical regions, mostly bilaterally, were activated, including the bilateral lingual gyrus, the right calcarine gyrus, the right middle occipital gyrus and the right fusiform gyrus. Right inferior frontal gyrus (*pars opercularis*), right insula lobe, right cuneus, left precuneus, right anterior cingulate cortex, left superior frontal gyrus and left middle frontal gyrus were also found to be significantly activated. Bilateral activations were found in the supplementary motor area. Moreover, clusters of activation were also observed in subcortical regions including left putamen, prefrontal thalamus, right caudate nucleus, right amygdala, left hippocampus.

The linguistic task was further analysed in order to investigate the specific effects associated to abstract and action-related categories, with a focus on each semantic category within each domain.

3.2.2.1 Abstract and action-related effects

The analysis of the main effect of concreteness (i.e., abstract-related vs. action-related categories) revealed the involvement of specific brain systems for the processing of abstract-related versus action-related sentences (Table 6). Abstract-related sentences elicited left hemispheric activations in the middle temporal gyrus, the inferior frontal gyrus (*pars triangularis*), the temporal pole and the medial temporal pole (Figure 4A). The ANOVA performed on the beta values extracted from literature-based ROIs in the left temporal pole and in the left middle temporal gyrus revealed a significant effect of semantic category ([ROI=Wang_etal_2010 (4)]: $F(5, 175) = 8.91$, $p < 0.001$; [ROI=Wang_etal_2010 (6)]: $F(5, 175) = 6.21$, $p < 0.001$), with abstract-related categories significantly differing from action-related categories (all $p < 0.05$) (for the ROI specification, see Table 4). The ANOVA performed on beta values extracted from the literature-based ROI corresponding to the activation in the left medial temporal pole indicated a significant main effect of semantic category ([ROI=Wang_etal_2010 (5)]: F(5, 175) = 4.36, $p < 0.001$); post-hoc comparisons revealed that only Em and Ms conditions were significantly different from action-related conditions.

Action-related sentences activated a left hemispheric fronto-parieto network including the inferior frontal gyrus (*pars triangularis*) and the inferior parietal lobule. In addition, the right supramarginal gyrus was found to be significantly activated (Table 6B, Figure 4B).

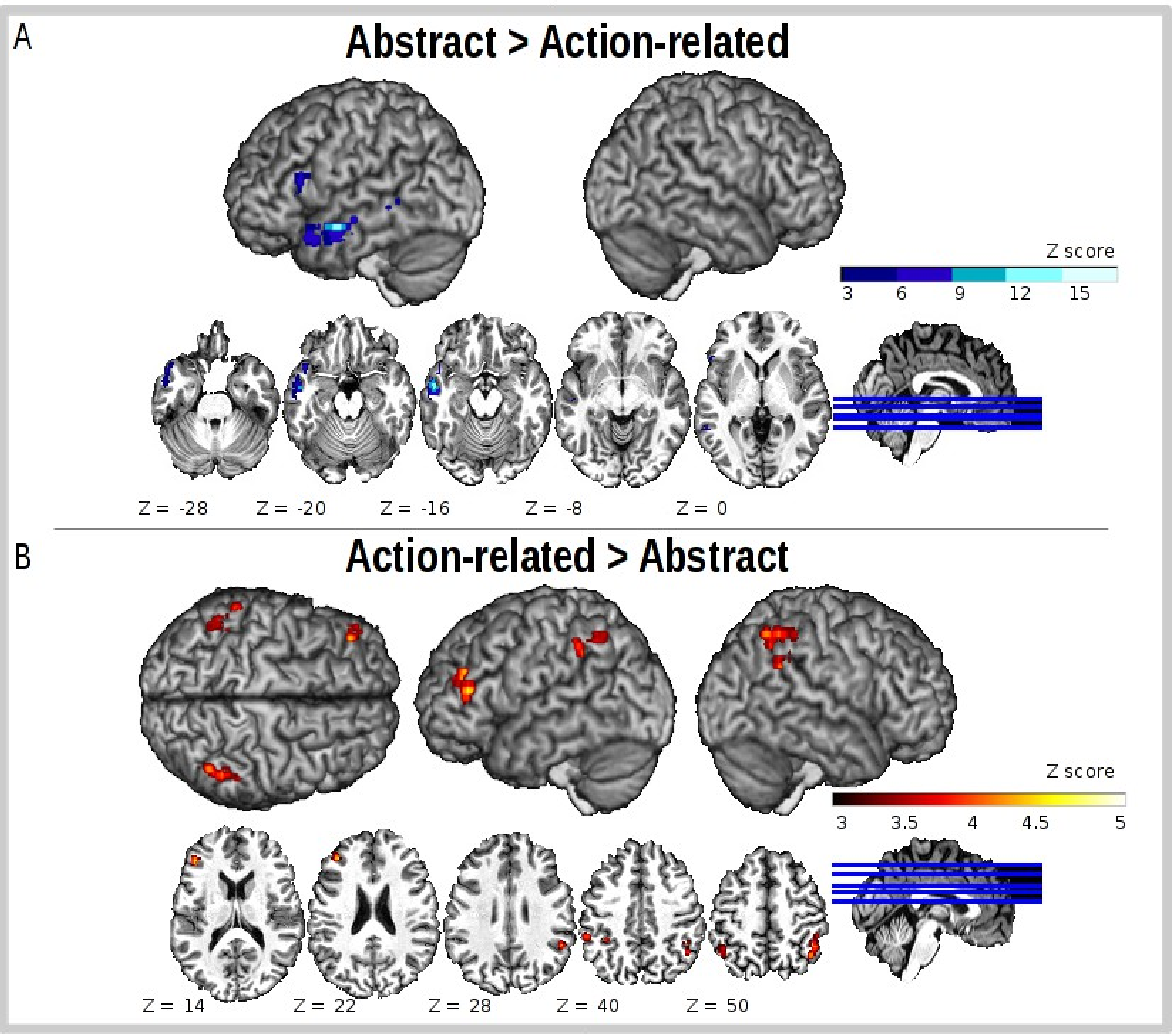


Figure 4. Functional localization of the concreteness effects. Significant activations ($p < 0.05$, FWE corrected for multiple comparisons) are displayed on cortical renderings and on axial (Z coordinate levels in mm) slices of the anatomical image of one of the participants (warped to the MNI coordinate space). **(4A)** Activations specific to the Abstract > Action-related contrast; **(4B)** Activations specific to the Action-related > Abstract contrast.

3.2.2.2 Semantic category-specific effects

For the abstract domain, the semantic category-specific effects were investigated by comparing a given category with the other abstract categories (Table 7 A-C) and by exclusively masking the contrast of a given semantic category vs. action-related categories with the corresponding contrasts for all the other abstract-related conditions (Table 8 A-C):

Ms (Table 7A, 8A) was not associated with any specific activation. No specific activations were found even based on the a priori hypothesis for the default mode network.

Em (Table 7B, Figure 5A) elicited left hemispheric activation in the paracentral lobule. We also found a trend toward significance for the activation in the left amygdala. The ANOVA performed on the beta values extracted from the anatomically-constrained functional ROI comprising the left amygdala revealed a significant effect of the semantic category (F (5, 175) = 3.20, p = 0.01), with Em significantly differing from Ma, Ms, and Le (all $p < 0.05$). In addition, based on the a priori hypothesis of an involvement of the motor areas for emotion-related meanings, we tested several brain regions of interest comprising motor-specific brain regions. Clusters of activation were found in the right precentral gyrus, in the left postcentral gyrus, in the bilateral supplementary motor area. ANOVAs performed on beta values extracted from anatomically-constrained functional motor and premotor cortex ROIs including mouth-specific, hand-specific, and leg-specific areas, showed a trend toward a significantly higher activation for Em compared to the other categories (Table 7B).

The analysis of specific effects (Table 8B) revealed that only left hemispheric motor regions were significantly involved. In addition, this analysis revealed significant activations in the left medial temporal pole and in the right insula lobe.

Ma (Table 7C, 8C) was not associated with any specific activation. No specific activations were found even with respect to the a priori hypothesis for the parietal cortex/intraparietal sulcus and the inferior frontal cortex.

For the concrete domain, the semantic category-specific effects were investigated by comparing a given category with the other action-related categories (Table 7 D-F) and by exclusively masking the contrast of a given semantic category vs. abstract-related categories with the corresponding contrasts for all the other action-related conditions (Table 8 D-F):

Mo (Table 7D, Figure 5B) elicited activations in the left inferior frontal gyrus (pars triangularis), in the bilateral supplementary motor area, in the left cerebellum and in the right temporal pole. Clusters of activation were found in the left precentral gyrus (area 44), in the left precentral gyrus (area 6) and the right inferior frontal gyrus (area 44). The analysis performed on the beta values extracted from the anatomically-constrained functional ROIs comprising the left area 44, the right area 44, and the left area 6 showed a significant difference for Mo vs. Ha and Mo vs. Le in area 6 (all $p < 0.05$), and a trend toward significant difference in area 44 left and right (all $p < 0.1$). In addition, specific activations were found in the left middle orbital gyrus and in the left fusiform gyrus.

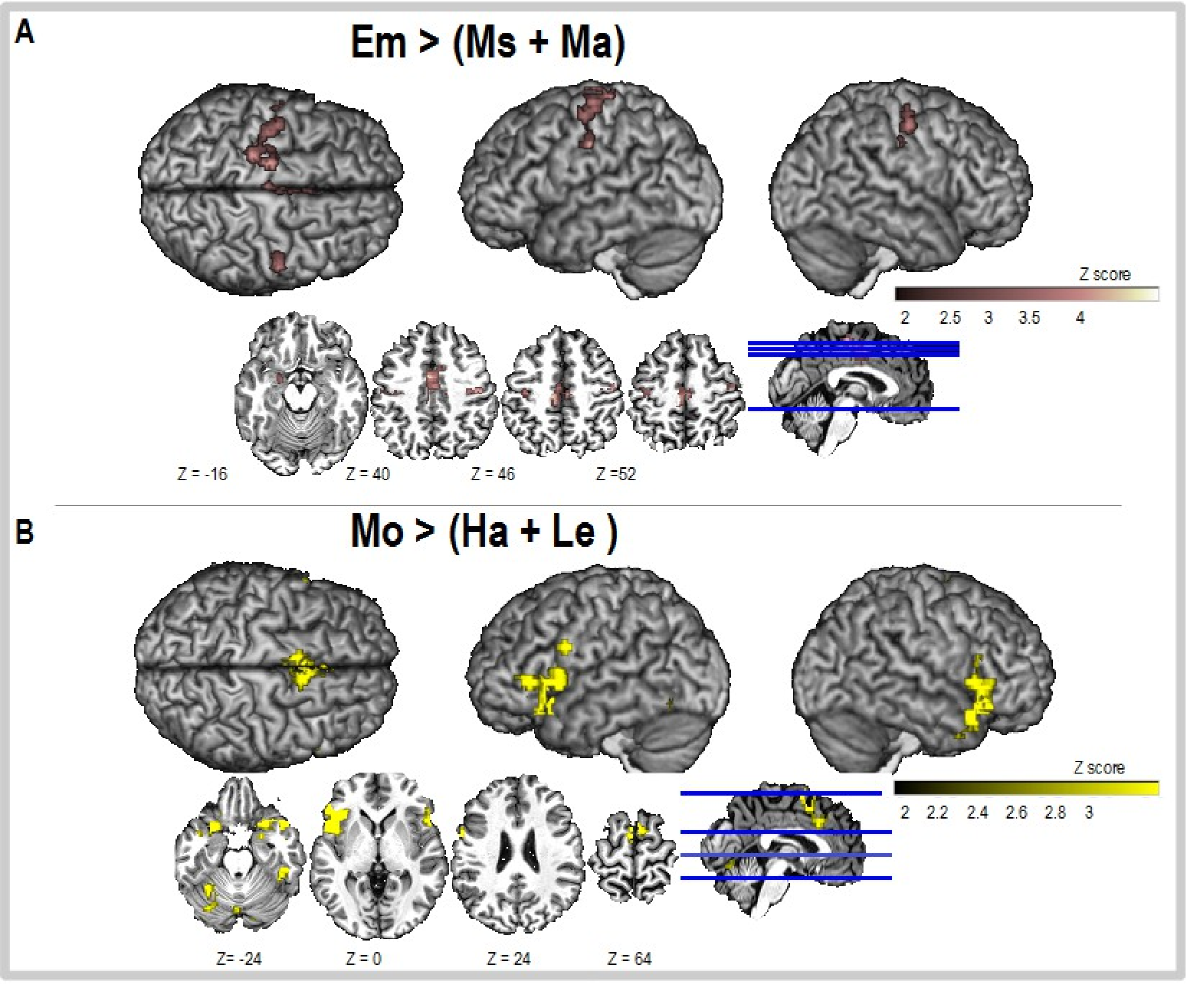


Figure 5. Functional localization of specific-category effects. Significant activations ($P < 0.05$, FWE corrected for multiple comparisons) are displayed on cortical renderings and on axial (Z coordinate levels in mm) slices of the anatomical image of one of the participants (warped to the MNI coordinate space). (**5A**) Activations specific to the Em > (Ms + Ma), (Em = emotion-related meanings; Ms = mental state-related meanings; Ma = Mathematics-related meanings); (**5B**) Activations specific to the Mo > (Ha + Le), (Mo = mouth-related meanings; Ha = hand-related meanings; Le = leg-related meanings).

The ANOVA analysis of the beta values extracted from the literature-based ROIs comprising these two regions indicate a significant effect of the semantic category ([ROI=Simmons et al., 2005 (2)]: $F (5, 175) = 5.18$, $p < 0.001$ and [ROI=Simmons et al., 2005 (5)]: $F = 2.65$, $p< 0.05$). Post-hoc comparisons revealed that Mo was associated with higher activations compared to Ha, Le, Em, Ma and Ms (all $p < 0.5$).

The analysis of the specific effects (Table 8D) mostly confirmed this pattern of activations, revealing significant clusters in the right inferior frontal gyrus, in the right supplementary motor area, in the middle orbital gyrus and in the left fusiform gyrus.

Ha (Table 7E) was not associated with any specific activation. However, the analysis of the specific effect (Table 8E) revealed significant activations in the left inferior parietal lobule.

Le (Table 7F, 8F) was not associated with any specific activation. No specific activations were found even by analysing the beta values extracted from the anatomically constrained functional ROIs comprising the leg-specific region of the primary motor and premotor areas.

3.2.3 Multivoxel pattern analyses results

3.2.3.1 Abstract/action-related classification analysis results

We performed a two-way discrimination among the abstract-related and the action-related sentences over all brain voxels. Between-subjects mean classification accuracy was high 94.4% as compared to a chance level of 50%. Examination of the confusion matrix (in which each cell provides a count of how many samples of the target class were correctly classified into the corresponding class, see Table 9) revealed that both abstract and concrete stimuli were successfully classified (Chi-square = 56.89, $p < 0.001$). The result of the Monte-Carlo permutation test, confirmed that the classification accuracy was significantly higher than chance level ($p < 0.001$) (Figure 6).

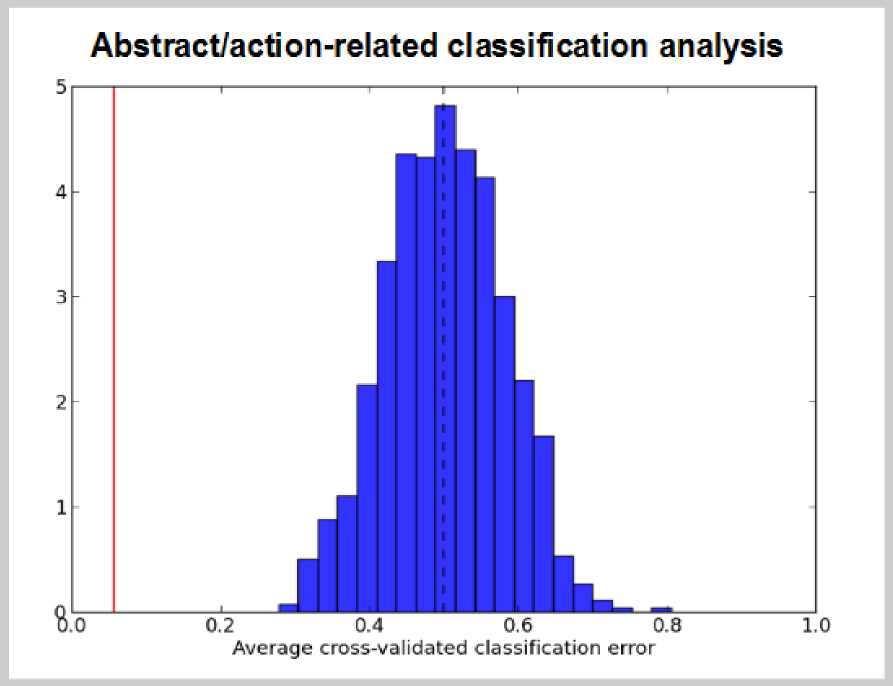


Figure 6. Abstract/action-related sensitivity analysis. Result of the Monte-Carlo permutation test. The target attributes within each fold (/subject) were resampled 1000 times with replacements, and the ensuing distribution of the classification accuracy values for each simulated sample is represented in the figure. The red line indicates the empirical result (i.e., the one computed from the original training dataset). X-axis represents the empirical error.

By applying the sensitivity analysis with recursive feature elimination, the classification accuracy remained at 94.4%. The confusion matrix revealed that the two classes were correctly classified. Furthermore, we also rendered the voxel-by-voxel sensitivity weight of the LinearCSVMC classifier onto a sensitivity map, in order to obtain spatial localization information on the brain regions that mostly contribute to the separation between the two classes. The brain regions identified through this map – filtered to exclude clusters smaller than 10 voxels – are reported in Table 10 and in Figure 7. As illustrated by the Figure 7, we found higher sensitivity weights in a left hemispheric fronto-parieto-temporal network, consistently with the results obtained in the standard whole brain analysis, and, in addition, in the right hemisphere. In particular, as shown in Table 10, a greater sparseness of the effects in both the hemispheres was found relative to the univariate analysis.

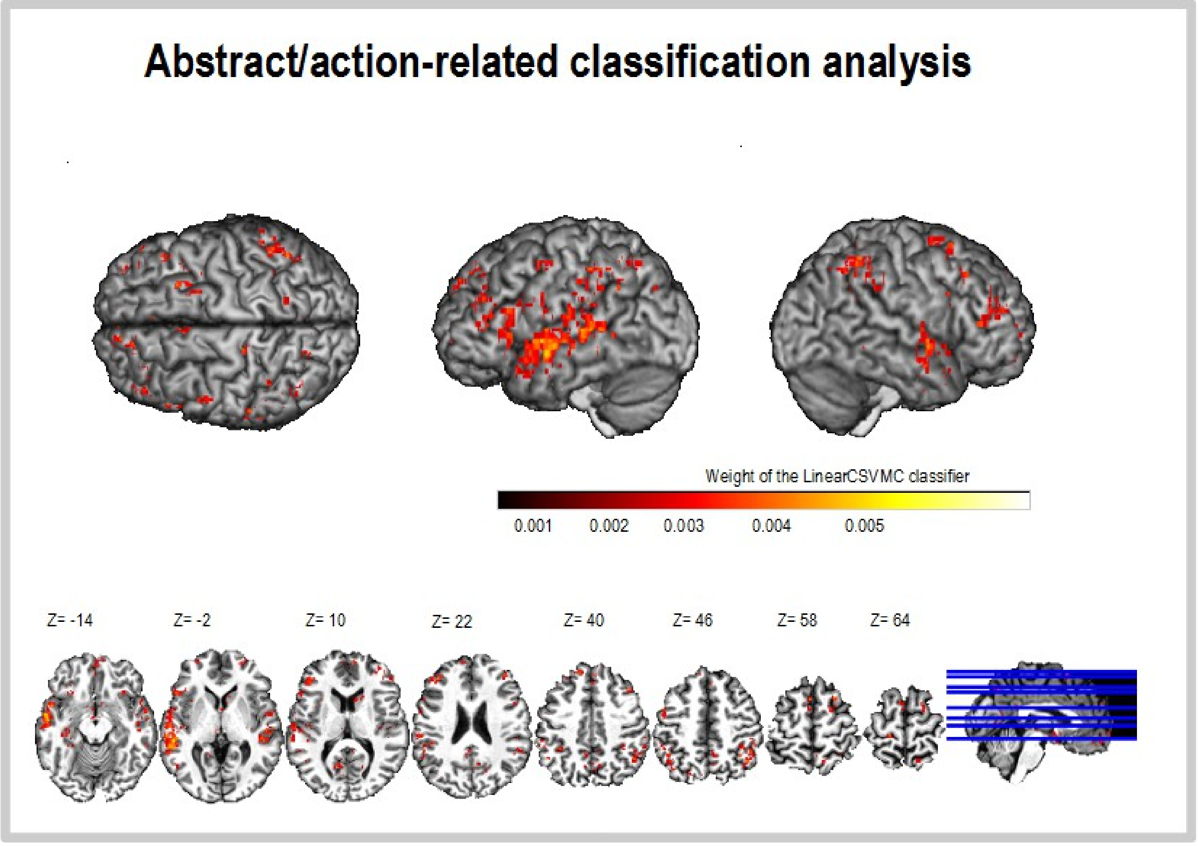


Figure 7. Abstract/action-related sensitivity analysis. LinearCSVMC weights using all brain mask voxels. The anatomical locations reported in the figure refer to clusters of higher sensitivity weight of the LinearCSVMC classifier, representing the degree of contribution or each cluster to the separation between the two classes. The brain regions were filtered to exclude clusters smaller than 10 voxels. Extent threshold k > 10 voxels. Anatomical locations are displayed on cortical renderings and on axial (z coordinate levels in mm) slices of the anatomical image of one of the participants (warped to the MNI coordinate space).

3.2.3.2 Semantic category-specific classification analysis results

The six-way discrimination performed using voxels from the whole brain yielded a mean classification accuracy of 37.0% across subjects, as compared to a chance level of 16.6%. The confusion matrix is reported in Table 11, Figure 8 (Chi-square = 143.67, p < 0.001). By applying the hypothesis test proposed by Olivetti et al. (2012) to the confusion matrix, we found that the most likely partition was [Ms Ma], [Em], [Mo], [Ha], [Le], discriminating between 5 classes.

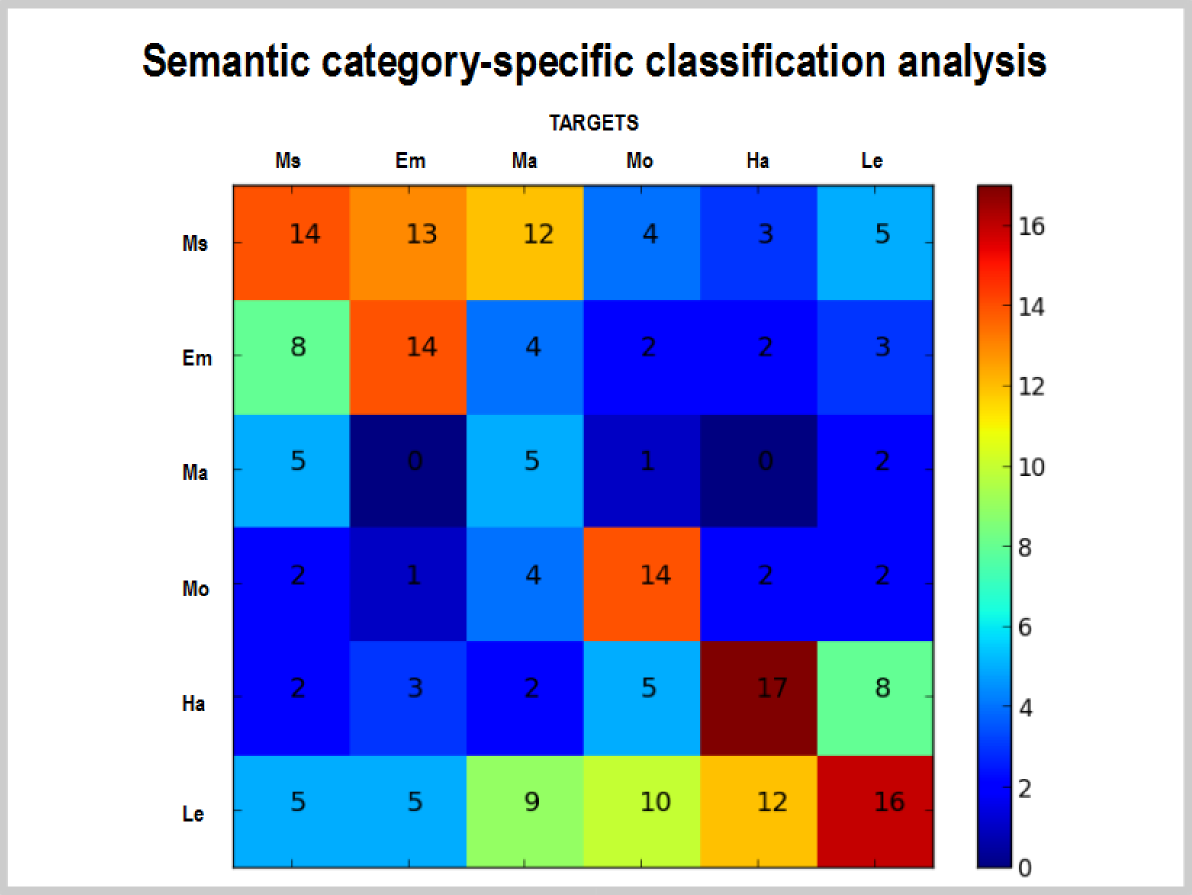


Figure 8. Semantic category-specific classification analysis. The Confusion matrix visualizes the generalization performance of the classifier. Each row of the matrix represents the instances in a predicted class, while each column represents the samples in an actual (target) class. Each cell provides a count of how many samples of the target class were classified into the corresponding class [(Ms) mental state-related meanings; (Em) emotion-related meanings; (Ma) mathematics-related meanings; (Mo) mouth-related meanings; (Ha) hand-related meanings; (Le) leg-related meanings).

The sensitivity analysis with the recursive feature elimination yielded an increase of the mean classification accuracy (41.7%). The Bayesian analysis of the confusion matrix revealed that the most likely partition was [Ms], [Em], [Ma], [Mo], [Ha], [Le]. These results suggested that our data contained enough information to discriminate these 6 semantic categories. Furthermore, we also rendered the voxel-by-voxel sensitivity weight of the LinearCSVMC classifier onto a sensitivity map, in order to obtain spatial localization information on the brain regions that mostly contribute to the separation between the six classes. The brain regions identified through this map – filtered to exclude clusters smaller than 10 voxels – are reported in Table

12 and in Figure 9. We found higher sensitivity weights distributed over both hemispheres and lobes, and subcortical structures.

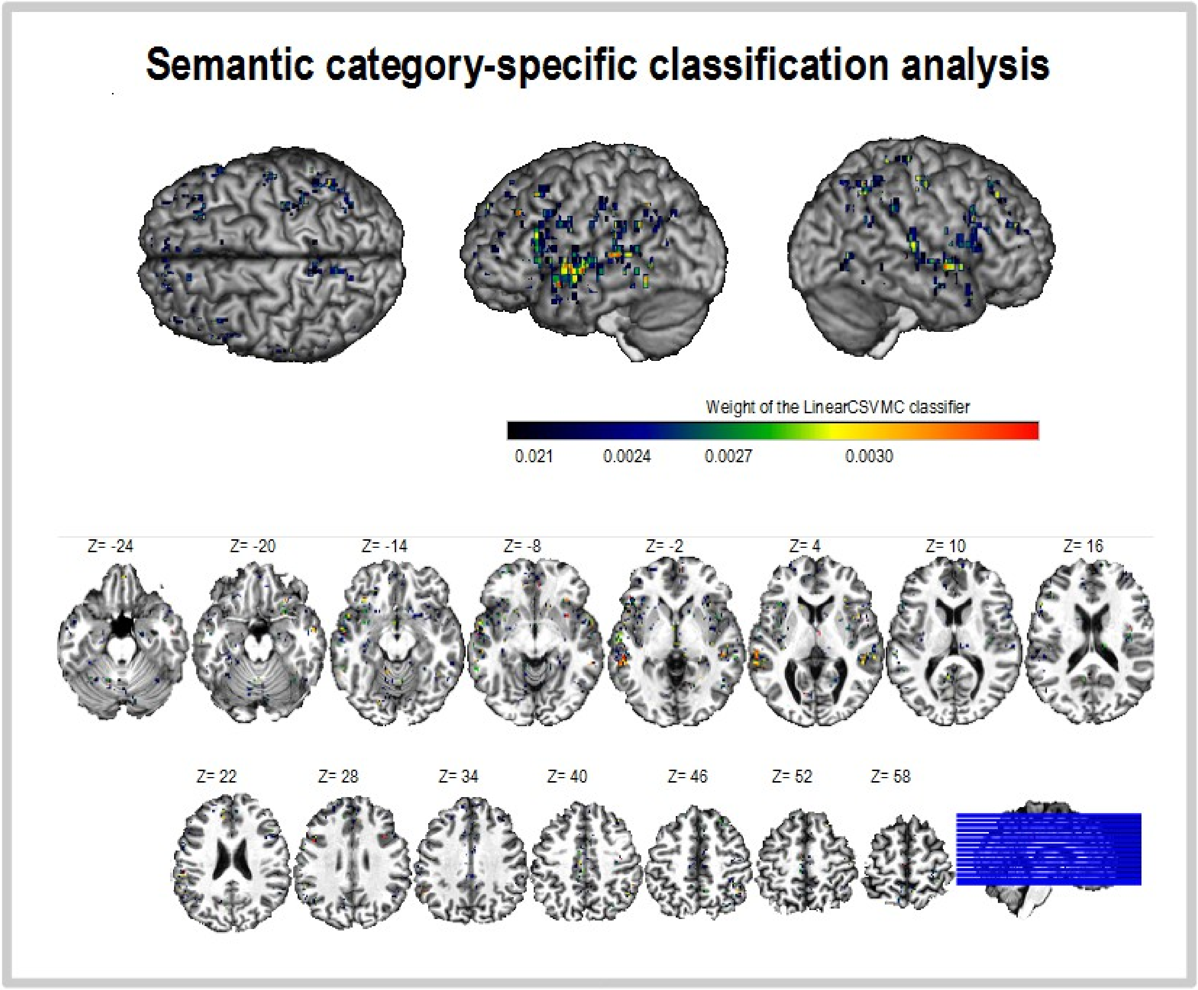


Figure 9. Semantic category-specific sensitivity analysis. LinearCSVMC weights using all brain mask voxels.The anatomical locations reported in the figure refer to clusters of higher sensitivity weight of the LinearCSVMC classifier, representing the degree of contribution or each cluster to the separation between the six classes. The brain regions were filtered to exclude clusters smaller than 10 voxels. Extent threshold k > 10 voxels. Anatomical locations are displayed on cortical renderings and on axial (z coordinate levels in mm) slices of the anatomical image of one of the participants (warped to the MNI coordinate space).

3.2.4 Parametric analysis results

Regression analyses were carried out in order to test the role of several psycholinguistic dimensions in processing sentences.

Parametric effects of concreteness. The positive effect of concreteness was associated with activations in the left parietal cortex and in the left inferior frontal gyrus (*pars triangularis*). Activations in the right inferior parietal lobule and the right supramarginal gyrus were also found (Table 13.1). The negative effect of concreteness was associated with bilateral activations in the temporal pole, the medial temporal pole and the middle temporal gyrus. Clusters of activations were found also in the left inferior frontal gyrus (Table 13.2).

Parametric effects of context availability. The positive effect of context-availability was associated with activations in the inferior parietal lobule and in the angular gyrus, bilaterally. A cluster of activation in the right middle frontal gyrus was also observed (Table 14.1). The negative effect of context-availability was associated with a left hemispheric network of areas, including the middle temporal gyrus, the postcentral gyrus, the precentral gyrus, and the inferior frontal gyrus (pars triangularis) (Table 14.2).

Parametric effects of association ratings. Emotion-association ratings (Table 15A) were associated with activations in the right fusiform gyrus and in the hippocampus. We found activations also in the left anterior and posterior area 4, and trend toward significance in the right middle frontal gyrus.

No significant activations were found for mathematics-association ratings (Table 15B) and mental-state association ratings (Table15C).

Parametric effects of body-part ratings. Mouth-related ratings were associated with a cluster of activation in the left middle orbital gyrus (Table 16A). No significant activations were found for hand-related and leg-related ratings (Table 16B-C).

Finally, we also examined the confounding effects of duration, frequency and familiarity. *Duration* (Table 17A). The positive effect of duration was associated with activations in the bilateral superior temporal gyrus, the right inferior parietal lobule, the right supramarginal gyrus, and the cerebellum. No significant activations were associated with the negative effect of duration. *Frequency* (Table 17B). The positive effect of frequency was not associated with any significant activations. The negative effect was associated with right hemispheric activations in the superior temporal gyrus, the inferior parietal lobule, the angular gyrus, the supramarginal gyrus, the inferior and the middle frontal gyrus. Left-hemispheric activations were found in the superior temporal gyrus and in the middle temporal gyrus. *Familiarity* (Table 17C). The positive effect of familiarity was associated with several areas in the left hemisphere, including the middle occipital gyrus, the middle temporal gyrus, the middle cingulate cortex, the angular gyrus, the fusiform gyrus, the lingual gyrus. Right-hemispheric activations in the middle occipital gyrus, the angular gyrus, the cuneus, the precuneus, the middle temporal gyrus, the parahippocampal gyrus, the fusiform gyrus, the lingual gyrus were also found. The negative effect of familiarity was associated with activations in the left middle temporal gyrus and in the right superior temporal gyrus.

3.3 Discussion

Contemporary cognitive accounts of conceptual knowledge seem to share the view that conceptual representations are essentially distributed, reflecting specific configurations of representational units. This general idea has been addressed by various approaches pursuing different intuitions of what exactly constitutes a distributed conceptual representation, including a pattern of values of a vector (computational linguistic models), a pattern of features (psycholinguistic distributional models), a pattern of multiple brain regions (cognitive neuroscience models) (for reviews, see McRae & Jones, 2013). In the latter approach, as formulated within the grounded framework, the representation and processing of a conceptual representation expressed by a word or a sentence is thought to rely on distributed neural networks. This involves the amodal perisylvian cortex devoted to the processing of linguistic stimuli and extending to modality-specific brain regions, according to the type of experience that is more relevant for that meaning (Barsalou, 2008; 2010; Pezzulo et al., 2013). The dispute as to whether the involvement of modality-specific brain regions support the processing of concrete conceptual representations as well as the processing of abstract conceptual representations is still open (Mahon & Caramazza, 2008; Binder & Desai, 2011; Meteyard et al., 2012; Kiefer & Pulvermueller, 2012).

In the present study, we attempted to address this controversy by developing a comprehensive approach in which both the abstract and the concrete domains were further specified by categories, including mental state-, emotion-, mathematics-related abstract categories, and mouth-, hand-, leg-related concrete categories. Characteristic linguistic and semantic traits for each category were also considered (Ghio et al., 2013). In the present study, moreover, we employed a continuous data acquisition technique in order to avoid the partial reduction of the statistical power due to the sparse data acquisition technique applied in Study B1. Notably, this procedure yielded higher fMRI statistical power as clearly revealed by the direct comparison between Study B1 and Study B2 with respect to the statistically significant effects associated with the control contrasts (i.e., the linguistic task vs. the cross-modal 1-back task, and the cross-modal 1-back task vs. linguistic task), and with the concreteness contrast (see below).

In order to take into account the multidimensional nature of the data space, various techniques of fMRI data analysis were applied, including the standard univariate analysis, the multivariate pattern analysis (MVPA), and the parametric analysis. On the one hand, the univariate and the multivariate analyses provided different views of the functional organization of mental processing. The former is more sensitive to global engagement in ongoing tasks, while the latter is more sensitive to distributed coding of information. For this reason, the MVPA was particularly appropriate to unravel the distributed nature of conceptual representations when considering the fine-grained categorization within the concrete and the abstract domains. On the other hand, the parametric analysis allowed investigating the concreteness/abstractness dimension beyond the a priori categorization of stimuli.

As for the abstract/concrete categorization, by using the univariate analyses we confirmed previous findings of distributed network of brain regions activated by concrete and abstract meanings. By applying the MVPA, we extended previous results showing that it is possible to accurately discriminate the abstract and the concrete semantic domain of stimuli from patterns of brain activity. In addition, not considering the a priori distinction of stimuli into abstract and concrete domains, evidence has been provided indicating that specific abstract and concrete distributed networks are revealed only by using the concreteness rating scores.

As for the fine-grained categorization into semantic-specific categories, the univariate standard analysis mainly failed to reveal activity in modality-specific networks of brain regions. Some significant mean activations in modality-specific areas were found for the processing of emotion-related sentences (within the abstract domain), and for the processing of mouth- and hand-related meanings (within the concrete domain), but not for the remaining categories. Such activations, however, were only partially specific as revealed by the ROI analysis, i.e. the beta values of each specific semantic category were not significantly higher than the beta values of all the other semantic categories. Nevertheless, by applying the MVPA we demonstrated that the neural activity underlying sentence processing contained sufficient information to successfully discriminate between all the six semantic-specific categories that we have tested. We thus provide a first evidence that the semantic category that a participant was processing can be accurately classified only considering how different people’s brains represent the same category, both within the abstract and the concrete domain. This finding suggests that each category might be represented by a specific distributed pattern of brain activity mainly reflected by small relative changes in activation across populations of voxels. Such relative small modulations of neural activity may not be revealed by conventional contrastive analysis, which assume homogeneous stimulus conditions and monotonic brain activations of a fixed scale and local topography (Akama et al., 2012).

In the following paragraphs, we first concentrate on the discussion of abstract-related and concrete-related neural systems as revealed by the univariate standard analysis. Then, we discuss how the shift of the focus both from the detection of activation to the quantification of information and from specific regions to large-scale networks, determined by the application of the MVPA, can shed more light on the representation of specific semantic-categories. Finally, we discuss the impact of the concreteness/abstractness dimension on brain activity beyond any classification.

3.3.1 Detecting and localizing the meaning

3.3.1.1 Abstract vs. Action-related categories

The univariate analysis revealed that the processing of abstract meanings compared to concrete meanings activated a set of left-hemispheric brain regions, including perisylvian regions (the middle temporal gyrus and the inferior frontal gyrus) and extrasylvian regions linked to semantic processing (the temporal pole and the medial temporal pole). As already discussed in Study B1 (section 2.3), such results are highly consistent with previous evidence (Wang et al., 2010). Activations in perisylvian, linguistic brain areas mainly reflect the processing of linguistic information (for a review, see Price, 2010). Activations in the extrasylvian regions can be differently interpreted. According to the seminal work of Patterson et al. (2007), the temporal pole may serve as a semantic hub, integrating modality-specific information into abstract, supramodal representations. However, a growing amount of evidence suggests that the temporal pole receives strong affective inputs from the ventral frontal lobe and amygdala and is better characterized as a modal region for processing emotional and social concepts (Olson et al., 2007; Zahn et al., 2007). Binder and Desai (2011) proposed a neuroanatomical model of semantic processing in which the temporal pole is part of the modality-specific emotion system, while supramodal hubs are located in the lateral and ventral temporal cortex, and in the parietal cortex. Moreover, some evidence showed that also the medial temporal pole plays a role in the emotional processing (see Adolph et al., 2001; Anderson et al., 2012) and in socio-emotional processing (Olson et al., 2007). Intriguingly, and in agreement with these data, the region of interest analysis revealed that the medial temporal lobe was specifically activated by emotion- and mental state-related sentences, but not by mathematics-related ones. This result suggests that the emotion- and mental state-related sentences might share emotional and social features/aspects that, on the the contrary, are not characteristic of the mathematics-related meanings. Interestingly, the rating study (Ghio et al., 2013) revealed that emotion- and mental state-related sentences resulted similar with respect to concreteness, context availability and familiarity, while they differed from mathematics-related sentences. Future research may further specify the properties accessed by different types of abstract-related sentences, mainly focusing on the emotional and social features.

The idea that abstract meanings rely on linguistic information is shared by classic semantic accounts (Paivio, 1971; 2007; Schwanenflugel et al., 1988) and by some (weak) accounts within the grounded framework (Barsalou, 2008; 2010; Simmons et al., 2008; Vigliocco et al., 2009), although it has been differently interpreted. Classic accounts claimed that abstract meanings are only verbally represented (Paivio, 1971; 2007; Schwanenflugel et al., 1988). In this perspective it may be difficult to explain the involvement of the extrasylvian brain regions. In turn, some embodied accounts posit that linguistic information is relevant for the representation of abstract meanings, alongside other types of experiential information, such as affective/emotional and social information (Simmons et al., 2008; Vigliocco et al., 2009; Binder & Desai, 2011). As further suggested by Della Rosa et al. (2010), the greater reliance of abstract meanings on linguistic information compared to concrete ones can also be interpreted as reflecting the "mode of acquisition" of abstract meanings, mainly acquired through linguistic experience (see also Goldberg et al., 2007). The current results seem to be consistent with modal theories proposing that abstract meanings primarily rely on both linguistic information and, to a varying degree, depending on the specific abstract category, on affective/social information (Barsalou 2008, 2010; Vigliocco et al., 2009; Binder & Desai, 2011; Kiefer & Pulvermueller, 2012). In particular, in our study, emotion- and mental state-related meanings seemed to rely on affective and social information more than mathematics meanings. As further detailed below, these results suggest the plausibility of further specifying the involvement of modality-specific brain areas with respect to the specific types of abstract categories.

As for the concrete domain, in accordance with the a priori hypothesis, we found that the processing of concrete sentences, compared to abstract-related ones, was associated with activations within the left fronto-parietal network, which has been shown to be activated also by action execution and observation (for a review, see Pulvermueller & Fadiga, 2010). The activation of the fronto-parietal network for processing action-related compared to abstract-related sentences is consistent with previous findings showing that areas in this circuit are specifically involved in processing action-related meanings expressed either by words, phrases or sentences (Hauk et al., 2004; Tettamanti et al., 2005; 2008, Aziz-Zadeh et al., 2006; Rueschemeyer et al., 2007; Kemmerer et al., 2008; Postle et al., 2008; Raposo et al., 2009; Desai et al., 2010; Carota et al., 2012). At the methodological level, this result is relevant for the comparison of Study B2 with Study B1, in which no significant activations were found in the motor cortices. Compared to Study B1, the statistically significant effects associated with the processing of the concrete sentences relative to the abstract ones speak in favor of an higher fMRI statistical power yielded by the continuous vs. sparse data acquisition procedure.

In the present experiment, the third-person singular form was employed in order to emphasize narrative linguistic comprehension and to test stronger grounded cognition accounts. Indeed, the use of the 1st vs. 3rd person perspective has been considered a useful test for investigating whether the activation of sensory-motor brain regions results from the semantic processing of action words *per se* or from post-semantic simulation strategies adopted by participants, which might be facilitated by the 1st person perspective (Tomasino et al., 2007; Papeo et al., 2011; Gianelli et al., 2011). At present, evidence is scarce, and quite controversial. A recent TMS study showed a motor facilitation during hand action verb processing when they were presented in 1st person, but not when the same verbs were presented in 3rd person (Papeo et al., 2011). Gianelli et al (2011) investigated this issue by employing a version of the Action-sentence Compatibility Effect behavioral paradigm, in which participants were asked to perform a movement compatible or not with the direction embedded in the sentence presented either in 1st or 3rd person. The results showed that the motor effects of language processing were modulated by the perspective of a specific agent with a specific body position in the space. On the contrary, fMRI evidence reported by Tomasino et al. (2007) showed that 1st person utterances (e.g., *I paint*) and 3rd person sentences (e.g., *He jumps*) equally elicited activity in the primary motor cortex. The results of the present study seem to be in line with Tomasino et al. (2007), showing an involvement of the left-hemispheric action representation system even when sentences were presented in 3rd person perspective instead of 1st person perspective. The differences in motor regions activations between the present study and Tomasino et al. (2007) may be due to the use of different stimuli (subject + verb vs. subject + verb + object complement), and to use of both the third and the first person vs. exclusive use of the third person in the present study.

It should be observed that the findings of the present experiment do not exclude the possibility that the use of the third person perspective might have modulated the activation of the motor representation system in a flexible manner that reflects the linguistic context (*She* vs. *I*). In this sense, we observed that the left fronto-parietal circuit was activated to a lesser extent compared to previous studies (Tettamanti et al., 2005; 2008; Pulvermueller & Fadiga, 2010), and an activation in the right supramarginal gyrus was found. Moreover, unlike previous studies, no clear evidence of somatotopic activations in the motor cortices was found (see also the next paragraph). However, given that the impersonal vs. personal perspective was not explicitly manipulated, any interpretations in this direction would be rather speculative. Further research should be done in order to explore how changing perspective in linguistic stimuli could affect their representation at the brain level.

In sum, the results of the present study seem to be compatible with some versions of embodiment according to which abstract meanings mainly rely on linguistic and affective/introspective information, while concrete meanings mainly rely on sensory and motor information (Barsalou, 2008; Simmons et al., 2008; Vigliocco et al., 2009). On the contrary, strong versions of the embodiment claiming that abstract meanings are exclusively represented in sensory-motor regions are not consistent with the present results.

3.3.1.2 Semantic-domain specific categories

As anticipated previously, the univariate standard analysis mainly failed to demonstrate modality-specific brain activations with respect to the fine-grained categorization of both the abstract and the concrete domains. Some evidence was found only for emotion-related abstract meanings, and for mouth- and hand-related concrete meanings, as briefly discussed in the following.

Emotion-related category. Some theoretical approaches within the grounded cognition framework generally posit that abstract meanings, including emotion-related ones, rely on affective/introspective, emotional, and social information (Barsalou, 2008; 2010; Meteyard et al., 2012; Kiefer & Pulvermueller, 2012). As we have discussed above, the processing of such information mainly activates emotion/social modality-specific brain regions, such as the temporal pole (Binder & Desai, 2011). In the present study we proposed to further specify the hypothesis concerning the representation of abstract meanings by considering different categories, among which the emotion-related category. In particular, for emotion-related related meanings we hypothesized the activation of the brain network involved in the actual processing of emotions, including several brain regions, such as the amygdala, the insula, and the cingulate cortex. Indeed, the results point to this direction. Listening to emotion-related sentences compared to mathematics- and mental state-related sentences activated the right paracentral lobule, a region known to be part of the somatosensory cortices (Mayka et al., 2006). The functional role of the somatosensory cortices in processing emotional expression has received little attention so far. Some neuropsychological and neuroimaging evidence showed that recognizing emotions from facial expressions requires right somatosensory-related cortices (Adolphs, 2000; Adolphs et al., 2002). These data were interpreted as suggesting that the recognition of another individual's emotional state is mediated by internally generated somatosensory representations that simulate how the other individual would feel when displaying a certain facial expression. More recently, fMRI studies on subjects with various deficits in the processing of emotions (such as, for example, the alexithymia, which refers to a cluster of deficits in the recognition, differentiation, and verbalization of emotions), showed that somatosensory cortices, including the paracentral lobule, are critically involved in the processing of emotional mimicry and simulation somatosensory (i.e., generation of emotional reactions and interoceptive processing) (Suslow et al., 2010; Reker et al., 2010). However, given the paucity of information on the role of the paracentral lobule in emotion processing, the interpretation of the present finding with respect to the hypothesis of a specific involvement of emotion brain areas in processing emotion-related sentences should be considered with caution.

Some hints in favor of this interpretation seem to come from further analyses we did to explore the emotion network based on the a priori hypothesis. Processing emotion-related meanings compared to mental state- and mathematics-related meanings significantly activated emotion-specific brain regions, including the right insula (for a review, Chang et al., 2012), and the left medial temporal pole (for a review, Olson et al., 2007). A trend toward significant activation was also found in the left amygdala, a region known to play a major role in emotion processing (for a recent review, Pessoa & Adolphs, 2010). Previous evidence showed that such brain regions are involved not only in the actual experience of emotions, but also when emotions are expressed through language (for recent reviews, see Citron, 2012; Kotz & Paulmann, 2011). As illustrated by

Citron (2012), however, most previous studies used the term “emotion word” to refer to any word denoting a specific emotion (e.g., *sadness*) as well as to any other word characterized by a more general emotional connotation (e.g., *flower*, *war*). In turn, we followed the approach of Moseley et al. (2012), in which abstract emotion verbs were used as stimuli (see also the Introduction). Consistently with Moseley et al. (2012), the current results, though only partially specific, seem to suggest an involvement of emotion brain regions in processing emotion-related sentences, carefully controlled for the abstractness dimension (Ghio et al., 2013). However, it should be noted that other well-known emotion areas were not found to be activated in the present study, including the anterior cingulate cortex, the orbitofrontal cortex, the medial prefrontal cortex, and subcortical regions (Citron, 2012). It has been shown that such regions are differently modulated by positive or negative words, and by the valence and arousal of items (Citron, 2012). The use of both positively and negatively valenced sentences in the present study may have obscured some positive-specific or negative-specific patterns of brain activation. Furthermore, given that neither the arousal nor the valence were manipulated in the present study (see also Moseley et al., 2012; Vigliocco et al., *in press,* for a similar approach), we cannot exclude effects due to differences in such affective dimensions. The role of these factors on processing abstract emotion meanings might be the matter of further investigations.

Motor information was also predicted to contribute in representing emotion-related meanings (Moseley et al., 2012; Wilson-Mendenhall et al., 2011). Relying on literature-based region of interest analysis (Moseley et al., 2012), we found significant emotion-related activations in some portions of the somato-sensory cortices specific for the mouth and the hand actions. This finding seems also to be consistent with the body-part ratings reported in our previous study, indicating an involvement of the motor information in the representation of emotion-related meanings (Ghio et al., 2013- Study A).

In sum, some hints were provided as for the involvement of modality-specific brain regions in processing abstract emotion-related meanings. This evidence seems to be in accordance with some weak version of embodiment, while it is in contrast with strong versions of embodiment (Meteyard et al., 2012; Kiefer & Pulvermueller, 2012). Nevertheless, such interpretation of the results must be taken with a measure of caution since the activations in both emotion and motor brain regions were only partially specific (see also Tables 7, 8, and Figure 5A).

Mouth- and hand-related categories. Category-specific activations were found for mouth-related meanings in the bilateral inferior frontal gyrus and in the bilateral supplementary motor areas. Such regions were identified in the mouth-localizer task as specific for mouth movements (see also Tomasino et al., 2010; Carota et al., 2012). It should be noted, however, that the finding of activations in the right hemisphere was not consistent with some previous findings showing left-hemispheric activations only (Tettamanti et al., 2005; 2008).

We observed that most mouth-related sentences included food words as the complement object (e.g., *She devours the biscuit*, *She bites the sandwich*). For this reason, further analyses were performed in brain regions known to be involved in processing food meanings/concepts (Simmons et al., 2005). Specifically, the region of interest analysis based on the brain regions found in Simmons et al. (2005) revealed that the left middle orbital gyrus and the left fusiform gyrus were significantly activated for the processing of mouth related meanings (see also Carota et al., 2012; Barros-Loscertales et al., 2011). The left middle orbital gyrus is a gustatory processing area, active during the tasting of actual food. Several studies demonstrated that this area is specifically involved in representing the reward values of tastes (Gottfried et al., 2003; Kringelbach, 2005). The activation of the left fusiform gyrus may reflect the visual properties of foods (Simmons et al., 2005; Carota et al., 2012). We argued that the involvement of such regions in processing mouth-related sentences is particularly interesting with respect to the idea that conceptual representations are flexibly retrieved depending on several contextual factors (Barsalou 2008; 2010; Hoenig et al., 2008). In this case, the linguistic context – represented by the object complement – seems to dynamically modulate the representation of the action meaning.

The processing of hand-related meanings activated the left inferior parietal lobule. This finding is in good accordance with Tettamanti et al. (2005), actually, at closely similar peak coordinates (present study: x = -44; y = -46; z = 44; Tettamanti et al. (2005): x = -46; y = -38; z = 44). As discussed in Tettamanti et al. (2005), it has been demonstrated that the inferior parietal regions are engaged in the action observation only when actions are performed towards an object or a goal (see also Buccino et al., 2001). This agrees with the use of hand-related sentences describing actions involving object manipulation (e.g., *She embroiders the handkerchief*). Again, this finding may be interpreted as a case of flexible representation of the action meaning, modulated by the presence of the object complement. Finally, it should be observed that no brain activations in motor cortices were found. There is some fMRI evidence that hand-related meanings sometimes are not clearly associated with motor activity in the left hand/arm region, in particular when also mouth- and leg-related meanings are included in the study (Hauk & Pulvermueller, 2004; Pulvermueller, 2001; Shtyrov et al., 2004; Postle et al., 2008; Pulvermueller et al., 2012). Pulvermueller et al. (2012) suggested that weak activations in the motor system associated to arm-related items might be due to the fact that arm/hand representations are close to both the face/mouth and the leg. Accordingly, overspilling activity may partly keep these networks active during the experiments, thus leading to relatively stronger adaptation in arm-motor systems than elsewhere (Pulvermueller et al., 2012). Nevertheless, the involvement of the motor cortex in processing hand-related meanings has been clearly documented in previous research (for a review, see Pulvermueller & Fadiga, 2010). We thus believe that, in the present study, the absence of significant neural activity within the motor cortices should be interpreted with caution, also in the light of the null results obtained for the processing of the leg-related meanings.

In sum, evidence found for mouth- and hand-related meanings was only partially consistent with previous studies and with the a priori predictions. The specific results obtained seem to point to a flexible representation of the action meanings as modulated by linguistic context, with activations that mainly reflect the meaning of the object complement.

Mental state-, mathematics-, leg-related categories. Results showed no specific activations associated with the processing of mental state-, mathematics, and leg-related categories. This is a null result, for which strong interpretations are not possible (Henson, 2006). On the one hand, it indicates that the global mean activity in predicted modality-specific brain areas for each of these categories was not sufficiently different from the activity elicited by other categories to be detectable in this paradigm. On the other hand, this finding cannot in itself rule out the possibility that there are specific distributed patterns of brain activity subserving the semantic processing of these categories. In order to explore this possibility, we applied the MVPA. This approach has been shown to be suitable to examine distributed coding of task-relevant information, even in the absence of mean activation (Mur et al., 2009). The results are discussed in the next paragraphs.

3.3.2 Decoding the meaning

3.3.2.1 Decoding abstract and action-related categories

In the present study, we provided novel data about the processing of concrete and abstract meanings by demonstrating the possibility of decoding the semantic domain (abstract vs. concrete) from fMRI spatial patterns across subjects by means of MVPA. In this procedure, a classifier learned to distinguish the neural patterns evoked by abstract and concrete meanings based on the data from a sub-group of the subjects and was then tested on data from an individual that was not part of that sub-group (Kaplan & Meyer, 2012). We found the predicted performance to be significantly above chance when using voxels from the whole brain. In other words, we demonstrated that the category of stimuli (concrete vs. abstract) that the target participant was listening to can be accurately identified based only on the brain activations measured when the other participants were processing the same category of items (Shinkareva et al., 2008; 2011). Other studies, employed this procedure to demonstrate the possibility of discriminating between different type of visual object categories (e.g., animals vs. tools), auditory object categories (e.g., cat vs. guitar), natural scene categories (e.g., beach vs. buildings) (for reviews, see Vindiola & Wolmetz, 2011; Kriegeskorte, 2011; Kaplan & Meyer, 2012). Some studies also succeeded in identifying the cognitive states associated with different object categories by using single concrete words (e.g., *leg*, *chair*, *car*) (Mitchell et al., 2004; Just et al., 2010). In the present study, we extend those findings by using abstract and concrete meanings expressed by sentences.

Furthermore, the sensitivity analysis allowed the localization of brain regions that were involved in successful decoding across subjects. The data revealed better than chance classification of concrete and abstract sentences primarily on the basis of the pattern of activity in a left fronto-parieto-temporal network. This finding is highly consistent with the results obtained in the univariate whole-brain analysis. However, we also observed a greater sparseness of the effects for abstract and concrete meanings relative to the univariate analysis. Specifically, we found higher sensitivity weights also in several regions of the right hemisphere and in the subcortical structures. Crucially, the sensitivity analysis only reveals brain regions that contain enough information for discriminating between abstract and concrete meanings. In this sense, it is not possible to specify abstract- or concrete-related brain networks. By using this technique, we nevertheless demonstrated the commonality of the neural bases of the conceptual representations of abstract and concrete meanings across participants (Shinkareva et al., 2011).

3.3.2.2 Decoding semantic-specific categories

Even more interesting for the present research, we showed that fMRI data contain sufficient information to separate all the semantic-specific categories that we have tested. In other words, the a priori fine-grained categorization within both the abstract and the concrete domain validated by means of behavioral data (Ghio et al., 2013) seems to be supported also by brain data. By applying the same across-subject procedure sketched above, we demonstrated that content-specific activity patterns were consistent across subjects. Specifically, we showed that a classifier trained on the data of a subgroup of subjects successfully discriminated between semantic-specific categories when tested on novel data of a different subject. Previous MVPA research provided scarce evidence on the representation of action-related knowledge. Among other object categories, a limited number of studies included either the object category of “tools” (Reddy et al., 2010) or a specific tool category, such as “scissors” (Haxby et al., 2001; O’Toole et al., 2005). Classification analyses were mainly conducted within-participant in specific brain regions (such as the ventral temporal cortex), showing the possibility of accurately classify tool concepts among others. As for abstract-related meanings, there is a lack of evidence. Anderson et al. (2012) reported only preliminary results indicating that WordNet style taxonomic categories for abstract concepts (e.g., social role, event, communication, attribute) can be distinguished, to a certain extent, from brain activity through the MVPA. The results of the present study extend previous findings suggesting that (i) neural activity patterns can be identified throughout the brain specifically reflecting the content of abstract-related meanings and action-related meanings; (ii) these content-specific representations share similarities across subjects.

Furthermore, we showed that information that was common across individuals was not uniformly distributed throughout the brain. The voxel weights from the whole-brain analysis reveal an information-containing neural pattern of activations distributed over both hemispheres and lobes. Such sparseness of the neural representation of meanings is not univocally interpretable. It possibly reflects the distributed representation across cortical areas that are specialized for various types of category-specific information, according to the embodied framework (Shinkareva et al., 2011). In particular, abstract meanings are considered relational structures resulting from the integration of many different concepts in a situated conceptualization (Barsalou, 2010; Wilson-Mendenhall et al., 2011; 2013). Crucially, however, the MVPA permits to individuate brain regions in which the neural activity contains enough information for discriminating across categories, without ascribing a set of brain regions to the processing of a specific semantic category. In future, more fine-grained analyses of our data, we may restrict the MVPA to specific brain regions in order to test more specific hypothesis. For the time being, a discussion in terms of modality-specific networks subserving each semantic category is not appropriate. Furthermore, depending on the sentence type, different mechanisms may have evoked spatial representations during conceptual processing. This question deserves future empirical attention.

In conclusion, the results of the multivariate analysis seem to suggest that the representation of fine-grained meanings expressed by different types of sentences may be coded in the brain through specific patterns of activity rather than through overall activations of brain regions. The discussion on the distribute nature of conceptual representation may be enriched by also considering the degree to which neural signals relevant to semantic processing are coded in distributed patterns.

3.3.3 Beyond the categorization

Disregarding the a priori categorization of items into the abstract and the concrete domain, we demonstrated a clear effect of the concreteness/abstractness dimension on brain activations. When assessing concreteness-dependent modulations for all sentences by using the concreteness rating scores collected in our previous rating study, we found a pattern of activations highly consistent with the results of the univariate analysis. Specifically, sentences that are judged as more abstract were associated with activations in the left perisylvian regions and in extrasylvian regions, including the temporal pole and the medial temporal pole. In turn, sentences that are judged more concrete were associated with activations in the left fronto-parietal network for action representation, and in the right supramarginal gyrus. These findings are particularly interesting since they are obtained only by considering the concreteness/abstractness judgments expressed by language users, regardless of experimental manipulations.

In addition, we also controlled the impact of various psycholinguistic dimensions on brain activations. In particular, in the rating study we showed that concreteness was highly correlated with context availability and familiarity (Ghio et al., 2013). The present analysis helps disentangling these dimensions. We observed that such psycholinguistic variables activated multiple distinct brain areas, only marginally overlapping with the networks associated with the concreteness dimension. First, sentences higher in context availability were specifically associated with the inferior parietal lobule and the angular gyrus, bilaterally. Several previous studies reported activations in these brain regions for the processing of concrete meanings compared to abstract ones (for a review, see Wang et al., 2010). The current results suggest that the involvement of the inferior parietal lobule and the angular gyrus in the semantic processing of meanings might be primarily related to the context availability dimension. Second, highly familiar words correlated with activations in a bilateral occipito-parietal-temporal network, including the middle occipital gyrus, the angular gyrus, the cuneus, the precuneus, the fusiform gyrus, the parahippocampal gyrus. Previous functional imaging studies showed an involvement of a similar network in episodic retrieval of personally familiar places and objects compared to unfamiliar ones (Sugiura et al., 2005). In sum, the effect of the concreteness/abstractness of items does not seem to be reducible to differences in context availability and familiarity. At the same time, these dimensions have a clear impact on the brain activations in accordance with previous evidence (Binder et al., 2003; Carreiras et al., 2006; Fiebach et al., 2007; Hauk et al., 2008).

4. Conclusions

Altogether, the present study confirmed that the abstract/concrete distinction is, at a broad level, a basic semantic organizational principle of conceptual knowledge. Indeed, distributed networks of brain regions specific for the processing of both abstract and concrete meanings clearly result either considering the a priori distinction of stimuli into abstract and concrete semantic domains or not. The findings are consistent with the view that abstract meanings mainly rely on linguistic and affective/emotion information and are represented in modal emotional and social areas, to a varying degree, depending on the specific abstract category, while concrete meanings mainly rely on sensory-motor information and are represented in modality-specific sensory-motor systems. At the finer level, only some hints about the involvement of modality-specific systems in processing specific semantic categories were provided. However, the present study found clear evidence that the application of multivariate techniques to fMRI can reveal the existence of semantically organized structure in the pattern of fMRI-measured brain activation during the semantic processing of different categories both within the concrete and the abstract domain. It can also reveal a degree of commonality across subjects with respect to this semantic organization. This finding contributes to the current debate in the field, by suggesting that fine-grained categories might be represented by specific distributed patterns of brain activity mainly reflected by small relative changes in activation across populations of voxels, rather than by global activation in specific brain areas. The degree to which neural signals relevant to semantic processing are coded in either localized increases of activity or distributed patterns is still debated, and should be the matter of further investigations.

5. Tables

Table 1.
Study B1. Specific effects for the (Linguistic task – cross-modal 1-back task, **A.**) and for the (cross-modal 1-back task – Linguistic task, **B.**) contrasts. P < 0.05, FWE corrected at the whole brain level, Extent threshold k > 10 voxels.
[§] According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability[§])	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
A. Linguistic task > Cross-modal n-back task						
L Heschl's gyrus (TE 1.2 40%)	623	Inf	0.000	-54	-12	8
L superior temporal gyrus (TE 1.0 90%)	“	7.57	0.000	-48	-20	8
L superior temporal gyrus (TE 3 70%)	“	7.41	0.000	-64	-24	12
R superior temporal gyrus (TE 1.0 50%)	652	7.69	0.000	60	-10	4
R superior temporal gyrus (TE 3 10%)	“	7.53	0.000	64	-22	8
R Heschl's gyrus (TE 1.0 90%)	“	7.35	0.000	50	-18	8
L Cerebellum (Lobule VIIa Crus II Hem 70%)	101	6.85	0.000	-36	-60	-44
L Cerebellum (Lobule VIIb Hem 18%)	“	5.86	0.000	-34	-44	-44
R inferior frontal gyrus (pars orbitalis) (area 45 10%)	37	5.96	0.000	48	28	-12
R inferior frontal gyrus (pars orbitalis)	“	5.58	0.001	40	24	-16
L superior orbital frontal gyrus	14	5.71	0.000	-18	44	-16
L Cerebellum (Lobule VIIa Crus I Hem 74%)	47	5.60	0.000	-34	-68	-24
L Cerebellum (Lobule VI Hem 92%)	“	5.40	0.002	-22	-68	-28
L angular gyrus (IPC (PGp) 70%)	16	5.50	0.001	-44	-68	28
L rectal gyrus	“	5.14	0.007	-8	20	-20
L superior orbital gyrus	“	4.89	0.023	-16	22	-20
R angular gyrus (IPC (PGa) 50%)	13	5.20	0.005	52	-56	32
R angular gyrus (IPC (PGa) 40%)	15	5.13	0.000	-46	-58	32
B. Cross-modal n-back task > Linguistic task						
L Pallidum	1022	11.31	0.000	-18	6	-4
(Thalamus 6%)	141	10.85	0.000	-30	-24	4
L insula lobe (area 44 30%)	“	4.90	0.022	-40	16	4
R fusiform gyrus (hOC4v (V4) 20%)	407	7.35	0.000	34	-66	-12
R inferior occipital gyrus (hOC3v (V3v) 30%)	“	7.30	0.000	34	-92	-4
R lingual gyrus (area 18 60%)	“	7.08	0.000	24	-88	-12
L lingual gyrus (hOC4v (V4) 20%)	550	8.88	0.000	-20	-62	-8
L inferior occipital gyrus (hOC4v (V4) 30%)	“	8.45	0.000	-36	-88	-8
L inferior occipital gyrus (hOC3v (V3) 50%)	“	8.30	0.000	-22	-96	-12
R calcarine gyrus (area 17 40%)	174	8.71	0.000	16	-66	12
R calcarine gyrus (area 17 50%)	“	6.96	0.000	24	-60	8
R lingual gyrus (area 18 20%)	“	6.17	0.004	20	-54	-4
R insula lobe	153	6.58	0.000	32	24	4
L thalamus (Thalamus 33%)	41	6.12	0.000	-4	-24	0
L precentral gyrus (area 44 40%)	26	5.56	0.000	-58	10	28
R thalamus (Thalamus 50%)	13	5.49	0.000	6	-12	8
L paraHippocampal Gyrus (Hipp (SUB) 20%)	4	4.95	0.019	-20	-42	-4

Table 2.
Study B1. Specific effects for the (Abstract-related – Action-related) and opposite (Action-related – Abstract-related) contrasts. P < 0.05, FWE corrected at the whole brain level. Coordinates belonging to the same activation cluster are grouped by “.
[§] According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability[§])	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
A. Abstract-related > Action-related						
L temporal pole	1117	6.91	0.000	-52	8	-16
L superior temporal gyrus	“	6.11	0.000	-62	-24	-4
L middle temporal gyrus	“	5.81	0.000	-66	-32	0
R superior temporal gyrus	329	6.05	0.000	58	0	-8
R temporal pole	“	4.96	0.017	50	16	-20
B. Action-related > Abstract-related						
No specific activations						

Table 3.
Study B1. Specific effects for the (Affirmative – Negative) and opposite (Negative – Affirmative) contrasts. P < 0.05, FWE corrected at the whole brain level.
[§] According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability[§])	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
A. Affirmative > Negative						
No specific activations						
B. Negative > Affirmative						
No specific activations						

Table 4.
Study B1. Literature-based and functional regions of interest (ROIs) used in the analysis. For ROIs created by applying a combined functional and anatomical approach we indicate by the symbol "∩" the intersection between functionally-defined and anatomically-defined regions (Ms = mental state, Em = emotion; Ma = Mathematics; Mo = Mouth; Ha = Hand; Le =Leg; L = left; R = right).

Literature-based ROIs	Functional ROIs (∩ anatomically constrained)	Peak MNI coordinates		
		x	y	z
A. Abstract-related ROIs				
[ROI=Wang_etal_2010 (1)]		-48	18	-2
[ROI=Wang_etal_2010 (2)]		-50	20	4
[ROI=Wang_etal_2010 (3)]		-42	20	-4
[ROI=Wang_etal_2010 (4)]		-52	10	-18
[ROI=Wang_etal_2010 (5)]		-52	8	-32
[ROI=Wang_etal_2010 (6)]		-58	-42	-4
[ROI=Wang_etal_2010 (7)]		-48	18	-10
[ROI=Wang_etal_2010 (8)]		-48	10	-18
B. Emotion-related ROIs				
	[ROI=Em_Localizer_study1∩Amygdala CM_LB_SF_L]			
	[ROI=Em_Localizer_study1∩Fusiform L]			
[ROI=Moseley_etal_2012 (1)]		-10	56	12
[ROI=Moseley_etal_2012 (2)]		8	52	8
[ROI=Moseley_etal_2012 (3)]		-60	-34	2
[ROI=Moseley_etal_2012 (4)]		50	-32	24
[ROI=Moseley_etal_2012 (5)]		36	-64	2
[ROI=Moseley_etal_2012 (6)]		-40	-40	-14
[ROI=Moseley_etal_2012 (7)]		-48	16	-26
[ROI=Moseley_etal_2012 (8)]		-48	-12	40
[ROI=Moseley_etal_2012 (9)]		-56	0	40
[ROI=Moseley_etal_2012 (10)]		-56	-8	44
[ROI=Moseley_etal_2012 (11)]		16	-50	64
[ROI=Moseley_etal_2012 (12)]		-56	4	24
[ROI=Moseley_etal_2012 (13)]		-28	24	6
[ROI=Moseley_etal_2012 (14)]		-34	-16	22
C. Mathematics-related ROIs				
	[ROI=Ma_Localizer_studyB1] (left middle frontal gyrus)	-40	32	32
	[ROI=Ma_Localizer_studyB1] (left inferior parietal lobule)	-48	-36	44

D. Mental state-related ROIs				
	[ROI=Ms_Localizer_studyB1] (posterior cingulate cortex)	2	-52	20
	[ROI=Ms_Localizer_studyB1] (medial superior frontal gyrus)	2	56	28
	[ROI=Ms_Localizer_studyB1] (L angular gyrus)	-34	18	44
	[ROI=Ms_Localizer_studyB1] (R angular gyrus)	42	18	32
	[ROI=Ms_Localizer_studyB1] (L middle frontal gyrus)	-34	18	44
E. Mouth-related ROIs				
	[ROI=Mo_Localizer_studyB1∩ area 44L]			
	[ROI=Mo_Localizer_studyB1∩ area 44R]			
	[ROI=Mo_Localizer_studyB1∩area 4aL]			
	[ROI=Mo_Localizer_studyB1∩area 4aR]			
	[ROI=Mo_Localizer_studyB1∩area 4pL]			
	[ROI=Mo_Localizer_studyB1∩area4pR]			
	[ROI=Mo_Localizer_studyB1∩area 6 L]			
	[ROI=Mo_Localizer_studyB1∩area 6 R]			
[ROI=Simmons_etal_2005 (1)]		-18	45	-6
[ROI=Simmons_etal_2005 (2)]		-21	33	-18
[ROI=Simmons_etal_2005 (3)]		36	-6	9
[ROI=Simmons_etal_2005 (4)]		48	-45	-12
[ROI=Simmons_etal_2005 (5)]		-48	-60	-18
[ROI=Simmons_etal_2005 (6)]		48	-66	-9
F. Hand-related ROIs				
	[ROI=Ha_Localizer_studyB1∩area 4aL]			
	[ROI=Ha_Localizer_studyB1∩area 4pL]			
	[ROI=Ha_Localizer_studyB1∩area 6 L]			
	[ROI=Ha_Localizer_studyB1∩area 6 R]			
G. Leg-related ROIs				
	[ROI=Le_Localizer_studyB1∩area 4aL]			
	[ROI=Le_Localizer_studyB1∩area 4aR]			
	[ROI=Le_Localizer_studyB1∩area 4pL]			
	[ROI=Le_Localizer_studyB1∩area 6L]			
	[ROI=Le_Localizer_studyB1∩area 4aR]			

Table 5.
Study B2. Specific effects for the (Linguistic task – cross-modal 1-back task, **A.**) and opposite (Cross-modal 1-back task – Linguistic task, **B.**) contrasts. P < 0.05, FWE corrected at the whole brain level, Extent threshold k > 10 voxels. Coordinates belonging to the same activation cluster are grouped by ".
§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
A. Linguistic task > cross-modal n-back task						
R superior temporal gyrus (TE 1.1 60%)	1545	Inf	0.000	48	-24	8
R superior temporal gyrus (TE 1.2 20%)	"	Inf	0.000	62	-8	0
R superior temporal gyrus (TE 3 40%)	"	Inf	0.000	60	-2	-8
L superior temporal gyrus (TE 1.0 70%)	1658	Inf	0.000	-50	-16	4
L superior temporal gyrus (TE 3 30%)	"	Inf	0.000	-60	-10	0
L middle temporal gyrus	"	Inf	0.000	-62	-38	4
R inferior frontal gyrus (pars. Triangularis) (area 45 10%)	26	5.66	0.000	46	20	24
B. Cross-modal n-back task > linguistic task						
R fusiform gyrus (hOC3v(V3v) 40%)	262	Inf	0.000	30	-80	-12
R lingual gyrus (area 18 70 %)	"	6.39	0.000	16	-88	-8
R middle occipital gyrus	"	6.26	0.000	46	-78	4
L lingual gyrus (area 18 20 %)	2012	7.56	0.000	-22	-54	-8
L lingual gyrus (area 18 50 %)	"	7.33	0.000	-12	-60	-4
R calcarine gyrus (area 18 50%)	"	6.98	0.000	12	-66	16
L putamen	302	7.25	0.000	-22	6	8
R inferior frontal gyrus (pars opercularis) (area 44 30%)	96	6.92	0.000	50	8	8
R insula lobe	"	6.17	0.000	42	2	4
R cuneus	12	6.56	0.000	20	-84	40
R amygdala (Amyg SF 90%)	157	6.30	0.000	26	-2	-12
Thalamus (Th-Prefrontal 14%)	"	5.79	0.000	16	-2	-12
R caudate nucleus	"	5.69	0.000	20	16	8
L SMA (area 6 50%)	255	6.18	0.000	0	0	48
R SMA (area 6 50%)	"	5.90	0.000	8	4	64
L precuneus (SPL (7A) 60%)	13	5.49	0.001	-14	-70	56
R anterior cingulate cortex	68	5.48	0.001	6	14	28
L superior frontal gyrus (area 6 30%)	21	5.36	0.002	-22	-2	64
L Hippocampus (Hipp (FD) 60%)	15	5.22	0.004	-24	-30	-
L middle frontal gyrus	16	5.15	0.006	-42	40	24

Table 6.
Study B2. Specific effects for the (Abstract > Action, **A.**), and the opposite (Action > Abstract, **B.**) contrasts. P < 0.05, FWE corrected at the whole brain level, or Small Volume Corrected where indicated by $^{\#}$.
§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)
Results of the analysis on the group mean beta values extracted from the Regions Of Interest (ROIs) (see Table 4) are summarized in the last column. Each analyzed ROIs approximately corresponds to the activation coordinates reported in the left columns of the table. Post-hoc comparisons P < 0.05 are indicated by *, P < 0.1 are indicated by (*).

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)			ROI analysis
				x	y	z	
A. Abstract > Action							
L middle temporal gyrus	214	6.67	0.000	-54	-6	-16	
L inferior frontal gyrus (pars. Triangularis) (Brodmann area 45 60%)	58	4.26	0.075	-52	20	16	
L temporal pole	38	5.32	0.000	-52	8	-16	[ROI=Wang_etal_2010 (4)] Em > Ms > Ma > *Mo > *Le > *Ha
L medial temporal pole	11	3.98	0.000$^{\#}$	-50	12	-28	[ROI=Wang_etal_2010 (5)] Em > Ms > *Mo > *Ma > *Ha > *Le
L middle temporal gyrus	39	6.94	0.000$^{\#}$	-56	-40	0	[ROI=Wang_etal_2010 (6)] Ms > Em > Ma > *Mo > *Ha > *Le
B. Action > Abstract							
L Inferior Frontal Gyrus (p. Triangularis) (area 45 30%)	92	4.36	0.014	-44	36	12	
R SupraMarginal Gyrus (IPC-PFm 50% IPC-PGa 50%)	155	4.09	0.001	56	-48	28	
L Inferior Parietal Lobule	95	3.91	0.012	-54	-42	36	

Table 7.

Study B2. Specific effects for: **A.** Ms – (Em + Ma); **B.** Em – (Ms + Ma); C. Ma – (Ms + Em); **D.** Mo – (Ha + Le); **E.** Ha – (Mo + Le); **F.** Le – (Mo + Ha). P < 0.05, FWE corrected at the whole brain level, or Small Volume Corrected where indicated by $^{\#}$.

§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left).

Results of the analysis on the group mean beta values extracted from the Regions Of Interest (ROIs) (see Table 4) are summarized in the last column. Each analyzed ROIs approximately corresponds to the activation coordinates reported in the left columns of the table. Post-hoc comparisons P < 0.05 are indicated by *, P < 0.1 are indicated by (*).

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)			ROI analysis
				x	y	z	
A. Ms – (Em + Ma)							
No significant activations							
B. Em – (Ms + Ma)							
L paracentral lobule (area 4a 40%)	411	4.4	0.003	-6	-26	52	
L Amygdala (Amyg SF 90%)	64	3.04	$0.067^{\#}$	-20	-4	-16	[ROI=Em_Localizer_studyB1∩Amygdala CM_LB_SF_L] Mo > Em > Ha > *Ma > *Ms > *Le
L Postcentral gyrus (area 3b 70%)	26	3.54	$0.006^{\#}$	-48	-16	36	[ROI=Moseley_etal_2012 (8)] Em > Ms > (*)Ha > (*)Ma > (*)Mo > *Le
R precentral gyrus (area 4a 70%)	30	2.99	$0.041^{\#}$	50	-14	44	[ROI=Mo_Localizer_studyB1∩area 4aR] Em > Ha > Ms > (*)Mo > *Le > *Ma
R precentral gyrus (area 4p 70%)	29	3.47	$0.009^{\#}$	48	-14	40	[ROI=Mo_Localizer_studyB1∩area 4pR] Em > Ha > (*)Mo > *Ms > *Le > *Ma
L middle cingulate Cortex (area 4 40%)	16	3.56	$0.013^{\#}$	-8	-20	48	[ROI=Ha_Localizer_studyB1∩area 4aL] Em > Ha > Ms > (*)Mo > *Le > *Ma
L precentral gyrus (area 4p 50)	30	3.4	$0.016^{\#}$	-36	-18	48	[ROI=Ha_Localizer_studyB1∩area 4pL] Em > Ha > Ms > *Le > *Mo > *Ma
L SMA (area 6 70%)	140	3.82	$0.026^{\#}$	-6	-22	52	[ROI=Ha_Localizer_studyB1∩area 6 L] Em > Mo > Ms > Ha > *Ma > *Le
L paracentral lobule (area 4a 70%)	1	3.17	$0.044^{\#}$	-4	-26	56	[ROI=Le_Localizer_studyB1∩area 4aL] Em > Ha > Ms > (*)Ma > *Mo > *Le
L paracentral lobule (area 6 80%)	120	3.57	$0.038^{\#}$	-4	-22	56	[ROI=Le_Localizer_studyB1∩area 6 L] Em > Mo > Ms > Ha > Ma > *Le
C. Ma – (Ms + Em)							
No significant activations							
D. Mo – (Ha + Le)							
L inferior frontal gyrus (p. Triangularis) (area 45 10%)	601	4.89	0.000	-42	22	0	
R SMA (area 6 60%)	395	4.17	0.004	4	8	64	
L cerebellum (Lobule VI (Hem) 77%)	321	4.1	0.014	-32	-58	-28	
R temporal pole	412	4.05	0.003	38	20	-28	
L precentral gyrus (area 44 40%)	56	4.15	$0.001^{\#}$	-58	6	24	[ROI=Mo_Localizer_studyB1∩area 44L] Mo > Ms > Ma > (*)Em > (*)Le > *Ha
R inferior frontal gyrus (area 44	48	3.35	$0.029^{\#}$	56	14	0	[ROI=Mo_Localizer_studyB1∩area 44R]

40%)							Mo > (*)Ha > (*)Ms > *Le > *Em > *Ma
L precentral gyrus (area 6 30%)	10	4.15	0.007#	-58	4	24	[ROI=Mo_Localizer_studyB1∩area 6 L] Em > Mo > Ms > *Ha > (*)Ma > *Le
L middle orbital gyrus	27	3.82	0.002#	-24	34	-16	[ROI=Simmons_etal_2005 (2)] Mo > *Ha > *Em > *Le > *Ms > *Ma
L fusiform gyrus	29	3.43	0.008#	-46	-56	-20	[ROI=Simmons_etal_2005 (5)] Mo > *Ma > *Ms > *Ha > *Em > *Le
E. Ha – (Mo + Le)							
No significant activations							
F. Le – (Mo + Ha)							
No significant activations							

Table 8.
Study B2. Specific effects for: **A.** [Ms – (Mo + Ha + Le)] exclusively masked by [Em – (Mo + Ha + Le)] and [Ma – (Mo + Ha + Le)] ; **B.** [Em – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Ma – (Mo + Ha + Le)]; **C.** [Ma – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Em – (Mo + Ha + Le)]; **D.** [Mo – (Ms + Em + Ms)] exclusively masked by [Ha – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]; **E.** [Ha – (Ms + Em + Ms)] exclusively masked by [Mo – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]; **F.** [Le – (Ms + Em + Ms)] exclusively masked by [Ha – (Ms + Em + Ms)], [Mo – (Ms + Em + Ms)]. P < 0.05, FWE corrected at the whole brain level, or Small Volume Corrected where indicated by $^{\#}$. § According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left) Results of the analysis on the group mean beta values extracted from the Regions Of Interest (see Table 4) are summarized in the last column. Each analyzed ROIs approximately corresponds to the activation coordinates reported in the left columns of the tab. Post-hoc comparisons P < 0.05 are indicated by *, P < 0.1 are indicated by (*).

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)			ROI analysis
				x	y	z	
A. [Ms – (Mo + Ha + Le)] exclusively masked by [Em – (Mo + Ha + Le)], [Ma – (Mo + Ha + Le)]							
No significant activations							
B. [Em – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Ma – (Mo + Ha + Le)]							
L postcentral gyrus (area 3b 70%)	38	3.18	$0.017^{\#}$	-50	-16	36	[ROI=Moseley_etal_2012 (8)] Em > Ms > *Ha > *Ma > *Mo > *Le
L medial temporal pole	3	3.14	$0.019^{\#}$	-46	12	-28	[ROI=Moseley_etal_2012 (7)] Em > Ms > *Ma > *Mo > *Le > *Ha
L postcentral gyrus (area 1 70%)	24	2.87	$0.039^{\#}$	-56	-10	40	[ROI=Moseley_etal_2012 (10)] Em > Ms > Mo > *Ma >*Ha > *Le
R insula lobe (OP3 80%)	12	3.03	$0.017^{\#}$	38	-12	20	[ROI=Moseley_etal_2012 (14)] Em > Ms > Le > Ma > *Ha > *Mo
L postcentral gyrus (area 4a 30%)	15	3.0	$0.035^{\#}$	-42	-14	52	[ROI=Mo_Localizer_studyB1∩area 4aL] Em > Ms > Ha > Mo > Ma > *Le
L precentral gyrus (area 4p 50%)	36	3.15	$0.027^{\#}$	-36	-18	48	[ROI=Mo_Localizer_studyB1∩area 4pL] Em > Ha > *Ms > *Le > *Ma > *Mo [ROI=Ha_Localizer_studyB1∩area 4pL] Em > Ha > Ms > *Le > *Mo > *Ma
R superior parietal lobule (SPL (5M) 40%)	28	2.57	$0.079^{\#}$	14	-48	64	[ROI=Moseley_etal_2012 (11)] Em > Ms > Ha > Ma > *Ha > *Mo > *Le
C. [Ma – (Mo + Ha + Le)] exclusively masked by [Ms – (Mo + Ha + Le)], [Em – (Mo + Ha + Le)]							
No significant activations							
D. [Mo – (Ms +Em +Ms)] exclusively masked by [Ha – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]							
L inferior frontal gyrus (pars. Opercularis) (area 44 30%)	2	2.99	$0.07^{\#}$	-50	14	0	[ROI=Mo_Localizer_studyB1∩area 44L] Mo > Ms > Ma > (*)Em > (*)Le > *Ha
R inferior frontal gyrus (p. Opercularis) (area 45 10%)	47	3.57	$0.015^{\#}$	56	14	0	[ROI=Mo_Localizer_studyB1∩area 44R] Mo > (*)Ha > (*)Ms > *Le > *Em > *Ma
R SMA (area 6 50%)	117	4.18	$0.005^{\#}$	8	8	64	[ROI=Mo_Localizer_studyB1∩area 6 R] Mo > Em > (*)Ha > (*)Ms > *Ma > *Le
L middle orbital gyrus	28	4.78	$0.000^{\#}$	-22	34	-16	[ROI=Simmons_etal_2005 (2)] Mo > *Ha > *Em > *Le > *Ms > *Ma
L fusiform gyrus	24	3.2	$0.016^{\#}$	-46	-56	-20	[ROI=Simmons_etal_2005 (5)] Mo > *Ma > *Ms > *Ha > *Em > *Le

E. [Ha – (Ms +Em +Ms)] exclusively masked by [Mo – (Ms + Em + Ms)], [Le – (Ms + Em + Ms)]							
L inferior parietal lobule (hIP2 30%)	34	3.59	$0.007^{\#}$	-44	-46	44	[ROI=AIPS_IP2] Ha > Mo > *Ms > *Le > *Ma > *Em
F. [Le – (Ms +Em + Ms)] exclusively masked by [Ha – (Ms + Em + Ms)], [Mo – (Ms + Em + Ms)]							
No specific activations.							

Table 9.

Study B2. Abstract/action-related classification analysis. The Confusion matrix illustrates the generalization performance of the classifier. Each row of the matrix represents the instances in a predicted class, while each column represents the samples in an actual (target) class. Each cell provides a count of how many samples of the target class were classified into the corresponding class. Abstract-related class included Ms, Em, Ma; Action-related class included Mo, Ha, Le.

	Abstract	Action
Abstract	34	2
Action	2	34

Table 10.

Study B2. Abstract/action-related sensitivity analysis. LinearCSVMC weights using all brain mask voxels.
The anatomical locations reported in the table refer to clusters of higher sensitivity weight of the LinearCSVMC classifier, representing the degree of contribution or each cluster to the separation between the two classes. The brain regions were filtered to exclude clusters smaller than 10 voxels. Extent threshold k > 10 voxels. § According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left).

Brain region (cytoarchitectonic probability§)	**Voxels**	**LinearCSVMC weight**	**MNI coordinates (mm)**		
			X (mm)	Y (mm)	Z (mm)
L Inferior Frontal Gyrus p. Triangularis (area 45 50%)	80	0.00195	-52	22	16
L Middle Frontal Gyrus	79	0.0022	-38	38	24
L Middle Frontal Gyrus	13	0.00153	-36	36	32
L Superior Frontal Gyrus	56	0.00209	-10	54	36
L Rectal Gyrus	30	0.00177	0	52	-16
L Rectal Gyrus	16	0.00146	-6	50	-20
L Superior Orbital Gyrus	17	0.00186	-28	60	-4
L SMA (area 6 30%)	25	0.00197	-2	14	60
L Precentral Gyrus (area 6 40%)	14	0.00165	-58	0	28
L Precentral Gyrus (area 6 70%)	18	0.00169	-50	2	48
L Precentral Gyrus (area 44 10%)	12	0.00178	-44	4	32
L Insula Lobe (area 44 10%)	23	0.0019	-40	6	4
L Insula Lobe (Insula Ig2 60%)	12	0.00176	-38	-16	8
L Insula Lobe	22	0.00168	-28	24	-8
L Insula Lobe	10	0.00139	-40	8	-4
L Middle Temporal Gyrus	670	0.00309	-56	-36	4
L Middle Temporal Gyrus	11	0.00152	-48	-58	12
L Superior Temporal Gyrus (IPC Pfcm 40%)	28	0.00195	-56	-36	16
L Fusiform Gyrus (Hipp CA 10%)	20	0.00167	-36	-28	-20
L Fusiform Gyrus	43	0.00192	-34	-42	-20
L Superior Parietal Lobule (SPL 7A 40%)	10	0.00173	-18	-68	52
L Inferior Parietal Lobule (hIP1 50%)	78	0.00188	-36	-44	40
L Inferior Parietal Lobule (IPC PF 90%)	19	0.00197	-58	-34	44

L Paracentral Lobule (area 4a 50%)	13	0.00193	-16	-28	64
L SupraMarginal Gyrus (IPC Pfcm 80%)	10	0.00167	-44	-36	24
L SupraMarginal Gyrus (IPC Pfm 50%)	17	0.0018	-60	-52	32
L SupraMarginal Gyrus (IPC Pft 60%)	10	0.00169	-60	-24	36
L Precuneus (SPL 7P 30%)	12	0.00228	-4	-64	44
L Precuneus	12	0.00144	-6	-50	40
L Posterior Cingulate Cortex	10	0.00157	-2	-46	24
L Middle Occipital Gyrus (IPC Pgp 20%)	13	0.00213	-34	-76	32
L Calcarine Gyrus (area 18 10%)	52	0.00173	-10	-62	12
L Thalamus (Th-Prefrontal 49%)	14	0.00154	-4	-22	0
L Cerebellum (Lobule VI Hem 87%)	11	0.00153	-24	-50	-20
R Inferior Frontal Gyrus p. Opercularis (area 44 40%)	16	0.0017	54	14	28
R Inferior Frontal Gyrus p. Triangularis (area45 40%)	17	0.00204	52	36	8
R Middle Frontal Gyrus	42	0.00179	40	42	20
R Middle Frontal Gyrus	15	0.00145	34	0	56
R Middle Frontal Gyrus	10	0.00177	46	26	40
R Superior Frontal Gyrus (area 6 30%)	14	0.00164	22	2	60
R Superior Frontal Gyrus	20	0.00218	26	16	60
R Superior Frontal Gyrus	12	0.00181	24	56	0
R Superior Medial Gyrus (area 6 10%)	26	0.00165	4	32	48
R Superior Medial Gyrus	13	0.00195	8	58	8
R Mid Orbital Gyrus	12	0.00154	8	62	-12
R Insula Lobe (Insula Ig2 70%)	11	0.00172	36	-20	4
R Insula Lobe	15	0.00192	30	22	-16
R Inferior Temporal Gyrus	16	0.00163	52	-54	-20
R Superior Temporal Gyrus (IPC Pfcm 20%)	10	0.00159	52	-30	16
R Superior Temporal Gyrus (TE 3 30%)	142	0.00247	62	0	-4
R Superior Temporal Gyrus	99	0.00228	54	-22	0
R Medial Temporal Pole	28	0.00182	46	12	-32
R Inferior Parietal Lobule (IPC PFm 90%)	178	0.00215	50	-48	48
R SupraMarginal Gyrus (IPC PF 90%)	14	0.00156	62	-36	32
R Cuneus (area 18 40%)	13	0.00131	6	-70	20
R Precuneus (SPL 7A 80%)	11	0.00153	14	-58	60
R Precuneus (SPL 7M 50%)	27	0.00157	4	-70	44
R Precuneus	12	0.00153	6	-48	20
R Middle Cingulate Cortex	22	0.00198	12	-46	36
R Calcarine Gyrus (area 18 60%)	25	0.00172	14	-62	16
R Caudate Nucleus	25	0.0017	8	16	8
R Olfactory cortex	10	0.00148	10	12	-16

Table 11.

Study B2. Semantic category-specific classification analysis. The Confusion matrix illustrates the generalization performance of the classifier. Each row of the matrix represents the instances in a predicted class, while each column represents the samples in an actual (target) class. Each cell provides a count of how many samples of the target class were classified into the corresponding class.

	Ms	Em	Ma	Mo	Ha	Le
Ms	14	13	12	4	3	5
Em	8	14	4	2	2	3
Ma	5	0	5	1	0	2
Mo	2	1	4	14	2	2
Ha	2	3	2	5	17	8
Le	5	5	9	10	12	16

Table 12.

Study B2. Semantic category-specific sensitivity analysis. LinearCSVMC weights using all brain mask voxels.
The anatomical locations reported in the table refer to clusters of higher sensitivity weight of the LinearCSVMC classifier, representing the degree of contribution or each cluster to the separation between the six classes. The brain regions were filtered to exclude clusters smaller than 10 voxels. Extent threshold k > 10 voxels. [§]According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left).

Brain region (cytoarchitectonic probability[§])	**Voxels**	**LinearCSVMC weight**	**MNI coordinates (mm)**		
			X (mm)	Y (mm)	Z (mm)
L Inferior Frontal Gyrus (p. Opercularis) (area 44 50%)	19	0.00313	-46	10	28
L Inferior Frontal Gyrus (p. Orbitalis)	24	0.00292	-26	32	-16
L Inferior Frontal Gyrus (p. Triangularis) (area 44 40%)	37	0.00302	-48	16	8
L Inferior Frontal Gyrus (p. Triangularis) (area 45 10%)	133	0.0035	-48	30	0
L Inferior Frontal Gyrus (p. Triangularis) (area 45 30%)	12	0.00266	-42	32	16
L Middle Frontal Gyrus (area 44 10%)	12	0.0029	-44	22	44
L Middle Frontal Gyrus	12	0.00303	-40	34	32
L Middle Frontal Gyrus	12	0.00246	-44	44	16
L Superior Frontal Gyrus	34	0.00283	-10	54	36
L Superior Frontal Gyrus	18	0.0025	-14	38	48
L Mid Orbital Gyrus	39	0.00299	-6	52	-4
L Mid Orbital Gyrus	15	0.00334	-12	42	-8
L Middle Orbital Gyrus	23	0.00259	-34	54	-8
L SMA (area 6 60%)	48	0.00323	-2	-20	60
L SMA (Area 6 70%)	16	0.00291	-4	-10	52
L Insula Lobe (Ig2 30%)	12	0.00249	-34	-14	8
L Insula (Ig1 50%)	19	0.00282	-36	-24	0

L Middle Temporal Gyrus	389	0.00339	-54	-10	-12
L Middle Temporal Gyrus	16	0.00293	-56	-46	-12
L Middle Temporal Gyrus	10	0.00341	-60	-54	-4
L Hippocampus (Hipp SUB 40%)	19	0.00287	-16	-10	-20
L Fusiform Gyrus (Hipp CA 50%)	11	0.00242	-32	-28	-20
L Fusiform Gyrus	35	0.00294	-22	-36	-16
L Postcentral Gyrus (area 3b 30%)	13	0.0028	-58	-2	20
L Postcentral Gyrus (IPC Pft 70%)	11	0.0032	-58	-20	32
L Inferior Parietal Lobule (IPC Pfm 70%)	27	0.00325	-56	-52	36
L SupraMarginal Gyrus (IPC Pfop 40%)	11	0.00255	-52	-32	24
L Angular Gyrus (IPC Pgp 30%)	20	0.0027	-44	-62	28
L Anterior Cingulate Cortex	23	0.00285	-10	38	24
L Middle Cingulate Cortex (area 6 10%)	11	0.00289	-6	-18	40
L Middle Cingulate Cortex	10	0.00248	-10	22	32
L Precuneus (SPL 5L 50%)	23	0.00367	-14	-46	64
L Precuneus (SPL 5M 10%)	11	0.00247	-6	-46	48
L Precuneus (SPL 7A 60%)	23	0.00297	-8	-60	56
L Precuneus	29	0.00293	-2	-54	16
L Lingual Gyrus (area 18 80%)	15	0.00301	-10	-72	4
L Calcarine Gyrus (area 18 30%)	22	0.00301	-4	-68	20
L Cerebellum (Lobule V 84%)	10	0.00287	-10	-50	-12
L Cerebellum (Lobule V 86%)	62	0.00316	-4	-62	-16
L Cerebellum (Lobule VI Hem 85%)	26	0.00315	-34	-54	-28
L Caudate Nucleus	12	0.00256	-12	8	16
L Putamen	25	0.00291	-20	14	-8
L Putamen	12	0.00307	-28	-2	8
R Inferior Frontal Gyrus (p. Opercularis)	31	0.00329	38	18	32
R Inferior Frontal Gyrus (p. Orbitalis)	54	0.00299	28	26	-20
R Inferior Frontal Gyrus (p. Triangularis) (area 45 70%)	23	0.00265	56	20	20
R Middle Frontal Gyrus	23	0.00284	40	46	20
R Middle Frontal Gyrus	13	0.00281	36	34	32
R Superior Frontal Gyrus	10	0.00274	26	28	52
R Superior Medial Gyrus (area 6 10%)	15	0.00328	8	20	44
R Superior Medial Gyrus	29	0.00289	6	58	0
R Mid Orbital Gyrus	15	0.00365	4	40	-8
R SMA (area 4a 60%)	18	0.00242	2	-24	56
R SMA (area 6 100%)	48	0.00288	4	-14	60
R Precentral Gyrus (area 4p 40%)	13	0.00315	44	-12	40
R Precentral Gyrus (area 6 90%)	27	0.00298	34	-24	64

R Insula (Ig2 90%)	25	0.00268	38	-20	0
R insula lobe	77	0.00322	38	12	-4
R Rolandic Operculum (OP 3 20%)	10	0.00299	42	-4	16
R Middle Temporal Gyrus (hOC5 V5 20%)	28	0.00271	56	-66	0
R Middle Temporal Gyrus	19	0.00335	54	-8	-24
R Middle Temporal Gyrus	12	0.00293	60	-40	-8
R Superior Temporal Gyrus	83	0.00319	48	-30	0
R Superior Temporal Gyrus (TE 1.0 10%)	69	0.00312	52	-12	-4
R Fusiform Gyrus	25	0.00288	40	-48	-20
R Fusiform Gyrus	11	0.00258	30	-48	-12
R Temporal Pole	23	0.00291	44	16	-28
R Hippocampus (Hipp FD 70%)	18	0.00264	24	-38	0
R Hippocampus (Right Hipp CA 40%)	13	0.00262	16	-12	-20
R ParaHippocampal Gyrus (Hipp SUB 100%)	10	0.00291	28	-18	-24
R Postcentral Gyrus (Area 1 100%)	27	0.00283	36	-40	64
R Postcentral Gyrus (Area 1 80%)	16	0.00293	56	-14	48
R Inferior Parietal Lobule (IPC Pfm 50%)	68	0.00303	50	-56	48
R SupraMarginal Gyrus (IPC Pfm 100%)	12	0.00301	58	-44	40
R Angular Gyrus (IPC Pga 90%)	12	0.00259	60	-54	24
R Anterior Cingulate Cortex	25	0.00254	4	40	28
R Anterior Cingulate Cortex	13	0.00281	8	34	28
R Middle Cingulate Cortex	16	0.0032	6	16	32
R Precuneus	27	0.00275	4	-50	44
R Lingual Gyrus (area 18 70%)	11	0.00322	12	-60	0
R Cerebellum (Lobule VIIa Crus I Hem 89%)	16	0.00299	14	-84	-24
Cerebellar Vermis (Lobule V 84%)	14	0.00278	4	-60	-12
R Caudate Nucleus	13	0.00261	14	14	0
R Caudate Nucleus	16	0.0028	14	18	-8
R Thalamus (Th-Temporal 67%)	26	0.00313	4	-8	4

Table 13.
Study B2. Parametric effect of concreteness: **1.** positive correlation; **2.** negative correlation. P < 0.05, FWE corrected at the whole brain level. Coordinates belonging to the same activation cluster are grouped by “.
§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
1. Positive correlation						
L inferior frontal gyrus (p. Triangularis) (area 45 20%)	164	4.87	0.000	-42	36	12
L inferior frontal gyrus (p. Triangularis)	“	3.88		-40	40	4
L middle frontal gyrus	“	3.62		-44	46	0
L inferior parietal lobule (SPL 7PC 20%)	92	4.30	0.008	-44	-50	48
L inferior parietal lobule (IPC (PFm) 60%)	“	3.77		-52	-48	44
L inferior parietal lobule (IPC (PF) 70%)	“	3.60		-54	-42	36
R inferior parietal lobule (IPC (PFm) 90%)	165	4.29	0.000	50	-50	48
R supramarginal gyrus (IPC (PF) 60%)		4.13		52	-40	44
R inferior parietal lobule (IPC (PGa) 50%)		3.64		48	-58	48
2. Negative correlation						
L middle temporal gyrus	805	7.88	0.000	-56	-42	4
L middle temporal gyrus	“	7.71		-56	-8	-8
L temporal pole	“	7.42		-54	6	-12
R temporal pole	150	4.86		60	2	-8
R middle temporal gyrus		4.50		52	-10	-20
R medial temporal pole	48	4.22	0.096	46	8	-28
R temporal pole	“	3.71		52	14	-20

Table 14.
Study B2. Parametric effect of context availability: **1.** positive correlation; **2.** negative correlation. P < 0.05, FWE corrected at the whole brain level. Coordinates belonging to the same activation cluster are grouped by “.
§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
1. Positive correlation						
L inferior parietal lobule (IPC PFm 90%)	255	4.74	0.000	50	-50	48
R angular gyrus (IPC (PGa) 50%)	“	3.86		50	-48	24
R inferior parietal lobule (hIP2 50%)	“	3.54		42	-42	44
L inferior parietal lobule (IPC PF 50%, hIP2 20%)	73	4.00	0.024	-48	-48	52
L angular gyrus (IPC PGa)	“	3.76		-42	-60	48
L inferior parietal lobule (hIP2 30%)	“	3.68		-44	-46	44
R middle frontal gyrus	56	4.01	0.064	30	16	56
2. Negative correlation						
L middle temporal gyrus	614	5.55	0.000	-54	-44	4
L middle temporal gyrus	“	5.52		-56	-8	-8
L middle temporal gyrus	“	5.33		-56	-18	-4
L postcentral gyrus (area 1 40%)	58	4.21	0.057	-50	-12	52
L precentral gyrus (area 6 100%)	“	3.66		-50	-2	48
L inferior frontal gyrus (p. Triangularis) (area 45 90%)	50	4.42	0.092	-56	24	12

Table 15.

Study B2. Parametric effects of association ratings: **A.** Emotion-association rating (positive correlation); **B.** Mathematics-association rating (positive correlation); **C.** Mental state-association rating (positive correlation). $P < 0.05$, FWE corrected at the whole brain level, or Small Volume Corrected where indicated by $^{\#}$. Coordinates belonging to the same activation cluster are grouped by ".

§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Results of the analysis on the group mean beta values extracted from the Regions Of Interest (ROIs) (see Table 4) are summarized in the last column. Each analyzed ROIs approximately corresponds to the activation coordinates reported in the left columns of the tab. Post-hoc comparisons $P < 0.05$ are indicated by *, $P < 0.1$ are indicated by (*).

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)			ROI analysis
				x	y	z	
A. Emotion-association rating (positive correlation)							
R fusiform gyrus	27	3.62	0.037$^{\#}$	-32	-56	-20	Em_Localizer_studyB1∩Fusiform L *
L (Hippocampus (CA) 30%)	2	3.07	0.021$^{\#}$	-36	-36	-12	[ROI=Moseley_etal_2012] (6) (*)
L postcentral gyrus (area 4p 60%)	12	3.25	0.044$^{\#}$	-34	-18	44	[ROI=Mo_Localizer_studyB1∩area 4pL] (*) [ROI=Ha_Localizer_studyB1∩area 4pL]*
L postcentral gyrus (area 4a 40%)	13	3.46	0.021$^{\#}$	-40	-14	48	[ROI=Mo_Localizer_studyB1∩area 4aL] / (*) [ROI=Ha_Localizer_studyB1∩area 4aL]*
R middle frontal gyrus	3	3.15	0.053$^{\#}$	40	-12	44	[ROI=Mo_Localizer_studyB1∩area 4aR]*
L postcentral gyrus (area 4a 40%)	2	2.52	0.099$^{\#}$	-44	-12	44	[ROI=Moseley_etal_2012 (8)] (*)
B. Mathematics-association rating (positive correlation)							
No significant activations							
C. Mental state-association rating (positive correlation)							
No significant activations							

Table 16.
Study B2. Parametric effects of body-part ratings: **A.** Mouth rating (postive correlation); **B.** Hand rating (positive correlation); **C.** Leg rating (positive correlation). P < 0.05, FWE corrected at the whole brain level, or Small Volume Corrected where indicated by #. Coordinates belonging to the same activation cluster are grouped by ".
§ According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)
Results of the analysis on the group mean beta values extracted from the Regions Of Interest (ROIs) (see Table 4) are summarized in the last column. Each analyzed ROIs approximately corresponds to the activation coordinates reported in the left columns of the tab. Post-hoc comparisons P < 0.05 are indicated by *, P < 0.1 are indicated by (*).

Brain region (cytoarchitectonic probability§)	K	Z-score	P-value	MNI coordinates (mm)			ROI analysis
				x	y	z	
A. Mouth rating (positive correlation)							
L middle orbital frontal gyrus	20	3.43	0.008#	-24	36	-16	[ROI=Simmons_etal_2005 (2)]*
B. Hand rating (positive correlation)							
No significant activations.							
C. Leg rating (positive correlation)							
No significant activations.							

Table 17.

Study B2. Parametric effects of: **A.** duration (A1 = positive correlation; A2 = negative correlation); **B.** frequency (B1 = positive correlation; B2 = negative correlation); **C.** familiarity (C1 = positive correlation; C2 = negative correlation). P < 0.05, FWE corrected at the whole brain level. Coordinates belonging to the same activation cluster are grouped by “.

[§] According to www.fz-juelich.de/ime/spm_anatomy_toolbox (R = Right; L = Left)

Brain region (cytoarchitectonic probability[§])	K	Z-score	P-value	MNI coordinates (mm)		
				x	y	z
A1. Duration (positive correlation)						
R superior temporal gyrus (TE 3 40%)	1285	6.59	0.000	64	-16	0
R superior temporal gyrus (TE 3 40%)	“	6.27		66	-20	8
R superior temporal gyrus (TE 1.0 60%)	“	5.80		50	-12	0
L superior temporal gyrus (TE 1.1 50%)	860	6.48	0.000	-42	-34	12
L superior temporal gyrus (TE 1.0 60%)	“	6.39		-48	-16	4
L superior temporal gyrus (TE 1.1 70%)	“	6.17		-42	-26	8
L cerebellum (Lobule VI)	287	4.37	0.000	-12	-72	-16
L cerebellum (Lobule VI)	“	4.21		-26	-60	-20
L cerebellum (Lobule VIIa Crus I 98%)	“	4.07		-36	-70	-24
R inferior parietal lobule (hIP3 50%)	103	4.18	0.006	40	-56	48
R supramariginal gyrus (IPC (PFt) 30%)	“	3.87		48	-34	44
A2. Duration (negative correlation)						
No significant activations						
B1. Frequency (positive correlation)						
No significant activations						
B2. Frequency (negative correlation)						
R superior temporal gyrus (TE 3 10%)	415	4.53	0.003	64	-28	8
R superior temporal gyrus (TE 3 20%)	“	4.27		64	-14	-4
R superior temporal gyrus	“	4.04		58	-18	0
L middle temporal gyrus (TE 3 30%)	360	4.18	0.007	-60	-10	-4
L middle temporal gyrus (IPC (PF) 30%)	“	3.97		-60	-32	0
L superior temporal gyrus (TE 1.0 50%)	“	3.90		-50	-12	0
R inferior parietal lobule (IPC (PFm) 30%)	233	3.95	0.064	48	-38	52
R angular gyrus (IPC (PFm) 40%)	“	3.26		52	-48	32
R supramarginal gyrus (IPC (PF) 100%)	“	3.04		62	-38	36
R inferior frontal gyrus (p. Opercularis)	507	3.82	0.001	52	10	28
R middle frontal gyrus	“	3.66		42	40	12
C1. Familiarity (positive correlation)						
L middle occipital gyrus (IPC (PGp) 60%)	200	5.72	0.000	-36	-78	36
L middle temporal gyrus (IPC (PGp) 40%)	“	3.64		-50	-68	20
	“	3.16		-42	-64	36
L middle cingulate cortex (SPL (5M) 20%)	4	5.17	0.000	-6	-32	40

L angular gyrus (IPC (Pa) 6%)	1014	5.17	0.000	-6	-32	40
Rcuneus	“	5.11		8	-62	20
R precuneus	“	5.02	0.000	12	-52	12
R middle temporal gyrus (IPC (PGp) 20%)	235	4.91	0.000	48	-58	16
R angular gyrus (IPC (PGp) 70%)	“	4.87		44	-68	40
R middle occipital gyrus (IPC (PGp) 50%)	“	4.15		42	-74	28
R parahippocampal gyrus	183	4.87	0.000	18	-34	-12
R fusiform gyrus	“	4.27		28	-34	-16
R lingual gyrus (Hipp (SUB) 50%)	“	3.99		12	-40	0
L middle occipital gyrus (IPC (PGp) 20%)	73	4.62	0.016	-48	-78	12
L middle occipital gyrus (hOC3v (V3v) 10%)	“	4.20		-38	-88	12
L middle occipital gyrus (hOC5 (V5) 10%)	“	3.76		-40	-68	8
L fusiform gyrus	58	4.24	0.040	-36	-64	-16
L lingual gyrus (hOC4v (V4) 10%)	“	4.09		-26	-64	-12
L fusiform gyrus	“	3.74		-32	-56	-16
C2. Familiarity (negative correlation)						
L middle temporal gyrus (TE 3 60%)	106	4.94	0.002	-64	-34	8
R superior temporal gyrus	59	4.25	0.037	62	-28	4
	“	3.73		50	-30	0

Part II

Study C

Differentiating among pragmatic uses of words through timed sensicality judgments: Metaphor, metonymy and approximation[4]

Abstract

Pragmatic and cognitive accounts of figurative language posit a difference between metaphor and metonymy in terms of underlying conceptual operations. Recently, other pragmatic uses of words have been accounted for in the Relevance Theory framework, among which approximation, described in terms of conceptual adjustment that varies in degree and direction with respect to the case of metaphor. Despite the theoretical distinctions, there is very poor experimental evidence addressing the metaphor/metonymy distinction, and none concerning approximation. Here we used tightly normed materials to investigate the interpretation mechanisms of these three phenomena through timed sensicality judgments. Results revealed that interpreting metaphors and approximations differs from literal interpretation both in accuracy and reaction times, with higher difficulty and costs for metaphors than for approximations. This suggests similar albeit gradual interpretative costs, in line with the latest account of Relevance Theory. Metonymy, on the contrary, almost equates literal comprehension and calls for a theoretical distinction from metaphor. Overall, this work represents a first attempt to provide an empirical basis for a theory-sound and psychologically-grounded taxonomy of figurative and loose uses of language.

---

[4] Bambini, V., Ghio, M., Moro, A., Schumacher, P. (*submitted*). Differentiating among pragmatic uses of words through timed sensicality judgments: Metaphor, metonymy and approximation.

1. Introduction

Word meaning is often modified in use, giving rise to a number of loose and figurative uses. These modulations of meaning are thoroughly context-dependent and their description has fallen under the domain of pragmatics. While it is useful to group together non-literal uses, potential differences in the representation and the underlying interpretative mechanisms should be accounted for and considered experimentally. This paper is concerned with the characterization of three pragmatic phenomena, namely metaphor (e.g., 'Some theses are marathons'), metonymy (e.g., 'No comments from Buckingham Palace') and approximation (e.g., 'Her face is oval'), and whether they exhibit different interpretation costs, which might support and sharpen theoretical distinctions.

1.1 Theoretical accounts on metaphor, metonymy and approximation

Existing pragmatic and cognitive accounts differ on whether they treat different types of pragmatically enriched meanings as distinct operations or not. In the Gricean framework, figurative expressions such as metaphor, irony, meiosis and hyperbole are grouped together as cases of flouting the first Maxim of Quality ("Do not say what you believe to be false"), and require the derivation of an implicature (Grice, 1975). The nature of metonymy is not explored, but presumably also metonymic expressions would be described inferentially, either as another case of flouting the first Maxim of Quality or as a tool to adhere to the Maxim of Manner (Egg, 2004).

In more recent times, Relevance Theory has deepened the study of the inferential processes underlying the comprehension of the lexical items, suggesting that grasping the intended meaning of a word requires a process of adjusting the linguistically encoded concept to construct an *ad hoc* concept, i.e., a concept inferentially derived for that occasion of use, whose denotation is broader (i.e., more inclusive) or narrower (i.e., less inclusive) than the denotation of the lexical concept. Consider the utterance 'Boris is a man': in most contexts the lexically encoded concept MAN would result underinformative and would require, for instance, to be narrowed down to the *ad hoc* concept MAN* as 'ideal man', in order to reach the intended interpretation. Conversely, in 'This policy will bankrupt the farmers', the encoded concept BANKRUPT could be taken literally, but in certain contexts is likely to require an adjustment that goes in the opposite direction, namely to be broadened in order to include cases in which the farmers are close enough to bankruptcy (Carston, 2009).

In this view, most words require an adjustment process resulting in an *ad hoc* concept, and the difference depends on the direction of the adjustment (broadening or narrowing with respect to the denotation), and also on the degree of it, ranging from less to more context-dependent and occasion-specific uses (Wilson, 2003; Sperber & Wilson, 2005; Wilson & Carston, 2006). Well studied examples of different degrees of broadening include language uses known in the literature with the labels of approximation, hyperbole and metaphor. Approximation is a variety of broadening that includes a relatively marginal adjustment of the encoded concept, to cover just a "penumbra" of cases that only strictly speaking fall outside the linguistically-specified denotation (e.g., in 'The house is empty', 'empty' is used to communicate that the house is lacking of furniture). Hyperbole involves a more substantial adjustment of the encoded concepts (e.g., 'empty' in the previous example is used to communicate that the house, although furnished, does not have as much furniture as desired). Metaphor is a use of language based on an even more radical broadening of the lexical concept (e.g., 'empty' can be used metaphorically to indicate that the house lacks of emotional content).

Recently, Carston and Wearing (2011) proposed a finessing of the relevance-theoretic account, by positing a stronger distinction between metaphor and hyperbole: while concept broadening is required both in metaphor and hyperbole understanding, metaphorical uses would require also additional concept narrowing. Consider the following utterances: 'My evening jog with Bill turned into a marathon' and 'Writing a thesis was a marathon Jane didn't want to repeat'. When intended hyperbolically, as in the first case, the denotation of the *ad hoc* concept MARATHON* is simply more inclusive (broader) than that of the original lexical concept, involving a relaxing of the length of the episode of running. When intended metaphorically, as in the second example, the word goes through a broadening (in order to include instances of activities that are psychologically demanding and exhausting) combined with narrowing (in order to exclude professional marathons; Carston & Wearing, 2011: 286-296). Following this idea, it seems reasonable to assume that also approximation differs from metaphor in requiring only (and marginally) concept broadening: if a separation holds between the case of metaphor and the case of a substantial yet not radical broadening such as hyperbole, the separation should hold also between metaphor and a marginal broadening like that required by approximation.

As concerns metonymy, a full description in Relevance Theory terms is still lacking (but see Papafragou, 1996 for a preliminary account), and is indeed considered as an interesting challenge for pragmatics (Carston, 2010). Following Nunberg's (1995) distinction between reference transfer (e.g., 'The ham sandwich wants to pay', where an NP is used to refer to another NP) and predicate transfer (e.g., 'Nixon bombed Hanoi', where the shift concerns the whole predicate 'bombed Hanoi'), Wilson & Carston (2007) suggested that both cases seem to require the construction of an *ad hoc* concept, but only the latter case involves broadening and narrowing, while the former should be accommodated in terms of genuine reference substitution or a real world association (see also Carston, 2010: p. 160). Here we focus on the reference transfer case, and on the idea that this type of metonymy, while still representing a pragmatic use that requires the construction of an *ad hoc* concept, does not involve the same kind of conceptual adjustment observed for metaphor. The view that at least some cases of metonymy can be described as cases of reference transfer – presumably involving a meaning shift mechanism – is in line with other pragmatic accounts distinguishing between loosening, meaning shift, and free enrichment (Recanati, 2010).

Turning to the framework broadly known as Cognitive Linguistics, both metaphor and metonymy are thought of as conceptual phenomena grounded in general cognition. Yet a difference is assumed to characterize the two. Metaphor is described in terms of mapping between two distinct cognitive domains (e.g., in 'Love is a journey', the source domain JOURNEY is mapped onto the target domain LOVE). Metonymy, on the contrary, is based on mapping within the same cognitive domain (e.g., in 'He is reading Shakespeare', the source domain SHAKESPEARE provides access to its sub-domain SHAKESPEARE'S WRITINGS, which is the target domain; Lakoff & Johnson, 1980; Ruiz de Mendoza Ibáñez, 2007). Furthermore, the mapping is taken to be based on different associative relations: resemblance for metaphor and contiguity for metonymy. Standard types of metonymic mappings are, among others, *part for whole*, *producer for product, place for institution*, *object used for user* (Panther & Thornburg, 2007). This list suggests the routinized status of many metonymic mappings, and points in the direction of a close relation between metonymy and grammar. There are indeed grammatical structures that seem to be sensitive to metonymically induced interpretations. For example, in 'The author began the book', the verb's logical structure coerces an interpretation in which a part of an event, *the book*, denotes the whole event, *writing the book*, a phenomenon known as "logical metonymy" (Pustejovsky, 1995; Lascarides & Copestake, 1998).

1.2 Formulating empirical predictions

Do the differences among pragmatic uses brought about in the theoretical literature find experimental support? Evidence from direct comparison of metaphor and metonymy is sparse. In a self-paced reading study, Gibbs (1990) showed that metaphorical referential descriptions are understood more easily than metonymic ones. Developmental psychology, however, points in a different direction: with respect to metaphor, metonymy is acquired earlier and processed more accurately (Rundblad & Annaz, 2010). Similarly, patients seem to have more difficulties with metaphorical processes than metonymic ones (Semenza et al., 1980; Klepousniotou & Baum, 2005).

In considering how different pragmatic uses are processed, we can nevertheless rely on the extensive literature on metaphor and – to a lesser extent – metonymy. It has been shown that metaphor processing is influenced by many factors, such as context, familiarity, difficulty, novelty (Gibbs, 1994; Giora, 2003; Cardillo et al., 2010). However, when placed in a minimal context and controlled for the other factors, processing metaphorical expressions still requires additional effort measured both at the behavioral level (Cacciari &

Glucksberg, 1994; Noveck et al., 2001; Bosco et al., 2009) and in terms of brain response (De Grauwe et al., 2010; Bambini et al., 2011). Metonymy, on the contrary, has produced mixed results. Neurophysiological and neuroimaging evidence for a difference between metonymy and literal processing has been reported, although with no visible effects in terms of behavioral response (Schumacher, 2011; Rapp et al., 2011). Eye-tracking studies showed no differences compared to literal interpretations when metonymies are licensed by the context and mediated by a common metonymic convention (Frisson & Pickering, 1999; Frisson, 2009). As for approximation, to the best of our knowledge it hasn't received empirical consideration up to now. However, it has been shown that other types of loose use, such as hyperbole, are read faster than metaphor (Deamer et al., 2010), lending support to the gradient of meaning extension.

From the perspective of experimental pragmatics (Noveck & Reboul, 2008), we attempted to formulate empirical predictions out of the cognitive pragmatic accounts discussed above. Following Relevance Theory, metaphor and approximation are the result of the same conceptual adjustment process, differing solely in the degree and in the direction of this process. Approximation involves only a marginal broadening, while metaphor involves a wider broadening (Wilson, 2003) or broadening coupled with narrowing (Carston & Wearing, 2011). In this sense, metaphor seems to require more meaning modulation than approximation. It may be hypothesized that more meaning modulation would require higher costs of interpretation, thus predicting: (i) higher interpretation costs for metaphor and approximation in comparison with literal expressions; (ii) higher interpretation costs for metaphor in comparison with approximation.

Metonymy, on the contrary, seems to rely on specific mechanisms. The hypothesis – to a certain extent shared by Relevance Theory, Cognitive Linguistics and other frameworks – is that metonymy is supported by conceptual processes different from those involved in metaphor processing, such as meaning shift operations within the same conceptual domain. Re-formulating in processing terms, our prediction is that (iii) metonymy could exhibit different interpretation costs with respect to metaphor and approximation. More specifically, based on previous developmental and neuropsychological evidence, metonymic interpretation could come with no extra cost with respect to literal comprehension, at least when the transfer type is routinized (e.g., *producer for product*).

In order to provide empirical evidence in favor of either a distinction or a unified view of the three phenomena, we compared interpretation availability and costs for metaphor, metonymy, and approximation through a timed sensicality judgment paradigm, where participants are asked to decide quickly if a sentence is meaningful or not, and their performance is measured in terms of accuracy and reaction times. This experimental paradigm seems especially suitable to explore meaning modulation, as it requires subjects not only to access the linguistic items but also to elaborate and interpret their meanings at the level of detail that would distinguish different senses (Klein & Murphy, 2001).

Below we will first present detailed background on the construction of the stimulus material, i.e., a *de novo* built set of Italian metaphors, metonymies and approximations with corresponding literal and anomalous counterparts. Then, we describe a rating study in which all sentences were normed for the major psycholinguistic properties, namely meaningfulness, difficulty, cloze probability and familiarity, in order to obtain a pool of stimuli especially controlled for their interpretability. Finally, we go back to the kinds of conceptual adjustments and how they might result in different sensicality judgment responses.

2. Rating study

In building an experimental set of different pragmatic uses, two major issues emerge: first the need to rule out possible confounding effects due to sentential and contextual environment, and second the need to control for a number of psycholinguistic variables that are well known to influence figurative language.

As for the first point, for each of the three pragmatic phenomena under investigation (metaphor, metonymy and approximation) we constructed a set of Italian sentences of the form 'That Y verb X', where X was the word triggering the pragmatic interpretation and was taken as the target word for the experimental measures. Given the well-documented role of context in facilitating figurative language processing (Gibbs, 1994; Schumacher, 2012), across sets context was set to a minimal yet sufficient level for interpretation, in order to allow distinct pragmatic mechanisms to emerge neatly. Literal and anomalous counterparts were created for each set by selecting different subject nouns 'Y' or different verbs. All the anomalies contained a world knowledge violation. For the metaphor and approximation sets, the anomalous condition resulted from the clash of two incompatible semantic fields, while in the metonymy set the anomalies are related to selectional properties of the lexicon, and specifically of the verbs (see below for set specific criteria). It is also possible to create world knowledge anomalies where all the elements pertains to the same semantic fields and what is wrong is the combination of them (for example, due to causal effects, i.e., 'to dry with water'). However, we left this type of anomaly aside and we concentrated on most common anomalies based on semantic clash.

Among the many variables involved in figurative language, the three sets were rated for meaningfulness, difficulty, cloze probability and familiarity. The importance of these variables has been extensively described for metaphor processing (Kintsch & Bowles, 2002; Cardillo et al., 2010), partially addressed for metonymy processing (Frisson & Pickering, 1999), and never explored for approximation. In the perspective of the timed sensicality judgments to be collected afterwards, the main purpose of this rating study was to assess the meaningfulness of the experimental items, i.e., the interpretability of the sense of the utterances, along with their difficulty, i.e., the overall ease of interpretation.

As for familiarity, we aimed at setting a medium level of familiarity for the pragmatic uses, in order to exclude both fully conventionalized expressions – that could be processed as idioms rather than through pragmatic adjustment – and highly creative expressions – that could demand special pragmatic processes or even result senseless. The familiarity dimension was differently operationalized for each pragmatic phenomenon under investigation. Following the main literature on metaphor, familiarity was assessed by asking participants to rate frequency of experience for each metaphorical sentence. For the metonymy set, we devised a world-knowledge task to control for both the familiarity of the names used as target words and the familiarity of the metonymic transfer for those names. For the approximation set, a typicality task was used, where participants rated how appropriately the adjectives used as target words X qualify the subject words Y.

Lastly, to ensure that context was kept minimal and equal across conditions, we tested the contextual expectancy of each target word X for each sentence in the triplet and for all sets through a cloze probability task.

2.1 Materials and Methods

2.1.1 Participants

Eighty-five native speakers of Italian (42 F /43 M, mean age = 26.85 ± 3.80, mean schooling years = 18.02 ± 2.04 years of education) completed the questionnaire. Participants were unaware of the aim of the questionnaire and were not informed about the inclusion of figurative language. They gave written consent to participate after receiving an explanation of the procedures, according to the Declaration of Helsinki.

2.1.2 Stimuli

For each phenomenon under consideration (metaphor, metonymy and approximation) we constructed a set of forty-eight triplets including sentences with the pragmatic use, literal and anomalous counterparts (henceforth, for the sake of brevity, the label 'pragmatic sentences' will be used to refer to the pragmatic use condition for each set). The triplets were designed according to the criteria below, resulting in a total pool of 432 sentences. Table 1 shows an example of triplets from each set (Metaphor set, Metonymy set, Approximation set).

Metaphor set

We constructed nominal metaphors where a noun X is the vehicle for the metaphorical meaning (e.g., 'Those dancers are butterflies'). For each noun, one literal sentence (e.g., 'Those insects are butterflies') and one anomalous sentence (e.g., 'Those bottles are butterflies') was created. Literal sentences were obtained by using semantically compatible terms, while anomalous sentences resulted from the clash of two semantically non homogeneous terms. Each sentence was constructed in such a way that the first NP was a subject and the second was a predicate, that is only canonical copular sentences in the sense of Moro (1997). This was made to exclude inverse copular constructions which would have shifted the focus on the post copular noun phrase, unbalancing the stimuli. Plural forms were used in order to avoid predictability effects carried by the gender-marked articles required in the singular forms.

Metonymy set

We built a set of producer-for-product metonymies, where proper names of well-known Italian people were metonymically used to refer to objects. Different types of *producer for product* shift were used, such as author for book (e.g., 'That student reads Camilleri'), musician for song, designer for manufacture, painter for painting. In terms of Ruiz de Mendoza Ibáñez's taxonomy (2007), all metonymies were of the type target-in-source, i.e., the product is a subdomain of the producer. Each proper name X was also combined with different subject nouns Ys and different verbs, resulting once in a literal sentence (e.g., 'That journalist interviews Camilleri') and once in an anomalous sentence (e.g., 'That chef cooks Camilleri'). In order to confine metonymic interpretation to the target word, subject nouns and verbs were syntactically and semantically congruent in all conditions. Anomalous sentences resulted from the violation of selectional properties of the lexicon, i.e., animate objects were used for verbs selecting inanimate objects. Names of presently popular Italian people (e.g., Camilleri, Vasco) were chosen instead of very famous people from the past (e.g., Dante, Verdi) in order to reduce conventionality, as it has been suggested that the use of famous names (e.g., Dickens) in the metonymic form might have become lexicalized in ordinary language (Frisson & Pickering, 2007).

Approximation set

Among the different cases of approximate uses (Wilson, 2003; Wilson & Carston, 2006), we focused on adjectives. Following the examples provided by Wilson & Carston (2007), four main types of adjectives used in an approximate fashion were included: sense-related (e.g., 'Those tires are smooth'), geometric-related (e.g., 'Those sunglasses are rectangular'), color-related (e.g., 'Those clouds are black'), and negative-related adjectives (e.g., 'Those strawberries are tasteless'). For each target word X, we created a literal sentence by selecting a prototypical exemplar having the property described by the adjective (e.g., 'Those marbles are smooth'), and an anomalous sentence (e.g., 'Those restaurants are smooth'). As in the metaphor set, anomalous condition resulted from the clash of two semantically incompatible terms. As in the metaphor set, all sentences were copular constructions in order to reduce morphological factors and get both the subject and the predicate implemented with the same category.

2.1.3 Tasks

Meaningfulness and *Difficulty tasks*. We asked participants to rate on a five-point Likert scale how meaningful each sentence was (1 = meaningless; 5 = very meaningful). Each sentence was presented one at a time, and participants selected the value of the scale representing their judgment. Next, participants were asked to rate how difficult it was to rate the meaningfulness for that item, on a scale from 1 (very easy) to 5 (very difficult). All sets were tested.

Familiarity task. For each item in the Metaphor set, participants were instructed to indicate the frequency of experience with the sentence on a Likert scale from 1 (very unfamiliar) to 5 (very familiar).

World Knowledge task. For each proper name used in the Metonymy set, participants were instructed to associate the proper name with the corresponding product, choosing between four options. The options vary according to the type of metonymy (e.g., for Camilleri, the options were: book / song / movie / painting). This should account for both the familiarity of the proper names and the familiarity of the producer-for-product transfer.

Typicality judgments task. We asked participants to indicate how appropriate a given adjective (e.g., 'smooth') is to qualify three different nouns (e.g., 'marble', 'tires' and 'restaurants'), which corresponded to the nouns used in the triplet. A 5-point Likert scale (1 = very inappropriate; 5 = very appropriate) was available for each noun. This should assess both the familiarity of the approximate use and the literal use.

Cloze probability task. Each sentence was truncated before the target word, and participants were asked to complete with the first word that came to mind. All sets were tested.

2.1.4 Procedure

To preserve a high level of attention and avoid fatigue, two different questionnaires were created. Questionnaire 1 included three tasks: meaningfulness coupled with difficulty, world knowledge and typicality. Questionnaire 2 included cloze probability and familiarity tasks. For each questionnaire, the pool of 432 sentences was inserted into six different lists. Number of pragmatic, literal and anomalous sentences from each set was equally subdivided in the different lists and tasks. The lists were rotated among tasks so that each sentence was judged only once by each participant. The order of the tasks was counterbalanced across participants using a Latin Square procedure. Within each task the order of sentences was randomized. One group of the participants completed one of the six lists of Questionnaire 1, the other group completed one of the six lists of Questionnaire 2 (number of data points per item per task ≥ 6).

Ratings were administered online through Survey Monkey software (SurveyMonkey.com, LCC, Palo Alto, California, USA, www.surveymonkey.com). Each participant completed the questionnaire on a computer console, after giving informed consent through the same on-line procedure and reading online instructions. Each questionnaire lasted approximately 30 minutes.

2.2 Results

Inclusion criteria

Since the main aim of the ratings was to ensure the interpretability of the pragmatic sentences in untimed conditions for the purpose of the timed sensicality judgment task, we excluded pragmatic sentences for which both the following criteria were satisfied: (i) median score equal to 1 for the meaningfulness scale and median score ≥ 3 for the difficulty scale; (ii) depending on the set, for metaphors: median score equal to 5 or to 1 for the familiarity scale; for metonymies: cases in which less than 80% of participants correctly associated the producer with the corresponding product; for approximation: approximations for which the adjective-nouns pair scored < 2 (Mdn) on the typicality judgment scale. Literal and anomalous counterparts of the excluded pragmatic sentences were dropped as well.

From the original pool of sentences, 6 triplets were eliminated from each set. Final stimuli comprised 42 triplets for each of the 3 sets, resulting in a total of 378 sentences. In the following, we only report rating results for the final pool of sentences, to be further employed in the timed sensicality task.

Linguistic measures

Since the target word X was constant in the pragmatic, literal, and anomalous sentences of each triple, length and frequency were exactly balanced within each set. Length of the target words was also balanced across sets (mean number of characters: metaphor = 7.07; metonymy: = 7.16; approximation: = 7.49; $F_{(2,123)} = 1.42$, $p = 0.24$). Frequency of the target words were controlled for metaphor and approximation (mean log frequency: metaphor = 1.51; approximation = 1.54; $F_{(1,82)} = 0.43$, $p = 0.83$) based on a 3 million words database of written Italian, fully lemmatized and annotated (Corpus e Lessico di Frequenza dell'Italiano Scritto, CoLFIS, Bertinetto et al., 2005), available through the web interface EsploraCoLFIS (Bambini & Trevisan, 2012). No values were available in the database for the proper names used in the metonymy sets. Ratings collected in the world knowledge task should suffice as a measure of subjective frequency (see below). Overall, average frequency based on the values of all content words of the sentence was balanced across sets (mean log frequency: metaphor = 1.92; metonymy: = 1.85; approximation: = 2.04; $F_{(2,123)} = 1.47$, $p = 0.23$).

Rating results

We applied nonparametric methods since the assumptions underlying the use of parametric tests were violated in our sets (Knapp, 1990; Jamieson, 2004; Carifio & Perla, 2008; Norman, 2010). For descriptive statistics, we used median as a measure of central tendency and interquartile range as a measure of dispersion. For each set, Kruskal-Wallis test was performed on raw data to assess whether there were overall differences across pragmatic, literal and anomalous sentences (meaning modulation factor). We used post hoc Mann–Whitney U tests with Bonferroni correction for multiple comparisons (true alpha level = 0.0167) to determine which of the three types of sentences differed from each other. We also conducted parametric statistics on rank transformed data (Conover & Iman, 1981). In all cases, the results confirmed those obtained with the nonparametric procedure, and will not be reported in the results section. Table 2 presents descriptive statistics of the key psycholinguistic variables computed for each set.

Meaningfulness and difficulty data were further explored across sets through correspondence analysis (PASW Statistics 18.0.0 and R 2.13.0, “languageR” package, Baayen, 2011), an explorative computational method for interpreting categorical variables (Greenacre, 1993). The correspondence analysis tests the association between two variables organized into a contingency table and seeks to provide a low dimensional map of the association between rows and columns of the contingency table. By means of this analysis, the dimension of the original space is reduced, and an optimal subspace – closest to the cloud of points in the chi square-metric – is found. The information’s loss due to the dimension reduction is represented by the inertia explained by the axes of the map. The number of dimensions (hereafter called ‘factors’, according to Baayen, 2011) needed to explain the variation in the data were determined by using the scree plot. In the scree plot the factors’ eigenvalues were plotted in order of magnitude from largest to smallest, and the point where there was a marked drop in the amount of variation explained was considered: only those factors with inertia contribution higher than this point were selected for the interpretation. Hence, the coordinates of both row and column points of the contingency table were projected onto the selected low-dimensional subspace. In this spatial map, row and column points that are close together are more alike than points that are far apart. For each factor, we also plotted the mean coordinates of the points of each sentence type by means of bar plots in order to describe the distribution of points with respect to the different types of sentences across sets. Mean coordinates were then statistically compared.

Metaphor set

Both metaphorical and literal sentences scored median 4 on the meaningfulness scale (metaphorical: Mdn = 4, iqr = 2-4; literal: Mdn = 4, iqr = 4-5), while anomalous sentences were rated as meaningless (Mdn = 1, iqr = 1-2). The effect of meaning modulation (three levels: pragmatic, literal, anomalous) was found significant ($\chi^2_{(2)} = 432.84$, $p < 0.001$). Metaphorical and literal sentences significantly differed from anomalous counterparts (metaphorical vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). Although both metaphorical and literal scores were in the upper end of the scale, statistically literal sentences resulted more meaningful than metaphorical sentences (literal vs. metaphorical, $p < 0.001$), probably due to a greater dispersion for metaphors. In all cases, participants formulated their judgments about the sense/nonsense of the sentences with no difficulty (metaphorical: Mdn = 2, iqr = 1-2; literal: Mdn = 1, iqr = 1-2; anomalous: Mdn = 1, iqr = 1-2), although there was an effect of meaning modulation ($\chi^2_{(2)}$ = 29.59, $p < 0.001$), due to higher scores for metaphor (metaphorical vs. literal, $p < 0.001$; metaphorical vs. anomalous, $p < 0.001$). Familiarity ratings showed that metaphorical sentences received medium values (Mdn = 3, iqr = 1-4). Literal sentences scored higher (Mdn = 4, iqr = 3-5), and the difference was significant ($\chi^2_{(2)}$ = 495.71, p < 0.001; literal vs. metaphorical, $p < 0.001$). Cloze probability was very low throughout the set, scoring 0.00% for metaphorical and anomalous sentences, and 0.39% for literal sentences.

Metonymy set

Metonymic and literal sentences received high scores on the meaningfulness scale (metonymic: Mdn = 4, iqr = 3-5; literal: Mdn = 5, iqr = 4-5), while anomalous sentences scored median 1 (iqr = 1-2). Meaning modulation yielded a significant effect ($\chi^2_{(2)}$ = 446.02, $p < 0.001$), with metonymic and literal items more meaningful than anomalies (metonymic vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). As in the metaphor set, scores for both the metonymic and the literal items were at the upper end of the scale, but the comparison was significant (literal vs. metonymy, $p < 0.001$). Difficulty was very low across conditions (in all cases, Mdn = 1, iqr =1-2). Nevertheless, we observed an effect of meaning modulation ($\chi^2_{(2)}$ = 14.69, $p < 0.001$), and the comparison between metonymies and literal sentences was significant ($p < 0.001$). World knowledge task showed that participants correctly associated the producer with the product (accuracy = 90.29%). Cloze probability was 0.00% for any version of any item.

Approximation set

On the meaningfulness scale, both approximations and literal sentences received high scores, while anomalous sentences received low scores (approximate: Mdn = 4, iqr = 3-5; literal: Mdn = 5, iqr = 4-5; anomalous: Mdn = 2, iqr = 1-2). A significant effect of meaning modulation was found ($\chi^2_{(2)}$ = 451.60, $p < 0.001$): both approximation and literal sentences were judged more meaningful than anomalies (approximate vs. anomalous, $p < 0.001$; literal vs. anomalous, $p < 0.001$). Consistently with findings on metaphor and metonymy, literal sentences were more meaningful than approximation sentences (p < 0.001). Difficulty was low throughout the set (approximate: Mdn = 1, iqr = 1-2; literal: Mdn = 1, iqr = 1-1; anomalous: Mdn = 2, iqr = 1-2), although the comparison between approximation and literal sentences reached significance ($\chi^2_{(2)}$ = 101.89, p < 0.001). Results of the typicality task showed that the adjectives were judged moderately appropriate when referred to the nouns used in the approximations (Mdn = 4, iqr = 3-4), and fully appropriate when referred to the nouns used in the literal sentences (Mdn = 5, iqr = 4-5), while they were rated inappropriate in combination with the nouns from the anomalous sentences (Mdn = 1, iqr = 1-1). All comparisons were significant ($\chi^2_{(2)} = 560.61$, $p < 0.001$; *p*'s < 0.001). Cloze probability remained below the threshold of 12%, with averaged values of 1.66% for approximations, 11.44% for literal and 0.39% for anomalous expressions.

Meaningfulness and difficulty across sets

A synthetic view of the similarity among the different types of sentences across sets with respect to meaningfulness and difficulty is provided through the correspondence analysis. For meaningfulness, a significant model was generated ($\chi2(1508) = 3437.10$, $p < 0.001$). The scree plot revealed a marked decrease in the proportion of inertia explained by the second and subsequent eigenvalues; accordingly, only the first factor was considered to be interpreted. The first factor, accounting for 49.87% of the total inertia, roughly revealed a segregation of the different types of sentences into two clusters: literal and pragmatic sentences belonging to the three sets are on the left side of the map, whereas all anomalous sentences are on the right side (Figure 1). Hence, the first factor seems to indicate that all pragmatic and literal sentences were similarly scored on the meaningfulness scale. Anomalous sentences differed from pragmatic and literal ones, being collocated apart. It can also be observed that, among meaningful sentences, on the one hand literal sentences were clustered together, on the other hand approximate, metonymic and metaphorical sentences were close together. By statistically comparing the mean coordinates along the first factor, a significant difference was found between anomalous and both pragmatic and literal sentences ($F(2,375) = 987.40$, $p < 0.001$; Bonferroni post hoc comparisons, $p < 0.001$). Significant difference was also found between literal and pragmatic sentences (Bonferroni post hoc comparisons, $p = 0.002$). Considering the first factor with respect to the Likert scores, we observed that it was organized from right to left according to the ascending order of the Likert scale values.

The correspondence analysis was also applied to difficulty ratings. A significant model was generated ($\chi2(1508) = 2067.82$, $p < 0.001$). The resulting scree plot revealed a marked decrease in the proportion of inertia explained by the third and subsequent eigenvalues; hence, the first factor (explaining the 33.41% of the total inertia) and the second factor (explaining the 26.61 % of the total inertia) were interpreted.

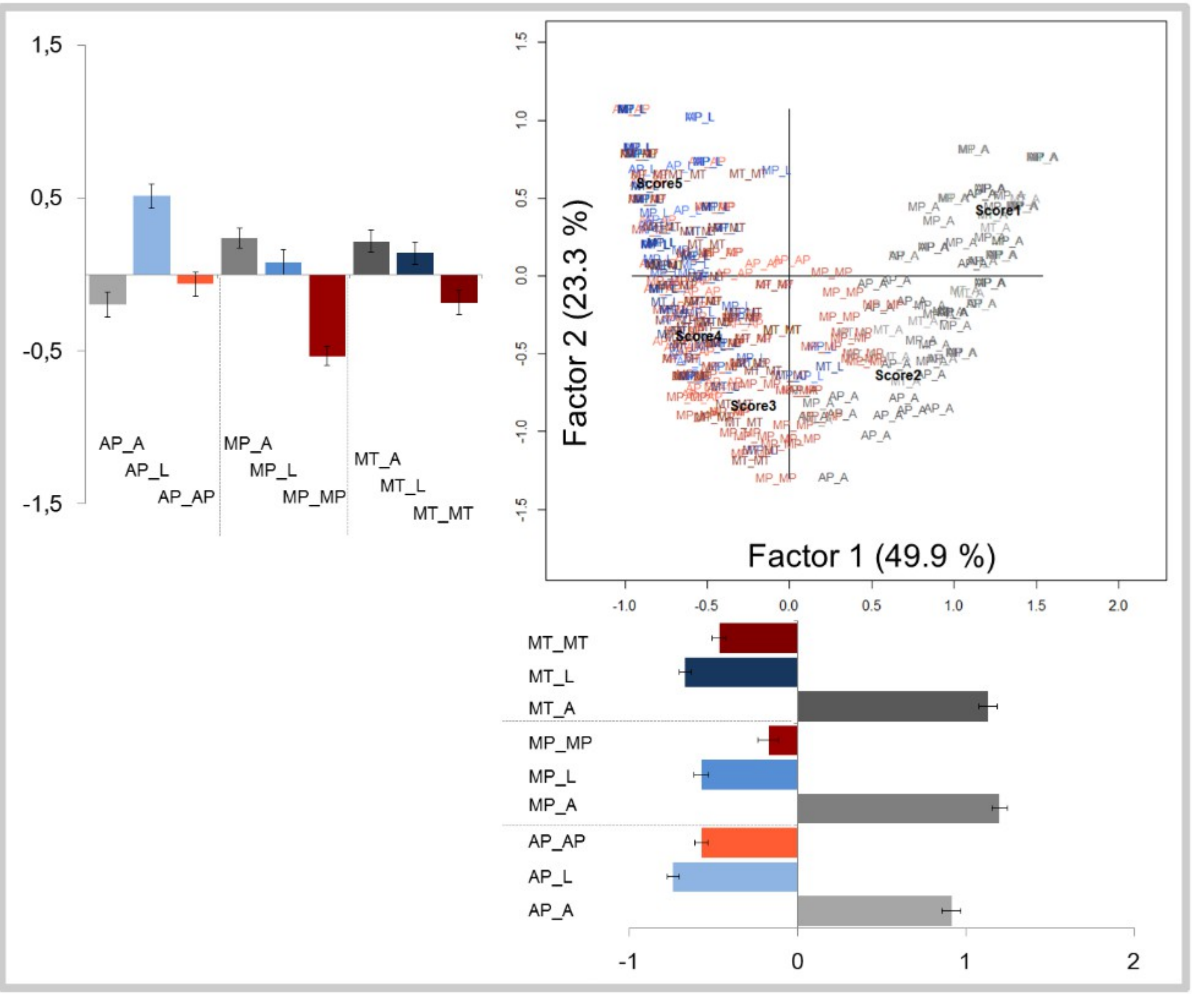


Figure 1. Correspondence analysis for meaningfulness ratings. The 378 sentences belonging to the three sets (Metaphor set = MP; Metonymy set = MT; Approximation set = AP) and the 5 Likert points are plotted at their corresponding coordinates. MP_MP: Metaphor set – metaphorical sentences; MP_L: Metaphor set – literal sentences; MP_A: Metaphor set – anomalous sentences; MT_MT: Metonymy set – metonymic sentences; MT_L: Metonymy set – literal sentences; MT_A: Metonymy set – anomalous sentences; AP_AP: Approximation set – approximate sentences; AP_L: Approximation set – literal sentences; AP_A: Approximation set – anomalous sentences. Pragmatic sentences are shown in magenta (MP_MP = magenta; MT_MT = dark magenta; AP_AP = light magenta); literal sentences are shown in blue (MP_L = blue; MT_L = dark blue; AP_L = light blue); anomalous sentences are shown in grey (MP_A = grey; MT_A = dark grey; AP_L = light grey). Barplots indicate mean coordinates for each factor and sentence types; error bars indicate standard error means.

As shown in Figure 2, the majority of sentences were close together, clustered in the upper part of the plot, suggesting that sentences were almost perceived as similarly difficult to be understood. However, upon a closer inspection, the first factor seems to reveal a distinction between literal sentences (on the right side) and anomalous ones (on the left side). Pragmatic sentences showed a more dispersed distribution, with metaphors mainly distributed on the left side and approximations mainly distributed on the right side of the map. By statistically comparing the coordinates along the first factor, literal sentences significantly differ from both pragmatic and anomalous sentences ($F(2,375) = 27.50$, $p < 0.001$; Bonferroni post hoc comparisons:

pragmatic vs. literal, p < 0.001; anomalous vs. literal, p < 0.001). With respect to the Likert scores, we observed a separation between score 1 and all other score values.

By visually inspecting the second factor, we observed a rough separation of literal and pragmatic sentences (in the upper half of the map) from anomalous sentences (on the bottom half of the map), thus most likely reflecting the similarity between literal and pragmatic sentences as compared to anomalous ones. This interpretation was further supported by the statistical analysis on the mean coordinates on the second factor (F(2,375) = 27.16, p < 0.001; Bonferroni post hoc comparisons: pragmatic vs. anomalous, p < 0.001; literal vs. anomalous, p< 0.001; pragmatic vs. literal, p = 0.38).

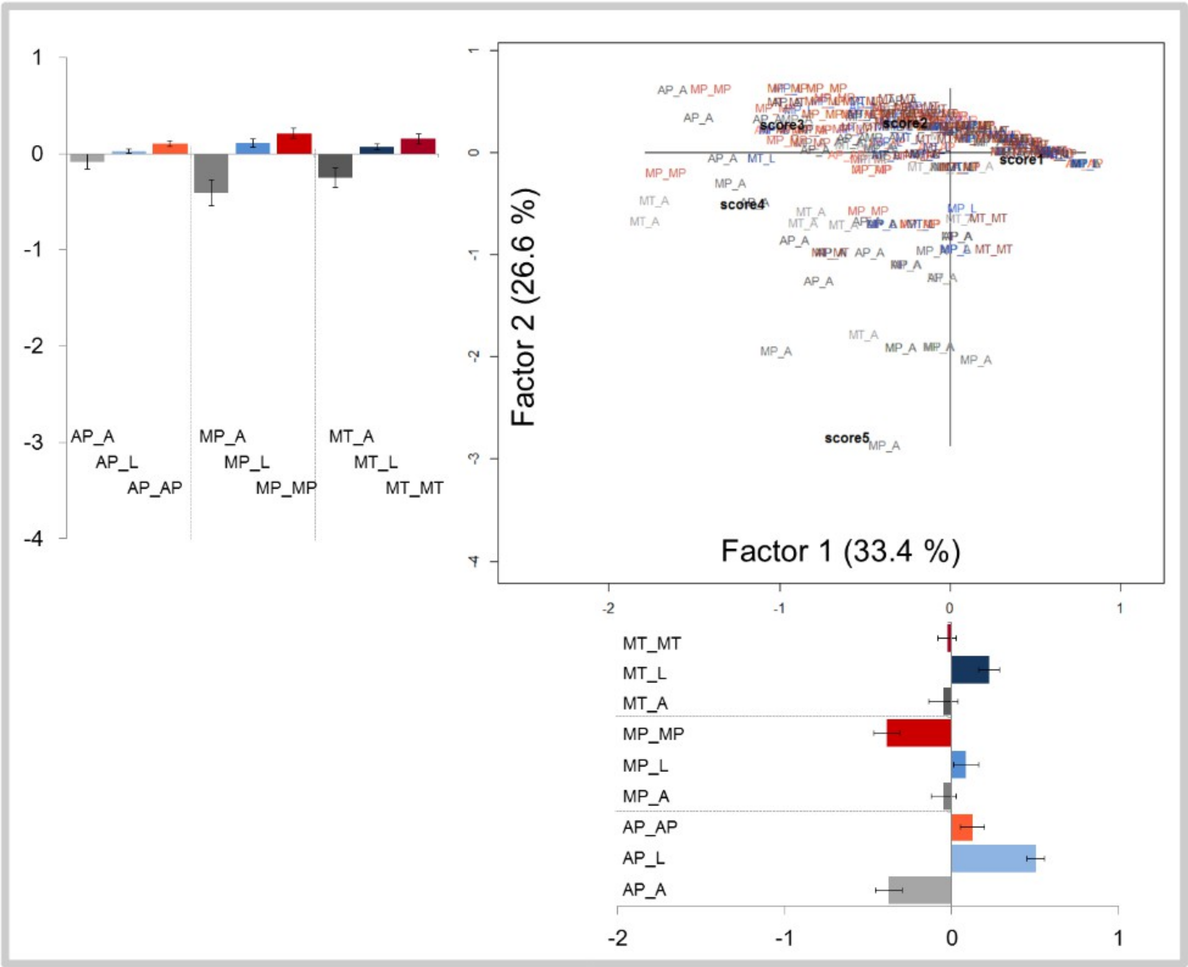


Figure 2. Correspondence analysis for difficulty ratings. The 378 sentences belonging to the three sets (Metaphor set = MP; Metonymy set = MT; Approximation set = AP) and the 5 Likert points are plotted at their corresponding coordinates. MP_MP: Metaphor set – metaphorical sentences; MP_L: Metaphor set – literal sentences; MP_A: Metaphor set – anomalous sentences; MT_MT: Metonymy set – metonymic sentences; MT_L: Metonymy set – literal sentences; MT_A: Metonymy set – anomalous sentences; AP_AP: Approximation set – approximate sentences; AP_L: Approximation set – literal sentences; AP_A: Approximation set – anomalous sentences. Pragmatic sentences are shown in magenta (MP_MP = magenta; MT_MT = dark magenta; AP_AP = light magenta); literal sentences are shown in blue (MP_L = blue; MT_L = dark blue; AP_L = light blue); anomalous sentences are shown in grey (MP_A = grey; MT_A = dark grey; AP_L = light grey). Barplots indicate mean coordinates for each factor and sentence types; error bars indicate standard error means.

2.3 Discussion

Through a rating procedure, we built a set of tightly controlled metaphorical, metonymic and approximate uses (and literal and anomalous counterparts) based on the same sentence structure ‘That X verb Y’, characterized in terms of meaningfulness and difficulty. For each set, inferential analysis revealed that both pragmatic and literal sentences were rated as meaningful and differently from anomalous sentences, thus excluding the possibility that pragmatic items are interpreted as anomalous. Inferential analysis also pointed out differences within the group of meaningful sentences. Although pragmatic sentences were judged interpretable (with high median scores for meaningfulness and low median score for difficulty), literal uses scores higher in meaningfulness and lower in difficulty than pragmatic uses. This is probably related to the fact that the context prompting pragmatic mechanisms to emerge is a minimal one. However, these differences do not seem to jeopardize the general consistency of the sets. Accordingly, the correspondence analysis on meaningfulness shows that literal and pragmatic sentences of the three sets clustered together and were judged similarly for meaningfulness, clearly differing from anomalous sentences. It also showed the similarity among all pragmatic uses, as opposed to all literal counterparts. Furthermore, all sentences were judged similarly on the difficulty scale, suggesting that our stimuli were all easily interpretable although some specific characteristics are revealed. All literal sentences, regardless the specific set, resulted overall easy to be interpreted. More importantly, literal and pragmatic uses were similarly for difficulty and different from anomalous sentences.

All pragmatic uses were also controlled for familiarity. Metaphorical sentences received medium scores on the familiarity scale, which suggests that they were perceived neither as fully conventionalized nor as extremely creative. Likewise, for metonymy, participants correctly associated the proper names of the producers to the corresponding product, thus implicitly demonstrating the familiarity of the names and of the metonymic transfer, although not fully lexicalized. For approximation, data suggested that the selected adjectives where judged appropriate when referred to the nouns used in the approximations, yet less typical than when used literally, providing first quantitative evidence for the definition of the category of approximation.

The set of stimuli also appears to be well controlled for the contextual expectancy: cloze probability was very low across sets, never above 12% for any condition of any sets. Interestingly, participants never created metaphor nor approximation in completing the sentences. Strictly speaking, also for metonymy the cloze probability was equal to zero. However, we observed 7 cases in which the final word reported by the participants was a proper name, albeit different from the one used in the corresponding stimulus (e.g., original stimulus: ‘That writer translates *Fruttero*’; cloze probability results: ‘That writer translates… *Hesse / Sartre*’). These results highlighted that there were some verbs spontaneously used in their metonymic sense, while the probability of creating a metaphor or an approximation was not verified, suggesting that metonymy is somehow more prone to routinization.

Overall, the result of the ratings allows us to assume that potential differences in processing pragmatic and literal sentences in the timed sensicality judgments will not depend on significant psycholinguistic differences of materials, but will truly reflect distinct interpretations for the three phenomena.

As a final note, one may argue that target words had different syntactic functions across sets, being used predicatively in metaphor and approximation, and referentially in the case of metonymy. However, we believe that it was important to focus on standard uses of the three pragmatic phenomena, rather than maintaining the same target word at the price of less clear and prototypical pragmatic types.

3. Timed sensicality judgment study

The timed sensicality judgment task has been used as a valuable paradigm to explore interpretation assignment, at different levels of the linguistic structure. This paradigm has been widely employed in investigations targeting conceptual operations, including polysemy (Klein & Murphy, 2001) and compounds (Gagné, 2001). At the sentence level, sensicality judgments have been used to explore pragmatic interpretation of conjunctions (Bott et al., 2009) and processing literal, metaphorical and idiomatic expressions containing actions verbs (Cacciari & Pesciarelli, 2013). The advantage of the sensicality judgment task is that it requires not only to access but also to elaborate the meaning of the expression. Information can be gathered both on the availability of the correct interpretation under time pressure (measured in terms of accuracy, i.e., proportion of correct responses - judging a sensible expression to be sensical or a nonsense expression to be non-sensical) and on the costs of interpretation (measured in terms of latencies). Interestingly, sensicality judgments often recur as behavioral task in a number of experimental paradigms targeting figurative language processing, from Speed-Accuracy Tradeoff (McElree & Nordlie, 1999) to neurophysiological and neuroimaging studies (Arzouan et al., 2007; Lai et al., 2009; Rapp et al., 2011; Subramaniam et al., 2012).

Here we used timed sensicality judgments to explore interpretation assignment of different types of pragmatic uses (i.e., metaphor, metonymy and approximation) compared to literal counterparts. It could be hypothesized that: (i) the greater meaning modulation required by loose uses compared to literal uses might reflect in higher interpretation costs for both metaphorical and approximate expressions compared to literal expressions; (ii) in turn, the greater meaning modulation required by metaphor compared to approximation (broadening + narrowing vs. only narrowing) might reflect in higher interpretation costs for metaphorical relative to approximate expressions; (iii) as for metonymy, being based on a different conceptual operation with respect to metaphor and approximation, and being more subject to routinization, it is possible that no additional interpretation costs are required compared to literal interpretation.

3.1 Materials and Methods

3.1.1. Participants

Twenty-five native speakers of Italian (12 M /13 F; mean age = 25.32 ± 3.02 years; mean schooling years = 18.3 ± 3.03) participated in the study. Participants were unaware of the aim of the study, and not experts in linguistics or psycholinguistics. None of them had participated in the rating study. They gave written consent to participate after receiving an explanation of the procedures, according to the Declaration of Helsinki, and received a monetary reimbursement for their participation.

3.1.2 Stimuli

The final pool of sentences described in the previous section was used as stimuli, i.e., 42 triplets for each set (Metaphor, Metonymy, Approximation). Additional 42 anomalous sentences were included for each set, with the purpose of having a similar ratio of sense and non-sense items. In order to minimize potential effects related to the repetition of the target words X in the triplets, the additional items recombined other words in the set, partly by repeating the subject nouns Y (e.g., 'Those insects are tables', where 'insects' is the subject noun in the literal version of one triplet in the metaphor set; see Table 1) and partly by repeating the last word of the additional item (e.g., 'Those trousers are tables', where 'tables' is the last word of the additional item obtained as above). Furthermore, to reduce the proportion of pragmatically used words and avoid metalinguistic awareness on figurative language, the experimental items were intermixed with 594 fillers (66% sense, 33% non-sense), consisting of four word sentences, like the experimental stimuli. In total, there was a sense:non-sense ratio of 1.44:1, and pragmatic sentences represented 12% of the stimuli (4% metaphors, 4% metonymies, and 4% approximations).

3.1.3 Procedure

Each participant was tested individually. Stimulus presentation and response collection were all carried out on a personal computer, using Presentation© software (Version 14.9, www.neurobs.com). Each trial began with a fixation cross presented in the middle of the screen for 500 ms. Next, the sentence was presented word by word at a fixed rate (300 ms). After the final word, YES/NO appeared on the screen to indicate that participants could give their response. Participants were instructed to respond as quickly and accurately as possible, and to make a sensicality judgment by pressing the green button when the string was meaningful and the red button when the sentence was meaningless on an RB 530 response pad (SuperLab Pro, Cedrus Corporation). The assignment of red and green to the left and right keys was counterbalanced across participants. After response or time-out (4000 ms), there was a blank inter-trial interval of 1000 ms. Response times were measured from the offset of the target word X.

Each subject was presented with all sentences. To avoid fatigue, three experimental blocks were created. An equal number of pragmatic, literal, and anomalous sentences from each set were included in each block, along with an equal number of fillers. We assigned the members of each triplet and the additional anomalous counterpart to distinct blocks, in order to avoid long-distance priming effects. Within each block, sentences were presented in a random order, while the order of the block was pseudo-randomized across participants. Mandatory stops between experimental blocks were fixed. A training session including ten items preceded the experiment. Furthermore, two practice trials (not included in the analysis) were administered at the beginning of each block. Overall the experimental session lasted 1 hour.

3.2 Results

Responses faster than 250 ms and slower than 1750 ms were excluded from the analysis (10.4 % of the data). We also excluded data by two participants with overall accuracy rate lower than 80%, and by one participant with 40% of responses faster than 250 ms. We observed that for the Metaphor set one participant never answered correctly for metaphors. However he was not excluded from the analysis as his overall level of accuracy was higher than the 80% threshold. Similarly, one metaphorical item was never judged accurately, but it was not excluded from the analysis based on the results of the rating study.

In Table 3 accuracy rates and mean reaction times for correct responses for each experimental condition are reported.

Accuracy

A Univariate General Linear Model (PASW Statistics 18.0.0) with meaning modulation (three levels: pragmatic, literal, anomalous) and set type (three levels: metaphor set, metonymy set, approximation set) as fixed factors was carried out on accuracy rates, treating either subjects (F_1) or items (F_2) as a random factor. Results showed that both the meaning modulation factor ($F_{1\ (2,168)} = 60.74$, $p < 0.001$; $F_{2\ (2,328)} = 68.98$, $p < 0.001$) and the set type factor ($F_{1\ (2,168)} = 10.74$, $p < 0.001$; $F_{2\ (2,328)} = 12.41$, $p < 0.001$) were significant. Also their interaction was significant ($F_{1\ (4,168)} = 32.33$ $p < 0.001$; $F_{2\ (4,328)} = 37.47$, $p < 0.001$), indicating that the effect of one factor depends on the level of the other factor. We therefore explored the effect of meaning modulation set by set, focusing on the comparison between the pragmatic and the literal sentences. In the Metaphor set, this factor yielded significant effects ($F_{1\ (2,42)} = 68.43$, $p < 0.001$; $F_{2\ (2,82)} = 100.77$, $p < 0.001$), with metaphorical sentences being less accurate than literal sentences (Bonferroni post hoc comparisons, $p < 0.001$ both by subjects and by items). Also in the Metonymy set meaning modulation was significant ($F_{1\ (2,42)} = 12.73$, $p < 0.001$; $F_{2\ (2,82)} = 8.10$, $p = 0.001$). Differing from the Metaphor set, however, accuracy doesn't seem to vary for metonymic and literal sentences: post-hoc comparisons revealed only a marginal difference between the two conditions (Bonferroni post hoc comparisons, $p = 0.05$ in the by subject analysis, $p = 0.18$ in the by item analysis). In the Approximation set, again we observed a main effect of meaning modulation ($F_{1\ (2,42)} = 3.94$, $p = 0.02$; $F_{2\ (2,82)} = 8.49$, $p = 0.001$). Accuracy for approximation was significantly lower than for literal sentences in the by item analysis (Bonferroni post hoc comparisons, $p = 0.006$), although the difference was not significant in the by subject analysis (Bonferroni post hoc comparisons, $p = 0.12$). Overall, the data suggest a higher availability of literal uses as compared to metaphor and – to a lesser degree – approximation, but not for metonymy.

Latencies

Following the standard in analyzing response times, only trials in which participants responded correctly were included in the analysis.

The effect of meaning modulation (three levels: pragmatic vs. literal vs. anomalous) and set type (three levels: Metaphor set, Metonymy set, Approximation set) on response times were examined with Univariate General Linear Model treating either subjects (F_1) or items (F_2) as a random factor. We observed a significant effect of meaning modulation ($F_{1\ (2,\ 167)} = 12.23$, $p < 0.001$; $F_{2\ (2,\ 327)} = 24.97$, $p < 0.001$), as well as a significant effect of set type in the by item analysis and marginally significant in the by subject analysis ($F_{1\ (2,\ 167)} = 2.95$, $p = 0.05$; $F_{2\ (2,\ 327)} = 9.46$, $p < 0.001$). A significant interaction between meaning modulation and set type was found ($F_{1\ (4,\ 167)} = 3.77$, $p = 0.006$; $F_{2\ (4,\ 327)} = 8.70$, $p < 0.001$), as shown in Figure 3.

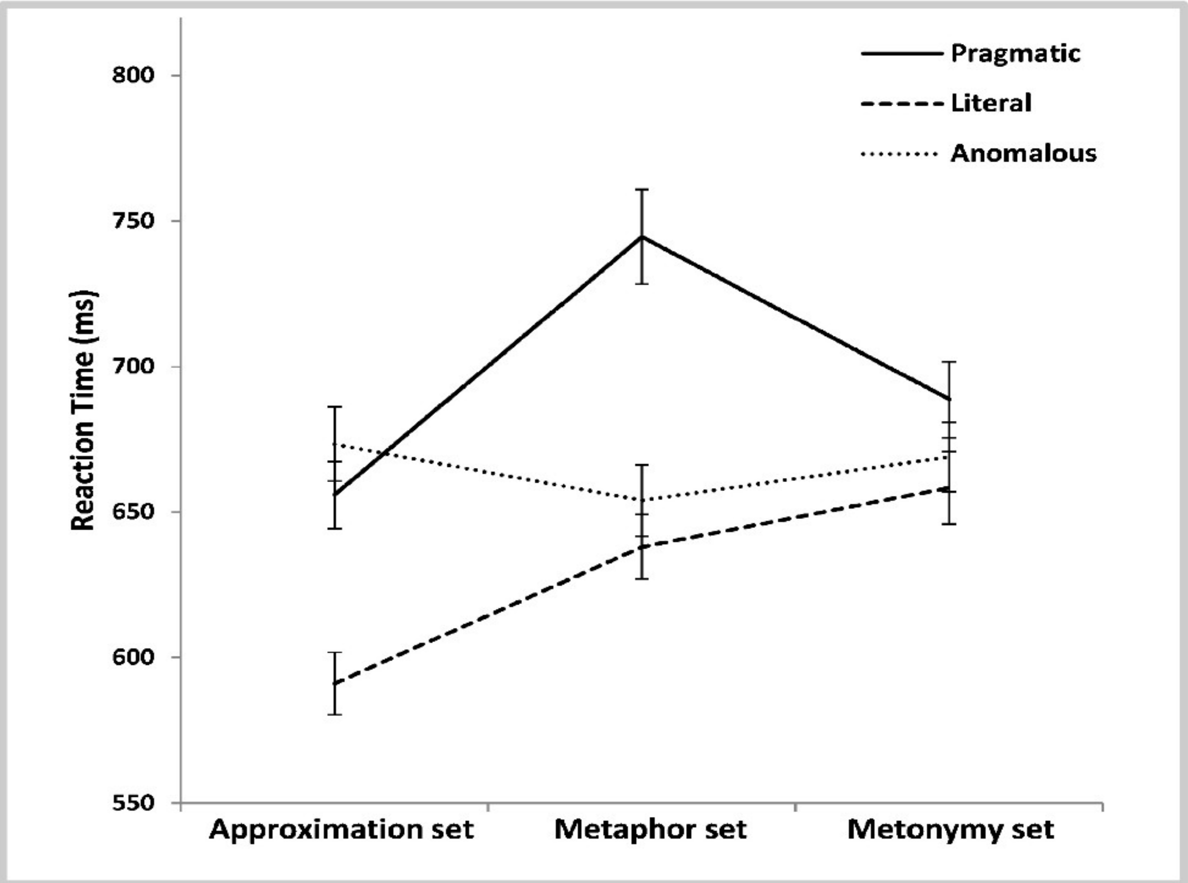


Figure 3. Mean reaction times (ms) for the Approximation set, the Metaphor set and the Metonymy set as a function of the meaning modulation factor. Pragmatic level is represented by the solid line, literal level by the dashed line, and anomalous level by the dotted line. Error bars indicate standard error.

In order to explore the interaction of meaning modulation and set type, simple effect analyses were conducted. As concerns the meaning modulation factor, in the Metaphor set we observed that metaphorical sentences were interpreted slower than literal counterparts ($F_{1\,(2,\,41)} = 5.85$, $p = 0.006$; $F_{2\,(2,\,81)} = 25.51$, $p < 0.001$; Bonferroni post hoc comparisons, $p < 0.05$ in the by subjects, $p < 0.001$ in the by items analysis). On the contrary, in the Metonymy set there were no differences across conditions ($F_{1\,(2,\,42)} = 0.74$, $p = 0.48$; $F_{2\,(2,\,82)} = 2.06$, $p = 0.13$), indicating that metonymic interpretation was reached as rapidly as literal interpretation. Similarly to the Metaphor set, the Approximation set showed a significant effect of the meaning modulation factor ($F_{1\,(2,\,42)} = 10.66$, $p < 0.001$; $F_{2\,(2,\,82)} = 8.37$, $p < 0.001$), with approximations interpreted slower than literal sentences (Bonferroni post hoc comparisons, $p = 0.006$ in the by subjects, $p = 0.003$ in the by items analysis).

We also assessed whether the type of pragmatic use has an effect on response times. Since pragmatic sentences differ in some respects – as needed to preserve clear pragmatic types, we avoided direct comparisons of metaphors, metonymies and approximations across sets. Rather, we measured the latency difference between the pragmatic condition and the literal condition for the corrected pairs of each set: (metaphor – literal), (metonymy – literal) and (approximation – literal), as represented in Figure 4.

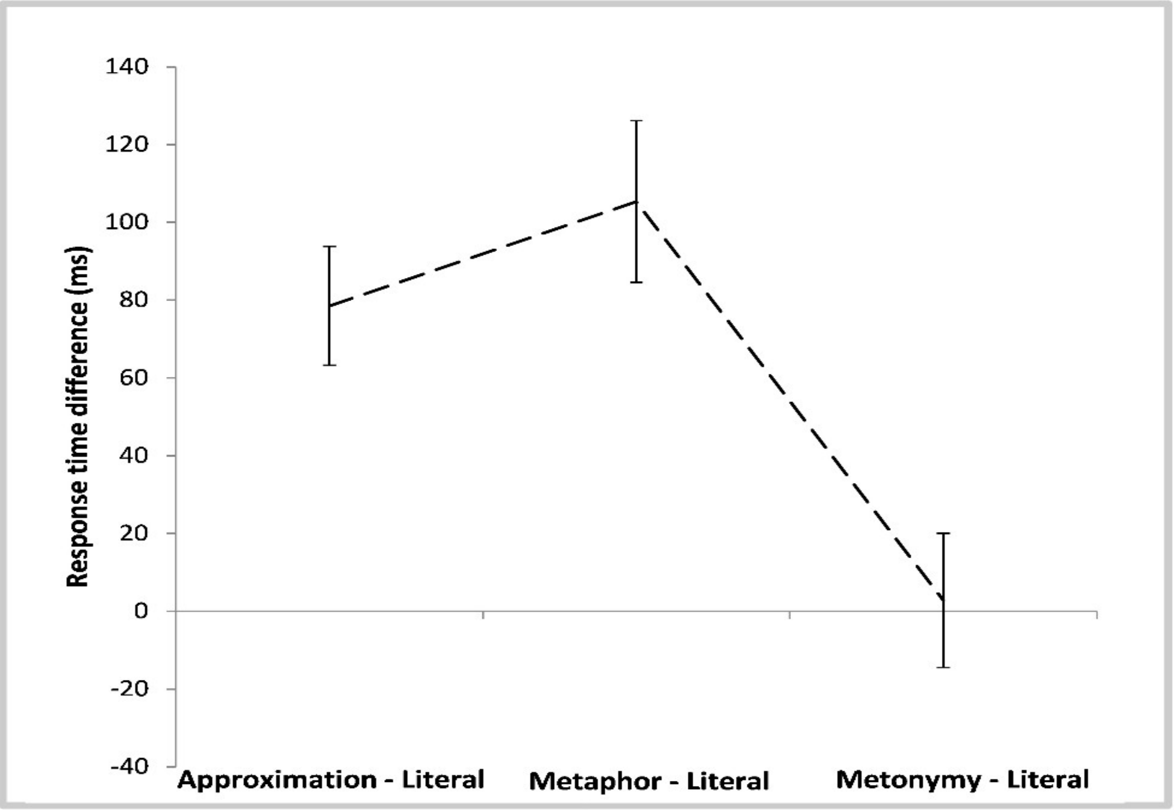


Figure 4. Reaction times differences (pragmatic minus literal) for the Approximation set, the Metaphor set and the Metonymy set. Error bars correspond to the standard error of the difference for each set.

The highest latency difference was obtained for metaphor ($M_{\text{metaphor – literal}}$ = 105.34 ms), followed by approximation ($M_{\text{approximation – literal}}$ = 78.50 ms), while a minimal latency difference was observed for metonymy ($M_{\text{metonymy – literal}}$ = 2.76 ms). The comparison reveals an effect of the type of pragmatic use ($F_{(2, 41)} = 14.14$, $p < 0.001$), with metaphor and approximation significantly different from metonymy, but not different from each other (Bonferroni post hoc comparisons: metaphor/approximation vs. metonymy, $p \leq 0.001$; metaphor vs. approximation, $p = 0.92$).

4. General discussion

The results of the rating study indicated that the three pragmatic uses were easily interpreted as meaningful in a manner similar to the corresponding literal constructions. However, timed sensicality judgments revealed that there are differences across metaphor, metonymy and approximation, reflected both in accuracy rates and latencies of interpretation. Extending the current theoretical models into experimental predictions, we expected (i) metaphors and approximations to exhibit higher interpretation costs with respect to literal interpretation; (ii) metaphors to be more costly with respect to approximations; and (iii) metonymies to behave differently, possibly in the direction of no additional costs with respect to literal interpretation. Our results seem to confirm these predictions and to show differences across the three pragmatic phenomena, reflected both in accuracy and response times.

First, meaning modulation, i.e. whether the expression was pragmatic, literal or anomalous, significantly affects the accuracy of the response. Assigning a pragmatic interpretation under the pressure of time seems to be more difficult than constructing a literal interpretation. There are, however, notable differences across pragmatic uses. For metaphor, the percentage of correct responses was around 50%. Although judged as meaningful in the offline rating, in timed sensicality judgments metaphors proved significantly more difficult to interpret than literal expressions. This piece of evidence is consistent with previous literature employing sensicality judgments in a variety of paradigms, from Speed Accuracy Tradeoff (McElree & Nordlie, 1999) to neurophysiological recording (Arzouan et al., 2007) and neuroimaging (Subramaniam et al., 2012). Accuracy around chance seems thus a common performance associated with the interpretation of nominal metaphor in speeded condition. Participants performed much better when they were presented with approximations and metonymies, reaching 87% and 83% respectively. This suggests higher availability for approximate and metonymic uses with respect to metaphorical use. Interestingly, we observed that metaphors and approximations – although less clearly – are interpreted less accurately than their literal counterparts, while there is a marginal or no difference between metonymies and literal expressions. This points in the direction of similarities in the interpretation style of metaphor and approximation, although associated with different degrees of availability. In turn, the availability of metonymy seems to equate that of literal interpretation. It should be noted that in the Metonymy set literal sentences reached only 89% accuracy, which is lower than accuracy literal sentences in other sets. This is probably related to the costs of processing proper names, which require the retrieval of stored knowledge (see Gorno-Tempini et al., 1998), also when they are well-known by the participants (as assessed in the rating study). However, this should not affect the comparison between metonymies and literal controls, as proper names were included in both conditions and their presence should not block interpretative differences to emerge, if any.

Second, pragmatic interpretation also reflects in time. Analyzing error-free trials, i.e. limiting the analysis to those cases where pragmatic sentences were judged to be sensical, we observed that interpreting pragmatic uses is not always slower than literal interpretation. This is the case for metaphor and approximation, but not for metonymy. These findings accord with behavioral literature on metaphor showing that, in minimal context and for not conventionalized expressions, metaphor processing requires extra costs compared to literal processing (Noveck et al., 2001). Here, we carefully controlled the sentential environment, by providing each pragmatic use with a minimal context, and the familiarity of the expressions, by avoiding lexicalized cases. Thus, the higher reaction times for metaphors seem to truly reflect extra costs required by the interpretation of metaphorically used words as compared to literal uses.

Results are also consistent with eye-tracking studies reporting no differences between metonymic and literal expressions (Frisson & Pickering, 1999, Pickering et al., 2004). It cannot go unnoticed, however, that, at the neural level, metonymic expressions elicit robust differences from literal comprehension (Schumacher, 2011; Rapp et al., 2011). It is up to future studies to elucidate whether this discrepancy is motivated by differences in the materials, either in conventionality or supportive context, or in the methodological techniques and the type of information they offer. Our view is that, when metonymy – like in our case – is based on common shifts such as *producer for product*, no matter the relative conventionality of the specific lexical items, in a minimal yet sufficient context, interpretation costs in speeded conditions closely mirror literal comprehension, and radically differ from those required by metaphor processing. Besides, this view is consistent with findings reported for other types of routinized meaning shift. For example, sensicality judgments on logical metonymy showed no differences in accuracy nor in latencies between the coerced and control conditions, yet again evoking neural differences (Brennan & Pylkkänen, 2008). By contrast, our results seem to be conflicting with the higher reading times for resolving metonymic referring expressions as compared to metaphorical referents presented in Gibbs (1990). However, those data are controversial, as potentially affected by the plausibility of the items (Frisson & Pickering, 1999), and obscuring some comparisons of interest (Noveck et al., 2001).The different behavior observed for metaphor and metonymy gains support from the results on approximation. We showed that adjectives used approximately are interpreted slower than the same adjectives used literally. This piece of evidence seems to place approximation closer to metaphor than to metonymy. This result seems to strengthen the distinction between the processing styles for metaphor and for metonymy, by introducing a third case that patterns alike the former but differently from the latter. Consonant with this are also the latency differences between the pragmatic and the literal conditions across sets. When we disentangle the costs of interpreting each type of pragmatic use, we see that approximation and metaphor are associated with extra costs, while metonymy doesn't prompt extra effort. Accuracy data, with metaphor and approximation as well (although to a lesser degree) departing from literality, and metonymy tending to equate it, are in harmony.

Collectively, this pattern of results carries importance for discussing theoretical accounts of the nature of pragmatic phenomena. The relevance-theoretic claim that metaphor and approximation both require conceptual adjusting of the linguistically encoded concept but in different degrees (Wilson, 2003) seems to be supported by our data, and specifically by the gradient observed in availability and latency. Also the direction of the conceptual adjustment may contribute to the different gradient observed in the sensicality judgments. According to Carston & Wearing (2011), hyperbole, is considered as a case of marginal broadening as opposed to metaphor involving a broadening coupled with narrowing. Extending this proposal further, also approximation could be taken as case of marginal broadening (possibly more marginal than in hyperbole) as opposed to the combination of broadening and narrowing supporting metaphor. According to this interpretation, higher difficulty and costs for metaphor compared to approximation might stem from a more complex operation – broadening and narrowing – with respect to the marginal broadening required for approximation processing. For the type of task used here, we do not have direct evidence to discriminate whether the difference between approximation and metaphor lays in marginal versus radical broadening or in marginal broadening versus a combination of radical broadening and narrowing. Intuitively, our data fit well with the degree claim posited by Relevance Theory, while the direction claim is less straightforwardly answerable. More sophisticated designs will be needed that manipulate the (degree of) direction of the adjustment, possibly exploring the temporal dynamics of the process or the conceptual properties that undergo manipulation. Granted this caveat, the general idea of a modulation in the underlying conceptual adjustment process seems to be well supported by our findings. Converging evidence comes from Deamer et al.'s (2010) reading time study, where hyperbolic uses were compared to metaphorical uses, showing that even a more substantial type of broadening such as hyperbole is distinct from metaphor. This study actually failed in finding a difference between hyperboles and literal expressions. This discrepancy is possibly related to contextual modulation: Deamer et al. (2010) used supportive contexts that might have facilitated hyperbole resolution and reduced the broadening, while we used a minimal sentential environment that allowed for the marginal extra costs required by approximation to emerge.

Also the hypothesis put forward in Carston (2010), i.e. that metonymy is not straightforwardly reducible to narrowing or broadening but involves some kind of shift seems to fit with our data. The different pattern of results observed for metonymy as opposed to metaphor might reflect different conceptual operations. Some support for this interpretation comes from acquisition data showing that metonymy not only is acquired at a faster rate than metaphor, but it also processed more accurately throughout childhood to adulthood (Rundblad & Annaz, 2010). In order to explain this finding, Rundblad and Annaz (2010) hypothesized a more basic type of conceptual operation for metonymy as opposed to metaphor. The results described for metonymy can also be reconciled with the Cognitive Linguistics account, to the extent that metaphor and metonymy are ascribed to distinct types of mappings: across domains for metaphor, within the same domain for metonymy. A greater cognitive distance between concepts can be assumed for metaphor (see also Rundblad & Annaz 2010) and might be reflected in higher difficulty and costs. As it is still difficult to translate the different types of mappings in terms of processing costs, we leave this for further research to develop.

Our data also point to reduced efforts for metonymy, and to the routinization of some types of metonymic shifts, such as *producer for product*. It might be of some interest here to report some qualitative insights from the post-experiments session: despite the very low percentage of metaphors in the sentence pool, some participants noticed their presence, while none seemed to notice metonymy, as if metonymic uses were more integrated in the lexical knowledge and less prominent in the speakers' metalinguistic awareness. Interestingly, in a developmental study, Annaz et al. (2009) observed a correlation of metonymic comprehension with the expansion of receptive vocabulary that might suggest that in some cases metonymic meanings might be part of the lexicon.

According to these observations, a possible distinction might be sketched between the combination of broadening and narrowing on the metaphor side, less pre-configured in direction and degree of the conceptual adjustment of the lexical concept, and conceptual shift on the metonymy side, based on more routinized patterns. Highly creative metonymic uses are possible as well (consider, for instance, 'The best pencils of the world gather together for the annual drawing convention'), and this might call upon higher interpretation costs. However, it seems psychologically implausible to posit different elaboration procedures for the same class of phenomena, as the difference between routinized *producer for product* cases and less typical *tool for worker* cases could probably be made not by different types of conceptual adjustment processes, but rather by the role of context.

As a final consideration, sensicality judgments are a good measure of the availability and difficulty of correct interpretation, but are limited to stages where the sense has already been construed, and do not account for online processing nor for the type of process involved (Frisson, 2009). Thus, our results shed light on the costs of interpretation assignment, and provide insights into interpretative style, but further investigations are needed to explore the temporal dynamics and the nature of the conceptual operations involved.

5. Conclusions

Behind a label such as figurative language, many different mechanisms are grouped. Although we assume that all require pragmatic inferencing to be interpreted, interpretation might come with different procedures, linked to different operations at the conceptual level. Through timed sensicality judgments recorded for different pragmatic uses in minimal context condition, we found that there are significant differences in the interpretation availability and costs of metaphor, metonymy and approximation. The findings support a theoretical distinction between metaphor and approximation, which seem to vary in degree and possibly in the direction of the underlying adjustment process, compatible with Relevance Theory, and an even more marked separation with metonymy, whose meaning shift might be subject to routinization.

With these data, we hope to have strengthened the empirical basis available on figurative language, by providing the first evidence in favor of the psychological reality of the phenomenon of approximation, and with a first attempt to answer the challenge raised by metonymy. We believe that deepening the understanding of the phenomena included under the realm of pragmatics, by pinpointing potential differences for the parser and elaborating on whether natural classes of cases can be identified on this basis, is one promising line of research for the experimental pragmatics enterprise.

6. Tables

Table 1. Examples of stimulus triplets for the Metaphor set, the Metonymy set and the Approximation set. Original Italian; English translation in italics.

	Pragmatic	**Literal**	**Anomalous**
Metaphor set	Quelle ballerine sono farfalle	Quegli insetti sono farfalle	Quelle bottiglie sono farfalle
	Those dancers are butterflies	*Those insects are butterflies*	*Those bottles are butterflies*
Metonymy set	Quello studente legge Camilleri	Quel giornalista intervista Camilleri	Quel cuoco cucina Camilleri
	That student reads Camilleri	*That reporter interviews Camilleri*	*That chef cooks Camilleri*
Approximation set	Quelle gomme sono lisce	Quel marmo è liscio	Quei ristoranti sono lisci
	Those tires are smooth	*That marble is smooth*	*Those restaurants are smooth*

Table 2. Descriptive statistics of rating scores for the Metaphor set, the Metonymy set and the Approximation set. Median and interquartile range (in brackets) are reported for meaningfulness, difficulty, and typicality tasks. Cloze probability and world knowledge results are reported in percentage.

		Pragmatic	**Literal**	**Anomalous**
Metaphor set	*Meaningfulness*	4 (2-4)	4 (4-5)	1 (1-2)
	Difficulty	2 (1-2)	1 (1-2)	1 (1-2)
	Familiarity	3 (1-4)	4 (3-5)	1 (1-1)
	Cloze Probability	0.00%	0.39%	0.00%
Metonymy set	*Meaningfulness*	4 (3-5)	5 (4-5)	1 (1-2)
	Difficulty	1 (1-2)	1 (1-2)	1 (1-2)
	World Knowledge	90.29%	-	-
	Cloze Probability	0.00%	0.00%	0.00%
Approximation set	*Meaningfulness*	4 (3-5)	5 (4-5)	2 (1-2)
	Difficulty	1 (1-2)	1 (1-1)	2 (1-2)
	Typicality	4 (3-4)	5 (4-5)	1 (1-1)
	Cloze Probability	1.66%	11.44%	0.39%

Table 3. Accuracy rates and mean reaction times (ms) for correct responses as a function of pragmatic modulation (pragmatic, literal, anomalous conditions) and set type (Metaphor set, Metonymy set, Approximation set). Standard deviations in parentheses.

	Metaphor set		**Metonymy set**		**Approximation set**	
	Accuracy	*Reaction time*	*Accuracy*	*Reaction time*	*Accuracy*	*Reaction time*
Pragmatic	0.52 (0.24)	744.66 (339.02)	0.83 (0.10)	688.70 (338.17)	0.87 (0.09)	655.96 (308.82)
Literal	0.95 (0.04)	638.10 (305.05)	0.89 (0.07)	658.32 (326.55)	0.95 (0.04)	591.13 (293.18)
Anomalous	0.97 (0.05)	653.98 (339.44)	0.95 (0.06)	669.01 (332.04)	0.85 (0.16)	673.37 (333.20)

Study D

The time course of figurative language processing: A speed-accuracy tradeoff study on metaphor and metonymy[5]

Abstract

Psycholinguistic accounts propose general models of figurative language processing, regardless the variety of figurative language uses. Pragmatic theory, however, makes finer distinctions, differentiating metaphor and metonymy on the basis of different processes of meaning enrichment. The present study tested general psycholinguistic predictions about the costs of processing metaphors and metonymies by employing a multi-response Speed Accuracy Trade-Off procedure. Results showed that both metaphor and metonymy processing resulted in lesser availability than their literal counterparts, whereas only metonymy also exhibited slower processing speed. The results are interpreted by extending and integrating the psycholinguistic discussion on figurative language with the theoretical issues concerning the differentiation of lexical enrichment phenomena.

---

[5] Ghio, M., Bott, L., Schumacher, P., Bambini, V. ,The time course of figurative language processing: A speed-accuracy tradeoff study on metaphor and metonymy (*in preparation*).

1. Introduction

Figurative language is a wide label usually used to refer to different types of meanings departing from what is literally said and requiring a pragmatic enrichment. This umbrella term includes several figures of speech (often called “tropes”) including, among others, metaphors, similes, ironies, idioms, proverbs, indirect requests, metonymies and hyperboles. Despite such a variety in linguistic distinctions, psycholinguistic accounts have generally tested only one type of figurative language, metaphor. In this study we asked whether conclusions about metaphor processing should be generalized to processing of other figurative tropes. Specifically, we tested whether metaphor and metonymy are processed differently. In doing so, we evaluated whether the standard psycholinguistic models should be augmented so that they make distinctions between tropes.

1.1 Figurative language processing

Classically, two classes of models have been proposed (for reviews, see Bambini & Resta, 2012; Gibbs & Colston, 2012). On the one hand, the comprehension of figurative language is assumed to require serial stages, including the analysis of the literal meaning, the rejection of the literal meaning because not appropriate in the context, and the construction of an alternative meaning appropriate in the context (Grice, 1975; Searle, 1979). This theoretical proposal was translated by the Standard Pragmatic model into experimental hypotheses about the time of processing: inasmuch as the comprehension of figurative meaning requires more steps than the interpretation of literal meaning, the processing of the former should take more time than the processing of the latter (Glucksberg & Keyser, 1993; Glucksberg, 2003). On the other hand, direct, one-stage models posit that figurative meaning is accessed without intermediate stages. For example, the Direct Access model emphasizes that, provided an appropriate context, figurative meanings are accessed directly (Gibbs, 2001). Specifically, the comprehension of figurative meaning does not mandatorily and automatically require the analysis of the complete literal meaning. Accordingly, processing figurative meanings does not necessarily take longer than processing literal meanings. More recently, other psycholinguistic models have been proposed sharing the assumption that figurative language is not always (or never) more difficult to process than non-figurative language, but rather different linguistic and non-linguistic constraints combine in parallel to construct the communicated meaning (Graded Salience Hypothesis, Giora, 2003; Underspecification view, Frisson, 2009).

The general consensus on figurative language processing is that the Standard Pragmatic model is not wholly correct (Gibbs & Colston, 2012). While there are disagreements about the techniques used in individual studies, it seems highly unlikely that people always (and completely) process a literal meaning before deriving the figurative meaning. Nevertheless, also the Direct Access model is not definitely confirmed by experimental evidence (Bambini & Resta, 2012). A more pertinent concern is that most figurative studies only test a single trope (generally metaphor; see Gibbs & Colston, 2012). But can the results of studies on metaphor be generalized to the processing of other figurative language uses? For example, if the Standard Pragmatic model should be rejected when considering metaphor, does this mean that it should be rejected when considering metonymy? There are at least two reasons to suggest that generalization from one trope to another is problematic.

First, cognitive pragmatic accounts, such as Relevance Theory (Sperber & Wilson, 1986/1995; Wilson & Carston, 2007; Carston, 2010) and Cognitive Linguistics (Lakoff, 1980; Evans & Green, 2006), increasingly point to a fine-grained description of the pragmatic processes involved in figurative language interpretation. Within the Relevance Theory framework, a difference between metaphor and metonymy has been discussed with respect to the conceptual operations leading to the construction of the communicative meaning. Assuming a continuum of cases ranging from literal to metaphorical expressions, metaphor has been described as a conceptual operation based on broadening and narrowing of the lexical concepts (e.g., in 'John is a shark', the concept SHARK is broadened in order to include some human beings; Wilson & Carston, 2007; Carston, 2010). By contrast, at least some cases of metonymy (i.e., reference transfer, e.g., 'The saxophone walked out') seem to be not reducible to broadening/narrowing operations, but to involve a reference substitution operation (Carston, 2010). Similarly, Cognitive Linguistics posits different cognitive operations underlying metaphor and metonymy (Ruiz de Mendoza Ibáñez, 2007). Metaphor consists in a mapping between two different conceptual domains (e.g., in 'Life is a journey', the target domain 'Life' is mapped on the source domain 'Journey'), whereas metonymy relies on a mapping within the same conceptual domain (e.g., in 'Shakespeare is on the top shelf', the source domain 'Shakespeare' provides access to its sub-domain 'Shakespeare's writings', which is the target domain; Lakoff & Johnson, 1980; Gibbs et al., 1999; Ruiz de Mendoza Ibáñez, 2007). Furthermore, metonymy seems to affect the grammatical structure, and the selection of lexical properties (Panther & Thornburg, 2003).

Second, the literature provides evidence, albeit fragmentary, that the processing costs of figurative meanings interpretation varies according to the specific figurative use (Gibbs & Colston, 2012). For example, Gibbs (1990) directly compared metaphor and metonymy processing in a reading time experiment. The results showed that readers spent more time on processing metonymical expressions (e.g., 'scalpel' to refer to a surgeon) than metaphorical ones (e.g., 'butcher'), although both expressions were more costly than literal ones (e.g., 'doctor'). Developmental and neuropsychological data, however, suggest a different picture. Acquisition studies showed that metaphor comprehension evolves later during development than metonymy resolution (Annaz et al., 2009; Rundblad & Annaz, 2010). Patients studies demonstrated that right hemisphere damaged patients performed worse in metaphor processing relative to metonymy and homonymy processing (Klepousniotou & Baum, 2005a,b; see also Semenza et al., 1980). Consistently, in a standard reaction time study employing a sensicality judgment task, we found a difference between metaphorical sentences (compared to their literal counterparts) and metonymical expressions (compared to their literal counterparts) with respect to the interpretation costs (Bambini et al., *submitted*). Potential differences also emerge when considering side by side studies focusing on either of the two phenomena. On one side, a huge amount of psycholinguistic evidence has been accumulated suggesting that metaphor comprehension, compared to literal language, is often associated with higher difficulty in retrieving and integrating semantic and pragmatic information needed for computing a metaphorical interpretation (Glucksberg, 2003; Gibbs & Colston, 2012). Interestingly, some studies within the experimental pragmatics framework revealed a more complex scenario, pointing to the conceptual load required by metaphor (Rubio-Fernández, 2007) and suggesting that metaphor processing is associated with extra processing costs as well as with benefits contingent on context (Noveck et al., 2001). Differences between metaphorical relative to literal expression processing, context being equal, are also visible in terms of brain activations (Bambini et al., 2011), with metaphors activating a bilateral fronto-temporal network (for an extensive review on fMRI studies on metaphor, see Rapp et al., 2012).

On the other side, evidence for metonymy is relatively scarce. A series of eye tracking experiments investigating the time course of processing different types of familiar (e.g., ‘reading Dickens’, ‘during Vietnam’) and unfamiliar (e.g., ‘reading Needham’, ‘during Finland’) metonymies showed that: (i) familiar metonymical expressions were as easy to process as literal ones, as indicated by equal first pass and gaze duration measures; (ii) unfamiliar metonymies were more difficult to process than literal expressions, as indicated by second pass times, and by measures of later regions of interest (Frisson & Pickering, 1999; Pickering et al., 2004; Frisson & Pickering, 2007). The authors argued that the type-shifting process required by metonymy is not costly *per se*, but the difficulty lies in the online generation of non-lexicalized senses (Pickering et al., 2004). In addition, no processing difficulties were observed when unfamiliar metonymies were provided with a supportive context (e.g., ‘My great-grandmother has all the novels written by Needham in her library. I heard that she often read Needham when she had the time’), thus suggesting that readers can use contextual information immediately to compute a metonymical interpretation. Differently, brain data provide evidence that metonymy processing is associated with specific neural correlates relative to literal language. In particular, an ERPs study investigating metonymies like “The hepatitis called….” showed that the process of enriching the lexical information in order to compute a sensible interpretation is associated with particular neurophysiological signatures, i.e., the N400 component – interpreted as reflecting lexical/semantic mechanisms – and a late positivity component – interpreted as reflecting pragmatic mechanisms (Schumacher, 2011).

In general, however, it should be noted that the comparison between studies on metaphor and studies on metonymy is questionable, because the experiments employed different experimental paradigms, tasks, participants and types of materials in terms of psycholinguistic properties.

To summarize, we have argued that most psycholinguistic theories of figurative language processing have been based on only one trope. Yet, linguistic analyses and the few studies that have compared multiple tropes indicate that are important differences. In this study we investigate the time course of metaphor and metonymy within a single experiment. We present predictions about the time course of processing and then test these predictions using a Speed Accuracy Trade-Off procedure (SAT; Reed, 1973; for a review, see Meyer et al., 1988). The SAT is a chronometric paradigm developed in the late ’70 that provides a higher temporal resolution than a standard reaction time task. The two main advantages are that: (i) it provides information about the early phases of processing; (ii) it provides a means of directly measuring processing time in the presence of concomitant differences in accuracy, thus allowing discrimination between time and the likelihood of reaching a sensible interpretation (McElree & Nordlie, 1999; Bornkessel et al., 2004). We next provide a general introduction to SAT and then provide a more detailed explanation of our experiment.

1.2 The Speed-Accuracy Trade-Off methodology

In experimental psycholinguistics, a widely used technique for investigating the time-course of information processing is to measure the response time and accuracy. Importantly, it has been observed that response time and accuracy are not independent from each other, but tend to vary together. Faster responses tend to produce more errors, and vice versa. This phenomenon – called speed-accuracy trade-off – is mainly related to the decision criteria that can vary across conditions (Reed, 1973; McElree & Dosher, 1989; Carrasco et al., 2006; Traxler, 2011). Moreover, even when decision criteria are assumed to remain constant, differences in response time may be due to differences in stimuli discriminability and/or information availability (McElree & Dosher, 1989; McElree & Nordlie, 1999; Carrasco et al., 2006). For example, in psychophysical experiments on visual processing, if a stimulus is less discriminable, the response threshold will be reached at a later point in time, even if the speed of information accrual is the same as for a more discriminable stimulus (Carrasco et al., 2006). Similarly, in psycholinguistic studies on figurative language, if figurative items are inherently more difficult compared to literal items, people may delay committing to an interpretation until they are more confident of their responses. In this case, longer response times for figurative/pragmatic interpretation may reflect the difficulty in interpreting words in figurative contexts rather than the time needed to retrieve and process the information (Bott et al., 2012; McElree & Nordlie, 1999). McElree and Nordlie (1999) claimed that the large part of evidence accumulated for figurative language processing has been obtained with experimental paradigms that do not control for the speed-accuracy effect, thus not allowing a clear interpretation of longer processing times.

In order to control for the speed-accuracy trade-off, different strategies have been applied (Liu & Smith, 2009), among which the response signal or Speed-Accuracy Trade-Off (SAT) paradigm. In the SAT procedure, the experimenter specifies the time at which a response has to be executed. Two variants of the SAT procedure exist. In the standard SAT paradigm, a single response cue (typically a tone) is provided in each trial, varying across a range of times (e.g., 100-3000 ms), and participants respond within 250-350 ms of the response cue (McElree & Nordlie, 1999; Bott et al., 2012). In the multiple-response variant of the SAT procedure (MR-SAT), a series of response cues is provided (e.g., a tone is presented 14-17 times at 300-350 ms interval), and subjects are asked to perform the task after each cue (Bornkessel et al., 2004; Foraker & McElree, 2007; McElree et al., 2006; Martin & McElree, 2008). The main advantage of the MR-SAT procedure is that it requires a relatively small number of trials compared to the standard SAT.

In both SAT variants, a full time course function that measures how the accuracy of processing varies with processing time is derived. Accuracy (measured in d') is the dependent variable and it is modeled as an exponential function of the response time by the equation:

$$d' = \lambda[1-e^{-\beta(t-\delta)}],\ t > \delta,\ \text{else } 0$$

This function is characterized by three parameters: λ – related to the accuracy of processing; β and δ – related to the speed of processing. In particular, the λ parameter serves to estimate the asymptotic level of performance; the β and δ parameters provide joint measures of the speed of processing, indexing how quickly accuracy accrues to its asymptotic level. More specifically, the β parameter estimates the rate at which accuracy grows from chance to asymptote; the δ parameter estimates the intercept of the function, or the point at which participants are first sensitive to the information necessary to make an accurate discrimination (i.e., d' departs from 0, chance performance). Two SAT curves may differ with respect to the asymptote alone (proportional dynamics), the speed parameters alone (disproportional dynamics), or both (Bornkessel et al., 2004; McElree et al., 2006).

The SAT procedure has been used to measure the accuracy and the speed of processing in a wide range of linguistic processes, including syntactic processing (Bornkessel et al., 2004; Martin & McElree, 2011) and pragmatic enrichment (Bott et al., 2012). As far as figurative language is concerned, McElree and Nordlie (1999) investigated metaphor processing in two standard SAT experiments. In the first one, participants were asked to judge whether metaphorical (e.g., 'Some mouths are sewers'), literal (e.g., 'Some tunnels are sewers') and anomalous sentences (e.g., 'Some lamps are sewers') were meaningful. Results showed that metaphorical and literal sentences were processed in equal time, with no differences in the speed parameters, although asymptotic level of performance was lower for metaphorical than for literal sentences. In the second experiment, participants judged whether the same sentences were literally true. No differences in speed processing between figurative and nonsense items were found, while there was a difference in the asymptotic accuracy. Thus, literal and figurative interpretations were computed in equal time or in parallel, contrary to the Standard Pragmatic model. To our knowledge, there are no SAT studies on standard metonymy. Nonetheless, there are indications that other forms of metonymy (i.e., logical metonymy) may affect the retrieval speed. In a MR-SAT study, McElree et al. (2006) asked participants to express sensicality judgments on sentences like 'The carpenter built the table' (conventional form) or 'The carpenter begun the table' (logical metonymy). In the latter case, the logical structure of the verb 'to begin' coerces an interpretation in which a part of an event (i.e., the table) denotes the whole event (i.e., building the table; Pustejovsky, 1995; Lascarides & Copestake, 1998). The time course revealed that logical metonymies are processed less accurately and more slowly than the conventional expressions, suggesting that enriching the meaning in order to compute the communicated meaning requires time-consuming operation.

1.3 The present study

In the present study we applied MR-SAT methodology to investigate the time-course of processing different types of figurative meanings in Italian. We compared nominal metaphors, e.g., 'Those dancers are butterflies,' against comparative literal sentences, e.g., 'Those insects are butterflies'; and standard metonymic sentences, e.g., 'That student reads Camilleri', against comparative literal sentences, e.g., 'That journalist interviews Camilleri'. The sentences were carefully controlled for several psycholinguistic variables that are known to affect the initial stages of figurative language processing, such as meaningfulness, difficulty, familiarity and cloze probability (Frisson & Pickering, 1999; Glucksberg, 2003; Frisson, 2009; Cardillo et al., 2010). In particular, meaningfulness and difficulty were measured in order to ensure the interpretability and the ease of interpretability of the sentences (Libben & Titone, 2008; Kintsch & Bowles, 2002). The familiarity of metaphorical expressions was controlled in order to exclude both fully conventionalized and extremely creative expressions. Similarly, for metonymy, we employed proper nouns for which a metonymical interpretation was not fully lexicalized (e.g., Camilleri vs. Dante). At the same time, we ensured that proper nouns were familiar enough to make a metonymical interpretation possible through a world knowledge task. Finally, cloze probability was measured in order to exclude that differences between figurative and literal expressions were due to the probability of the final word (Kutas & Federmeier, 2011).

We had a number of goals. Our first goal was to replicate McElree and Nordlie's study (1999) using a different set of materials and a potentially more powerful SAT methodology. McElree and Nordlie's study has been highly influential in the field but there are several reasons why a replication is warranted. First, the conclusions are based on a null effect: McElree and Nordlie found no significant difference between metaphor and literal sentences for the rate and intercept. These findings might simply be due to a lack of power (although McElree and Nordlie note that other SAT tasks using near identical procedures have detected effects of less than 50ms). Second, the familiarity of the metaphors was not measured. Some of the metaphors seem very familiar, e.g., 'Some jobs are jails' whereas others far less so, e.g., 'Some smiles are rubber'. Familiar metaphors might be encoded as idioms, or some other lexicalized form, which may not require the same type of processing as other metaphors. This would then reduce the difference between literal and metaphoric processing and explain the similar rates and intercepts. In our study we measured familiarity of the metaphors in a pre-test and analysed whether familiar metaphors are processed differently to unfamiliar metaphors.

The hypotheses relating to metaphor are those identified by McElree and Nordlie (1999). According to the indirect, serial stage view, metaphor should exhibit a different time course than literal sentences. This should be reflected in differences in the speed parameters of the SAT curves. More specifically, if figurative meaning is computed only after a literal interpretation has been processed and rejected, intercepts should be earlier or rates slower in the metaphor condition than the literal condition. Conversely, the direct, one-stage view would predict similar rate and intercept measures across figurative and literal conditions. Traditional psycholinguistic models do not allow for clear predictions with respect to the accuracy parameter. As suggested by McElree and Nordlie (1999), the accuracy parameter reflects the probability of successfully retrieving semantic knowledge and pragmatic information in order to compute a sensible interpretation. In this sense, it is plausible that it is higher for metaphor than for literal processing.

Our second goal was to investigate whether, and if so how, processing of metonymy differs to processing of metaphor. On the one side, it should be highlighted that the predictions about the temporal dynamics formulated on the basis of the psycholinguistic models (Standard Pragmatic and Direct Access) are applicable to metaphor and to metonymy as well, since no specific features are formulated to predict differences between the two phenomena. On the other side, the two non-literal phenomena might show different behavior from one another, motivated at the level of the differences posited by cognitive pragmatic theories (Relevance Theory and Cognitive Linguistics). In other words, the conceptual operations in metaphor processing (based on broadening and narrowing of lexical concepts or mapping between two different domains) might influence the elaboration of meaning in a way that differs from the reference transfer process required by metonymy. One hypothesis is that, assuming that reference transfer process is not costly per se – as suggested by some previous evidence – and being metonymy more subject to routinization (here reflected by the *producer for product* rule), metonymical interpretation might not require additional costs relative to literal interpretation. This would be reflected by similar temporal dynamics for metonymical and literal SAT functions. However, it has also been shown that the online construction of non-lexicalized sense or of extra-lexical sense is difficult, as in the case of unfamiliar metonymies without a supportive context and in the case of logical metonymy, respectively. Thus, it is also possible that the construction of the appropriate metonymical interpretation is indeed costly, and that this reflects in the time-course of processing.

2. Materials and Methods

2.1 Participants

Eighteen native speakers of Italian (right-handed, 10M/8F, mean age = 22.5, SD = 1.9) participated in the study. Each participant completed 1-h practice session for familiarization with the SAT procedure, and three 1-h experimental sessions. Participants were paid for their participation.

2.2 Materials

Two sets of Italian sentences of the form "That Y verb X" were employed. The Metaphor set included 48 nominal metaphors (e.g., 'Quelle ballerine sono *farfalle*' / Engl.: 'Those dancers are *butterflies*'), with their corresponding literal (e.g., 'Quegli insetti sono *farfalle*' / Engl.: 'Those insects are *butterflies*') and anomalous counterparts (e.g., 'Quelle bottiglie sono *farfalle*' / Engl.: 'Those bottles are *butterflies*'. Literal sentences were obtained by using semantically compatible terms, while anomalous sentences resulted from the clash of two semantically non homogeneous terms. Each sentence was constructed in such a way that the first NP was a subject and the second was a predicate, that is only canonical copular sentences in the sense of Moro (1997). This was made to exclude inverse copular constructions which would have shifted the focus on the post copular noun phrase, unbalancing the stimuli. The plural form was chosen in order to avoid predictability effects carried by the Italian gender-marked articles required in the singular form.

The Metonymy set included 48 producer-for-product metonymies (e.g., 'Quello studente legge *Camilleri*' / Engl.: 'That student reads *Camilleri*'), with their corresponding literal (e.g., 'Quel giornalista intervista *Camilleri*' / Engl.: 'That reporter interviews *Camilleri*') and anomalous counterparts (e.g., 'Quel cuoco cucina *Camilleri*' / Engl.: 'That chef cooks *Camilleri*'). In order to reduce the conventionality of the metonymical expressions, the use of names of very famous people from the past (e.g., Dante, Verdi) was avoided in favor of the use of presently popular Italian writers, singers, painters and designers (e.g., Camilleri, Armani). Indeed, it has been suggested that the use of famous names (e.g., Dickens) in the metonymic form might have become lexicalized in ordinary language (Frisson & Pickering, 2007). It should be noted, nonetheless, that the familiarity of each proper name and of the metonymical transfer was controlled through a rating procedure, as illustrated below (see also the Introduction).

All stimuli were construed so that participants could not judge the meaningfulness of the sentence prior to the onset of the final target word. It should be noted that the target word (e.g., *butterflies, Camilleri*) was held constant in the triplet and forced either a pragmatic, literal, or nonsensical interpretation in different sentential contexts.

In order to have a similar ratio of sense and nonsense items (i.e, pragmatic + literal vs. anomalous sentences), additional 48 anomalous sentences were included in each set. The additional sentences had the same structure of the experimental stimuli. They were created by recombining other words in the set, with the purpose of minimizing potential effects related to the repetition of the target noun in the triplets (e.g., consider the example above: 'Those insects are *butterflies*'. The subject noun 'insects' was reused to create the additional anomalous sentence 'Those insects are tables').

Furthermore, the experimental items were intermixed with 594 filler items, thus reducing the proportion of metaphorical and metonymical sentences in the set and avoiding possible metalinguistic awareness with respect to figurativeness. The fillers were 4-word sentences of the same form of the experimental stimuli (e.g., 'She kicks the ball'). Moreover, 48 cases of approximate use - e.g. 'Those sunglasses are rectangular' - with their literal (48 cases) and anomalous (96 cases) counterparts were considered as fillers (for more details see Bambini et al., *submitted*), as no specific hypothesis on the temporal dynamics of approximation processing could be formulated. Adding the fillers (total = 786), sense sentences constituted the 58% of stimuli.

Materials norming

All experimental stimuli were characterized at the psycholinguistic level for meaningfulness, difficulty, familiarity and cloze probability in a previous rating study. The results of the rating study guided the selection of the stimuli included in the analysis in the present experiment. Comprehensive details on the norming and the inclusion criteria of the experimental stimuli can be found in Bambini et al. (*submitted*). Six triplets were excluded from the original Metaphor and Metonymy sets in Bambini et al. (*submitted*), resulting in a total of 42 triplets for each set. In the following, we only report results for the pool of selected sentences.

Within each set, the length and the frequency of target words were exactly balanced among the pragmatic, literal and anomalous conditions as the target noun was held constant in each triplet. The mean number of characters of the target words was balanced across sets (Metaphor set = 7.07, Metonymy set = 7.16, t(82) = 0.236, *ns*). The lexical frequency of the target words in the Metaphor set was controlled on the basis of the available frequency norm of Italian Corpus and Frequency lexicon of written Italian (ColFIS, Bertinetto et al., 2005), available through the web interface EsploraCoLFIS (Bambini & Trevisan, 2012). The mean log frequency was equal to 1.51 (SD = 0.55). For the Metonymy set frequency values were not available in the ColFIS since we used proper names as target words. Frequency was roughly controlled across sets by using a combined measure of the lexical item frequency: for metaphorical sentences, we summed the lexical frequency of the subject nouns and the target nouns (e.g., in 'Those dancers were butterflies', the lexical frequency of 'dancer' and 'butterfly' was considered); for metonymical sentences, we summed the lexical frequency of the subject nouns and the verbs (e.g., in ‘That student reads *Camilleri*’, the lexical frequency of ‘student’ and ‘to read’ was considered). No significant frequency differences were found across sets (mean log frequency: Metaphor set = 1.85; Metonymy set = 1.92, t(82) = 0.66, *ns*).

For each sentence, meaningfulness, difficulty, familiarity, and cloze probability ratings were collected from a minimum of 6 and a maximum of 10 participants using online questionnaires. With respect to meaningfulness and difficulty, ratings on five-point Likert scales showed that both metaphorical and literal sentences were rated as significantly more meaningful and less difficult than anomalous sentences (metaphorical sentences: Mdn = 4, iqr = 2-4; literal sentences: Mdn = 4, iqr = 4-5; anomalous sentences: Mdn = 1, iqr = 1-2; all *p*s < 0.001). Similar results were obtained for the Metonymy set (metonymical sentences: Mdn = 4, iqr = 3-5; literal sentences: Mdn = 5, iqr = 4-5; anomalous sentences: Mdn = 1, iqr =1-2; all *p*s < 0.001). This excluded the possibility that pragmatic sentences were interpreted as anomalous. However, we also observed that, when directly compared, pragmatic sentences were also less meaningful than their corresponding literal counterparts (metaphorical vs. literal sentences, $p < 0.001$; metonymical vs. literal sentences, $p < 0.001$). This may suggest a potential difference in the overall accuracy. This, however, is not a drawback for the present experiment because the SAT procedure separates accuracy from speed.

With respect to familiarity, ratings on a 5-point Likert scale revealed that metaphorical sentences were neither fully familiar nor extremely novel (median score = 3, range 1-4). For the Metonymy set, both the familiarity of the proper names and the familiarity of the producer-for-the-product transfer were assessed by asking participants to associate each producer’s name with the corresponding product, choosing between four options (e.g., for Camilleri, a writer, the options were: book / song / movie / painting). We found that participants correctly associated the proper name of the producers to the corresponding products (accuracy = 90.29%), thus implicitly demonstrating the familiarity of the proper names and of the metonymic transfer.

As for contextual expectancy, results showed that all conditions had cloze probability 0% or slightly higher (0.39 % for the literal counterparts of metaphors), ensuring that context was kept minimal and equal across conditions.

2.3 SAT Procedure

Stimulus presentation and response collection were carried out using Presentation® software (Version 14.9, www.neurobs.com). Each trial was structured as displayed in Figure 1. Each sentence was presented word by word at a fixed rate (300 ms). A tone (50 ms, 1000 Hz) served as the signal to cue participants’ response. The first response cue occurred at the onset of the last word (i.e., the target word), followed by 14 more response cues (inter-cue interval: 350 ms). Participants were trained to respond after each tone within

350 ms and to give their current sensicality judgment by pressing one of two keys on an RB 530 response pad (SuperLab Pro, Cedrus Corporation). As the first cue response occurred at the onset of the last word, participants responded before the sentence processing was complete (Bornkessel et al., 2004; Martin & McElree 2008; 2009). Therefore, one group of participants (n = 9) was instructed to start with the “sense” key as an initial, undecided response and to continue on the same key until they formulated their judgment about the meaningfulness of the sentence. Then, in order to express their judgment they could either continue on the same key – indicating that the sentence was meaningful – or switch to the other key – indicating that the sentence was meaningless. Vice versa, a second group of participants (n = 9) was instructed to start with the “nonsense” key and to express their judgment either by continuing on the same key – indicating that the sentence was nonsense – or switching to the other key – indicating that the sentence was meaningful. Participants were also encouraged to modulate their responses if their judgment changed during the trial by switching from the sense key to the nonsense key (or vice versa), as long as the response tones were still sounding. After the last response, a feedback was given on the participant’s performance in pressing the buttons in the rhythm of the tones (the feedback did not concern the correctness of the judgment). In particular, they received an error message when they started by not pressing the initial key they had been instructed to press, or by pressing the correct key but twice, or by pressing the correct key after a delay greater than 350 ms. They also received a negative feedback when they gave either too many or too few responses, or when they responded out of sync. Multiple error message occurred for different combinations of the errors. Otherwise, they received an “Ok!” message associated with a \tä-ˈdä\ sound. Feedback messages were displayed for 1000 ms (simple messages) or 1500 ms (combined messages) before the end of the trial. Between-trial intervals were participant controlled: participants initiated the next trial by pressing the central key of the response pad. There were six mandatory breaks in each session.

The experiment consisted of one training session and three experimental sessions completed over a period of two days. Participants first completed a 1-hour practice session in order to familiarize with the procedure (first day, morning session). They were trained on pressing a key in the rhythm of the tones and switching responses by pressing the other key in order to be able to modulate their responses. Participants practiced until they feel comfortable with the procedure. The training session included four-word sentences different from those of the experimental set. The experimental sessions consisted of three 1-h sessions, one in the same day of the training session (first day, afternoon session) and the other two on the subsequent day (one in the morning and one in the afternoon). Each session consisted of 128 experimental trials and 262 filler trials (total of 390 trials per session, with critical trials in which the last word is used metaphorically or metonymically constituting 9% of each session). An equal number of pragmatic, literal and anomalous sentences were included in each session. The member of each triplet was assigned to distinct sessions, thus participants saw every target item in every sentential context (pragmatic, literal and anomalous), but in different experimental sessions.

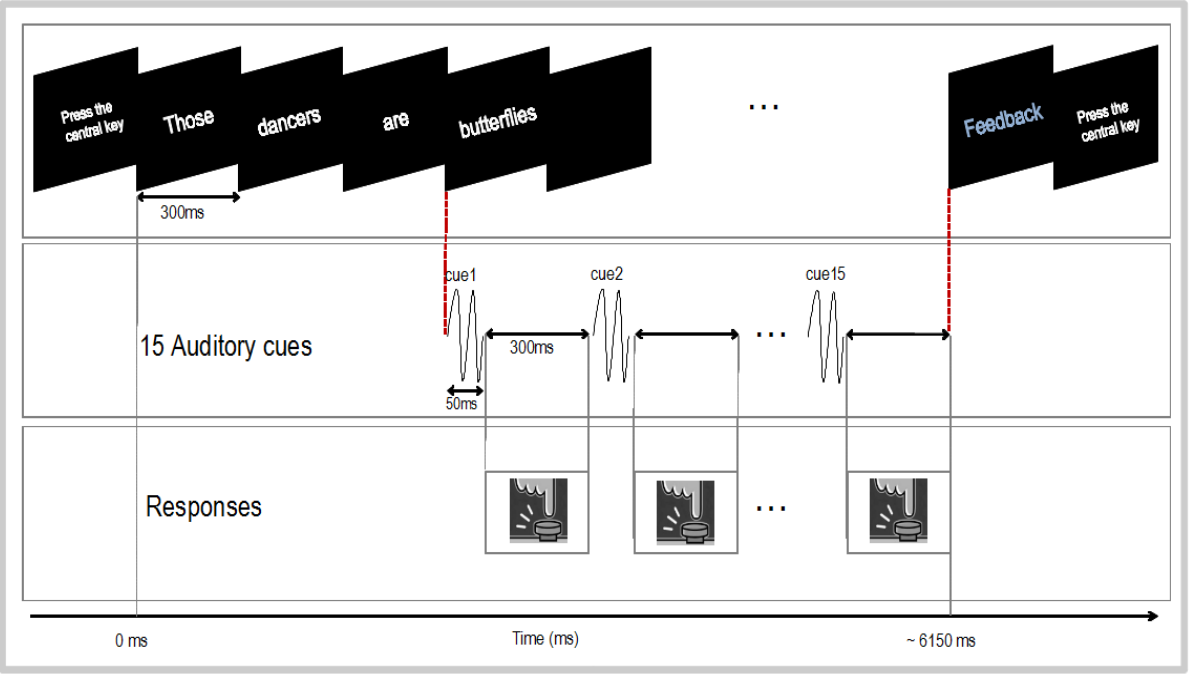


Figure 1. Schematic representation of an SAT experimental trial.

2.4 Data analysis

Metaphor and Metonymy sets were analyzed separately, following the same procedure. We removed one participant who obtained an average d' of less than 1.0 on the longest time lag. Practice data were not included in the analysis.

2.4.1 Computing d'

Accuracy was measured in terms of d' corresponding to the difference between the z-score of proportion hits and the z-score of proportion false alarms. Hits were the “sense” responses to the figurative and literal sentences; false alarms were the “sense” responses to the anomalous sentences. In order to ensure that z-scores were always finite, we applied the correction proposed in Bott et al. (2012): proportions hits were calculated as the total number of correct figurative/literal trials plus 0.5, divided by the total number of figurative/literal trials plus 1. The same correction was applied for calculating the proportion false alarms. We calculated d's in fifteen 350-ms time bins surrounding the response cues (McElree et al., 2006).

2.4.2 Analysis

We derived SAT function by measuring accuracy (d' units) as a function of processing time (time of the response cue plus response latency measured in ms) with an exponential approach to limit (see also the Introduction, section The Speed Accuracy Trade-Off methodology). For parameter estimation we applied the maximum likelihood estimation (MLE) method which consists in seeking the parameter values that are most likely to have generated the data (Liu & Smith, 2009). The analyses were performed using Matlab 7.6.0 (The MathWorks, Inc., Natick, Massachusetts, United States).

A key question in analyzing time-accuracy curves is whether the analyses should be performed on data from each participant or on averaged data. A recent methodological paper warned against fitting nonlinear functions (such as exponential SAT function) to averaged data since such procedure “may yield distorted estimates that do not reflect the true underlying process for any individual” (Liu & Smith, 2009: 191). Indeed, several authors underlined the importance of analyzing individual curves since the variances in the parameters of the time-course functions across participants may exceed the variance between conditions (McElree & Nordlie, 1999). A close examination of our data revealed high subject variability, as illustrated in Figure 2 for the metaphor processing and in Figure 3 for the metonymy processing. In particular, we observed that 6 out of 17 participants had an average intercept that was very late, as shown by high values of the intercept parameter. In this case, a high variability of the intercept parameter value across participants was observed (min = 0.40, max = 2.50, mean = 1.25, SD = 0.77). Accordingly, we performed the analyses on the individual participants’ data in order to avoid averaging artifacts in the shape of the functions.

For the analysis of individual participants’ data we adopted the procedure described in Bott et al. (2012). First, SAT functions were optimized for each participant separately. In particular, for each participant we fitted the fully parameterized SAT model in which separate asymptote, beta and intercept parameters were set for each condition. More specifically, this 2λ-2β-2δ model has six parameters: λ(pragmatic), β(pragmatic), δ(pragmatic), λ(literal), β(literal), δ(literal). Then, at the group level, we compared individual fitted parameter values between the pragmatic and the literal conditions by using nonparametric inferential statistics (Wilcoxon Signed Rank test). This approach was selected as we are primarily interested in how the pragmatic and literal curves may differ in any way, avoiding the assumption that some parameter values may be constrained to be identical (as in the model selection approach) (Liu & Smith, 2009). Non parametric test has been used in order to avoid the effect of outlier values (Bott et al., 2012).

A similar procedure was employed for the item analysis. First, for each item separately we fitted the fully parameterized 2λ-2β-2δ model. Then, we compared parameter values between the pragmatic and the literal conditions by means of nonparametric inferential statistics (Wilcoxon Signed Rank test).

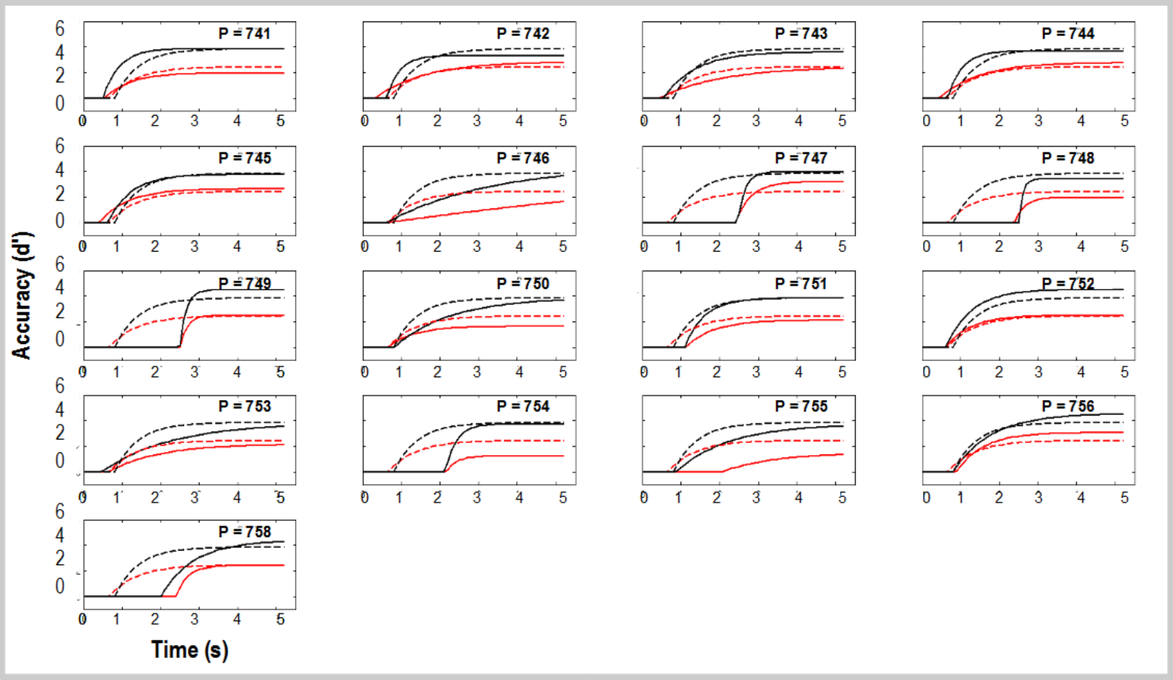


Figure 2. Time-course of metaphor processing. For each Participant (P = participant's ID), the best fitting SAT functions (continuous lines) for the metaphorical (red) and the literal (black) conditions are illustrated. Dotted lines represent the best fitting SAT functions for the median data (n participant = 17).

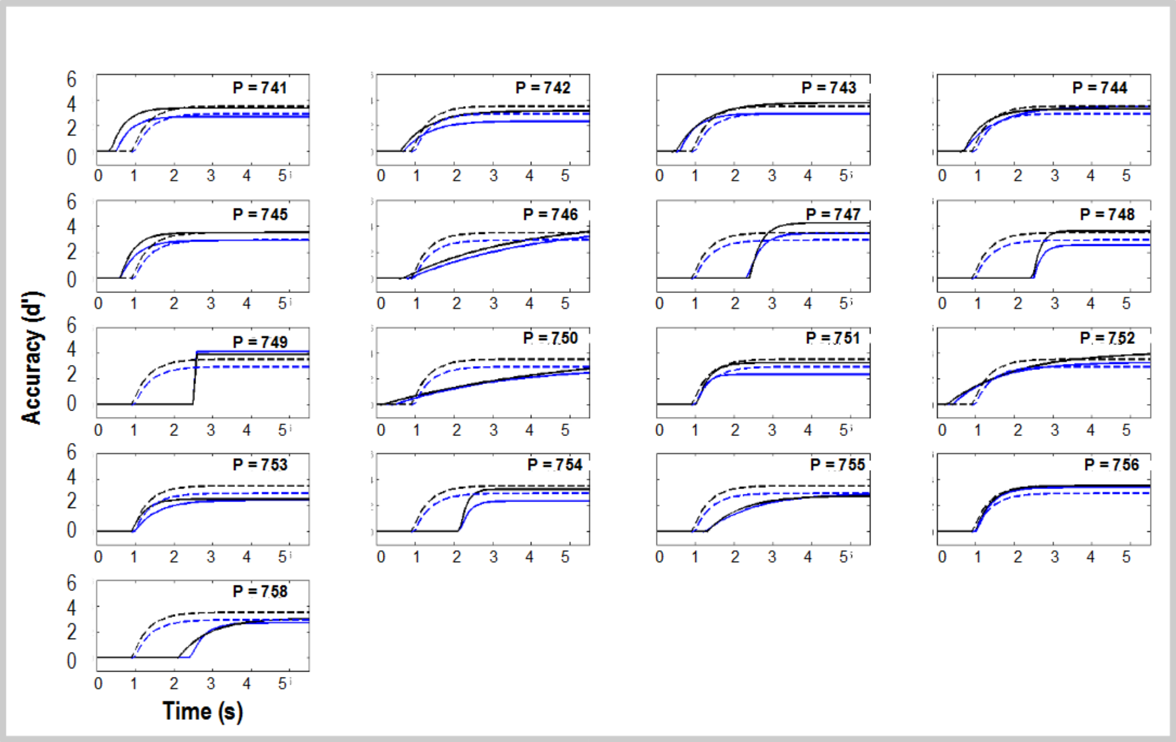


Figure 3. Time-course of metonymy processing. For each Participant (P = participant's ID), the best fitting SAT functions (continuous lines) for the metonymical (blue) and the literal (black) conditions are illustrated. Dotted lines represent the best fitting SAT functions for the median data (n participant = 17).

3. Results

3.1 Metaphorical vs. literal sentences

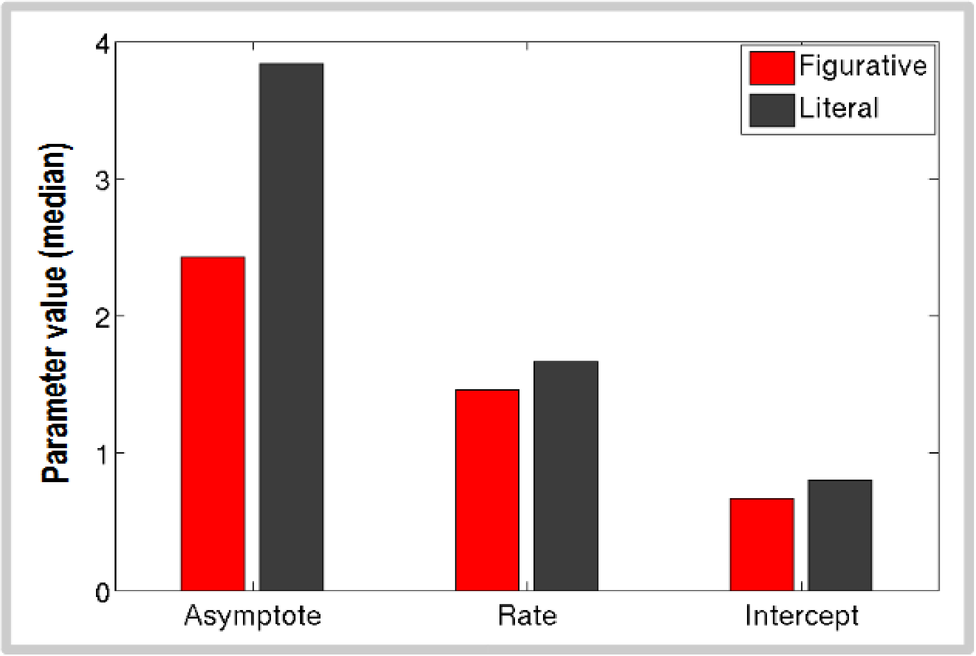
To obtain an empirical measure of asymptotic recognition accuracy we averaged the d's for the last three time-lags (McElree et al., 2006; Öztekin & McElree, 2007; van Dyke & McElree, 2011). The asymptotes were higher for literal than for metaphorical sentences (median difference = 1.61, $V_1 = 0$, $n = 17$, $p < 0.001$; median difference = 1.50, $V_2 = 13.5$, $n = 42$, $p < 0.001$).

We fitted the six-parameter models (2λ-2β-2δ) to each individual and compared estimated parameter values at the group level (Figure 4A). Consistently with the analysis on d', by comparing estimated λ parameter values we found a difference between the literal and the metaphorical conditions ($V_1 = 9$, $n = 17$, $p = 0.0005$; $V_2 = 94$, $n = 42$, $p < 0.001$). This indicates that readers were less likely to compute a meaningful interpretation of the metaphorical sentences than of the literal sentences. As for the speed of processing, there were no significant differences in rate ($V_1 = 43$, $n = 17$, ns; $V_2 = 556$, $n = 42$, *ns*) nor in intercept ($V_1 = 73$, $n = 17$, *ns*; $V_2 = 590$, $n = 42$, *ns*).

As noted above, for a subset of participants ($n = 6$) there was evidence of a late-intercept function ($\delta > 2$). Although we controlled the inter-individual variability by fitting the model on individual data, we ensured that the group-level result is not due to individual strategies by performing the analysis excluding the participants with the delay-strategy. Consistently with the results reported above, we found a difference in the asymptote. No reliable differences in either the rate or the intercept were found between conditions (asymptote: $V_1 = 7$, $n = 11$, $p = 0.018$; $V_2 = 74$, $n = 42$, $p < 0.001$; rate: $V_1 = 17$, $n = 11$, *ns*; $V_2 = 511$, $n = 42$, *ns*; intercept: $V_1 = 22$, $n = 11$, *ns*; $V_2 = 499$, $n = 42$, *ns*).

Given that there is much evidence on the role of familiarity in processing metaphors (Glucksberg, 2003), we also performed the analysis taking into account the familiarity ratings. We distinguished between high familiar and low familiar metaphors, splitting the data into two subsets. For high familiar metaphors we found significant differences in accuracy, with literal more accurate than metaphor ($V_1 = 5$, $n = 17$, $p < 0.001$) but not in speed processing (rate: $V_1 = 40$, $n = 17$, *ns*; intercept: $V_1 = 52$, $n = 17$, *ns*). The same results have been obtained for low familiar metaphors (asymptote: $V_1 = 1$, $n = 17$, $p < 0.001$; rate: $V_1 = 96$, $n = 17$, *ns*; intercept: $V_1 = 90$, $n = 17$, *ns*).

A. Metaphor



B. Metonymy

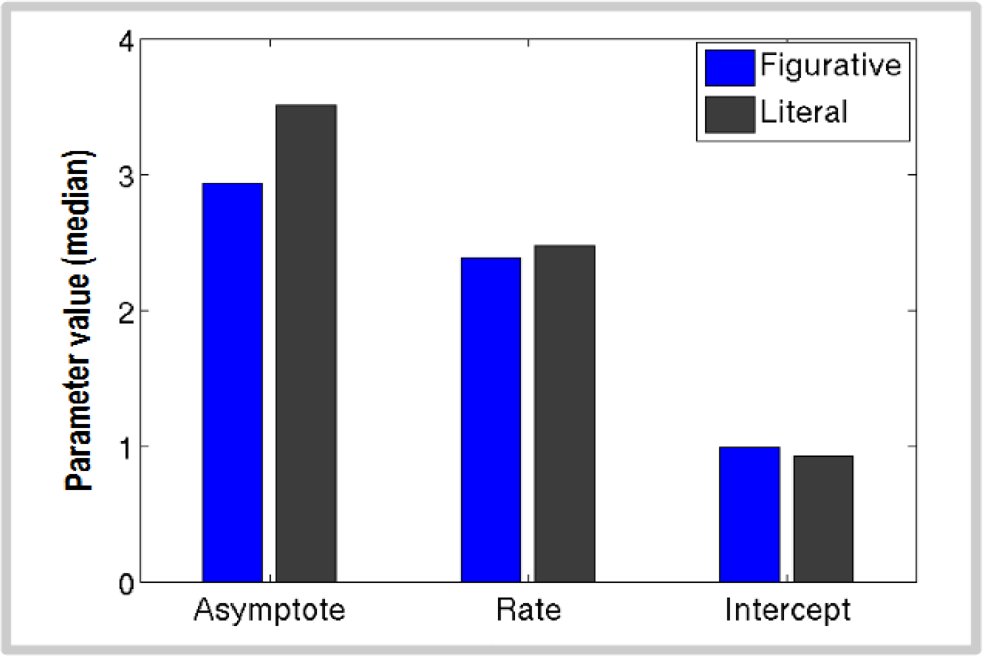


Figure 4. Exponential parameters estimates. (**A**) Metaphor processing. Median exponential parameter estimates (asymptote, rate, intercept) of the fully parameterized model are represented. Red = Metaphorical condition; Grey = Literal condition. (**B**) Metonymy processing. Median exponential parameter estimates (asymptote, rate, intercept) of the fully parameterized model are represented. Blue = Metonymical condition (blue); Grey =Literal condition.

3.2 Metonymical vs. literal sentences

We computed the average accuracy (d' units) across the three last time-lags in order to provide an empirical measure of asymptotic performance. The asymptotes were higher for literal than for metonymical sentences (median difference = 0.44, V_1 = 13, n = 17, *p* = 0.001; median difference = 0.14, V_2 = 239, n = 42, *p* = 0.01).

We fitted the six-parameter models (2λ-2β-2δ) to each individual and compared estimated parameter values at the group level (Figure 4B). Consistently with the d' measure, we found a difference in the estimated λ parameters between the literal and the metonymical conditions (V_1= 15, n = 17, *p* = 0.002). As for speed processing, readers were slower in interpreting metonymical as compared to literal sentences. Although there were no significant differences in the rate estimated values (V_1 = 65, n = 17, ns), significant differences in the intercept estimated values were found (V_1 = 130, n = 17, *p* = 0.009). The same results were obtained by excluding participants (n = 6) who adopted a delay strategy in responding (asymptote: V_1 = 4, n = 11, *p* = 0.006; rate: V_1 = 38, n = 11, *ns*; intercept: V_1 = 65, n = 11, *p* = 0.001).

However, we did not find significant differences in the items analysis (asymptote: V_2 = 365, n = 42, *ns*; rate: V_2 = 487, n = 42, *ns*; intercept: V_2 = 491, n = 42, *ns*). A potential explanation for this is that we did not have enough data to accurately model individual curves for each item (we had 17 data points per cell for each item, but 42 per cell for each subject). We therefore removed items which displayed implausible parameter values, i.e., asymptote estimated value less than 1 or greater than 5, and/or rate estimated value greater than 10 (in total, 15 items were removed). The results of this analysis were consistent with the subject analysis with respect to the asymptote values (asymptote: V_2 = 63, n = 27, *p* = 0.001) but neither the rate nor the intercept values were significantly different across conditions (rate: V_2 = 252, n = 27, *ns*; intercept: V_2 = 214, n = 27, *ns*).

4. Discussion

The present study aimed at investigating the processing of two different types of figurative language use, i.e., metaphor and metonymy, by applying an experimental paradigm specifically suitable for disentangling the time-course and the accuracy of processing. In the following paragraphs, we first concentrate on the discussion of the results obtained for metaphor processing, then we discuss results for metonymy processing.

4.1 Metaphor processing

The results on the metaphor set replicate previous findings reported in McElree and Nordlie (1999), showing that metaphorical meanings resulted in lesser availability (lower asymptotic accuracies) than their literal counterparts, while they were processed at the same speed (same intercept and same rate). Importantly, our experimental data complement the results obtained by McElree and Nordlie (1999) by applying a more powerful variant of the SAT procedure (MR-SAT vs. standard SAT), allowing to perform a lower number of trials, and by extending to a different language (i.e., Italian). In addition, our stimuli were controlled for several psycholinguistic variables, ruling out the possibility that the results reflect confounding factors, including frequency, meaningfulness, difficulty, familiarity and cloze probability. In particular, further analyses taking into account the familiarity dimension revealed the same pattern of results for both high and low familiar metaphors (i.e., significant difference in the accuracy parameter, no significant differences in the speed parameters). These results might be explained by considering that, though varying in familiarity, all metaphorical sentences were highly comprehensible as revealed by meaningfulness and difficulty ratings (Bambini et al., *submitted*). There is evidence that factors such as comprehensibility (strongly correlated also with aptness) mediate the metaphorical processing (Jones & Estes, 2006), and thus might have reduced the effects of familiarity.

Altogether, by controlling for several psycholinguistic variables and without confounding accuracy with processing speed, the present study confirms that the interpretation of metaphorical meanings is not delayed compared to the interpretation of literal meanings (McElree & Nordlie, 1999; Glucksberg, 2003). With respect to the psycholinguistic models discussed in the Introduction, the absence of speed differences between metaphorical and literal sentences seems to be contrary to the view that the computation of figurative language, compared to literal language, requires several stages of processing (Standard Pragmatic Model). Rather, these time-course functions seem to be consistent with accounts claiming that both metaphorical and literal interpretations might be directly accessed.

Importantly, as clearly stated by McElree and Nordlie (1999), time-course profiles *per se* do not provide information about the mental processes underlying figurative and literal language processing. In this respect, similar temporal dynamics might only suggest that both metaphorical and literal meanings rely on similar processes (McElree & Nordlie, 1999), without hints on what type of processes they are. Considering this premise, the present findings are briefly discussed in the light of pragmatic and cognitive theories of metaphor that explore the type of processes involved, with the goal of finding possible insights able to be integrated with the results on the temporal dynamics. In the framework of Relevance Theory, there is a continuum of cases from literal to metaphorical meanings (Wilson & Carston, 2007; Carston, 2010). All language uses rely on the same process, i.e., the construction of ad hoc concepts through the narrowing or the broadening of the encoded meaning on the basis of relevant contextual information (Wilson & Carston 2007; Carston 2010). This view is also compatible with the class-inclusion theory (Cacciari & Glucksberg, 1994), according to which all literal and metaphorical expressions of the form 'A is B' are interpreted by asserting that the subject noun (A, metaphor's topic) is a member of the conceptual category represented by the predicate (B, metaphor's topic). Interpretation involves, in both cases, the retrieval of some conceptual properties from the predicate to be ascribed to the subjects. Cognitive Linguistics largely defined metaphor as conceptual mapping across two different domains (Lakoff & Johnson, 1980; Gibbs, 1999). All these theories seem to share the assumption that both literal and metaphorical expressions mainly rely on conceptual operations, although different types of conceptual properties are to be retrieved and manipulated in each case. As suggested by McElree and Nordlie (1999), previous SAT studies showed that the retrieval and manipulation of different types of semantic/conceptual properties is reflected by similar dynamics (Corbett & Wickelngren, 1978; Ratcliff & McKoon, 1989), consistently with the finding of no speed differences between the metaphorical and the literal conditions.

Based on these theoretical frameworks, a possible interpretation of the difference in the overall accuracy between metaphorical and literal sentences can also be attempted. In particular, we found that metaphorical sentences were associated with lower accuracy parameter values relative to literal sentences, consistently with previous evidence in untimed sensicality judgments (Bambini et al., *submitted*). This finding indicates that readers were less likely to compute a meaningful interpretation for metaphorical expressions compared to literal ones (McElree & Nordlie, 1999; McElree et al., 2006). By considering the class-inclusion framework, McElree and Nordlie (1999) suggested that the differences in the overall accuracy might reflect the difficulty of recovering the key semantic properties associated with the metaphoric vehicle and ascribing those properties to the topic (see also Cacciari & Glucksberg, 1994). Similarly, based on the Relevance Theory framework, it might be hypothesized that different asymptotic values reflect the difficulty of construing an ad hoc concept for metaphorical meanings compared to literal ones (see also Noveck et al., 2001). Alternatively, this finding might be ascribed to the difficulty of the mapping between domains. Specifying the type of process underlying the computation of metaphorical and literal sentences should be the matter of future studies, but a greater difficulty for metaphor interpretation seems a quite robust result.

In sum, these results seem to indicate that, at least in the case of the nominal metaphors presented in a minimal context, the metaphorical comprehension results from conceptual elaboration that lowers the likelihood that an appropriate interpretation is computed, but is not associated with variations in the time-course profile. In this respect, the current findings contribute to shed light on what is meant by the claim that figurative language requires additional effort to be understood. According to McElree and Nordlie (1999), it might be argued that additional effort in processing metaphorical relative to literal sentences found in previous behavioral studies likely reflected a lower probability of computing a sensible interpretation for metaphorical compared to literal sentences rather than serial stages of processing (see also Bott et al., 2012).

4.2 Metonymy processing

Different time-course profiles for metonymical and literal expressions were found, with metonymical sentences being processed more slowly (late intercept) relative to literal sentences. It should be noted, however, that the item analysis did not confirm such difference. Hence, the interpretation of this finding should be cautious.

With respect to the psycholinguistic models, this result may appear to be in accordance with the serial, indirect view as formulated by the Standard Pragmatic model, and against the direct, one-stage view. Specifically, the speed difference between metonymical and literal meanings might reflect the fact that the reader first computed a literal interpretation of the sentence but failed to derive an acceptable meaning. Hence, participants engaged additional processes to compute a figurative meaning. McElree et al. (2006), however, argued that serial models typical predict larger dynamics effects than those observed here (intercept differences = 5 ms) (see also McElree & Dosher, 1993). In this light, a clear-cut indirect explanation seems untenable. However, some temptative interpretations can be formulated and need to be explored.

One possible interpretation is that speed difference might reflect the online construction of non-lexicalized sense. Pickering et al. (2004) reviewed a series of behavioral studies on metonymy, arguing that the reference transfer process required by metonymy *per se* is not costly. Eye-tracking studies investigating either familiar or novel metonymies provided with a supportive context showed similar costs of interpretation for literal and metonymical expressions. Pickering et al. (2004) also demonstrated, however, that the metonymical interpretation might be costly when it requires the online computation of a non-lexicalized/unfamiliar sense without a supportive context, regardless the routinization of the underlying process (e.g., reading Needham). Metonymical expressions used in the present study might work similarly to the unfamiliar metonymy described by Pickering et al. (2004). Although we used a rating procedure to control that each proper name of producers (e.g., a writer, Camilleri) was easily associated with the corresponding products (e.g., a book), our stimuli displayed proper names for which a metonymical interpretation was not fully lexicalized (e.g., Camilleri vs. Dante) and were not provided with a supportive context.

In addition, Pickering et al. (2004) argued that also the generation of the appropriate sense in the case of complement coercion/logical metonymy is associated with processing costs likely due to the generation of unstated semantic content triggered by the mismatching properties between the noun phrase and the verb. This process, involving the computation of additional semantic material, extends the sense of an expression in order to yield a plausible interpretation (see also McElree et al., 2006; McElree, Frisson & Pickering, 2006). All the metonymical expressions used in the present experiment included a mismatch between the verb – which requires an inanimate argument – and the target noun – which refers to a person (e.g., 'That student reads Camilleri'). Similar to the case of logical metonymy, the selectional mismatch between the noun phrase and the predicate, based on the argumental structure of the lexical items, might have triggered the construction of the metonymical sense. Importantly, also in this case the process of generating the metonymical meaning is contingent on context, and the additional cost might indeed reflect the construction of the appropriate sense.

In this respect, however, an alternative explanation should also be considered. Based on the present study, we cannot rule out the possibility that speed differences might reflect time-consuming reanalysis of the syntactic structure of the sentence. For example, a previous SAT study showed intercept differences for thematic role structure violations (e.g., 'Some people amuse books') but not for subcategorization violations (e.g., 'Some people agree books') (McElree & Griffith, 1995), thus suggesting that the thematic violation of verbal requirements is costly. Moreover, Bornkessel et al. (2004), by employing the SAT methodology, demonstrated that additional costs required by some types of syntactic reanalysis, such as case reanalysis, are reflected by differences in the intercept. This point needs to be further investigated in future research, maybe specifically targeting the role of animacy violation in metonymic resolution.

As for the accuracy parameter, we found that metonymical sentences yielded lower overall accuracy than literal sentences. As discussed for metaphor, this finding indicates that participants were less likely to compute a sensible metonymical interpretation of the target noun. A possible interpretation might be that readers are less likely to recover and/or compose the conceptual information that is needed to compute a metonymical meaning (see also McElree & Nordlie, 1999). In particular, the differences in the overall accuracy might reflect the search of general world knowledge associated with the proper names (e.g., the search of the information associated with Camilleri, in order to determine if he is a writer). An alternative interpretation goes in the direction of the grammatical explanation of the speed difference result, as mentioned above: differences in the accuracy parameter might reflect the difficulty of the sentence's re-analysis triggered by the lexical-semantic properties of the linguistic items (see also Bornkessel et al., 2004): re-analysis might increase the likelihood of an error.

In sum, metonymy exhibited lower availability compared to literal expressions, and slower processing speed (though not confirmed in the item analysis). Whether the additional costs associated with metonymy relative to literal processing should be ascribed to process of computing extra- or new semantic content in the context or to the syntactic re-analysis triggered by the lexical properties of the expressions is still an open question. Future experimental investigations should focus on disentangling grammatical processes triggered by lexical/semantic properties of linguistic items and interpretative/pragmatic processes in computing metonymical expressions.

5. Conclusions

Combining linguistic-pragmatic distinctions and experimental evidence, the present study showed that different uses of figurative language, such as metaphor and metonymy, exhibited different processing dynamics relative to literal uses. Specifically, fits of the time-course functions showed that both metaphorical and metonymical meanings resulted in lesser availability than their literal counterparts. Metonymy also exhibited slower processing speed as compared to literals, but metaphor did not. These findings shed light on the current psycholinguistic debate on figurative language showing that general psycholinguistic models, mainly concerning the time-course of figurative language processing regardless the variety of figurative uses, might be quite misleading and need to be further specified with respect to specific phenomena. In particular, results for metaphor are contrary to the traditional formulation of the Standard Pragmatic model, and seem to suggest that metaphor processing is mainly based on conceptual elaboration. Also the findings for metonymy do not clearly support any general psycholinguistic model and suggest that metonymical processing might entail different processes, such as lexical/grammatical processes and/or conceptual processes. These differences between pragmatic phenomena, not specified by current psycholinguistic models, are indeed consistent with the theoretical assumption that metaphor and metonymy are two different cases of nonliteral language use. Although no explicit implications for psycholinguistic models have been formulated by cognitive pragmatic models, the present study suggests the importance of considering, and even integrating, theoretical issues concerning the differentiation of lexical enrichment phenomena in the psycholinguistic discussion on figurative meaning processing, in harmony with current trends in experimental pragmatics.

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169–177.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience*, *20*(7), 2683–2690.

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

Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Frontiers in Systems Neuroscience*, *5, 2*.

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete Words. *American Journal of Psychology*, *117*(3), 389-410.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31*(4), 578–602.

Anderson, A. J., Tao, Y., Murphy, B., & Poesio, M. On discriminating fMRI representations of abstract WordNet taxonomic categories. Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III), COLING 2012, Mumbai, December 2012:21–32.

Andres, M., Olivier, E., & Badets, A. (2008). Actions words and numbers. A motor contribution to semantic processing? *Current Directions in Psychological Science, 17*(5), 313–317.

Annaz, D., Van Herwegen, J., Thomas, M., Fishman, R., Karmiloff-Smith, A., & Rundblad, G. (2009). Comprehension of metaphor and metonymy in children with Williams syndrome. *International Journal of Language Communication Disorders, 44,* 962–978.

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *60*, 69–81.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818–1823.

Baayen, R. (2011). LanguageR: Data sets and functions with ‘Analyzing linguistic data: A practical introduction to statistics’. R package version 1.4 http://CRANRprojectorg/ package=languageR.

Bambini, V., 2010, "Neuropragmatics: A foreword", *Italian Journal of Linguistics,* 22(1): 1-20.

Bambini, V., & Bara, B.G., (2010). "What is neuropragmatics? A brief note", *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore 9(1)*.

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, *86*, 203–216.

Bambini, V., & Resta, D. (2012). Metaphor and experimental pragmatics: When theory meets empirical investigation. *Humana Mente Journal of Philosophical Studies*, 23, 37–60.

Bambini, V., & Trevisan, M. (2012). EsploraCoLFIS: Un’interfaccia web per le ricerche sul Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS). *Quaderni del Laboratorio di Linguistica*, *11,* 1–16.

Bambini, V., Ghio, M., Moro, A., & Schumacher, P.B. Differentiating among pragmatic uses of words through timed sensicality judgments: Metaphor, metonymy and approximation, *submitted*.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 424–434.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., et al., (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral*

Cortex, *22*(11), 2554–2563.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In Pecher, D., & Zwaan, R.A., (Eds.), *Grounding Cognition: The role of perception and action in memory, language and thinking*. New York: Cambridge University Press, 129-163.

Barsalou, Lawrence W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724.

Bartoli, E., Tettamanti, A., Farronato, P., Caporizzo, A., Moro, A., Gatti, R., Perani, D., & Tettamanti, M. (2013). The disembodiment effect of negation: negating action-related sentences attenuates their interference on congruent upper limb movements. *Journal of Neurophysiology*, *109*(7), 1782–1792.

Bastiaansen, J. A. C. J., Thioux, M., & Keysers, C. (2009). Evidence for mirror systems in emotions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1528), 2391–2404.

Beilock, S. L., Lyons, I. M., Mattarella-Micke, A., Nusbaum, H. C., & Small, S. L. (2008). Sports experience changes the neural processing of action language. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*(36), 13269–13273.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-Related fMRI of the auditory cortex. *NeuroImage*, *10*(4), 417–429.

Ben-David, B. M., van Lieshout, P. H. H. M., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *25*(2), 206–220.

Benzécri, J. P. (1973). *L'Analyse des Données. Volume II L'Analyse des Correspondances.* Paris France: Dunod.

Bertinetto, P. M. (1981). *Strutture prosodiche dell'Italiano*. Firenze: Accademia della Crusca.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., et al. (2005). Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS). http://linguisticasnsit/CoLFIS/Homehtm.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

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 (New York, N.Y.: 1991)*, *19*(12), 2767–2796.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., & Buchanan, L. (2003). Neural correlates of lexical access during visual word recognition. *The Journal of Cognitive Neuroscience*, *15*(3), 372–393.

Binder, J.R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *The Journal of Cognitive Neuroscience*, *17*(6), 905–917.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, (5), 9/10*, 341–345.

Bonner, M. F., & Price, A. R. (2013). Where is the anterior temporal lobe and what does it do? *The Journal of Neuroscience*, *33*(10), 4213–4215.

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495–522.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*(2), 185–189.

Bosco, F. M., Vallana, M., & Bucciarelli, M. (2009). Comprehension of communicative intentions: The case of figurative language. *Journal of Cognitive Science*, 10, 245-277.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*, 123–142.

Bott, L., Frisson, S., & Murphy, G.L. (2009). Interpreting conjunctions. *Quarterly Journal of Experimental Psychology*, 62, 681–706.

Boulenger, V., Hauk, O., & Pulvermueller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review*, *15*(6), 1035–1053.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain & Language, 106*, 132–143.

Bryant, P. (1988). Children's counting and concepts of number. *British Journal of Developmental Psychology,* 6(4), 395–397.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*, 400–404.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study.*Cognitive Brain Research*, *24*(3), 355–363.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526.

Butterworth, B. (1999). A head for figures. *Science (New York, N.Y.)*, *284*(5416), 928–929.

Cacciari, C. (2001). *Psicologia del Linguaggio,* Bologna, Il Mulino.

Cacciari, C., & Glucksberg, S. (1994). Understanding figurative language. In Gernsbacher, M. A. (ed.). *Handbook of Psycholinguistics*. New York: Academic Press, 447–477.

Cacciari, C., & Pesciarelli, F. (2013). Motor activation in literal and non-literal sentences: does time matter? *Frontiers in Human Neurosciences*, *7*, 1–10.

Canessa, N., Borgo, F., Cappa, S. F., Perani, D., Falini, A., Buccino, G., et al., (2008). The different neural correlates of action and functional knowledge in semantic memory: an FMRI study. *Cerebral Cortex*, *18*(4), 740–751.

Cappa, S. F. (2008). Imaging studies of semantic memory. *Current Opinion in Neurology*, *21*(6), 669–675.

Cappa, S. F. (2012). Imaging semantics and syntax. Neuroimage, 61(2), 427–431.

Cappa, S. F., & Pulvermueller, F. (2012). Cortex special issue: language and the motor system. *Cortex*, *48*(7), 785–787.

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

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*(8), 354–361.

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*, 651–564.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150–1152.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *The Journal of Cognitive Neuroscience*, *15*(5), 704–717.

Carota, F., Moseley, R., & Pulvermueller, F. (2012). Body-part-specific representations of semantic noun categories. *The Journal of Cognitive Neuroscience*, *24*(6), 1492–1509.

Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: feature and conjunction searches. *Vision Research*, *46*, 2028–2040.

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, *27*(12), 963–972.

Carston, R. (2009). Explicit/implicit distinction. In Cummings, L. (ed.). *The Pragmatics Encyclopedia*. London: Routledge.

Carston, R. (2010). Lexical pragmatics, *ad hoc* concepts and metaphor: A Relevance Theory perspective. *Italian Journal of Linguistics/Rivista di Linguistica*, 22, 153–180.

Carston, R., & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, *3*, 283–312.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition*, *106*(2), 579–593.

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2012). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex*, *23*(3), 739–749.

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition*, *2*(1), 79–116.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin,* 25(2), 215–224.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16*, 22–29.

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211–226.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Conover, W. J., & Iman, R. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124–129.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage*, *22*(1), 11–21.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2011). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience*, *7*(7), 764–770.

Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: Analysis by speed–accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology, 30*, 1–15.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(3), 615–627.

D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia*, *35*, 725–730.

Davison, A. C., & Hinkley, D. V. (1997). Confidence intervals. In: Davison, A. C., & Hinkley, D. V., (editors). Bootstrap methods and their application, Cambridge: Cambridge University Press, 191–251.

De Grauwe, S. De, Swain, A., Holcomb, P. P. J., De Grauwe, S., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. Neuropsychologia, *48*, 1965–1984.

De Winter, J.C. F., & Dodou, D. (2010). Five-point Likert items: t test versus Mann Whitney Wilcoxon. *Practical Assessment Research & Evaluation, 15(11)*, 1–16.

Deamer, F., Pouscoulous, N., & Breheny, R. (2010). A contrastive look at metaphor and hyperbole. *UCL Working Papers in*

Linguistics, 22.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3), 487–506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science (New York, N.Y.)*, *284*(5416), 970–974.

Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract-concrete dichotomy: mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, *42*(4), 1042–1048.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2), 145–177.

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*(2), 284–292.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7(1)*, 1–26.

Egg, M. (2004). Metonymie als Phänomen der Semantik-Pragmatik-Schnittstelle. *Metaphorik*, *6*, 36–53.

Elliott, M. R., Bowtell, R. W., & Morris, P. G. (1999). The effect of scanner sound in visual, motor, and auditory functional MRI. *Magnetic Resonance in Medicine*, *41*(6), 1230–1235.

Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: comment on Larsen, Mercer, and Balota (2006). *Emotion*, *8*(4), 441–444.

Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547.

Fiebach, C. J., Friederici, A. D., Smith, E. E., & Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual—semantic representations in verbal working memory. *The Journal of Cognitive Neuroscience*, *19*(12), 2035–2049.

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex*, *44*(4), 386–392.

Fischer, M. H., & Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology*, *2*, 260.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language,* 56, 357–383.

Foroni, F., & Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Frontiers in Human Neuroscience*, *7*(209).

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, *8*(9), 700–711.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–824.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, *64*(1-2), 74–92.

Frijda, N. H. (2009). Emotions individual differences and time course: Reflections. *Cognition & Emotion*, *23(7)*, 1444–1461.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, *3*, 111–127.

Frisson, S., & Pickering, M.J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *25*, 1366–1383.

Frisson, S., & Pickering, M.J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*, 595–613.

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. *NeuroImage*, *16*(2), 484–512.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al., (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry & Neuroscience*, *34*(6), 418–432.

Gage, N., & Hickok, G. (2005). Multiregional cell assemblies, temporal binding and the representation of conceptual knowledge in cortex: a modern theory by a "classical" neurologist, Carl Wernicke. *Cortex*, *41*(6), 823–832.

Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 236–254.

Gainotti, G. (2004). A metanalysis of impaired and spared naming for different categories of knowledge in patients with a visuo-verbal disconnection. *Neuropsychologia*, *42*(3), 299–319.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3), 455–479.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, *15*(11), 512–519.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology. General*, *113*(2), 256–281.

Ghio, M., & Tettamanti, M. (2010). Semantic domain-specific functional integration for action-related vs. abstract concepts. *Brain and Language*, *112*(3), 223–232.

Ghio, M., Vaghi, M. M. S., & Tettamanti, M. (2013). Fine-grained semantic categorization across the abstract and concrete domains. *PLoS ONE*, *8*(6), e67090.

Gianelli, C., Farnè, A., Salemme, R., Jeannerod, M., & Roy, A. C. (2011). The agent is right: When motor embodied cognition is space-dependent. *PLoS ONE*, *6*(9), e25036.

Gibbs, R. W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *16*, 56–66.

Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding.* Cambridge: Cambridge University Press.

Gibbs, R. W. (2001). Evaluating contemporary models of figurative language understanding. *Metaphor and Symbol*, 16, 317–33.

Gibbs, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind & Language*, *21*(3), 434–458.

Gibbs, R. W., & Colston, H.L. (2012). *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Gibbs, R. W., & Steen, G.J. (eds.) (1999). *Metaphor in Cognitive Linguistics*. Amsterdam: John Benjamins Publishing.

Giora, R. (2003). *On our Mind: Salience, Context, and Figurative language*. New York: Oxford University Press.

Glenberg, A. M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology*, *18*(7), 290–291.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*, 92–96.

Glucksberg, S., & Keyser, B. (1993). How metaphors work. In Ortony, A. (ed.) *Metaphor and Thought*. Cambridge: Cambridge University Press, 401–424.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, *26*(18), 4917–4921.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*(2), 906–912.

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science*, *306*(5695), 496–499.

Gorno-Tempini, M. L., Price, C.J., Josephs, O.E., Vandenberghe, R., Cappa, S.F., Kapur, N., & Frackowiak, R.S. (1998). The neural systems sustaining face and proper-name processing. *Brain*, *121*, 2103–2118.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104–1107.

Greenacre, M. (1993). *Correspondence Analysis in Practice*. London: Academic Press.

Grice, H. P. (1975). Logic and conversation. In Cole, P., & Morgan, J. (eds.) *Syntax and Semantics*. New York: Academic Press, 41–58.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., & Gee, J. (2002). The neural basis for category-specific knowledge: An fMRI study. *NeuroImage*, *15*(4), 936–948.

Hale, S. C. (1988). Spacetime and the abstract/concrete distinction. *Philosophical Studies,* 53(1), 85–102.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*(2), 149–156.

Hanke, M., Halchenko, Y., Sederberg, P., Hanson, S., Haxby, J., & Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53.

Hanson, S. J., & Halchenko, Y. O. (2008). Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. *Neural Computation*, *20*(2), 486–503.

Hauk, O., & Pulvermueller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201.

Hauk, O., Davis, M. H., & Pulvermueller, F. (2008). Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. *NeuroImage*, *42*(3), 1185–1195.

Hauk, O., Johnsrude, I., & Pulvermueller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hauk, O., & Tschentscher, N. (2013). The body of evidence: What can neuroscience tell us about embodied semantics? *Frontiers in Cognitive Science*, 4:, 50.

Haxby, J. V. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534.

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences*, *10*(2), 64–69.

Herlofsky, S. M., & Edmonds, L. A. (2013). Activating situation schemas: the effects of multiple thematic roles on related verbs in a continuous priming paradigm. *Journal of Psycholinguistic Research*, *42*(1), 1–19.

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. *The Journal of Cognitive Neuroscience*, *20*(10), 1799–1814.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721-742.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: evidence from the SNARC effect. *Memory & cognition*, *32*(4), 662–673.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*, 1217–1218.

Jansma, J. M., Ramsey, N. F., Coppola, R., & Kahn, R. S. (2000). Specific versus nonspecific brain activity in a parametric n-back task. *NeuroImage*, *12*(6), 688–697.

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: evidence for dual coding and context availability. *Brain Lang*, *74*(1), 103–112.

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language 55*, 18–32.

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE*, *5*(1), e8622.

Kaplan, J. T., & Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *NeuroImage*, *60*(1), 204–212.

Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition*, *29*(7), 960–967.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of experimental psychology. Learning, memory, and cognition*, *29*(3), 439–446.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355.

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, *107*(1), 16–43.

Kiefer, M., & Barsalou, L. W. (2012). Grounding the human conceptual system in perception action internal states. In Prinz, W., Beisert, M., Herwig, A., (eds). *Action Science: Foundations of an Emerging Discipline,* Cambridge, MA: MIT Press.

Kiefer, M., & Pulvermueller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825.

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, *28*(47), 12224–12230.

Kiehl, K. A., Liddle, P. F., Smith, A. M., Mendrek, A., Forster, B. B., & Hare, R. D. (1999). Neural pathways involved in the processing of concrete and abstract words. *Human Brain Mapping*, *7*(4KLS+99), 225–233.

Kintsch, W., & Bowles, A. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol*, 17, 249–262.

Klein, D. E., & Murphy, G.L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*, 259–282.

Klein, E., Moeller, K., Willmes, K., Nuerk, H.-C., & Domahs, F. (2011). The influence of implicit hand-based representations on mental arithmetic. *Frontiers in Psychology*, *2*, 197.

Klepousniotou, E., & Baum, S. R. (2005a). Unilateral brain damage effects on processing homonymous and polysemous words.

Brain and Language, *93*, 308–326.

Klepousniotou, E., & Baum, S. R. (2005b). Processing homonymy and polysemy: Effects of sentential context and time-course following unilateral brain damage. *Brain and Language*, *95*, 365–382.

Knapp, T. R. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing Research, 39*, 121–123.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain: Emotional speech and language comprehension. *Language and Linguistics Compass*, *5*(3), 108–125.

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, *56*(2), 411–421.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews. Neuroscience*, *6*(9), 691–702.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *11*, 145–155.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By.* Chicago: University of Chicago Press.

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

Lang, P. J. (1993). The three system approach to emotion. In: Birbaumer, N., & Ohman, O., (eds). *The Organization of Emotion.*Toronto: Hogrefe-Huber.

Lascarides, A., & Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics, 34*, 387–414.

LeDoux, J. (1996). Emotional networks and motor control: a fearful view. *Progress in Brain Research*, *107*, 437–446.

Leys C., & Schumann, S. (2010). A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology,* 46(4), 684–688.

Lindemann, O., Abolafia, J. M., Pratt, J., & Bekkering, H. (2008). Coding strategies in number space: memory requirements influence spatial-numerical associations. *Quarterly Journal of Experimental Psychology (2006)*, *61*(4), 515–524.

Linnaeus, C. (1735). *System naturae sive regna tria naturae systematice proposita per classes ordines genera & species*. Leiden: Haak, 1–12.

Liu, C. C., & Smith, P.L. (2009). Comparing time–accuracy curves: Beyond goodness-of-fit measures. *Psychonomic Bulletin and Review*, *16*, 190–203.

Liuzza, M. T., Candidi, M., & Aglioti, S. M. (2011). Do not resonate with actions: sentence polarity modulates cortico-spinal excitability during action-related sentence reading. *PloS one*, *6*(2), e16855.

Liversedge, S. P., Pickering, M. J., Clayes, E. L., & Branigan, H. P. (2003). Thematic processing of adjuncts: evidence from an eye-tracking experiment. *Psychonomic Bulletin & Review*, *10*(3), 667–675.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., & Baron-Cohen, S. (2009). Shared neural circuits for mentalizing about the self and others. *The Journal of Cognitive Neuroscience*, *22*(7), 1623–1635.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research*

Methods, Instruments, and Computer, 28(2): 203–208.

Macoir, J. (2009). Is a plum a memory problem? *Neuropsychologia*, *47*(2), 518–535.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport 5, 2000*, *11*(8), 1729–1733.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1-3), 59–70.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action, *Science*, 102–105.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*(1), 25–45.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language, 58*, 879–906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1231–1239.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language, 64*, 327–343.

Martin, C. B., McLean, D. A., O'Neil, E. B., & Köhler, S. (2013). Distinct familiarity-based response patterns for faces and buildings in perirhinal and parahippocampal cortex. *The Journal of Neuroscience*, *33*(26), 10915–10923.

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: A meta-analysis. *NeuroImage*, *31*(4), 1453–1474.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322.

McElree, B., & Dosher, B.A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *4*, 346–373.

McElree, B., & Dosher, B.A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, *122*, 291–315.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 134–157.

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*, 486–494.

McElree, B., Frisson, S., & Pickering, M. (2006). Deferred Interpretations: why starting dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*, 181–192.

McElree, B., Pylkkänen, L., Pickering, M., & Traxler, M. (2006). The time course of enriched composition. *Psychonomic Bulletin & Review, 13*, 53–59.

McKiernan, K. A., D'Angelo, B. R., Kaufman, J. N., & Binder, J. R. (2006). Interrupting the "stream of consciousness": an fMRI investigation. *NeuroImage*, *29*(4), 1185–1191.

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. Psychology Press.

McRae, K., & Jones, M. (2013). Semantic Memory. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology*. USA: Oxford University Press.

Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport*, *9*(5), 803–808.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804.

Meyer, D. E., Osman, A.M., Irwin, D.E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3–67.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, *99*(23), 15238–15243.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*(1-2), 145–175.

Moro, A. (1997). *The Raising of Predicates. Predicative Noun Phrases and The theory of Clause Structure*. England: Cambridge University Press.

Moseley, R., Carota, F., Hauk, O., Mohr, B., & Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex*, *22*(7), 1634–1647.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109.

Murtagh, F. (2005). *Correspondence Analysis and Data Coding with JAVA and R*. CRC Press.

Nanna M. J., & Sawilowsky S. S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. *Psychological Methods*, *3(1)*, 55–67.

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: enhanced sensitivity to negative word valence. *Emotion*, *9*(5), 609–618.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual review of neuroscience*, *32*, 185–208.

Noppeney, U., & Price, C. J. (2004). Retrieval of abstract semantics. *NeuroImage*, *22*(1), 164–70.

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education: Theory and Practice, 15*, 625–632.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357.

Noveck, I. A. & Reboul, A. (2008). Experimental Pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*, 425–431.

Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, 16, 109–121.

Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics*, *12*, 109-32.

O’Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590.

O’Toole, A. J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of cognitive neuroscience*, *19*(11), 1735–1752.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Olivetti, E., Veeramachaneni, S., & Nowakowska, E. (2012). Bayesian hypothesis testing for pattern discrimination in brain decoding. *Pattern Recognition*, *45*(6), 2075–2084.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, *130*(7), 1718–1731.

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, 57, 126–149.

Paivio A., Yuille J. C., Madigan, S. A. (1968). Concreteness imagery and meaningfulness values for 925 nouns. *Journal Experimental Psychology, 76*(1), 1–25.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah,NJ,US: Lawrence Erlbaum Associates Publishers.

Paivio, A., (1986). *Mental Representations: a Dual Coding Approach*. Oxford England: Oxford University Press.

Panther, K. U., & Thornburg, L. (2007). Metonymy. In Geeraerts, D., & Cuyckens, H. (eds.). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 236–263.

Papafragou, A. (1996). On metonymy. *Lingua*, *99*, 169–195.

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: the acquisition of belief verbs. *Cognition*, *105*(1), 125–165.

Papeo, L., Corradi-Dell'Acqua, C., & Rumiati, R. I. (2011). “She” is not like “I”: The tie between language and action is in our imagination. *The Journal of Cognitive Neuroscience*, *23*(12), 3939–3948.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Penny, W., & Holmes, A. P. (2004). Random-effects analysis. In *Human Brain Function*, Frackowiak, R.S.J., Friston, K. J., Frith, C. D., Dolan, R. J., Price, C. J., Zeki, S., Ashburner, J. & Penny, W (eds.). London: Elsevier Press, 843–850.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of verb and noun processing. A PET study. *Brain*, *122*(12), 2337–2344.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, *45*(1), S199–S209.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), 773–783.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *The Journal of Cognitive Neuroscience*, *19*(8), 1407–1419.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in psychology*, *3*, 612.

Pfeifer, R., & Scheier, C. (2001). *Understanding Intelligence*. Cambridge MA: MIT Press.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503.

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. (2004). Eye movements and semantic composition. In Carreiras, M. & Clifton, C. Jr. (eds.). *The On-Line Study of Sentence Comprehension: Eyetracking, Erps, and Beyond*. Hove: Psychology Press.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., et al., (2007). Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neuroscience*, *8*(1), 91.

Plaut, D. C., & Shallice, T. (1991). Effects of word abstractness in a connectionist model of deep dyslexia. *Proceedings of the 13th annual meeting of the Cognitive Science Society,* Chicago, 73–78.

Poldrack, R. A. (2008). The role of fMRI in Cognitive Neuroscience: where do we stand? *Current Opinion in Neurobiology*, *18*(2), 223–227.

Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*, *43*(3), 634–644.

Previtali, P., Rinaldi, L., & Girelli, L. (2011). Nature or nurture in finger counting: a review on the determinants of the direction of number-finger mapping. *Frontiers in Psychology*, *2*, 363.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88.

Pulvermueller, F. (1999). Words in the brain's language. *The Behavioral and Brain Sciences*, *22*(2), 253–279.

Pulvermueller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*(12), 517–524.

Pulvermueller, F. (2005). Brain mechanisms linking language and action. *Nature reviews. Neuroscience*, *6*(7), 576–582.

Pulvermueller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360.

Pulvermueller, F., Cook, C., & Hauk, O. (2012). Inflection in action: Semantic motor system activation to noun- and verb-containing phrases is modulated by the presence of overt grammatical markers. *NeuroImage*, *60*(2), 1367–1379.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge MA: MIT Press.

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. *The Journal of Cognitive Neuroscience*, *23*(10), 2636–2649.

R Core Team (2012) R: A language and environment for statistical computing R Foundation for Statistical Computing Vienna Austria ISBN 3-900051-07-0 URL http://wwwR-projectorg/.

Ranzini, M., Lugli, L., Anelli, F., Carbone, R., Nicoletti, R., & Borghi, A. M. (2011). Graspable objects shape number processing. *Frontiers in Human Neuroscience*, *5*, 147.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396.

Rapp, A. M., Erb, M., Grodd, W., Bartels, M., & Markert, K. (2011). Neural correlates of metonymy resolution. *Brain and language*, *119*, 196–205.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, *21*, 139–155.

Recanati, F. (2010). *Truth-conditional Pragmatics.* Oxford: Oxford University Press.

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825.

Reed, A. V. (1973). Speed–accuracy trade-off in recognition memory. *Science*, *181*, 574–576.

Reker, M., Ohrmann, P., Rauch, A. V., Kugel, H., Bauer, J., Dannlowski, U., et al. (2010). Individual differences in alexithymia and brain response to masked emotion faces. *Cortex*, *46*(5), 658–667.

Rotteveel, M., & Phaf, R. H. (2004). Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion*, *4*(2), 156–172.

Rubio-Fernández, P. (2007). Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *Journal of Semantics*, 24, 345–371.

Rueschemeyer, S.-A., Brass, M., & Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor sytems. *The Journal of Cognitive Neuroscience*, *19*(5), 855–865.

Rueschemeyer, S.-A., van Rooij, D., Lindemann, O., Willems, R. M., & Bekkering, H. (2010). The function of words: distinct neural correlates for words denoting differently manipulable objects. *Journal of Cognitive Neuroscience*, *22*(8), 1844–1851.

Ruiz de Mendoza Ibáñez, F. J. (2007). High-level cognitive models: In search of a unified framework for inferential and grammatical behavior. In Kosecki, K. (ed.) *Perspectives on Metonymy.* Frankfurt/Main: Peter Lang, 11-30.

Rundblad, G. & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*, 547–563.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39(6)*, 1161–1178.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage*, *27*, 188–200.

Sato, M., Cattaneo, L., Rizzolatti, G., & Gallese, V. (2007). Numbers within our hands: modulation of corticospinal excitability of hand muscles during numerical judgment. *Journal of Cognitive Neuroscience*, *19*(4), 684–693.

Sato, M., Mengarelli, M., Riggio, L., Gallese, V., & Buccino, G. (2008). Task related modulation of the motor system during language processing. *Brain and Language*, *105*(2), 83–90.

Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., Phillips, J., Richiardi, J., & Mourão-Miranda, J. (2013). PRoNTo: Pattern Recognition for Neuroimaging Toolbox. *Neuroinformatics*, 319–337.

Schumacher, P. B. (2011). The hepatitis called …: Electrophysiological evidence for enriched composition. In Meibauer, J., & Steinbach, M. (eds.) *Experimental Pragmatics/Semantics*. Amsterdam/Philadelphia: John Benjamins, 199–219.

Schumacher, P. B. (2012). Context in neurolinguistics: Time-course data from electrophysiology. In Finkbeiner, R., Meibauer, J., & Schumacher, P.B. (eds.). *What is a Context? Linguistic Approaches and Challenges*. Amsterdam/Philadelphia: John Benjamins, 33–53.

Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In Schwanenflugel P. J., (ed.). *The Psychology of Word Meanings*. Hillsdale NJ: Erlbaum, 223–250.

Schwanenflugel, P. J., Akin, C., & Luh, W. M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*, 499–520.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: comparing SNARC effects with saccadic and manual responses. *Perception & psychophysics*, *66*(4), 651–664.

Searle, J. (1979). Metaphor. In Ortony, A. (ed.) *Metaphor and Thought*. Cambridge: Cambridge University Press, 92–123.

Semenza, C., Denes, G., Lucchese, D., & Bisiacchi, P. (1980). Selective deficit of conceptual structures in aphasia: Class versus thematic relations. *Brain and Language*, *10*, 243–248.

Setti, A., & Caramelli, N. (2005). Different domains in abstract concepts. In *Proceedings of the XXVII Annual Conference of the Cognitive Science Society,* 1997–2002.

Shinkareva, S. V., Malave, V. L., Just, M. A., & Mitchell, T. M. (2012). Exploring commonalities across participants in the neural representation of objects. *Human Brain Mapping*, *33*(6), 1375–1383.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE*, *3*(1), e1394.

Shtyrov, Y., Hauk, O., & Pulvermueller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *The European Journal of Neuroscience*, *19*(4), 1083–1092.

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cognitive neuropsychology*, *20*(3), 451–486.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology, Paris*, *102*(1-3), 106–119.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*(10), 1602–1608.

Simmons, W. K., Reddish, M., Bellgowan, P. S. F., & Martin, A. (2009). The selectivity and functional connectivity of the anterior temporal lobes. *Cerebral Cortex*, *20*(4), 813–825.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19(5)*, 317–330.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (2005). Pragmatics. In Jackson, F., & Smith, M. (eds.). *Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press, 468–501.

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *The Journal of Cognitive Neuroscience*, *21*(3), 489–510.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, *19*(6), 498–502.

Staresina, B. P., Duncan, K. D., & Davachi, L. (2011). Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *Journal of Neuroscience*, *31*(24), 8739–8747.

Subramaniam, K., Faust, M., Beeman, M., & Mashal, N. (2012). The repetition paradigm: Enhancement of novel metaphors and suppression of conventional metaphors in the left inferior parietal lobe. *Neuropsychologia*, *50*, 2705–19.

Sugiura, M., Shah, N. J., Zilles, K., & Fink, G. R. (2005). Cortical representations of personally familiar objects and places: functional organization of the human posterior cingulate cortex. *The Journal of Cognitive Neuroscience*, *17*(2), 183–198.

Suslow, T., Kugel, H., Reber, H., Bauer, J., Dannlowski, U., Kersting, A., Arolt, W., Heindel, P., Ohrmann, & Egloff; B. (2010). Automatic brain response to facial emotion as a function of implicitly and explicitly measured extraversion. *Neuroscience*, *167*(1), 111–123.

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neuro-cognitive account. *Language and Cognitive Processes*, *26*(9), 1368–1401.

Taylor, K. I., Moss, H. E., & Tyler, L. K. (2007). The conceptual structure account: A cognitive model of semantic memory and its neural instantiation. In: Hart J., & Kraut M. (eds). *The Neural Basis of Semantic Memory*. Cambridge: Cambridge University Press, 265–301.

Tettamanti, M., & Moro, A. (2012). Can syntax appear in a mirror (system)? *Cortex*, *48*(7), 923–935.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Tettamanti, M., Manenti, R., Rosa, P. A. D., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2008). Negation in the brain: modulating action representations. *Neuroimage*, *43*(2), 358–367.

Tettamanti, M., Rognoni, E., Cafiero, R., Costa, T., Galati, D., & Perani, D. (2012). Distinct pathways of neural coupling for different basic emotions. *NeuroImage*, *59*(2), 1804–1817.

Tomasino, B., Weiss, P. H., & Fink, G. R. (2010). To move or not to move: imperatives modulate action-related verb processing in the motor system. *Neuroscience*, *169*(1), 246–258.

Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fMRI study of mental imagery and silent reading of action phrases. *NeuroImage*, *36*, T128–T141.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research*, *168*(3), 242–249.

Traxler, M. J. (2011). *Introduction to Psycholinguistics: Understanding Language Science*. New York: Wiley-Blackwell.

Tschentscher, N., Hauk, O., Fischer, M. H., & Pulvermüller, F. (2012). You can count on the motor cortex: finger counting habits modulate motor cortex activation evoked by numbers. *NeuroImage*, *59*(4), 3139–3148.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247–263.

Vigliocco, G., Kousta, S.-T., Rosa, P. A. D., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (*in press*). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition 1(2)*, 219-247.

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

Vindiola, M., & Wolmetz, M. (2011). Mental encoding and neural decoding of abstract cognitive categories: A commentary and simulation. *NeuroImage*, *54*(4), 2822–2827.

Vinson, D., Anderson, A., Ratoff, W., Bahrami, B., & Vigliocco, G. (2011) Slow to anger: emergence of emotionally loaded words and faces from interocular suppression. Poster presented at the Annual Conference of the Cognitive Science Society, Boston MA.

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Human Brain Mapping*, *31*(10), 1459–1468.

Whatmough, C., Verret, L., Fung, D., & Chertkow, H. (2004). Common and constrasting areas of activation for abstract and concrete concepts: an H2 15O PET study. *The Journal of Cognitive Neuroscience*, *16*(7), 1211–1226.

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, *29*(5), 719–736.

Wiemer-Hastings, K., Barnard, K. K., & Faelnar, J. (2003). Structural differences in abstract and concrete ítem categories. In Proceedings of 25th Annual Meeting of the Cognitive Science Society, Boston MA, USA. pp. 1253–1258.

Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraints, and abstractness. *Proceedings of 23rd Annual Meeting of the Cognitive Science Society*, Edimburgh, Scotland. Hillsdale NJ: Lawrence Erlbaum Associates, 1106–1111.

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychological Science*, *21*(1), 67–74.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. *The Journal of cognitive neuroscience*, *25*(6), 920–935.

Wilson, D. (2003). Relevance Theory and lexical pragmatics. *Italian Journal of Linguistics/Rivista di Linguistica*, 15, 273–291.

Wilson, D., & Carston, R. (2006). Metaphor, relevance and 'the emergent property' issue. *Mind & Language*, *21*, 404–433.

Wilson, D., & Carston, R. (2007).A unitary approach to lexical pragmatics: Relevance, inference and *ad hoc* concepts. In Burton-Roberts, N. (ed.). *Pragmatics*, Basingstoke and New York: Palgrave Macmillan, 230–259.

Wilson, D. (2011). Parallels and differences in the treatment of metaphor in relevance theory and cognitive linguistics. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, *128*, 195-213.

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H.-J., & Kotz, S. A. (2010). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex*, *20*(2), 383–392.

Zaehle, T., Schmidt, C. F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P., & Jancke, L. (2007). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *NeuroImage*, *37*(4), 1195–1204.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(15), 6430–6435.

Appendices

Appendix 1. Stimuli and ratings of Study A and Study B

Appendix 1.1. Rating study 1: Association rating

Association ratings for (Em) emotion-, (Ma) mathematics-, and (Ms) mental state-related sentences.

		Label	**Emotion association**			**Mathematics association**			**Mental State association**		
			Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
1	Lei affronta l'angoscia	Em	5.00	4.50	2.07	1.00	1.13	0.35	1.00	2.25	2.12
2	Lei agogna la fama	Em	5.50	4.63	2.33	1.50	2.00	1.20	2.00	2.75	2.05
3	Lei avverte l'eccitazione	Em	6.50	5.63	2.00	1.00	1.63	1.41	2.50	2.63	1.69
4	Lei brama l'amore	Em	4.00	3.88	1.13	1.00	1.25	0.71	2.00	2.50	1.20
5	Lei calma la rabbia	Em	7.00	7.00	0.00	1.00	1.63	1.06	3.50	3.75	1.98
6	Lei cela l'invidia	Em	5.00	5.25	1.28	1.00	1.00	0.00	2.00	2.63	1.85
7	Lei depreca la crudeltà	Em	7.00	7.00	0.00	1.00	1.63	1.77	1.00	2.13	1.89
8	Lei detesta la cattiveria	Em	6.00	5.38	2.00	1.00	1.25	0.46	2.00	1.75	0.71
9	Lei disdegna l'ostilità	Em	6.00	5.50	0.76	1.00	1.00	0.00	1.50	2.13	1.36
10	Lei disprezza la viltà	Em	5.00	5.00	1.51	1.00	1.25	0.71	1.00	1.63	1.41
11	Lei dissimula il disappunto	Em	5.50	5.75	1.16	1.50	1.63	0.74	1.50	1.88	1.13
12	Lei domina la passione	Em	6.50	6.25	0.89	1.00	1.00	0.00	1.00	1.50	1.07
13	Lei gradisce la sorpresa	Em	3.00	3.63	2.39	1.00	2.00	1.85	1.50	2.63	2.45
14	Lei inibisce l'ansia	Em	1.50	2.50	1.93	1.00	2.38	1.92	1.00	1.75	1.39
15	Lei invidia l'orgoglio	Em	4.00	3.63	2.45	1.00	1.50	1.07	2.00	2.75	2.12
16	Lei manifesta l'odio	Em	5.00	5.00	1.51	1.00	1.63	1.06	2.00	2.25	1.67
17	Lei mostra la contentezza	Em	6.00	5.50	1.69	1.00	1.63	1.77	1.50	2.13	2.03
18	Lei nutre l'affetto	Em	3.00	4.13	2.23	1.00	1.13	0.35	1.50	1.88	1.36
19	Lei odia la sofferenza	Em	6.50	6.13	0.99	1.00	1.63	1.77	2.50	2.88	2.10
20	Lei patisce la vergogna	Em	4.00	4.25	1.67	1.00	1.88	2.10	2.00	2.00	0.76
21	Lei percepisce lo stupore	Em	7.00	6.13	1.46	1.00	1.00	0.00	1.00	2.00	1.41
22	Lei placa l'ira	Em	2.50	3.13	1.89	1.00	1.50	1.41	2.00	2.88	2.36
23	Lei predilige l'allegria	Em	4.50	4.88	1.96	1.00	2.13	2.23	2.00	2.75	1.83
24	Lei prova il disgusto	Em	4.00	4.13	2.47	1.00	1.00	0.00	2.00	2.75	1.83
25	Lei reprime lo sconforto	Em	7.00	6.38	1.19	1.00	1.00	0.00	4.00	4.50	2.07
26	Lei serba il rancore	Em	5.00	4.38	2.45	1.00	1.00	0.00	1.00	1.00	0.00
27	Lei soddisfa il desiderio	Em	2.00	3.00	2.39	1.00	1.50	0.93	2.50	3.00	2.07
28	Lei soffre la solitudine	Em	5.00	4.63	1.30	1.00	1.14	0.38	1.00	1.71	1.25
29	Lei sogna la felicità	Em	4.00	3.75	1.75	1.00	1.50	1.07	1.00	1.75	1.39
30	Lei subisce la collera	Em	3.00	4.13	1.89	1.00	1.00	0.00	1.00	2.13	1.89
31	Lei svela l'imbarazzo	Em	3.50	3.00	1.51	2.50	2.88	2.17	2.00	2.38	1.51
32	Lei teme la tristezza	Em	6.00	5.00	1.77	1.00	1.13	0.35	1.50	2.13	1.36

		Label	**Emotion association**			**Mathematics association**			**Mental State association**		
			Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
33	Lei tollera la delusione	Em	5.00	4.88	0.99	1.00	1.86	2.27	3.50	4.00	2.20
34	Lei trasmette la gioia	Em	5.50	5.50	1.41	1.00	1.38	0.74	1.00	1.00	0.00
35	Lei trattiene l'esultanza	Em	2.50	3.00	1.85	1.00	1.00	0.00	1.50	2.25	1.75
1	Lei addiziona i gradi	Ma	1.00	2.00	1.93	5.00	4.88	1.13	3.00	2.50	0.76
2	Lei aggiunge il totale	Ma	1.00	1.63	1.77	5.00	5.00	1.07	3.00	3.25	1.98
3	Lei analizza i conti	Ma	1.00	1.13	0.35	6.50	6.00	1.41	1.00	2.50	2.33
4	Lei applica le formule	Ma	1.00	1.13	0.35	7.00	7.00	0.00	1.00	1.75	1.49
5	Lei arrotonda i valori	Ma	1.00	1.38	0.74	5.50	5.13	2.17	4.00	4.00	2.27
6	Lei calcola l'integrale	Ma	1.00	1.25	0.46	5.00	4.63	1.77	1.00	1.75	1.75
7	Lei centuplica la somma	Ma	1.00	1.13	0.35	5.50	4.88	1.96	1.00	2.00	1.41
8	Lei computa la costante	Ma	1.00	1.75	1.49	7.00	6.50	0.76	2.00	3.00	2.33
9	Lei conta gli insiemi	Ma	1.00	1.75	1.39	4.00	4.00	1.51	3.50	3.38	1.92
10	Lei conteggia le cifre	Ma	1.00	1.50	1.07	5.50	5.75	1.16	4.00	3.75	1.83
11	Lei correla le medie	Ma	1.00	1.13	0.35	4.50	4.63	1.92	1.00	1.50	1.07
12	Lei deriva la funzione	Ma	1.00	1.25	0.71	3.50	3.38	2.20	3.00	2.88	1.81
13	Lei dimezza il resto	Ma	1.00	1.88	1.46	7.00	6.75	0.46	1.00	1.75	1.75
14	Lei divide i decimali	Ma	1.00	1.63	1.77	6.50	6.00	1.20	1.00	1.63	0.92
15	Lei duplica il prodotto	Ma	1.00	1.75	2.12	5.00	4.00	2.56	1.00	1.63	1.06
16	Lei esegue la sottrazione	Ma	1.00	1.38	1.06	7.00	6.63	0.74	1.50	2.13	2.03
17	Lei fraziona il totale	Ma	1.00	1.75	1.16	5.50	4.13	2.64	1.50	2.63	2.20
18	Lei integra il logaritmo	Ma	1.00	1.88	2.10	4.50	4.88	1.64	1.00	1.25	0.71
19	Lei moltiplica l'unità	Ma	1.00	1.50	0.93	5.00	4.88	1.96	1.00	1.00	0.00
20	Lei numera gli insiemi	Ma	1.00	1.25	0.46	6.50	5.88	1.36	1.00	1.50	1.07
21	Lei permuta la matrice	Ma	1.00	1.63	1.41	6.00	5.25	1.98	1.50	2.13	1.73
22	Lei pondera le medie	Ma	1.00	1.13	0.35	5.00	4.75	1.83	4.00	4.13	1.46
23	Lei quadruplica i dati	Ma	1.00	2.25	2.38	6.50	6.13	1.13	1.00	1.86	1.86
24	Lei raddoppia il risultato	Ma	1.00	1.00	0.00	5.50	4.75	1.98	1.00	1.50	0.76
25	Lei ricava la soluzione	Ma	1.00	1.75	1.49	7.00	6.63	0.52	1.00	2.38	2.20
26	Lei riduce la varianza	Ma	1.00	2.25	2.38	6.00	5.75	1.39	1.00	2.00	1.77
27	Lei risolve l'equazione	Ma	1.00	1.63	1.77	3.50	3.63	1.51	1.00	1.25	0.46
28	Lei scompone la formula	Ma	1.00	2.25	2.12	6.00	5.38	2.07	1.00	1.00	0.00
29	Lei semplifica i calcoli	Ma	1.00	1.00	0.00	5.00	4.88	1.64	4.00	4.00	1.51
30	Lei somma gli addendi	Ma	1.00	1.63	1.06	7.00	5.63	2.26	1.00	1.13	0.35
31	Lei sottrae l'incognita	Ma	1.50	1.88	1.36	5.00	4.25	1.67	3.50	3.63	1.69
32	Lei sviluppa il teorema	Ma	1.00	1.13	0.35	7.00	5.75	2.19	6.00	5.75	0.71
33	Lei svolge il problema	Ma	1.00	1.25	0.46	7.00	6.75	0.46	1.50	1.88	1.13
34	Lei triplica la somma	Ma	1.00	1.50	0.93	6.50	5.50	2.14	2.00	2.75	2.25

		Label	**Emotion association**			**Mathematics association**			**Mental State association**		
			Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
35	Lei uguaglia le variabili	Ma	1.00	1.13	0.35	3.50	3.25	2.12	2.50	3.13	1.64
1	Lei apprende la nozione	Ms	1.00	2.00	1.93	3.50	3.88	2.64	5.50	5.38	2.00
2	Lei approva l'ideale	Ms	3.00	3.50	1.85	1.00	1.88	1.73	3.50	4.25	1.91
3	Lei arguisce lo scopo	Ms	1.50	2.75	2.25	1.00	2.00	1.60	3.00	3.25	1.39
4	Lei concepisce il dilemma	Ms	1.50	1.63	0.74	1.50	2.75	2.31	5.00	4.86	1.95
5	Lei condivide l'analisi	Ms	1.00	1.88	1.64	2.50	3.25	2.55	2.50	3.13	2.36
6	Lei congettura la tesi	Ms	1.00	1.00	0.00	1.00	1.88	1.36	5.50	4.88	1.96
7	Lei constata la finalità	Ms	1.00	2.00	2.07	1.00	1.14	0.38	4.00	3.75	2.31
8	Lei decreta la sorte	Ms	1.00	1.50	0.76	3.50	3.38	1.06	1.00	1.75	1.39
9	Lei desume la questione	Ms	2.50	3.13	2.47	1.00	1.63	1.41	6.50	6.00	1.20
10	Lei dimentica la promessa	Ms	1.00	2.13	1.89	1.00	1.38	0.74	5.00	4.75	1.39
11	Lei discerne l'opinione	Ms	2.00	2.50	1.77	3.50	3.50	1.93	3.00	3.75	1.39
12	Lei distingue i concetti	Ms	1.00	1.75	1.75	3.00	3.25	2.19	5.50	5.38	1.60
13	Lei distoglie la mente	Ms	4.00	4.13	1.55	2.00	3.13	2.47	1.00	1.75	1.16
14	Lei distrae l'attenzione	Ms	2.00	2.50	1.20	2.00	2.63	2.20	1.50	2.50	2.14
15	Lei esamina l'opzione	Ms	1.50	1.88	1.13	2.00	2.13	1.25	5.50	4.75	1.75
16	Lei ignora le fonti	Ms	1.00	2.13	1.81	1.00	1.13	0.35	1.50	2.50	2.14
17	Lei immagina il giudizio	Ms	2.50	2.38	1.30	1.00	2.25	2.38	4.00	4.25	1.39
18	Lei impara la lezione	Ms	1.00	1.38	1.06	3.50	3.50	2.20	3.00	3.63	2.45
19	Lei individua lo sbaglio	Ms	4.50	3.75	2.12	4.50	4.13	2.53	2.50	3.25	2.12
20	Lei influenza la scelta	Ms	4.50	3.88	2.36	4.00	3.88	2.85	4.50	4.13	1.36
21	Lei intuisce la novità	Ms	1.00	1.63	0.92	2.00	2.25	1.49	2.00	2.38	1.19
22	Lei medita l'alternativa	Ms	4.50	4.25	2.25	1.00	1.75	2.12	5.00	5.13	1.55
23	Lei memorizza la procedura	Ms	1.00	1.25	0.46	2.50	3.00	1.41	5.50	4.63	2.07
24	Lei nota l'imprecisione	Ms	1.00	1.13	0.35	1.00	1.38	0.52	6.50	5.88	2.03
25	Lei pianifica l'avvenimento	Ms	2.00	2.75	1.83	2.00	2.75	2.19	6.00	5.13	2.23
26	Lei rammenta l'episodio	Ms	3.00	3.25	2.19	1.00	2.50	2.51	3.00	3.50	2.00
27	Lei rievoca il ricordo	Ms	2.00	2.38	1.51	1.00	1.13	0.35	6.00	5.75	1.04
28	Lei rimugina i pensieri	Ms	4.00	3.88	1.64	1.00	1.63	1.77	6.00	5.63	1.51
29	Lei sbaglia il pronostico	Ms	1.00	1.13	0.35	3.50	3.75	1.91	5.00	4.75	2.05
30	Lei scorda l'intento	Ms	3.50	3.25	2.19	1.50	2.50	2.27	7.00	5.63	2.20
31	Lei simula l'interesse	Ms	3.00	3.13	1.55	1.00	1.38	1.06	1.50	2.13	1.55
32	Lei stima la franchezza	Ms	2.50	2.63	1.77	1.00	1.00	0.00	4.00	4.25	1.75
33	Lei suppone l'accaduto	Ms	1.00	1.50	0.76	1.00	1.38	1.06	7.00	6.38	0.92
34	Lei vaglia il parere	Ms	1.00	1.00	0.00	1.00	1.00	0.00	6.00	5.57	1.51
35	Lei valuta il merito	Ms	2.00	2.50	1.60	4.00	3.88	2.47	7.00	6.38	0.92

Appendix 1.2. Rating study 1: Body-part rating

Body-part ratings for (Em) emotion-, (Ma) mathematics-, (Ms) mental state-, (Mo) mouth-, (Ha) hand-, and (Le) leg-related sentences.

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
1	Lei affronta l'angoscia	Em	2.50	2.63	1.69	1.50	2.13	1.46	1.00	1.50	0.76
2	Lei agogna la fama	Em	1.00	2.50	2.51	1.00	1.50	0.93	1.00	1.25	0.71
3	Lei avverte l'eccitazione	Em	5.50	4.75	2.43	5.50	4.75	2.43	5.00	4.25	2.55
4	Lei brama l'amore	Em	1.00	2.25	1.83	1.00	1.63	1.41	1.00	1.13	0.35
5	Lei calma la rabbia	Em	5.50	4.88	2.23	5.00	4.13	2.42	2.00	3.00	2.27
6	Lei cela l'invidia	Em	5.00	4.88	2.17	2.50	3.13	2.36	1.00	2.13	1.89
7	Lei depreca la crudeltà	Em	5.00	5.13	1.55	3.00	3.50	2.14	1.50	2.25	2.05
8	Lei detesta la cattiveria	Em	1.50	2.25	1.75	1.00	1.63	1.19	1.00	1.25	0.71
9	Lei disdegna l'ostilità	Em	5.00	4.50	2.62	1.00	2.00	1.77	1.00	1.25	0.71
10	Lei disprezza la viltà	Em	3.50	3.63	2.07	1.50	2.00	1.20	1.00	1.50	0.93
11	Lei dissimula il disappunto	Em	5.00	5.50	1.41	3.00	3.25	2.05	1.00	2.00	1.60
12	Lei domina la passione	Em	5.00	4.25	2.87	2.50	3.50	2.83	2.50	3.13	2.42
13	Lei gradisce la sorpresa	Em	5.50	5.00	2.00	4.00	4.00	2.56	2.00	2.75	2.12
14	Lei inibisce l'ansia	Em	1.50	2.38	2.13	1.50	2.13	1.46	1.50	1.75	0.89
15	Lei invidia l'orgoglio	Em	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
16	Lei manifesta l'odio	Em	5.00	5.13	1.46	4.00	4.25	1.67	1.00	2.13	1.89
17	Lei mostra la contentezza	Em	6.00	5.50	1.69	3.00	3.63	2.72	1.00	2.00	1.60
18	Lei nutre l'affetto	Em	1.00	2.00	1.60	1.00	2.00	1.41	1.00	1.00	0.00
19	Lei odia la sofferenza	Em	1.00	1.50	1.07	1.00	1.00	0.00	1.00	1.00	0.00
20	Lei patisce la vergogna	Em	3.00	3.00	2.00	1.00	2.25	1.83	1.00	1.88	1.46
21	Lei percepisce lo stupore	Em	5.50	5.25	2.05	3.50	3.50	1.60	1.50	1.75	1.04
22	Lei placa l'ira	Em	3.50	3.50	1.93	2.50	3.13	1.89	1.50	2.38	2.13
23	Lei predilige l'allegria	Em	2.50	2.63	1.77	1.00	1.63	1.19	1.00	1.50	0.93
24	Lei prova il disgusto	Em	6.00	5.25	2.19	2.00	2.88	2.30	1.50	2.00	1.20
25	Lei reprime lo sconforto	Em	5.00	5.13	1.36	4.00	3.38	1.85	1.00	1.88	1.46
26	Lei serba il rancore	Em	1.00	2.00	1.77	1.00	2.13	2.23	1.00	1.38	0.74
27	Lei soddisfa il desiderio	Em	6.50	5.13	2.64	5.50	4.63	2.62	4.00	3.88	2.85
28	Lei soffre la solitudine	Em	1.00	2.38	2.20	1.00	2.50	2.33	1.00	1.63	0.92
29	Lei sogna la felicità	Em	3.00	3.50	2.62	2.00	2.75	2.19	1.00	1.63	1.19
30	Lei subisce la collera	Em	1.50	2.50	2.14	1.00	2.38	2.20	1.00	1.75	1.16
31	Lei svela l'imbarazzo	Em	6.50	5.75	1.75	6.00	5.38	1.92	3.50	3.75	2.38
32	Lei teme la tristezza	Em	4.00	3.38	1.69	2.00	2.38	1.51	1.00	1.00	0.00
33	Lei tollera la delusione	Em	3.50	4.13	2.30	2.00	2.50	1.85	1.00	1.25	0.71

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
34	Lei trasmette la gioia	Em	7.00	6.25	1.39	6.00	5.13	2.42	3.50	3.50	2.51
35	Lei trattiene l'esultanza	Em	6.00	5.13	2.23	6.00	5.00	2.39	5.00	4.63	2.20
1	Lei addiziona i gradi	Ma	1.50	1.63	0.74	2.00	2.25	1.39	1.00	1.00	0.00
2	Lei aggiunge il totale	Ma	1.50	1.75	1.04	2.00	2.00	1.07	1.00	1.00	0.00
3	Lei analizza i conti	Ma	4.00	3.63	2.13	4.00	3.63	2.39	1.00	1.00	0.00
4	Lei applica le formule	Ma	1.50	2.50	1.93	4.50	4.00	2.67	1.00	1.00	0.00
5	Lei arrotonda i valori	Ma	2.00	2.13	1.36	4.00	3.88	1.81	1.00	1.13	0.35
6	Lei calcola l'integrale	Ma	1.00	1.38	0.74	4.00	3.63	2.45	1.00	1.13	0.35
7	Lei centuplica la somma	Ma	5.00	4.25	2.55	5.00	4.25	2.05	1.00	1.13	0.35
8	Lei computa la costante	Ma	1.50	2.25	2.05	1.50	1.63	0.74	1.00	1.00	0.00
9	Lei conta gli insiemi	Ma	3.50	3.38	2.20	6.00	5.50	2.00	1.00	1.38	1.06
10	Lei conteggia le cifre	Ma	5.00	4.88	0.83	5.50	4.63	1.85	1.00	1.13	0.35
11	Lei correla le medie	Ma	1.00	1.13	0.35	2.00	2.50	1.60	1.00	1.00	0.00
12	Lei deriva la funzione	Ma	1.00	1.25	0.46	2.50	2.75	1.91	1.00	1.00	0.00
13	Lei dimezza il resto	Ma	2.00	1.75	0.71	3.50	3.88	2.30	1.00	1.25	0.46
14	Lei divide i decimali	Ma	1.50	1.50	0.53	1.50	2.13	1.46	1.00	1.13	0.35
15	Lei duplica il prodotto	Ma	1.00	1.38	0.74	3.00	2.75	1.28	1.00	1.00	0.00
16	Lei esegue la sottrazione	Ma	1.50	1.50	0.53	2.00	2.38	1.30	1.00	1.00	0.00
17	Lei fraziona il totale	Ma	1.00	1.50	1.07	2.00	2.50	1.60	1.00	1.00	0.00
18	Lei integra il logaritmo	Ma	1.50	2.00	1.20	3.00	3.13	1.25	1.00	1.38	0.74
19	Lei moltiplica l'unità	Ma	1.00	1.25	0.46	1.00	2.00	1.51	1.00	1.00	0.00
20	Lei numera gli insiemi	Ma	1.00	2.38	2.33	3.50	3.63	2.39	1.00	1.00	0.00
21	Lei permuta la matrice	Ma	1.00	1.63	1.06	2.00	2.50	1.69	1.00	1.13	0.35
22	Lei pondera le medie	Ma	1.00	1.25	0.46	1.00	1.50	1.41	1.00	1.00	0.00
23	Lei quadruplica i dati	Ma	2.50	3.13	2.10	3.00	3.13	1.73	1.00	1.00	0.00
24	Lei raddoppia il risultato	Ma	1.00	1.63	1.19	3.50	3.63	2.56	1.00	2.50	2.78
25	Lei ricava la soluzione	Ma	3.00	2.88	1.81	2.00	2.25	1.39	1.00	1.38	0.74
26	Lei riduce la varianza	Ma	1.50	1.88	1.13	1.50	2.25	1.75	1.00	1.13	0.35
27	Lei risolve l'equazione	Ma	1.00	1.25	0.46	4.00	3.13	1.81	1.00	1.00	0.00
28	Lei scompone la formula	Ma	2.50	3.13	2.10	3.50	3.25	1.49	1.00	1.00	0.00
29	Lei semplifica i calcoli	Ma	3.00	3.38	1.92	3.50	3.75	1.28	1.00	1.13	0.35
30	Lei somma gli addendi	Ma	3.50	3.75	1.75	3.50	3.75	0.89	1.00	1.00	0.00
31	Lei sottrae l'incognita	Ma	1.00	2.00	2.14	1.50	2.13	2.03	1.00	1.00	0.00
32	Lei sviluppa il teorema	Ma	1.00	1.50	0.76	1.00	2.25	1.75	1.00	1.00	0.00
33	Lei svolge il problema	Ma	3.50	3.38	1.60	4.00	3.50	1.77	1.00	1.13	0.35
34	Lei triplica la somma	Ma	2.00	3.13	2.47	5.00	4.50	1.93	1.00	1.00	0.00
35	Lei uguaglia le variabili	Ma	1.00	1.00	0.00	2.00	2.50	1.69	1.00	1.00	0.00

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
1	Lei apprende la nozione	Ms	3.50	3.50	2.07	1.50	2.13	1.46	1.00	1.38	0.74
2	Lei approva l'ideale	Ms	2.00	2.50	1.41	1.00	1.63	1.41	1.00	1.13	0.35
3	Lei arguisce lo scopo	Ms	1.50	2.38	2.00	1.00	2.00	1.60	1.00	1.25	0.46
4	Lei concepisce il dilemma	Ms	2.50	3.25	2.43	2.00	2.38	1.60	1.00	1.25	0.46
5	Lei condivide l'analisi	Ms	4.00	3.88	2.42	2.50	2.75	1.98	1.00	1.13	0.35
6	Lei congettura la tesi	Ms	3.00	3.63	2.88	1.50	2.13	1.55	1.00	1.13	0.35
7	Lei constata la finalità	Ms	2.50	3.00	2.20	1.50	1.75	1.04	1.00	1.00	0.00
8	Lei decreta la sorte	Ms	6.00	4.75	2.66	3.50	3.75	2.49	1.00	1.88	2.10
9	Lei desume la questione	Ms	2.50	2.75	1.83	1.00	1.50	1.07	1.00	1.00	0.00
10	Lei dimentica la promessa	Ms	1.00	1.75	1.16	1.00	1.75	2.12	1.00	1.00	0.00
11	Lei discerne l'opinione	Ms	6.00	4.75	2.71	1.00	1.38	0.52	1.00	1.38	1.06
12	Lei distingue i concetti	Ms	1.50	1.88	1.13	1.00	1.50	1.07	1.00	1.00	0.00
13	Lei distoglie la mente	Ms	1.00	2.00	1.60	1.00	1.88	1.46	1.00	1.13	0.35
14	Lei distrae l'attenzione	Ms	4.50	4.75	2.05	4.00	4.13	2.10	2.00	2.38	1.51
15	Lei esamina l'opzione	Ms	3.00	3.25	1.98	2.50	2.38	1.30	1.00	1.13	0.35
16	Lei ignora le fonti	Ms	1.00	1.50	0.76	1.00	1.25	0.46	1.00	1.00	0.00
17	Lei immagina il giudizio	Ms	1.00	1.50	0.76	1.00	1.00	0.00	1.00	1.00	0.00
18	Lei impara la lezione	Ms	3.00	3.38	2.56	1.00	2.50	2.33	1.00	1.63	1.41
19	Lei individua lo sbaglio	Ms	3.50	3.63	1.92	2.00	2.38	1.60	1.00	1.25	0.71
20	Lei influenza la scelta	Ms	7.00	5.88	2.10	4.50	4.00	1.93	1.00	1.88	1.73
21	Lei intuisce la novità	Ms	4.00	3.63	2.45	1.00	1.88	1.46	1.00	1.00	0.00
22	Lei medita l'alternativa	Ms	2.00	2.75	2.05	1.00	2.00	1.77	1.00	1.25	0.71
23	Lei memorizza la procedura	Ms	5.00	4.75	1.28	4.00	3.75	2.19	1.00	2.13	1.89
24	Lei nota l'imprecisione	Ms	1.00	2.00	2.14	2.00	2.63	2.13	1.00	1.75	2.12
25	Lei pianifica l'avvenimento	Ms	2.50	2.75	1.91	1.00	2.25	1.83	1.00	1.13	0.35
26	Lei rammenta l'episodio	Ms	1.00	2.13	1.81	1.00	1.13	0.35	1.00	1.00	0.00
27	Lei rievoca il ricordo	Ms	1.00	2.00	2.14	1.00	2.13	1.55	1.00	1.25	0.71
28	Lei rimugina i pensieri	Ms	1.00	2.63	2.33	1.00	2.00	1.77	1.00	1.25	0.71
29	Lei sbaglia il pronostico	Ms	3.50	3.25	1.83	1.50	1.75	1.04	1.00	1.25	0.46
30	Lei scorda l'intento	Ms	1.00	1.13	0.35	1.00	1.25	0.71	1.00	1.00	0.00
31	Lei simula l'interesse	Ms	6.00	6.00	1.07	6.00	5.63	1.19	1.00	2.00	1.60
32	Lei stima la franchezza	Ms	1.00	1.63	1.41	1.00	1.75	1.49	1.00	1.00	0.00
33	Lei suppone l'accaduto	Ms	2.00	2.50	2.07	1.00	1.50	0.76	1.00	1.00	0.00
34	Lei vaglia il parere	Ms	6.00	5.63	1.77	2.00	2.25	1.39	1.00	1.50	1.41
35	Lei valuta il merito	Ms	4.00	3.63	2.20	1.00	1.38	1.06	1.00	1.00	0.00
1	Lei addenta la merendina	Mo	7.00	7.00	0.00	3.50	3.50	1.51	1.00	1.00	0.00
2	Lei arriccia le labbra	Mo	7.00	7.00	0.00	1.00	1.13	0.35	1.00	1.13	0.35

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
3	Lei assaggia la pasta	Mo	7.00	6.75	0.71	4.00	3.50	1.69	1.00	1.00	0.00
4	Lei assapora il cibo	Mo	7.00	6.75	0.71	1.50	2.88	2.36	1.00	1.00	0.00
5	Lei bacia la guancia	Mo	7.00	6.88	0.35	5.00	4.38	2.00	2.50	2.50	1.41
6	Lei deglutisce il boccone	Mo	7.00	7.00	0.00	1.00	1.25	0.71	1.00	1.00	0.00
7	Lei degusta il vino	Mo	7.00	6.75	0.71	3.00	3.50	2.56	1.00	1.00	0.00
8	Lei digrigna i denti	Mo	7.00	6.75	0.71	1.00	1.13	0.35	1.00	1.00	0.00
9	Lei divora i biscotti	Mo	7.00	6.63	0.74	4.00	4.25	1.75	1.00	1.13	0.35
10	Lei fischietta la melodia	Mo	7.00	7.00	0.00	1.00	1.63	1.06	1.00	1.13	0.35
11	Lei gonfia il palloncino	Mo	7.00	6.75	0.46	4.50	4.50	1.20	1.00	1.13	0.35
12	Lei gusta lo champagne	Mo	7.00	6.88	0.35	3.50	3.38	1.51	1.00	1.00	0.00
13	Lei inghiotte la pillola	Mo	7.00	6.25	1.16	2.00	2.75	1.58	1.00	1.13	0.35
14	Lei ingoia la pastiglia	Mo	7.00	6.63	1.06	4.00	4.00	1.31	1.00	1.00	0.00
15	Lei ingurgita il cioccolato	Mo	7.00	7.00	0.00	4.00	3.75	2.12	1.00	1.25	0.71
16	Lei lecca il piatto	Mo	7.00	7.00	0.00	4.50	4.13	1.36	1.00	1.25	0.71
17	Lei mangia il pane	Mo	7.00	6.88	0.35	5.00	4.88	1.81	1.00	1.00	0.00
18	Lei mastica la carne	Mo	7.00	7.00	0.00	1.00	1.88	1.46	1.00	1.00	0.00
19	Lei mima la smorfia	Mo	7.00	7.00	0.00	2.50	2.75	1.75	1.00	1.38	0.74
20	Lei morde un frutto	Mo	7.00	7.00	0.00	5.00	4.13	1.89	1.00	1.13	0.35
21	Lei mordicchia la matita	Mo	7.00	7.00	0.00	4.50	4.00	1.20	1.00	1.00	0.00
22	Lei morsica il panino	Mo	7.00	6.75	0.71	5.50	5.13	1.36	1.00	1.13	0.35
23	Lei rigurgita il latte	Mo	6.50	6.13	0.99	1.00	1.63	1.19	1.00	1.38	1.06
24	Lei sbaciucchia l'orsacchiotto	Mo	7.00	7.00	0.00	4.00	3.75	1.39	1.00	1.00	0.00
25	Lei schiarisce la gola	Mo	7.00	6.63	0.74	2.50	2.50	1.31	1.00	1.00	0.00
26	Lei schiocca la lingua	Mo	7.00	7.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
27	Lei serra le labbra	Mo	7.00	6.50	1.07	1.00	1.13	0.35	1.00	1.00	0.00
28	Lei sgranocchia le patatine	Mo	7.00	6.75	0.46	3.00	3.00	1.69	1.00	1.13	0.35
29	Lei spalanca la bocca	Mo	7.00	6.88	0.35	1.00	1.50	0.76	1.00	1.25	0.46
30	Lei sputa l'oliva	Mo	7.00	6.88	0.35	3.00	2.75	1.67	1.00	1.25	0.71
31	Lei sputacchia il tabacco	Mo	7.00	6.88	0.35	1.00	1.38	0.52	1.00	1.00	0.00
32	Lei storce il labbro	Mo	7.00	7.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00
33	Lei succhia la caramella	Mo	7.00	7.00	0.00	2.00	2.50	1.69	1.00	1.25	0.46
34	Lei sussurra il nome	Mo	7.00	6.63	0.74	2.00	1.88	0.99	1.00	1.13	0.35
35	Lei trangugia la cena	Mo	7.00	6.63	1.06	4.50	4.50	2.07	1.00	1.00	0.00
1	Lei abbottona la camicia	Ha	1.00	1.00	0.00	7.00	6.88	0.35	1.00	1.13	0.35
2	Lei accarezza il cane	Ha	1.00	1.38	0.74	7.00	6.38	0.92	1.00	1.75	1.04
3	Lei afferra le forbici	Ha	1.00	1.13	0.35	6.50	6.00	1.41	1.00	1.13	0.35
4	Lei affetta la carota	Ha	1.00	1.25	0.71	7.00	6.88	0.35	1.00	1.38	0.52

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
5	Lei annoda la cravatta	Ha	1.00	1.25	0.46	7.00	6.75	0.71	1.00	1.25	0.46
6	Lei avvita il bullone	Ha	1.00	1.13	0.35	7.00	6.63	0.52	1.00	1.50	1.07
7	Lei carica l'orologio	Ha	1.00	1.00	0.00	7.00	6.63	0.74	1.00	1.00	0.00
8	Lei cuce il calzino	Ha	1.00	1.63	1.06	7.00	6.88	0.35	1.00	1.25	0.46
9	Lei digita il tasto	Ha	1.00	1.13	0.35	7.00	6.63	0.74	1.00	1.38	0.74
10	Lei dipinge il vaso	Ha	1.00	1.25	0.71	7.00	6.88	0.35	1.00	1.75	2.12
11	Lei grattugia il formaggio	Ha	1.50	2.13	2.03	7.00	6.50	0.76	1.00	1.63	1.19
12	Lei imbusta la lettera	Ha	4.00	3.75	2.12	7.00	5.88	2.10	3.00	3.25	2.25
13	Lei impasta il pane	Ha	1.00	1.25	0.71	7.00	6.88	0.35	1.00	1.38	0.74
14	Lei impugna la spada	Ha	1.00	1.00	0.00	7.00	6.88	0.35	1.50	1.63	0.74
15	Lei inietta il vaccino	Ha	2.00	2.00	0.76	6.50	6.25	1.04	1.00	1.25	0.46
16	Lei lega la fune	Ha	1.00	1.25	0.46	7.00	6.88	0.35	1.00	1.75	1.16
17	Lei lima le unghie	Ha	1.00	1.63	0.92	7.00	7.00	0.00	1.00	3.25	3.11
18	Lei martella il chiodo	Ha	1.00	1.38	0.52	7.00	6.63	0.74	2.00	1.88	0.64
19	Lei pettina i capelli	Ha	1.00	1.63	1.06	7.00	6.50	1.07	1.00	1.38	0.74
20	Lei pizzica le corde	Ha	1.00	1.38	0.74	7.00	6.75	0.71	1.00	1.63	0.92
21	Lei ricama la tovaglia	Ha	1.00	1.38	0.74	7.00	6.88	0.35	1.00	1.13	0.35
22	Lei ritaglia il disegno	Ha	1.00	1.25	0.46	7.00	6.50	1.07	1.00	1.13	0.35
23	Lei sbuccia il mandarino	Ha	1.00	1.25	0.46	7.00	6.88	0.35	1.00	1.13	0.35
24	Lei scrive l'indirizzo	Ha	1.00	1.25	0.46	7.00	6.88	0.35	1.00	1.13	0.35
25	Lei sfoglia il giornale	Ha	2.00	2.50	1.41	6.00	5.88	1.25	1.00	1.25	0.46
26	Lei spalma la marmellata	Ha	1.50	2.00	1.20	7.00	6.50	0.76	1.00	1.00	0.00
27	Lei spreme il limone	Ha	1.00	1.50	0.76	6.00	6.13	0.99	1.00	1.50	0.76
28	Lei stappa la bottiglia	Ha	1.00	1.88	2.10	7.00	6.88	0.35	1.00	1.38	0.74
29	Lei strappa il foglio	Ha	1.00	1.13	0.35	7.00	6.50	1.07	1.00	1.00	0.00
30	Lei strizza l'asciugamano	Ha	1.00	1.50	1.41	7.00	6.75	0.46	1.00	1.50	0.76
31	Lei strofina l'argenteria	Ha	1.00	1.63	1.19	7.00	6.75	0.71	1.00	1.13	0.35
32	Lei sventola il fazzoletto	Ha	1.00	1.00	0.00	7.00	6.63	1.06	1.00	1.50	1.07
33	Lei taglia la carne	Ha	1.00	1.00	0.00	7.00	7.00	0.00	1.00	1.25	0.46
34	Lei tocca il velluto	Ha	1.00	1.50	0.76	7.00	6.88	0.35	1.00	2.50	2.51
35	Lei trita la cipolla	Ha	1.50	2.00	1.41	7.00	5.88	2.10	1.00	1.63	0.92
1	Lei accavalla le gambe	Le	1.00	1.13	0.35	1.00	1.38	1.06	7.00	7.00	0.00
2	Lei balla il tip tap	Le	2.00	2.38	1.77	4.00	4.13	1.25	7.00	7.00	0.00
3	Lei batte il piede	Le	1.00	1.25	0.46	1.00	1.38	0.74	7.00	6.75	0.71
4	Lei calcia il pallone	Le	1.00	1.13	0.35	3.00	2.75	1.58	7.00	7.00	0.00
5	Lei calpesta il tappeto	Le	1.00	1.00	0.00	1.00	1.38	0.52	7.00	6.75	0.71
6	Lei calza le infradito	Le	1.00	1.00	0.00	2.50	2.88	1.96	7.00	6.13	1.46

		Label	**Mouth scale**			**Hand scale**			**Leg scale**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
7	Lei corre la maratona	Le	3.00	2.75	1.28	5.00	4.50	2.00	7.00	7.00	0.00
8	Lei distende le gambe	Le	1.50	2.13	1.46	2.00	2.63	1.92	7.00	6.88	0.35
9	Lei divarica le gambe	Le	1.50	2.13	1.73	1.50	2.00	1.31	7.00	6.75	0.71
10	Lei esegue i saltelli	Le	1.00	1.50	0.76	4.50	3.88	1.73	7.00	7.00	0.00
11	Lei flette le ginocchia	Le	1.00	1.13	0.35	2.00	2.38	1.19	7.00	7.00	0.00
12	Lei incrocia le gambe	Le	1.00	1.00	0.00	2.00	2.13	1.13	7.00	7.00	0.00
13	Lei infila gli zoccoli	Le	1.00	1.13	0.35	3.00	3.25	2.12	6.00	5.88	1.36
14	Lei palleggia il pallone	Le	1.00	1.38	0.74	7.00	6.13	1.64	7.00	6.75	0.71
15	Lei pesta il mozzicone	Le	1.00	1.00	0.00	1.00	1.63	0.92	7.00	6.50	0.76
16	Lei piega le ginocchia	Le	1.00	1.75	1.75	2.00	2.25	1.28	7.00	6.38	1.06
17	Lei pigia i grappoli	Le	1.00	1.38	1.06	3.50	4.13	2.53	5.50	4.50	2.83
18	Lei preme il pedale	Le	1.00	1.25	0.71	1.00	1.38	0.74	7.00	6.13	1.46
19	Lei sale la gradinata	Le	1.00	1.00	0.00	2.50	2.38	1.06	7.00	6.75	0.71
20	Lei salta l'ostacolo	Le	1.50	2.13	2.03	5.00	5.25	1.28	7.00	6.88	0.35
21	Lei sbatte i tacchi	Le	1.00	1.25	0.71	1.00	1.50	1.07	7.00	6.88	0.35
22	Lei scalcia il copriletto	Le	1.00	1.13	0.35	2.00	2.25	1.39	7.00	6.88	0.35
23	Lei scende la scalinata	Le	1.00	1.25	0.71	3.00	2.88	1.25	7.00	6.75	0.46
24	Lei schiaccia l'acceleratore	Le	1.50	1.88	0.99	2.00	2.00	0.93	7.00	6.88	0.35
25	Lei sferra un calcio	Le	1.00	1.13	0.35	2.50	2.38	1.06	7.00	7.00	0.00
26	Lei sfila le pantofole	Le	1.00	1.00	0.00	5.00	4.50	2.62	7.00	6.50	1.07
27	Lei solleva i talloni	Le	1.00	1.00	0.00	1.50	2.13	1.55	7.00	6.88	0.35
28	Lei sperona il cavallo	Le	3.00	2.63	1.69	4.00	3.88	1.81	7.00	6.63	0.52
29	Lei spiaccica lo scarafaggio	Le	1.00	1.63	1.06	3.00	3.38	2.20	6.50	6.00	1.41
30	Lei spicca un salto	Le	1.00	1.25	0.71	5.00	4.88	1.13	7.00	6.88	0.35
31	Lei spinge l'acceleratore	Le	1.00	1.38	1.06	3.00	3.75	2.87	7.00	5.50	2.78
32	Lei strascica i piedi	Le	1.00	1.00	0.00	1.00	1.50	0.76	7.00	6.50	0.76
33	Lei struscia le ciabatte	Le	1.00	1.00	0.00	2.00	2.13	1.25	7.00	6.63	0.74
34	Lei tira un calcio	Le	1.00	1.13	0.35	3.00	3.00	1.93	7.00	6.88	0.35
35	Lei toglie le ciabatte	Le	1.00	1.13	0.35	5.00	4.50	1.51	6.00	5.88	1.13

Appendix 1.3. Rating study 2: Concreteness, Context availability and Familiarity ratings

Concreteness, Context availability and Familiarity ratings for (Em) emotion-, (Ma) mathematics-, (Ms) mental state-, (Mo) mouth-, (Ha) hand-, and (Le) leg-related sentences.

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
1	Lei affronta l'angoscia	Em	2.50	2.88	1.81	2.00	2.13	1.17	6.00	5.25	1.73
2	Lei agogna la fama	Em	1.00	1.88	1.73	1.50	2.88	2.36	1.00	1.75	1.12
3	Lei avverte l'eccitazione	Em	3.00	3.00	1.31	2.50	3.38	2.20	5.50	5.63	1.32
4	Lei brama l'amore	Em	2.00	2.50	2.07	4.00	3.75	1.56	3.00	3.13	1.83
5	Lei calma la rabbia	Em	2.00	2.50	1.93	2.00	2.38	1.87	3.50	3.75	2.22
6	Lei cela l'invidia	Em	2.00	1.75	0.71	2.50	2.88	1.46	2.50	3.50	2.20
7	Lei depreca la crudeltà	Em	1.00	1.50	0.76	1.00	1.75	1.12	2.00	2.50	1.33
8	Lei detesta la cattiveria	Em	1.00	1.50	0.76	3.00	3.25	2.17	4.00	4.13	2.42
9	Lei disdegna l'ostilità	Em	1.50	1.50	0.53	2.00	2.00	1.00	2.50	2.88	1.46
10	Lei disprezza la viltà	Em	2.00	1.75	0.71	1.50	1.88	1.27	3.00	3.63	1.42
11	Lei dissimula il disappunto	Em	2.00	2.50	1.69	3.00	2.75	1.20	2.50	3.25	1.50
12	Lei domina la passione	Em	2.00	2.25	1.39	3.50	3.50	1.12	5.50	5.00	1.74
13	Lei gradisce la sorpresa	Em	2.50	3.13	1.36	2.00	3.00	2.03	7.00	6.63	0.50
14	Lei inibisce l'ansia	Em	3.00	2.75	1.28	2.00	1.88	0.60	3.00	2.75	1.48
15	Lei invidia l'orgoglio	Em	1.50	1.63	0.74	1.50	2.25	1.50	5.00	4.13	2.54
16	Lei manifesta l'odio	Em	3.00	3.25	2.05	2.50	3.13	1.98	5.50	5.00	1.88
17	Lei mostra la contentezza	Em	3.00	3.38	0.92	2.50	2.88	1.62	5.00	5.00	0.87
18	Lei nutre l'affetto	Em	2.50	2.63	1.30	3.00	3.00	1.22	7.00	6.50	1.01
19	Lei odia la sofferenza	Em	1.00	2.13	1.64	2.00	2.75	1.58	6.00	5.75	1.56
20	Lei patisce la vergogna	Em	2.00	1.88	0.83	2.00	2.25	1.86	3.00	3.38	2.29
21	Lei percepisce lo stupore	Em	1.50	1.50	0.53	3.00	3.25	1.79	2.50	2.75	1.86
22	Lei placa l'ira	Em	2.00	2.38	0.52	4.50	3.88	2.16	5.00	4.75	1.86
23	Lei predilige l'allegria	Em	1.50	1.63	0.74	1.50	2.25	1.73	3.00	4.00	2.20
24	Lei prova il disgusto	Em	2.00	2.50	0.76	3.00	3.75	2.45	5.00	5.25	1.30
25	Lei reprime lo sconforto	Em	1.00	2.13	1.81	2.00	1.63	0.50	2.00	2.50	1.33
26	Lei serba il rancore	Em	1.50	2.13	1.73	2.00	2.63	1.81	5.00	4.50	1.88
27	Lei soddisfa il desiderio	Em	3.50	3.63	1.77	4.50	4.38	2.12	6.00	5.88	1.05
28	Lei soffre la solitudine	Em	2.50	2.63	1.77	4.50	4.00	2.30	6.00	5.13	2.11
29	Lei sogna la felicità	Em	1.00	1.13	0.35	1.50	2.63	2.15	5.00	4.63	1.94
30	Lei subisce la collera	Em	1.50	1.75	1.04	2.50	3.00	1.59	3.50	3.63	1.32
31	Lei svela l'imbarazzo	Em	2.00	2.50	1.93	1.50	1.75	0.97	3.50	3.13	1.54
32	Lei teme la tristezza	Em	2.00	2.25	1.67	1.00	1.75	1.32	2.50	3.75	2.37

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
33	Lei tollera la delusione	Em	1.00	1.38	0.52	2.50	3.00	2.19	4.00	3.88	1.69
34	Lei trasmette la gioia	Em	3.50	3.25	2.12	1.00	2.00	1.96	3.50	4.13	2.16
35	Lei trattiene l'esultanza	Em	3.00	2.50	1.07	2.50	2.75	1.48	5.50	5.25	1.56
1	Lei addiziona i gradi	Ma	3.00	2.63	1.19	2.00	2.63	1.88	2.00	2.25	1.39
2	Lei aggiunge il totale	Ma	3.50	3.38	0.74	2.00	2.88	1.79	3.00	3.25	2.22
3	Lei analizza i conti	Ma	4.50	4.13	1.73	4.50	4.00	1.94	4.50	4.75	1.20
4	Lei applica le formule	Ma	3.00	3.38	1.60	5.00	5.25	0.97	7.00	6.13	1.20
5	Lei arrotonda i valori	Ma	2.00	2.88	1.81	3.50	3.75	1.72	4.50	4.38	1.80
6	Lei calcola l'integrale	Ma	5.50	4.50	2.73	3.00	3.25	2.22	3.50	3.63	1.87
7	Lei centuplica la somma	Ma	4.50	4.63	2.26	3.50	3.88	1.84	4.00	3.75	1.72
8	Lei computa la costante	Ma	2.50	2.50	1.20	1.50	2.75	2.09	1.00	1.00	0.00
9	Lei conta gli insiemi	Ma	5.00	4.63	1.60	3.50	3.75	1.86	4.00	4.25	1.30
10	Lei conteggia le cifre	Ma	4.00	4.75	1.67	5.00	4.88	1.36	4.00	3.75	1.30
11	Lei correla le medie	Ma	4.50	4.00	2.39	1.00	1.63	1.01	2.00	2.38	1.50
12	Lei deriva la funzione	Ma	3.00	3.00	1.93	4.00	4.38	1.73	5.50	4.50	2.32
13	Lei dimezza il resto	Ma	4.00	3.88	0.83	5.50	4.50	2.32	3.50	3.50	1.87
14	Lei divide i decimali	Ma	3.00	3.13	1.25	4.50	4.25	2.05	3.00	3.50	2.19
15	Lei duplica il prodotto	Ma	4.00	4.00	1.20	2.00	2.88	1.64	4.50	4.25	2.28
16	Lei esegue la sottrazione	Ma	4.00	3.88	1.46	5.50	5.38	1.32	5.50	4.75	2.00
17	Lei fraziona il totale	Ma	3.50	3.50	1.93	2.50	3.13	1.84	2.00	3.13	2.35
18	Lei integra il logaritmo	Ma	3.50	3.75	1.28	4.00	4.00	2.40	3.50	3.00	1.67
19	Lei moltiplica l'unità	Ma	3.00	3.25	2.25	3.00	3.00	1.32	3.50	3.75	1.39
20	Lei numera gli insiemi	Ma	4.00	4.00	1.20	4.00	3.75	1.20	5.00	4.25	2.12
21	Lei permuta la matrice	Ma	3.00	2.88	0.83	1.00	3.00	2.73	1.00	2.00	1.69
22	Lei pondera le medie	Ma	3.50	3.25	1.28	3.50	4.25	1.73	3.50	3.88	2.47
23	Lei quadruplica i dati	Ma	3.00	3.38	1.77	3.00	3.25	2.11	3.00	3.38	1.87
24	Lei raddoppia il risultato	Ma	4.50	4.13	2.30	3.00	3.25	1.92	5.00	4.63	1.32
25	Lei ricava la soluzione	Ma	3.00	3.25	1.04	4.00	3.88	1.96	6.00	5.88	0.78
26	Lei riduce la varianza	Ma	3.50	2.75	1.49	5.00	4.25	2.65	2.50	2.75	1.56
27	Lei risolve l'equazione	Ma	5.00	4.63	2.45	6.00	5.00	1.69	6.50	5.88	1.47
28	Lei scompone la formula	Ma	4.00	3.75	1.75	5.50	4.88	1.84	4.00	3.63	1.41
29	Lei semplifica i calcoli	Ma	3.50	3.88	1.46	5.00	4.75	1.79	6.50	5.75	1.50
30	Lei somma gli addendi	Ma	3.50	4.38	1.77	4.00	4.13	1.83	4.50	4.38	1.50
31	Lei sottrae l'incognita	Ma	2.50	2.75	1.49	3.50	4.13	1.98	3.00	3.13	1.69
32	Lei sviluppa il teorema	Ma	4.50	4.00	1.85	4.50	4.75	1.92	4.00	4.25	1.72
33	Lei svolge il problema	Ma	3.00	3.63	1.51	5.50	5.38	1.11	6.50	5.88	1.47
34	Lei triplica la somma	Ma	2.50	2.88	1.96	3.00	3.13	1.96	2.50	3.25	1.87

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
35	Lei uguaglia le variabili	Ma	3.50	3.25	2.19	3.00	3.25	2.22	3.50	3.50	1.66
1	Lei apprende la nozione	Ms	2.00	2.25	1.04	2.50	3.00	2.07	5.00	5.00	1.41
2	Lei approva l'ideale	Ms	1.50	1.75	1.04	1.50	2.25	1.50	4.50	4.38	2.12
3	Lei arguisce lo scopo	Ms	2.00	2.25	1.67	1.00	2.00	1.69	2.00	2.63	1.74
4	Lei concepisce il dilemma	Ms	1.00	1.38	0.74	2.00	2.75	1.73	3.00	3.13	2.03
5	Lei condivide l'analisi	Ms	2.00	2.50	1.41	2.50	3.00	1.74	2.00	2.50	1.59
6	Lei congettura la tesi	Ms	2.00	2.00	1.31	3.00	2.88	1.45	2.00	3.25	2.26
7	Lei constata la finalità	Ms	1.00	1.50	0.76	1.00	1.50	1.01	2.50	2.88	1.70
8	Lei decreta la sorte	Ms	2.00	2.75	1.67	2.00	2.75	2.00	2.00	2.13	1.17
9	Lei desume la questione	Ms	2.00	2.25	0.89	2.00	2.25	1.39	3.00	3.00	1.22
10	Lei dimentica la promessa	Ms	1.00	1.38	0.74	3.50	3.63	2.23	5.00	4.75	1.79
11	Lei discerne l'opinione	Ms	2.00	1.75	0.71	1.00	2.13	1.58	3.50	3.75	1.92
12	Lei distingue i concetti	Ms	1.00	1.38	1.06	3.50	3.75	2.28	5.00	5.00	1.80
13	Lei distoglie la mente	Ms	1.00	1.38	0.52	3.00	2.75	1.48	5.50	4.75	1.73
14	Lei distrae l'attenzione	Ms	3.00	3.50	1.51	3.00	3.38	1.58	5.50	5.38	1.32
15	Lei esamina l'opzione	Ms	1.50	2.25	2.05	2.50	3.38	2.02	4.50	4.75	1.20
16	Lei ignora le fonti	Ms	3.00	3.13	1.46	1.00	1.88	1.39	6.00	5.50	1.94
17	Lei immagina il giudizio	Ms	1.50	2.50	1.93	2.00	2.63	1.24	3.50	3.75	1.86
18	Lei impara la lezione	Ms	3.50	3.75	1.67	6.00	5.13	2.11	6.00	6.00	1.00
19	Lei individua lo sbaglio	Ms	3.00	3.38	1.77	2.50	3.13	1.55	5.50	4.75	2.06
20	Lei influenza la scelta	Ms	2.00	2.75	1.83	2.00	2.50	1.59	4.50	4.00	2.01
21	Lei intuisce la novità	Ms	1.50	2.50	2.14	1.00	1.75	1.66	3.00	3.25	1.56
22	Lei medita l'alternativa	Ms	2.00	2.25	1.67	2.00	3.00	2.15	4.00	4.25	1.56
23	Lei memorizza la procedura	Ms	3.50	3.63	2.07	3.00	3.13	1.36	5.50	5.38	0.70
24	Lei nota l'imprecisione	Ms	3.50	3.88	1.81	3.00	4.13	2.06	6.00	5.38	1.42
25	Lei pianifica l'avvenimento	Ms	3.00	3.25	0.71	3.50	3.88	2.09	5.50	5.13	1.70
26	Lei rammenta l'episodio	Ms	2.00	2.13	0.99	4.00	4.00	2.18	6.00	5.50	1.94
27	Lei rievoca il ricordo	Ms	1.50	2.50	2.14	5.00	4.25	2.45	5.50	4.63	1.89
28	Lei rimugina i pensieri	Ms	1.00	1.50	0.76	2.00	2.38	1.22	5.00	5.13	1.54
29	Lei sbaglia il pronostico	Ms	2.50	2.75	1.67	3.00	3.25	1.99	4.00	3.88	2.20
30	Lei scorda l'intento	Ms	1.50	2.00	1.41	2.50	2.88	1.70	5.00	4.50	1.67
31	Lei simula l'interesse	Ms	2.00	2.13	0.83	3.00	3.25	0.97	3.50	4.13	1.78
32	Lei stima la franchezza	Ms	2.00	2.63	2.20	1.50	2.00	1.13	3.50	3.75	2.05
33	Lei suppone l'accaduto	Ms	2.00	1.75	0.46	1.00	1.75	1.41	3.00	3.88	1.48
34	Lei vaglia il parere	Ms	2.50	2.50	0.93	2.00	2.25	1.39	2.00	2.50	1.67
35	Lei valuta il merito	Ms	3.00	3.25	1.39	2.50	3.00	1.74	4.00	4.38	1.32
1	Lei addenta la merendina	Mo	7.00	6.88	0.35	6.50	5.75	1.66	6.00	5.50	1.81

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
2	Lei arriccia le labbra	Mo	6.00	5.88	0.99	3.00	3.88	2.11	2.50	2.88	1.46
3	Lei assaggia la pasta	Mo	7.00	6.63	0.74	5.00	4.38	1.42	7.00	6.88	0.33
4	Lei assapora il cibo	Mo	5.00	4.88	1.25	5.50	5.25	1.56	6.00	5.63	1.41
5	Lei bacia la guancia	Mo	7.00	6.13	1.36	5.00	5.13	1.17	6.50	5.88	1.38
6	Lei deglutisce il boccone	Mo	6.50	6.38	0.74	6.50	6.13	1.06	4.00	4.00	1.58
7	Lei degusta il vino	Mo	5.00	5.38	1.19	5.00	4.88	1.83	4.50	5.00	1.24
8	Lei digrigna i denti	Mo	7.00	6.38	1.06	3.00	3.00	1.66	4.50	4.88	1.27
9	Lei divora i biscotti	Mo	6.50	6.13	1.36	5.50	5.38	1.50	7.00	6.50	0.73
10	Lei fischietta la melodia	Mo	7.00	6.50	0.76	5.00	4.63	1.94	4.00	4.38	1.12
11	Lei gonfia il palloncino	Mo	7.00	6.88	0.35	6.00	5.75	1.30	5.00	5.00	1.94
12	Lei gusta lo champagne	Mo	5.50	5.50	1.20	6.50	5.75	1.73	6.50	6.00	1.33
13	Lei inghiotte la pillola	Mo	7.00	7.00	0.00	5.00	5.00	1.58	7.00	6.38	1.01
14	Lei ingoia la pastiglia	Mo	7.00	6.88	0.35	5.00	4.88	1.62	5.50	5.75	1.09
15	Lei ingurgita il cioccolato	Mo	7.00	6.63	0.74	5.50	5.38	1.32	3.50	3.63	1.58
16	Lei lecca il piatto	Mo	7.00	6.75	0.46	6.00	5.38	1.42	4.50	4.13	2.89
17	Lei mangia il pane	Mo	7.00	6.88	0.35	6.00	5.75	1.20	7.00	6.38	1.13
18	Lei mastica la carne	Mo	7.00	6.75	0.71	6.50	5.63	1.60	7.00	6.38	1.67
19	Lei mima la smorfia	Mo	4.00	4.38	1.30	3.50	3.25	1.39	2.50	2.88	1.70
20	Lei morde un frutto	Mo	7.00	6.88	0.35	5.00	4.50	2.55	6.00	5.75	1.20
21	Lei mordicchia la matita	Mo	6.50	5.50	2.14	6.00	5.88	1.05	6.00	5.75	1.20
22	Lei morsica il panino	Mo	7.00	6.75	0.46	6.00	5.88	1.05	6.50	6.00	1.13
23	Lei rigurgita il latte	Mo	7.00	6.63	0.74	5.50	4.75	2.12	6.00	5.63	1.41
24	Lei sbaciucchia l'orsacchiotto	Mo	7.00	5.88	2.23	4.00	4.00	1.50	2.50	3.00	1.51
25	Lei schiarisce la gola	Mo	6.00	5.88	1.36	2.00	3.50	2.35	5.50	5.13	1.97
26	Lei schiocca la lingua	Mo	6.00	6.00	0.93	2.00	2.13	1.27	4.00	3.50	1.42
27	Lei serra le labbra	Mo	6.50	6.13	1.13	3.50	3.38	1.65	6.00	5.50	1.33
28	Lei sgranocchia le patatine	Mo	7.00	6.88	0.35	5.00	5.00	1.87	6.00	6.13	0.78
29	Lei spalanca la bocca	Mo	7.00	6.88	0.35	4.50	4.25	2.59	6.50	5.75	1.66
30	Lei sputa l'oliva	Mo	7.00	6.50	0.76	4.00	4.50	1.81	4.50	4.38	1.65
31	Lei sputacchia il tabacco	Mo	7.00	6.25	1.16	4.50	4.63	1.87	3.00	2.75	1.30
32	Lei storce il labbro	Mo	5.00	4.88	1.81	2.50	2.75	1.92	4.00	3.88	1.76
33	Lei succhia la caramella	Mo	7.00	6.00	1.77	5.00	4.75	1.56	4.00	3.88	1.62
34	Lei sussurra il nome	Mo	6.00	6.13	0.83	2.50	3.50	2.37	4.50	4.50	1.87
35	Lei trangugia la cena	Mo	7.00	6.63	0.74	6.00	5.25	1.73	5.50	4.63	1.89
1	Lei abbottona la camicia	Ha	7.00	6.00	1.85	5.50	4.75	2.00	6.00	6.13	0.78
2	Lei accarezza il cane	Ha	7.00	6.75	0.46	6.00	5.50	1.74	7.00	6.75	0.44
3	Lei afferra le forbici	Ha	7.00	6.75	0.46	4.50	4.50	2.24	7.00	6.75	0.67

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
4	Lei affetta la carota	Ha	7.00	6.88	0.35	6.50	5.75	1.66	4.00	4.13	2.03
5	Lei annoda la cravatta	Ha	7.00	7.00	0.00	5.00	4.63	2.00	4.50	4.13	1.46
6	Lei avvita il bullone	Ha	7.00	6.88	0.35	4.00	3.88	1.45	4.50	4.63	1.93
7	Lei carica l'orologio	Ha	6.50	5.88	1.55	6.00	5.38	1.59	4.00	3.63	2.00
8	Lei cuce il calzino	Ha	7.00	6.50	0.76	4.50	4.50	2.00	3.00	3.25	1.64
9	Lei digita il tasto	Ha	7.00	7.00	0.00	4.50	4.50	2.00	5.00	4.75	2.11
10	Lei dipinge il vaso	Ha	7.00	6.00	1.77	4.50	4.63	1.80	3.50	3.88	2.47
11	Lei grattugia il formaggio	Ha	7.00	7.00	0.00	6.00	5.50	1.51	6.50	6.25	0.83
12	Lei imbusta la lettera	Ha	6.00	6.25	0.71	5.50	5.50	0.87	4.00	4.00	1.87
13	Lei impasta il pane	Ha	7.00	7.00	0.00	7.00	6.38	0.88	4.50	4.38	2.12
14	Lei impugna la spada	Ha	6.00	5.75	1.39	6.00	5.50	1.67	7.00	5.38	2.40
15	Lei inietta il vaccino	Ha	7.00	6.88	0.35	7.00	5.63	2.22	7.00	6.38	1.33
16	Lei lega la fune	Ha	7.00	6.75	0.46	4.50	4.13	1.90	5.00	4.25	1.87
17	Lei lima le unghie	Ha	6.50	6.50	0.53	6.00	5.38	1.51	3.50	3.63	2.23
18	Lei martella il chiodo	Ha	7.00	7.00	0.00	6.50	5.75	1.66	6.50	6.13	1.06
19	Lei pettina i capelli	Ha	7.00	6.88	0.35	6.50	5.75	1.73	7.00	6.63	0.50
20	Lei pizzica le corde	Ha	7.00	6.88	0.35	6.00	6.13	0.93	5.50	4.50	2.09
21	Lei ricama la tovaglia	Ha	7.00	6.63	0.74	6.00	5.63	1.66	5.50	5.50	1.58
22	Lei ritaglia il disegno	Ha	7.00	7.00	0.00	5.50	4.75	2.06	4.00	3.75	1.92
23	Lei sbuccia il mandarino	Ha	7.00	6.63	0.74	6.00	5.50	1.67	7.00	6.63	0.50
24	Lei scrive l'indirizzo	Ha	7.00	6.38	1.06	4.50	4.25	2.22	6.00	6.13	0.78
25	Lei sfoglia il giornale	Ha	7.00	6.75	0.46	5.50	4.75	2.24	6.50	6.13	1.27
26	Lei spalma la marmellata	Ha	7.00	6.75	0.46	5.50	5.50	1.22	6.00	5.63	1.41
27	Lei spreme il limone	Ha	7.00	6.88	0.35	6.50	6.00	1.24	7.00	6.63	0.71
28	Lei stappa la bottiglia	Ha	7.00	6.50	1.07	5.00	5.25	1.09	7.00	6.50	0.73
29	Lei strappa il foglio	Ha	7.00	6.63	0.74	5.50	5.13	1.77	7.00	6.75	0.67
30	Lei strizza l'asciugamano	Ha	6.50	6.50	0.53	6.50	5.63	1.75	4.50	3.88	1.47
31	Lei strofina l'argenteria	Ha	7.00	6.63	0.74	6.50	6.13	0.94	5.50	5.00	2.19
32	Lei sventola il fazzoletto	Ha	7.00	6.63	0.74	6.50	5.75	1.66	6.50	5.88	1.63
33	Lei taglia la carne	Ha	7.00	6.75	0.71	5.50	5.50	1.22	7.00	6.88	0.33
34	Lei tocca il velluto	Ha	7.00	6.38	0.92	5.00	4.25	1.80	3.50	3.50	2.00
35	Lei trita la cipolla	Ha	7.00	6.88	0.35	7.00	6.13	1.39	5.00	4.75	1.92
1	Lei accavalla le gambe	Le	6.50	6.25	1.04	5.00	4.75	1.86	6.50	5.75	1.66
2	Lei balla il tip tap	Le	7.00	5.88	1.64	5.00	4.75	2.22	2.50	3.63	2.72
3	Lei batte il piede	Le	7.00	7.00	0.00	2.50	3.50	2.20	5.00	5.25	1.20
4	Lei calcia il pallone	Le	7.00	6.50	0.76	6.00	6.00	1.00	7.00	6.50	1.01
5	Lei calpesta il tappeto	Le	7.00	6.88	0.35	5.50	5.50	1.58	7.00	6.25	1.66

		Label	**Concreteness**			**Context availability**			**Familiarity**		
			Median	*Mean*	*SD*	*Median*	*Mean*	*SD*	*Median*	*Mean*	*SD*
6	Lei calza le infradito	Le	7.00	6.38	0.92	5.50	5.50	1.58	6.50	5.88	1.38
7	Lei corre la maratona	Le	7.00	6.00	2.14	5.50	5.50	0.87	5.00	4.63	1.87
8	Lei distende le gambe	Le	6.50	6.50	0.53	5.00	4.63	1.87	5.50	5.63	1.22
9	Lei divarica le gambe	Le	7.00	7.00	0.00	6.50	5.25	2.32	6.00	6.13	0.78
10	Lei esegue i saltelli	Le	6.00	6.25	0.71	4.50	4.75	1.56	2.00	2.00	0.87
11	Lei flette le ginocchia	Le	7.00	6.50	1.07	4.00	4.00	1.80	6.00	5.50	1.67
12	Lei incrocia le gambe	Le	6.00	6.25	0.71	5.00	4.75	1.56	7.00	6.13	1.39
13	Lei infila gli zoccoli	Le	7.00	6.88	0.35	4.00	4.00	2.40	4.00	4.63	2.07
14	Lei palleggia il pallone	Le	7.00	7.00	0.00	6.00	5.88	0.78	5.50	5.25	1.39
15	Lei pesta il mozzicone	Le	7.00	6.88	0.35	5.00	4.75	2.05	5.00	4.38	1.74
16	Lei piega le ginocchia	Le	7.00	6.75	0.46	4.00	4.13	2.03	7.00	6.00	1.54
17	Lei pigia i grappoli	Le	7.00	6.50	1.07	6.00	5.75	1.30	4.50	4.38	2.00
18	Lei preme il pedale	Le	7.00	6.75	0.46	7.00	6.25	1.41	7.00	6.88	0.33
19	Lei sale la gradinata	Le	7.00	6.63	0.52	5.50	5.00	1.42	5.00	5.00	1.66
20	Lei salta l'ostacolo	Le	6.50	5.75	1.49	5.00	4.63	1.66	5.50	5.38	1.50
21	Lei sbatte i tacchi	Le	7.00	5.63	2.20	4.00	4.00	1.58	3.00	3.50	1.74
22	Lei scalcia il copriletto	Le	7.00	6.00	1.93	5.00	4.63	1.80	3.00	3.50	1.88
23	Lei scende la scalinata	Le	7.00	6.88	0.35	5.00	4.63	2.06	7.00	5.88	1.80
24	Lei schiaccia l'acceleratore	Le	7.00	6.75	0.46	7.00	6.63	0.71	7.00	5.63	2.11
25	Lei sferra un calcio	Le	7.00	6.25	1.39	4.50	4.75	1.72	7.00	6.50	1.33
26	Lei sfila le pantofole	Le	7.00	7.00	0.00	5.50	5.25	1.48	4.50	4.13	1.37
27	Lei solleva i talloni	Le	7.00	6.25	1.75	3.00	3.88	2.11	6.00	5.25	1.73
28	Lei sperona il cavallo	Le	6.50	6.25	0.89	6.50	5.38	2.15	1.00	2.13	2.00
29	Lei spiaccica lo scarafaggio	Le	7.00	7.00	0.00	6.50	5.50	1.83	5.50	5.25	1.79
30	Lei spicca un salto	Le	6.50	6.38	0.74	4.50	5.00	1.42	4.00	4.25	1.56
31	Lei spinge l'acceleratore	Le	7.00	6.38	1.06	5.00	4.75	1.64	6.00	6.13	0.78
32	Lei strascica i piedi	Le	6.50	6.50	0.53	3.50	3.75	1.86	5.50	5.00	1.88
33	Lei struscia le ciabatte	Le	7.00	6.50	0.76	6.00	5.50	1.51	4.50	3.75	1.80
34	Lei tira un calcio	Le	7.00	7.00	0.00	6.50	5.25	2.15	6.50	5.75	1.41
35	Lei toglie le ciabatte	Le	7.00	7.00	0.00	6.00	4.75	2.47	5.00	4.75	1.92

Appendix 2. Stimuli and ratings of Study C and Study D

Appendix 2.1. Metaphor set

Meaningfulness, Difficulty, Familiarity and Cloze Probability (Cloze) ratings for metaphorical sentences (MP) and their literal (L) and anomalous (A) counterparts.

		Label	**Meaningfulness**			**Difficulty**			**Familiarity**			**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*Mean*	*Median*	*Sd*	*%*
	Metaphorical sentences											
1	Quegli aliti sono fogne	MP	3.00	3.00	1.00	2.38	2.00	1.41	3.10	3.00	1.37	0.00
2	Quegli avvocati sono squali	MP	4.00	4.50	1.55	1.67	2.00	0.52	3.86	4.00	1.46	0.00
3	Quegli eserciti sono dighe	MP	2.78	2.00	1.09	2.13	2.00	0.99	1.00	1.00	0.00	0.00
4	Quegli oratori sono treni	MP	2.86	3.00	1.21	1.71	2.00	0.76	2.00	2.00	1.10	0.00
5	Quegli scalatori sono scoiattoli	MP	3.67	4.00	1.03	2.00	2.00	0.89	2.20	2.00	1.23	0.00
6	Quegli schiaffi sono medicine	MP	3.50	4.00	0.84	2.00	2.00	0.63	2.33	2.00	1.22	0.00
7	Quegli uragani sono carrarmati	MP	2.33	2.50	1.21	2.33	2.50	1.21	2.11	2.00	1.05	0.00
8	Quei bambini sono fontane	MP	2.89	3.00	0.60	2.38	3.00	0.92	1.60	1.00	0.84	0.00
9	Quei banchieri sono vampiri	MP	3.83	4.00	1.17	1.50	1.50	0.55	4.17	4.50	1.17	0.00
10	Quei barbieri sono chirurghi	MP	2.67	3.00	1.51	2.00	2.00	0.63	2.83	3.00	1.17	0.00
11	Quei colleghi sono zecche	MP	3.67	4.00	1.03	1.50	1.50	0.55	2.89	3.00	1.05	0.00
12	Quei delitti sono rebus	MP	4.17	4.50	0.98	1.50	1.50	0.55	3.00	3.00	1.41	0.00
13	Quei ginnasti sono grilli	MP	3.50	4.00	0.84	1.67	1.50	0.82	2.33	2.00	0.82	0.00
14	Quei giocatori sono elefanti	MP	2.50	2.50	1.38	2.17	2.00	0.75	1.43	1.00	0.53	0.00
15	Quei giornalisti sono avvoltoi	MP	4.56	5.00	0.53	1.13	1.00	0.35	4.20	4.00	0.79	0.00
16	Quei libri sono bussole	MP	3.71	4.00	0.49	3.14	3.00	1.07	2.00	2.00	0.89	0.00
17	Quei maestri sono lanterne	MP	2.89	3.00	0.93	2.13	2.00	0.64	1.00	1.00	0.00	0.00
18	Quei politici sono calamite	MP	2.67	2.50	1.63	2.17	2.00	0.75	2.57	3.00	0.98	0.00
19	Quei portinai sono archivi	MP	2.33	2.00	1.51	2.33	2.00	0.52	1.33	1.00	0.82	0.00
20	Quei professori sono enciclopedie	MP	4.00	4.00	1.00	1.29	1.00	0.49	3.83	3.50	0.98	0.00
21	Quei ricordi sono spine	MP	3.78	4.00	0.83	1.63	2.00	0.52	4.10	4.00	0.88	0.00
22	Quei soldati sono leoni	MP	3.83	4.00	1.47	1.67	1.00	1.21	2.80	3.00	0.92	0.00
23	Quei viaggi sono terapie	MP	3.17	3.50	1.17	1.83	2.00	0.75	3.17	3.00	0.75	0.00
24	Quelle acconciature sono cespugli	MP	3.50	3.00	0.84	2.33	2.00	1.03	4.17	4.50	0.98	0.00
25	Quelle automobili sono frecce	MP	4.17	4.00	0.75	1.83	2.00	0.75	3.90	4.00	0.74	0.00
26	Quelle ballerine sono farfalle	MP	3.67	3.50	1.21	2.17	2.00	0.41	2.43	2.00	1.27	0.00
27	Quelle biblioteche sono miniere	MP	4.00	4.00	0.89	1.33	1.00	0.52	3.33	3.00	1.32	0.00
28	Quelle borse sono macigni	MP	3.33	3.50	0.82	1.33	1.00	0.52	4.00	4.00	0.63	0.00
29	Quelle canzoni sono droghe	MP	3.17	3.00	0.75	2.33	2.00	0.52	1.80	2.00	0.63	0.00
30	Quelle carezze sono balsami	MP	3.00	3.00	0.89	2.17	2.00	0.41	1.57	2.00	0.53	0.00

		Label	Meaningfulness			Difficulty			Familiarity			Cloze
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*Mean*	*Median*	*Sd*	*%*
31	Quelle città sono giungle	MP	3.86	4.00	0.38	2.14	2.00	1.35	4.17	4.50	1.17	0.00
32	Quelle fanciulle sono rose	MP	3.33	4.00	1.21	2.17	2.00	1.47	1.89	2.00	0.78	0.00
33	Quelle giacche sono forni	MP	2.50	2.00	1.38	2.17	2.50	0.98	2.70	3.00	1.06	0.00
34	Quelle guance sono pesche	MP	3.22	3.00	0.83	2.13	2.00	1.25	2.90	3.00	1.66	0.00
35	Quelle indossatrici sono bambole	MP	3.50	4.00	1.38	1.67	1.00	1.21	3.44	3.00	1.13	0.00
36	Quelle lezioni sono sonniferi	MP	4.29	4.00	0.76	1.29	1.00	0.49	3.33	3.50	0.82	0.00
37	Quelle malattie sono cecchini	MP	2.83	3.00	1.72	2.17	2.50	0.98	1.43	1.00	0.79	0.00
38	Quelle melodie sono camomille	MP	2.67	3.00	1.37	2.17	2.00	1.47	2.67	3.00	1.12	0.00
39	Quelle notizie sono terremoti	MP	3.57	4.00	0.98	1.86	1.00	1.21	2.67	2.00	1.51	0.00
40	Quelle tasche sono banche	MP	2.44	2.00	1.24	2.50	2.00	1.07	1.10	1.00	0.32	0.00
41	Quelle torte sono montagne	MP	3.00	2.50	1.26	2.67	3.00	0.52	2.00	2.00	0.89	0.00
42	Quelle voci sono trombe	MP	2.67	2.00	1.51	2.00	2.00	0.89	3.70	3.00	0.95	0.00
	Literal sentences											
1	Quegli scarichi sono fogne	L	3.83	4.00	0.75	2.33	2.00	0.52	4.70	5.00	0.67	0.00
2	Quei pesci sono squali	L	4.33	4.00	0.71	1.13	1.00	0.35	4.00	4.00	1.05	0.00
3	Quelle costruzioni sono dighe	L	4.17	5.00	1.60	1.33	1.00	0.52	4.00	4.00	0.94	0.00
4	Quei convogli sono treni	L	3.83	4.00	0.98	1.83	1.50	1.17	3.44	4.00	1.51	0.00
5	Quei roditori sono scoiattoli	L	4.17	4.50	1.17	1.33	1.00	0.52	3.71	3.00	0.95	0.00
6	Quelle sostanze sono medicine	L	4.17	4.00	0.75	1.67	2.00	0.52	4.67	5.00	0.52	0.00
7	Quei blindati sono carrarmati	L	3.50	4.00	1.38	1.83	1.50	0.98	3.17	3.00	1.17	0.00
8	Quegli impianti sono fontane	L	2.67	3.00	1.37	2.33	2.00	1.51	2.30	2.00	1.25	0.00
9	Quei mostri sono vampiri	L	4.43	4.00	0.53	1.14	1.00	0.38	3.17	3.00	1.72	0.00
10	Quei medici sono chirurghi	L	5.00	5.00	0.00	1.00	1.00	0.00	3.33	4.00	1.86	0.00
11	Quei parassiti sono zecche	L	4.50	4.50	0.55	1.17	1.00	0.41	2.50	2.00	1.38	0.00
12	Quegli indovinelli sono rebus	L	3.50	4.00	1.22	1.67	2.00	0.52	3.33	3.50	1.63	0.00
13	Quegli insetti sono grilli	L	4.14	4.00	1.07	1.29	1.00	0.49	2.67	2.50	1.63	0.00
14	Quei mammiferi sono elefanti	L	4.44	4.00	0.53	1.88	1.50	1.36	3.10	3.00	0.88	0.00
15	Quei predatori sono avvoltoi	L	4.50	4.50	0.55	1.67	2.00	0.52	4.20	4.00	0.79	0.00
16	Quei dispositivi sono bussole	L	4.17	5.00	1.60	2.00	1.00	1.67	3.56	4.00	0.88	0.00
17	Quelle lampade sono lanterne	L	3.67	3.50	1.21	1.50	1.00	0.84	3.30	3.50	1.34	0.00
18	Quegli oggetti sono calamite	L	4.56	5.00	0.73	1.38	1.00	0.52	3.90	4.00	1.20	0.00
19	Quegli schedari sono archivi	L	4.43	5.00	0.79	1.14	1.00	0.38	3.00	3.00	1.90	0.00
20	Quei volumi sono enciclopedie	L	3.83	4.00	0.75	1.83	2.00	0.75	4.11	4.00	1.05	0.00
21	Quelle punte sono spine	L	3.67	4.00	1.03	2.17	2.00	0.75	3.50	3.50	1.08	0.00
22	Quei felini sono leoni	L	4.83	5.00	0.41	1.50	1.50	0.55	4.00	4.00	1.00	0.00
23	Quei trattamenti sono terapie	L	4.43	5.00	0.79	1.57	2.00	0.53	3.83	4.50	1.60	0.00
24	Quelle siepi sono cespugli	L	3.57	4.00	1.40	1.86	2.00	0.90	3.00	3.00	1.90	0.00

		Label	**Meaningfulness**			**Difficulty**			**Familiarity**			**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*Mean*	*Median*	*Sd*	*%*
25	Quelle armi sono frecce	L	4.33	4.50	0.82	1.67	1.50	0.82	3.57	3.00	0.79	0.00
26	Quegli animali sono farfalle	L	4.11	5.00	1.17	1.50	1.50	0.53	3.30	3.50	1.77	0.00
27	Quelle cave sono miniere	L	3.00	3.50	1.26	2.50	2.50	1.05	3.33	3.50	1.63	0.00
28	Quelle pietre sono macigni	L	4.86	5.00	0.38	1.00	1.00	0.00	4.50	4.50	0.55	0.00
29	Quelle pasticche sono droghe	L	4.00	4.50	1.55	1.50	1.00	0.84	4.00	4.00	0.58	0.17
30	Quei detergenti sono balsami	L	3.11	3.00	0.78	2.38	2.00	0.74	3.20	3.00	1.03	0.00
31	Quelle foreste sono giungle	L	4.33	4.00	0.52	1.67	1.50	0.82	3.33	3.00	1.58	0.00
32	Quei fiori sono rose	L	4.33	5.00	1.63	1.00	1.00	0.00	4.83	5.00	0.41	0.00
33	Quegli elettrodomestici sono forni	L	3.67	4.00	1.37	1.50	1.00	0.84	3.71	4.00	1.11	0.00
34	Quei frutti sono pesche	L	4.83	5.00	0.41	1.00	1.00	0.00	4.80	5.00	0.42	0.00
35	Quei pupazzi sono bambole	L	3.67	4.00	1.03	2.00	2.00	1.10	3.50	3.50	1.05	0.00
36	Quelle pillole sono sonniferi	L	4.67	5.00	0.52	1.50	1.00	0.84	4.11	5.00	1.17	0.00
37	Quei militari sono cecchini	L	4.22	5.00	1.09	1.63	1.50	0.74	3.40	3.50	1.58	0.00
38	Quelle bevande sono camomille	L	3.50	4.50	1.97	1.50	1.50	0.55	4.00	4.00	1.10	0.00
39	Quelle scosse sono terremoti	L	4.17	4.50	0.98	1.83	1.50	0.98	2.78	3.00	0.67	0.00
40	Quegli istituti sono banche	L	4.83	5.00	0.41	1.33	1.00	0.52	3.80	3.50	1.14	0.00
41	Quei rilievi sono montagne	L	4.00	4.00	1.00	1.14	1.00	0.38	4.17	4.50	1.17	0.00
42	Quegli strumenti sono trombe	L	4.00	4.50	1.55	1.17	1.00	0.41	3.71	3.00	1.25	0.00
	Anomalous sentences											
1	Quegli spilli sono fogne	A	2.00	1.00	1.67	1.67	1.50	0.82	1.00	1.00	0.00	0.00
2	Quei fischietti sono squali	A	1.00	1.00	0.00	2.00	1.00	1.67	1.20	1.00	0.63	0.00
3	Quei rospi sono dighe	A	2.00	1.50	1.55	1.67	1.50	0.82	1.00	1.00	0.00	0.00
4	Quei carciofi sono treni	A	1.00	1.00	0.00	1.00	1.00	0.00	1.17	1.00	0.41	0.00
5	Quei cuscini sono scoiattoli	A	1.89	2.00	0.93	2.00	2.00	1.07	1.20	1.00	0.42	0.00
6	Quelle unghie sono medicine	A	1.43	1.00	0.53	1.43	1.00	0.79	1.00	1.00	0.00	0.00
7	Quelle case sono carrarmati	A	2.71	3.00	1.11	1.43	1.00	0.79	1.83	1.50	1.17	0.00
8	Quei picconi sono fontane	A	1.83	1.00	1.60	1.67	1.50	0.82	1.00	1.00	0.00	0.00
9	Quegli agrumi sono vampiri	A	1.17	1.00	0.41	2.67	2.00	1.97	1.00	1.00	0.00	0.00
10	Quei gelati sono chirurghi	A	1.00	1.00	0.00	2.33	1.00	2.07	1.00	1.00	0.00	0.00
11	Quei pennelli sono zecche	A	1.43	1.00	0.53	1.43	1.00	0.53	2.83	3.00	1.33	0.00
12	Quei pomodori sono rebus	A	1.43	1.00	1.13	2.14	1.00	1.68	1.00	1.00	0.00	0.00
13	Quelle ciotole sono grilli	A	1.17	1.00	0.41	2.50	1.50	1.97	1.00	1.00	0.00	0.00
14	Quelle zattere sono elefanti	A	1.17	1.00	0.41	2.33	1.50	1.75	1.10	1.00	0.32	0.00
15	Quei pedali sono avvoltoi	A	1.17	1.00	0.41	1.33	1.00	0.52	1.00	1.00	0.00	0.00
16	Quelle bistecche sono bussole	A	1.33	1.00	0.52	1.33	1.00	0.82	1.17	1.00	0.41	0.00
17	Quei merli sono lanterne	A	2.00	2.00	1.10	1.83	2.00	0.75	1.00	1.00	0.00	0.00
18	Quelle lumache sono calamite	A	1.83	2.00	0.75	2.33	2.00	1.51	1.00	1.00	0.00	0.00

		Label	**Meaningfulness**			**Difficulty**			**Familiarity**			**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*Mean*	*Median*	*Sd*	*%*
19	Quelle trote sono archivi	A	1.00	1.00	0.00	2.50	1.50	1.97	1.00	1.00	0.00	0.00
20	Quelle rane sono enciclopedie	A	1.00	1.00	0.00	1.33	1.00	0.82	1.00	1.00	0.00	0.00
21	Quei piazzali sono spine	A	1.17	1.00	0.41	1.17	1.00	0.41	1.00	1.00	0.00	0.00
22	Quelle gocce sono leoni	A	1.33	1.00	0.71	2.00	1.50	1.20	1.00	1.00	0.00	0.00
23	Quelle forchette sono terapie	A	1.33	1.00	0.82	3.17	3.50	2.04	1.00	1.00	0.00	0.00
24	Quei chiodi sono cespugli	A	1.67	1.50	0.82	2.67	2.50	1.63	1.00	1.00	0.00	0.00
25	Quelle miscele sono frecce	A	1.78	1.00	0.97	2.00	2.00	1.07	1.00	1.00	0.00	0.00
26	Quelle pentole sono farfalle	A	1.17	1.00	0.41	2.00	1.50	1.55	1.00	1.00	0.00	0.00
27	Quei timoni sono miniere	A	1.86	2.00	1.07	1.57	1.00	0.79	1.00	1.00	0.00	0.00
28	Quelle scintille sono macigni	A	1.67	2.00	0.52	3.33	3.50	1.63	1.11	1.00	0.33	0.00
29	Quelle chiavi sono droghe	A	1.78	2.00	0.97	2.38	2.00	0.92	1.00	1.00	0.00	0.00
30	Quei cucchiai sono balsami	A	1.17	1.00	0.41	2.33	2.50	1.21	1.00	1.00	0.00	0.00
31	Quelle carote sono giungle	A	1.67	1.00	1.63	1.33	1.00	0.52	1.33	1.00	0.82	0.00
32	Quegli aerei sono rose	A	1.86	1.00	1.46	2.86	3.00	1.35	1.00	1.00	0.00	0.00
33	Quelle etichette sono forni	A	1.67	2.00	0.71	2.00	2.00	0.93	1.00	1.00	0.00	0.00
34	Quei solchi sono pesche	A	1.33	1.00	0.52	1.50	1.50	0.55	1.00	1.00	0.00	0.00
35	Quelle pianure sono bambole	A	1.29	1.00	0.49	1.43	1.00	0.79	1.17	1.00	0.41	0.00
36	Quei rapaci sono sonniferi	A	1.17	1.00	0.41	1.17	1.00	0.41	1.00	1.00	0.00	0.00
37	Quelle olive sono cecchini	A	1.00	1.00	0.00	1.33	1.00	0.52	1.00	1.00	0.00	0.00
38	Quelle borchie sono camomille	A	1.29	1.00	0.49	1.71	1.00	1.25	1.00	1.00	0.00	0.00
39	Quelle tazze sono terremoti	A	1.33	1.00	0.52	1.00	1.00	0.00	1.17	1.00	0.41	0.00
40	Quelle castagne sono banche	A	1.67	1.00	1.21	1.00	1.00	0.00	1.00	1.00	0.00	0.00
41	Quelle finestre sono montagne	A	1.00	1.00	0.00	2.67	2.00	1.97	1.00	1.00	0.00	0.00
42	Quei ragni sono trombe	A	1.22	1.00	0.67	1.50	1.00	0.76	1.00	1.00	0.00	0.00

Appendix 2.2. Metonymy set

Meaningfulness, Difficulty, World Knowledge (WK) and Cloze Probability (Cloze) ratings for metonymic (MT) sentences and their literal (L) and anomalous (A) counterparts.

		Label	**Meaningfulness**			**Difficulty**			**WK**	**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*%*	*%*
	Metonymic sentences									
1	Quei ragazzi noleggiano Muccino	MT	3.17	3.50	1.47	1.83	2.00	0.75	100	0
2	Quel bambino incolla Totti	MT	3.14	3.00	0.69	1.86	2.00	0.90	100	0
3	Quel bibliotecario archivia Baricco	MT	3.17	3.50	1.47	1.83	2.00	0.75	100	0
4	Quel cabarettista canta Celentano	MT	4.11	5.00	1.27	1.13	1.00	0.35	100	0
5	Quel calciatore indossa Armani	MT	4.44	5.00	0.88	1.25	1.00	0.46	86	0
6	Quel cantante intona Bocelli	MT	4.33	4.50	0.82	1.83	1.50	1.17	100	0
7	Quel cinefilo restituisce Salvatores	MT	3.43	3.00	0.98	1.86	2.00	0.38	100	0
8	Quel collezionista cataloga Mina	MT	3.50	3.00	0.84	1.83	2.00	0.75	100	0
9	Quel commediante recita Fo	MT	4.83	5.00	0.41	1.67	1.50	0.82	100	0
10	Quel disc-jokey mixa Vasco	MT	3.83	4.50	1.60	1.83	1.00	1.60	100	0
11	Quel gallerista espone Toscani	MT	4.17	4.50	0.98	1.50	1.00	0.84	83	0
12	Quel lavandaio stira Missoni	MT	3.33	3.50	1.37	2.33	2.50	0.82	79	0
13	Quel lettore sottolinea Eco	MT	4.14	4.00	0.69	1.57	1.00	0.79	100	0
14	Quel libraio ordina Lucarelli	MT	3.67	4.00	1.37	1.67	1.50	0.82	86	0
15	Quel negoziante svende Fiorucci	MT	4.67	5.00	0.50	1.25	1.00	0.46	79	0
16	Quel pensionato legge Camilleri	MT	4.83	5.00	0.41	1.17	1.00	0.41	100	0
17	Quel presidente cancella Santoro	MT	3.86	4.00	0.69	1.71	2.00	0.76	83	0
18	Quel professore parafrasa Tabucchi	MT	3.17	3.50	1.47	1.83	2.00	0.75	100	0
19	Quel radioascoltatore spegne Fiorello	MT	3.29	3.00	0.95	1.71	2.00	0.49	100	0
20	Quel ragazzino ritaglia Cannavaro	MT	4.00	4.50	1.26	1.67	1.00	1.03	79	0
21	Quel recensore visiona Bertolucci	MT	3.33	3.00	1.22	2.38	2.50	1.06	57	0
22	Quel revisore corregge Carofiglio	MT	3.33	3.50	1.63	2.67	2.50	1.63	79	0
23	Quel tassista fischietta Baglioni	MT	4.00	4.00	1.10	1.33	1.00	0.52	100	0
24	Quel tecnico proietta Pieraccioni	MT	3.83	4.00	1.17	1.33	1.00	0.52	100	0
25	Quel tifoso sventola Maradona	MT	3.00	3.50	1.26	1.50	1.50	0.55	100	0
26	Quel tipografo ristampa Augias	MT	4.17	5.00	1.33	2.00	2.00	1.10	100	0
27	Quel vetrinista allestisce Ferrè	MT	4.00	4.00	0.87	1.75	2.00	0.71	71	0
28	Quella cameriera canticchia Elisa	MT	4.14	5.00	1.21	1.29	1.00	0.49	100	0
29	Quella contessa calza Prada	MT	4.67	5.00	0.82	1.50	1.50	0.55	100	0
30	Quel liceale colleziona Guccini	MT	4.33	4.50	0.82	1.50	1.00	0.84	100	0
31	Quella liceale legge Scarpa	MT	4.33	4.50	0.82	1.50	1.50	0.55	67	0

		Label	Meaningfulness			Difficulty			WK	Cloze
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	%	%
32	Quella maestra fotocopia Benni	MT	3.33	4.00	1.21	1.67	1.50	0.82	79	0
33	Quell'adolescente colleziona Ligabue	MT	3.89	4.00	0.93	1.88	2.00	0.64	100	0
34	Quelle quindicenni comprano Povia	MT	4.00	4.00	0.82	1.29	1.00	0.76	100	0
35	Quelle ragazzine trascrivono Moccia	MT	4.00	4.00	1.10	1.33	1.00	0.82	42	0
36	Quell'editore pubblica Saviano	MT	4.33	5.00	1.32	1.75	1.00	1.39	100	0
37	Quell'esperto analizza Tarantino	MT	3.83	4.00	0.75	1.67	2.00	0.52	100	0
38	Quell'indossatrice prova Valentino	MT	3.50	4.00	1.38	1.83	2.00	0.75	79	0
39	Quello sceneggiatore adatta Ammaniti	MT	4.00	4.00	0.58	1.43	1.00	0.53	100	0
40	Quello scrittore traduce Fruttero	MT	3.83	5.00	1.83	1.67	1.00	1.21	57	0
41	Quello studente acquista Cristicchi	MT	3.11	3.00	0.78	1.75	2.00	0.46	86	0
42	Quell'universitario masterizza Jovanotti	MT	3.67	3.50	1.21	1.67	1.50	0.82	100	0
	Literal sentences									
1	Quella valletta introduce Muccino	L	4.33	4.00	0.50	1.38	1.00	0.52		0
2	Quell'allenatore incita Totti	L	4.83	5.00	0.41	1.17	1.00	0.41		0
3	Quell'opinionista contraddice Baricco	L	4.86	5.00	0.38	1.14	1.00	0.38		0
4	Quella presentatrice saluta Celentano	L	5.00	5.00	0.00	1.33	1.00	0.52		0
5	Quel reporter intervista Armani	L	4.67	5.00	0.52	1.33	1.00	0.52		0
6	Quel maestro dirige Bocelli	L	3.83	4.50	1.60	1.33	1.00	0.52		0
7	Quella soubrette presenta Salvatores	L	3.67	3.50	1.21	2.17	2.00	1.47		0
8	Quel fotografo ritrae Mina	L	4.43	5.00	0.79	1.14	1.00	0.38		0
9	Quell'accademico accoglie Fo	L	4.17	5.00	1.33	1.33	1.00	0.52		0
10	Quel carabiniere perquisisce Vasco	L	4.00	4.50	1.55	1.33	1.00	0.52		0
11	Quel curatore chiama Toscani	L	3.33	4.00	1.21	2.00	2.00	0.63		0
12	Quella sarta interpella Missoni	L	4.00	4.00	0.71	1.75	1.50	1.04		0
13	Quel laureando aspetta Eco	L	4.00	4.50	1.26	2.00	1.50	1.26		0
14	Quel produttore assume Lucarelli	L	3.78	4.00	1.09	1.88	1.50	0.99		0
15	Quel disegnatore consulta Fiorucci	L	4.50	4.50	0.55	1.67	1.50	0.82		0
16	Quel letterato incontra Camilleri	L	4.33	4.50	0.82	1.33	1.00	0.82		0
17	Quel politicoQuerela Santoro	L	5.00	5.00	0.00	1.33	1.00	0.52		0
18	Quello studioso invita Tabucchi	L	4.43	4.00	0.53	1.00	1.00	0.00		0
19	Quel comico imita Fiorello	L	4.83	5.00	0.41	1.17	1.00	0.41		0
20	Quel fisioterapista massaggia Cannavaro	L	4.83	5.00	0.41	1.17	1.00	0.41		0
21	Quell'attore visita Bertolucci	L	3.83	4.00	0.75	2.00	2.00	0.89		0
22	Quel regista congeda Carofiglio	L	2.89	2.00	1.36	1.88	2.00	0.99		0
23	Quelle fans assalgono Baglioni	L	4.71	5.00	0.49	1.00	1.00	0.00		0
24	Quel critico aggredisce Pieraccioni	L	4.33	4.50	0.82	1.17	1.00	0.41		0
25	Quel medico cura Maradona	L	4.71	5.00	0.49	1.00	1.00	0.00		0

		Label	Meaningfulness			Difficulty			WK	Cloze
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*%*	*%*
26	Quel direttore licenzia Augias	L	4.33	4.50	0.82	1.17	1.00	0.41		0
27	Quella modella ispira Ferrè	L	3.83	4.00	1.17	2.67	2.50	1.21		0
28	Quel discografico ingaggia Elisa	L	4.67	5.00	0.52	1.17	1.00	0.41		0
29	Quel pubblicitario contatta Prada	L	4.17	4.50	1.17	1.50	1.50	0.55		0
30	Quel presentatore ringrazia Guccini	L	4.00	4.00	1.10	1.33	1.00	0.52		0
31	Quell'editore invita Scarpa	L	3.57	4.00	0.98	1.57	1.00	0.79		0
32	Quel rettore omaggia Benni	L	3.83	4.50	1.60	1.83	1.50	1.17		0
33	Quell'ammiratore interpella Ligabue	L	4.00	4.00	0.89	1.33	1.00	0.52		0
34	Quel paroliere aiuta Povia	L	4.17	4.50	1.17	1.67	1.00	1.21		0
35	Quelle ammiratrici importunano Moccia	L	4.44	5.00	0.88	1.13	1.00	0.35		0
36	Quel poliziotto protegge Saviano	L	5.00	5.00	0.00	1.17	1.00	0.41		0
37	Quell'operatore avvicina Tarantino	L	4.29	4.00	0.49	1.29	1.00	0.49		0
38	Quel paparazzo fotografa Valentino	L	4.67	5.00	0.52	1.67	1.00	1.21		0
39	Quel redattore convoca Ammaniti	L	4.17	4.00	0.75	1.67	1.50	0.82		0
40	Quel giornalista incalza Fruttero	L	4.00	4.00	1.00	1.63	1.50	0.74		0
41	Quell'inviato insulta Cristicchi	L	4.33	4.00	0.52	1.50	1.50	0.55		0
42	Quella guardia scorta Jovanotti	L	4.17	4.00	0.41	1.33	1.00	0.52		0
	Anomalous sentences									
1	Quel panettiere impasta Muccino	A	1.00	1.00	0.00	1.17	1.00	0.41		0
2	Quel pilota avvia Totti	A	1.33	1.00	0.52	1.67	2.00	0.52		0
3	Quell'alpinista scala Baricco	A	1.17	1.00	0.41	2.33	1.50	1.75		0
4	Quel giardiniere rastrella Celentano	A	1.33	1.00	0.82	1.33	1.00	0.52		0
5	Quel falegname costruisce Armani	A	1.17	1.00	0.41	1.67	2.00	0.52		0
6	Quella casalinga rassetta Bocelli	A	1.33	1.00	0.50	2.13	2.00	1.25		0
7	Quel calzolaio ripara Salvatores	A	1.17	1.00	0.41	1.50	1.50	0.55		0
8	Quel dentista estrae Mina	A	1.33	1.00	0.82	2.17	1.00	1.83		0
9	Quel portiere para Fo	A	2.11	2.00	1.05	2.75	2.00	1.39		0
10	Quell'avventore zucchera Vasco	A	1.57	1.00	0.79	1.86	2.00	0.69		0
11	Quel pagliaccio gonfia Toscani	A	2.86	2.00	1.07	3.43	4.00	1.13		0
12	Quell'agricoltore raccoglie Missoni	A	1.33	1.00	0.82	1.50	1.00	0.84		0
13	Quel netturbino spazza Eco	A	1.33	1.00	0.52	1.33	1.00	0.52		0
14	Quell'elettricista avvita Lucarelli	A	1.00	1.00	0.00	1.83	2.00	0.75		0
15	Quel cameriere sparecchia Fiorucci	A	1.67	2.00	0.52	1.83	1.50	0.98		0
16	Quel maggiordomo cucina Camilleri	A	2.14	2.00	1.21	1.43	1.00	0.79		0
17	Quel chitarrista accorda Santoro	A	1.83	1.00	1.60	1.17	1.00	0.41		0
18	Quel marinaio fuma Tabucchi	A	1.00	1.00	0.00	2.50	1.50	1.97		0
19	Quel boscaiolo sega Fiorello	A	2.17	1.50	1.60	1.17	1.00	0.41		0

		Label	**Meaningfulness**			**Difficulty**			**WK**	**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	%	%
20	Quel gioielliere lucida Cannavaro	A	1.56	1.00	0.73	1.63	1.00	0.92		0
21	Quel fioraio concima Bertolucci	A	1.00	1.00	0.00	1.33	1.00	0.52		0
22	Quell'imbianchino dipinge Carofiglio	A	1.83	1.50	0.98	2.50	2.50	1.52		0
23	Quell'iAiegato protocolla Baglioni	A	2.67	2.00	1.51	3.33	3.00	0.82		0
24	Quella pescivendola pulisce Pieraccioni	A	2.43	2.00	1.27	1.71	1.00	1.25		0
25	Quel gommista cambia Maradona	A	1.00	1.00	0.00	2.50	1.50	1.97		0
26	Quell'atleta salta Augias	A	2.00	2.00	0.82	2.14	2.00	1.21		0
27	Quel pastore munge Ferrè	A	1.00	1.00	0.00	1.00	1.00	0.00		0
28	Quel pizzaiolo inforna Elisa	A	1.67	1.00	1.63	1.00	1.00	0.00		0
29	Quel contadino semina Prada	A	1.14	1.00	0.38	1.29	1.00	0.76		0
30	Quel pasticciere sforna Guccini	A	1.44	1.00	0.73	1.88	1.00	1.46		0
31	Quel macellaio affetta Scarpa	A	1.33	1.00	0.82	2.83	2.50	1.83		0
32	Quello chef frigge Benni	A	1.78	2.00	0.67	2.00	1.50	1.20		0
33	Quella sarta cuce Ligabue	A	1.83	1.00	1.60	1.17	1.00	0.41		0
34	Quel meccanico aggiusta Povia	A	1.83	1.00	1.60	1.17	1.00	0.41		0
35	Quel sommelier beve Moccia	A	1.17	1.00	0.41	1.83	1.00	1.33		0
36	Quel cuoco condisce Saviano	A	1.83	1.00	1.60	1.17	1.00	0.41		0
37	Quel benzinaio miscela Tarantino	A	1.00	1.00	0.00	2.50	1.50	1.97		0
38	Quel gelataio mescola Valentino	A	1.33	1.00	0.71	2.63	2.50	1.41		0
39	Quel manovale smantella Ammaniti	A	1.17	1.00	0.41	1.33	1.00	0.52		0
40	Quel pescatore getta Fruttero	A	1.67	1.00	1.63	1.67	1.50	0.82		0
41	Quell'informatico programma Cristicchi	A	1.50	1.50	0.55	1.67	2.00	0.52		0
42	Quel muratore piastrella Jovanotti	A	1.57	1.00	0.79	1.57	1.00	1.13		0

Appendix 2.3. Approximation set

Meaningfulness, Difficulty, Typicality and Cloze Probability (Cloze) ratings for approximations (AP) and their literal (L) and anomalous (A) counterparts.

		Label	**Meaningfulness**			**Difficulty**			**Typicality**			**Cloze**
			Mean	*Median*	*SD*	*Mean*	*Median*	*SD*	*Mean*	*Median*	*Sd*	*%*
	Approximations											
1	Quegli occhiali sono rettangolari	AP	4.33	4.00	0.52	1.50	1.50	0.55	3.57	3.00	0.79	0.00
2	Quei capelli sono stirati	AP	3.67	3.50	0.82	2.00	2.00	0.89	3.43	3.00	0.98	0.00
3	Quei denti sono diritti	AP	4.11	4.00	1.05	1.63	2.00	0.52	3.29	3.00	1.25	0.00
4	Quel brodo è lungo	AP	3.50	4.00	1.64	1.67	1.50	0.82	2.86	3.00	1.07	0.00
5	Quel budino è duro	AP	3.67	3.00	0.87	1.75	2.00	0.71	2.14	2.00	1.21	0.00
6	Quel cappuccino è ghiacciato	AP	3.67	3.50	1.21	2.00	2.00	1.10	2.83	2.00	1.33	0.00
7	Quel centrino è ottagonale	AP	3.83	4.00	1.47	1.50	1.50	0.55	3.83	4.00	0.98	0.00
8	Quel ginocchio è appuntito	AP	3.17	4.00	1.72	1.67	2.00	0.52	3.00	3.00	1.15	0.00
9	Quel gomito è sbucciato	AP	4.17	5.00	1.60	1.17	1.00	0.41	4.17	4.00	0.41	0.00
10	Quel grappolo è piramidale	AP	3.86	4.00	0.69	1.86	2.00	0.90	3.00	3.00	0.89	0.00
11	Quel lago è verde	AP	3.43	3.00	1.40	1.14	1.00	0.38	2.00	2.00	0.89	0.00
12	Quel profumo è soffocante	AP	3.83	4.00	0.75	1.67	1.00	1.21	3.29	3.00	0.76	0.00
13	Quel quartiere è deserto	AP	4.33	4.00	0.50	1.25	1.00	0.46	4.17	4.00	0.41	0.00
14	Quel territorio è piatto	AP	4.00	4.00	0.58	2.14	2.00	1.35	4.17	4.50	0.98	0.00
15	Quel tessuto è inconsistente	AP	3.83	3.50	0.98	2.50	2.50	1.05	3.33	3.50	0.82	0.00
16	Quel vestito è trasparente	AP	4.00	4.00	0.87	1.38	1.00	0.52	4.50	4.50	0.55	0.00
17	Quel viso è triangolare	AP	4.00	4.00	1.10	1.50	1.00	0.84	2.83	2.50	1.17	0.00
18	Quel volto è immutato	AP	3.83	3.50	0.98	2.33	2.00	1.03	3.50	3.00	0.84	0.00
19	Quella casa è piena	AP	3.71	4.00	0.76	1.71	2.00	0.76	3.83	4.00	0.41	0.00
20	Quella cipolla è bianca	AP	4.83	5.00	0.41	1.17	1.00	0.41	3.57	4.00	0.98	0.00
21	Quella depilazione è indolore	AP	4.00	4.50	1.55	1.50	1.50	0.55	4.17	4.00	0.75	0.00
22	Quella faccia è gialla	AP	4.00	4.00	1.10	1.83	1.50	0.98	3.00	3.00	0.63	0.00
23	Quella minestra è insapore	AP	4.83	5.00	0.41	1.00	1.00	0.00	4.17	4.50	1.17	0.00
24	Quella pagnotta è rotonda	AP	3.50	4.00	1.38	1.00	1.00	0.00	3.43	4.00	0.79	0.00
25	Quella pannocchia è cilindrica	AP	3.67	4.00	1.51	2.17	2.00	1.47	3.29	3.00	1.11	0.00
26	Quella porzione è microscopica	AP	4.17	4.50	1.17	1.17	1.00	0.41	3.00	3.50	1.67	0.00
27	Quella testa è ovale	AP	3.56	3.00	1.01	1.75	1.50	1.04	3.83	4.00	1.17	0.00
28	Quella volpe è rossa	AP	4.83	5.00	0.41	1.00	1.00	0.00	2.86	3.00	1.07	0.17
29	Quell'alba è rosa	AP	4.00	4.00	0.71	1.88	2.00	0.64	3.50	3.50	0.55	0.00
30	Quell'albergo è vuoto	AP	5.00	5.00	0.00	1.00	1.00	0.00	4.71	5.00	0.49	0.00
31	Quell'arrosto è crudo	AP	4.67	5.00	0.71	1.50	1.00	0.76	3.83	4.00	1.47	0.00

32	Quell'attore è rifatto	AP	3.33	3.50	1.37	1.50	1.00	0.84	4.00	4.00	0.82	0.00
33	Quell'automobile è silenziosa	AP	4.43	4.00	0.53	1.14	1.00	0.38	3.50	3.00	0.84	0.00
34	Quelle gomme sono lisce	AP	4.67	5.00	0.52	1.17	1.00	0.41	3.86	4.00	1.07	0.20
35	Quelle guance sono concave	AP	3.33	3.50	1.63	1.83	1.50	0.98	2.14	2.00	0.90	0.00
36	Quelle labbra sono blu	AP	3.50	4.00	0.84	2.17	2.00	1.33	2.67	3.00	1.03	0.00
37	Quelle nuvole sono nere	AP	4.00	4.00	0.89	1.17	1.00	0.41	3.67	4.00	1.03	0.33
38	Quelle orecchie sono bollenti	AP	4.14	4.00	0.38	1.29	1.00	0.49	2.83	3.00	0.75	0.00
39	Quelle scarpe sono aperte	AP	4.67	5.00	0.52	1.17	1.00	0.41	4.00	4.00	0.82	0.00
40	Quelle sopracciglia sono arcuate	AP	4.50	5.00	0.84	1.33	1.00	0.82	4.50	4.50	0.55	0.00
41	Quelle spalle sono quadrate	AP	4.00	4.00	1.15	1.57	2.00	0.53	2.83	2.50	0.98	0.00
42	Quello schermo è accecante	AP	4.67	5.00	0.82	1.17	1.00	0.41	3.43	4.00	0.79	0.00
	Literal sentences											
1	Quella busta è rettangolare	L	4.00	4.50	1.26	1.17	1.00	0.41	4.86	5.00	0.38	0.00
2	Quella camicia è stirata	L	4.17	5.00	1.60	1.00	1.00	0.00	4.71	5.00	0.49	0.17
3	Quel righello è diritto	L	4.83	5.00	0.41	1.00	1.00	0.00	4.14	5.00	1.21	0.33
4	Quel cavo è lungo	L	5.00	5.00	0.00	1.00	1.00	0.00	4.86	5.00	0.38	0.43
5	Quel legno è duro	L	4.50	4.50	0.55	1.17	1.00	0.41	4.29	4.00	0.76	0.00
6	Quel fiume è ghiacciato	L	4.56	5.00	0.53	1.25	1.00	0.46	4.50	5.00	0.84	0.00
7	Quel castello è ottagonale	L	3.78	4.00	0.83	1.75	2.00	0.46	3.50	4.00	0.84	0.00
8	Quella freccia è appuntita	L	4.17	5.00	1.60	1.33	1.00	0.52	4.57	5.00	0.79	0.00
9	Quella patata è sbucciata	L	4.56	5.00	0.53	1.25	1.00	0.46	4.50	4.50	0.55	0.00
10	Quel monumento è piramidale	L	4.50	5.00	0.84	1.17	1.00	0.41	4.50	4.50	0.55	0.00
11	Quella foglia è verde	L	5.00	5.00	0.00	1.00	1.00	0.00	4.67	5.00	0.82	0.33
12	Quel gas è soffocante	L	4.67	5.00	0.52	1.33	1.00	0.52	4.43	4.00	0.53	0.00
13	Quel pianeta è deserto	L	4.67	5.00	0.52	1.17	1.00	0.41	4.00	4.00	1.10	0.00
14	Quella spatola è piatta	L	5.00	5.00	0.00	1.00	1.00	0.00	3.17	3.50	1.17	0.00
15	Quella cenere è inconsistente	L	2.89	3.00	1.27	2.00	2.00	0.76	3.17	3.50	1.47	0.00
16	Quel vetro è trasparente	L	5.00	5.00	0.00	1.00	1.00	0.00	4.83	5.00	0.41	0.00
17	Quel cartello è triangolare	L	4.17	5.00	1.60	1.17	1.00	0.41	4.83	5.00	0.41	0.00
18	Quell'orario è immutato	L	3.83	4.00	0.75	1.67	1.50	0.82	4.33	4.50	0.82	0.00
19	Quella valigia è piena	L	5.00	5.00	0.00	1.00	1.00	0.00	5.00	5.00	0.00	0.17
20	Quel foglio è bianco	L	4.83	5.00	0.41	1.00	1.00	0.00	4.14	5.00	1.46	0.50
21	Quel laser è indolore	L	3.89	4.00	0.78	1.63	1.50	0.74	3.67	4.00	0.82	0.00
22	Quel limone è giallo	L	4.78	5.00	0.44	1.13	1.00	0.35	5.00	5.00	0.00	0.22
23	Quella pillola è insapore	L	3.83	4.50	1.60	1.50	1.00	0.84	3.83	4.00	1.17	0.00
24	Quel piatto è rotondo	L	4.71	5.00	0.49	1.14	1.00	0.38	4.71	5.00	0.49	0.00
25	Quella lattina è cilindrica	L	4.17	5.00	1.60	1.17	1.00	0.41	4.86	5.00	0.38	0.00
26	Quell'organismo è microscopico	L	4.78	5.00	0.44	1.25	1.00	0.46	4.17	5.00	1.60	0.00
27	Quella vasca è ovale	L	4.50	4.50	0.55	1.17	1.00	0.41	3.50	3.50	1.05	0.00

28	Quel pomodoro è rosso	L	4.50	5.00	1.22	1.00	1.00	0.00	4.86	5.00	0.38	0.50
29	Quel tutù è rosa	L	5.00	5.00	0.00	1.00	1.00	0.00	4.17	4.00	0.75	0.17
30	Quel cassetto è vuoto	L	4.86	5.00	0.38	1.00	1.00	0.00	4.86	5.00	0.38	0.43
31	Quel peperone è crudo	L	4.67	5.00	0.52	1.17	1.00	0.41	4.67	5.00	0.52	0.00
32	Quell'orlo è rifatto	L	4.14	4.00	0.69	1.57	1.00	0.79	4.86	5.00	0.38	0.00
33	Quel cimitero è silenzioso	L	5.00	5.00	0.00	1.00	1.00	0.00	4.67	5.00	0.52	0.00
34	Quel marmo è liscio	L	4.57	5.00	0.53	1.14	1.00	0.38	4.86	5.00	0.38	0.00
35	Quella lente è concava	L	4.17	5.00	1.60	1.33	1.00	0.82	4.43	4.00	0.53	0.00
36	Quel cielo è blu	L	4.33	5.00	1.63	1.00	1.00	0.00	4.83	5.00	0.41	0.30
37	Quell'inchiostro è nero	L	4.17	5.00	1.60	1.17	1.00	0.41	4.67	5.00	0.52	0.50
38	Quella pentola è bollente	L	4.83	5.00	0.41	1.17	1.00	0.41	5.00	5.00	0.00	0.33
39	Quella finestra è aperta	L	5.00	5.00	0.00	1.00	1.00	0.00	4.86	5.00	0.38	0.43
40	Quella porta è arcuata	L	4.00	4.00	0.63	2.17	2.00	0.98	2.33	2.00	1.03	0.00
41	Quella piastrella è quadrata	L	5.00	5.00	0.00	1.00	1.00	0.00	4.67	5.00	0.52	0.00
42	Quella luce è accecante	L	4.33	5.00	1.21	1.33	1.00	0.82	5.00	5.00	0.00	0.20
	Anomalous sentences											
1	Quella noce è rettangolare	A	2.14	2.00	1.07	1.71	1.00	1.25	1.29	1.00	0.76	0.00
2	Quella giraffa è stirata	A	1.44	1.00	0.73	1.75	1.50	0.89	1.57	1.00	1.51	0.00
3	Quella camomilla è diritta	A	2.17	1.50	1.60	2.00	2.00	0.89	1.00	1.00	0.00	0.00
4	Quel mais è lungo	A	1.67	1.50	0.82	2.67	2.50	1.63	1.71	2.00	0.49	0.00
5	Quella fattoria è dura	A	2.50	2.50	0.55	2.83	3.00	1.17	1.29	1.00	0.49	0.00
6	Quel fuoco è ghiacciato	A	1.67	2.00	0.52	2.83	3.00	1.72	1.00	1.00	0.00	0.00
7	Quell'arancio è ottagonale	A	1.50	1.00	0.84	2.33	2.00	1.03	1.17	1.00	0.41	0.00
8	Quella crema è appuntita	A	1.44	1.00	0.73	2.13	2.00	1.13	1.00	1.00	0.00	0.00
9	Quel caffè è sbucciato	A	1.50	1.00	1.22	2.00	1.50	1.26	1.00	1.00	0.00	0.00
10	Quel cerchietto è piramidale	A	1.00	1.00	0.00	1.50	1.00	0.84	1.00	1.00	0.00	0.00
11	Quel pulcino è verde	A	1.33	1.00	0.52	1.50	1.00	0.84	1.17	1.00	0.41	0.00
12	Quella sedia è soffocante	A	2.22	2.00	0.83	2.38	2.00	1.19	1.00	1.00	0.00	0.00
13	Quell'albicocca è deserta	A	1.17	1.00	0.41	1.33	1.00	0.52	1.00	1.00	0.00	0.00
14	Quella mucca è piatta	A	2.00	1.50	1.55	1.50	1.00	0.84	1.00	1.00	0.00	0.00
15	Quella collina è inconsistente	A	2.00	2.00	0.63	2.83	3.00	0.41	1.17	1.00	0.41	0.00
16	Quello scoglio è trasparente	A	1.50	1.50	0.55	1.83	1.50	1.17	1.17	1.00	0.41	0.00
17	Quella pallina è triangolare	A	1.43	1.00	1.13	1.71	1.00	1.25	1.00	1.00	0.00	0.00
18	Quell'unghia è immutata	A	2.56	2.00	1.01	2.50	2.50	0.93	1.00	1.00	0.00	0.00
19	Quel torsolo è pieno	A	1.67	1.50	0.82	2.83	3.00	0.75	1.50	1.50	0.55	0.00
20	Quel sangue è bianco	A	2.22	2.00	1.48	2.38	2.00	1.51	1.29	1.00	0.49	0.00
21	Quella tapparella è indolore	A	1.17	1.00	0.41	2.00	1.50	1.55	1.00	1.00	0.00	0.00
22	Quel cane è giallo	A	3.00	2.50	1.26	2.33	2.50	0.82	1.67	1.50	0.82	0.17
23	Quella spina è insapore	A	1.86	2.00	0.38	1.43	1.00	0.53	1.50	1.50	0.55	0.00

24	Quel passaporto è rotondo	A	1.50	1.50	0.55	2.50	2.00	1.64	1.00	1.00	0.00	0.00
25	Quel vocabolario è cilindrico	A	1.86	2.00	0.69	1.86	2.00	0.90	1.00	1.00	0.00	0.00
26	Quel vulcano è microscopico	A	2.67	3.00	1.03	1.83	2.00	0.41	1.50	1.50	0.55	0.00
27	Quella spada è ovale	A	1.67	2.00	0.52	2.17	2.00	1.17	1.00	1.00	0.00	0.00
28	Quel pistacchio è rosso	A	2.57	2.00	1.13	1.43	1.00	0.53	1.57	1.00	1.13	0.00
29	Quegli spinaci sono rosa	A	2.50	2.00	1.64	2.50	2.50	1.05	1.00	1.00	0.00	0.00
30	Quell'ago è vuoto	A	2.00	1.50	1.26	2.33	1.50	1.75	1.29	1.00	0.76	0.00
31	Quella sciarpa è cruda	A	1.50	1.00	1.22	1.33	1.00	0.82	1.17	1.00	0.41	0.00
32	Quell'elefante è rifatto	A	2.17	1.50	1.60	2.33	2.00	1.51	1.14	1.00	0.38	0.00
33	Quella carota è silenziosa	A	1.17	1.00	0.41	1.33	1.00	0.52	1.00	1.00	0.00	0.00
34	Quel ristorante è liscio	A	1.17	1.00	0.41	2.17	1.00	1.83	1.29	1.00	0.49	0.00
35	Quella musica è concava	A	1.78	2.00	0.67	1.75	1.50	0.89	1.14	1.00	0.38	0.00
36	Quel girasole è blu	A	2.43	2.00	1.51	1.57	1.00	0.79	1.17	1.00	0.41	0.00
37	Quel latte è nero	A	2.22	2.00	0.67	2.88	3.00	1.25	1.33	1.00	0.52	0.00
38	Quel gelato è bollente	A	1.50	1.50	0.55	1.83	1.50	1.17	1.00	1.00	0.00	0.00
39	Quel tagliere è aperto	A	1.33	1.00	0.52	2.83	2.50	1.47	1.00	1.00	0.00	0.00
40	Quel giubbotto è arcuato	A	2.14	2.00	0.90	2.00	2.00	1.15	1.00	1.00	0.00	0.00
41	Quella mongolfiera è quadrata	A	1.67	1.50	0.82	1.50	1.50	0.55	1.00	1.00	0.00	0.00
42	Quella frittata è accecante	A	1.43	1.00	0.53	1.57	1.00	0.79	1.14	1.00	0.38	0.00

Acknowledgments

First of all, I am profoundly indebted to Pier Marco Bertinetto for his continuous and careful supervision, and for providing me with insightful suggestions on the entire work. I would like also to thank him as a head of the “Laboratorio di Linguistica” that represented an ideal scientific environment for developing my research.

I am immensely grateful to Valentina Bambini for determining my entering the field of pragmatics, for her guidance, and her precious advice during all stages of my work. Without her supervision, my research in experimental pragmatics would have not been possible.

I owe a profound gratitude to Marco Tettamanti for supporting me and for essentially contributing to my scientific development in the neuroimaging research. I am greatly indebted to him for his stimulating guidance, his trust, and his openness.

Thanks Valentina and Marco for your extraordinary expertise, your personal effort and your patience.

To Lewis Bott I owe many thanks for introducing me to the methodology of SAT, and for welcoming me to the School of Psychology, Cardiff University, for a research period. I wish to thank him also for taking part in my doctoral committee.

For their contribution to the work presented in the second part of the present thesis, thanks are also due to Andrea Moro and Petra Schumacher.

Many colleagues (and friends) have contributed to my research in various ways. In particular, I wish to express my gratitude to my co-author, Matilde Vaghi, for her essential collaboration in the cognitive neuroscience work presented in the first part the present thesis. My thanks also to Donatella Resta for challenging discussions and for offering me her unreserved support in many occasions.

My thanks are owed to the Irene Ricci and Chiara Bertini for their technical assistance, and for supporting me every time I had a problem.

I thank all the participants of either the behavioral or the neuroimaging studies.