University School for Advanced Studies IUSS Pavia

Vita-Salute San Raffaele University Milan

PH.D. PROGRAM IN COGNITIVE NEUROSCIENCE AND PHILOSOPHY OF MIND

XXXVI Cycle

Doctoral Dissertation

A NEUROPRAGMATIC ACCOUNT OF MULTIMODAL EXPERIENCE

FROM LITERAL TO FIGURATIVE LANGUAGE PROCESSING IN TYPICAL AND ATYPICAL POPULATIONS

Candidate: Federico Frau

Supervisor: Prof. Valentina Bambini

Academic year 2022-2023

«How might the process involved in comprehension and
spontaneous expression of the word "bell" be explained? […]
The acoustic message must stimulate the ***memory images*** of a bell»

(Carl Wernicke, *Recent works on aphasia*, 1885)

This image was generated using DALL-E, with the following prompt:
"A portrait of the study of mental imagery from the perspective of neuroscience in Magritte's style"
(generated on January 27, 2024)

ABSTRACT

Current up-to-date theoretical models of semantic memory claim that word meaning is organized in multiple modalities, including "amodal" information (e.g., lexical, orthographic, and grammatical features) and properties derived from non-linguistic, modality-specific experience, capitalizing on perceptual/motor simulations, bodily sensations, and mental imagery. The extent to which such multimodal organization of semantic representations is functionally relevant to language processing, including pragmatic inferencing, is still debated. In my thesis, I will present the results of five studies based on the application of psycho/neurolinguistic experimental approaches (i.e., behavioral, computational, and neurophysiological methods) to test the involvement of experience-related information during different linguistic tasks (including figurative meaning processing) in healthy and atypical adult populations.

In particular, *Study 1* investigated the effects of sensory-motor and bodily experience on the processing of words (e.g., verbs) in healthy individuals. The results showed that information associated with such experience influences lexical and semantic, but not grammatical, processes, opening up considerations of cross-task variations of modality-related effects.

Study 2 measured the involvement of motor-related semantic information in literal and figurative processing in individuals with motor neuron diseases (Amyotrophic Lateral Sclerosis and Hereditary Spastic Paraplegia positive for *SPG4* gene mutation). Difficulties in processing action meaning were observed in 20-50% of the samples at literal level, but did not emerge at the figurative level. These findings challenge the predictions – derived from the Embodied Cognition Framework – of a widespread disruption of action language grounding in motor impairment conditions and emphasize the need of an individual-based approach.

Study 3 and *Study 4* addressed the "linguistic roots" of concretism in schizophrenia, namely the extent to which semantic properties of linguistic stimuli might underlie impairment in figurative language processing and pragmatics in this population. In *Study 3*, a semi-automated approach was applied to investigate semantic perceptual-based properties of words (e.g., concreteness) in patients' verbal explanations of figurative expressions (e.g., idioms, metaphors, and proverbs). This approach showed that word concreteness not only characterized the verbal explanations provided by patients but was also indicative of patients' ability to

provide correct interpretations of figurative expressions. Going from concreteness to imagery, in *Study 4*, we tested the activation of sensory-motor properties of mental images related to the literal meaning of metaphors, based on the hypothesis that such images might be intrusive and an obstacle to pragmatic processing in individuals with schizophrenia. This experiment showed higher activation of visual properties of metaphor-evoked mental images in patients compared to control individuals, with such properties being negatively associated with the patients' ability to understand metaphoric meanings.

Finally, *Study 5* addressed the involvement of mental imagery during metaphor comprehension using neurophysiological methods (i.e., EEG recording) in healthy individuals, testing whether images evoked during metaphoric sentence comprehension would differ from images evoked by literal description or in a pure (i.e., non-linguistic) imagistic task. In particular, the experiment focused on the modulation of the neural response during the perceptual processing of a visual stimulus following different types of tasks, involving picture, imagistic, and linguistic (both literal and metaphoric) stimuli: the modulation of the P300/N200 ERP responses showed that the mental representations activated by metaphoric sentences were different from those activated by literal sentences, indicating that were richer and more open. Moreover, metaphor-evoked images were also different from pure imagery, as characterized by reduced amplitude for Late Positive Potentials and N400 ERP responses, indicating that images supported by propositional representations required less integration efforts. These results confirm that metaphor processes cannot be reduced to fully abstract operations but need to account for imagistic representations.

Overall, the results of this dissertation support theoretical models predicting the effects of multimodal semantic properties on language processing and extend the description of such effects to the pragmatic level, where the involvement of non-propositional operations has always been neglected by classic theoretical models. Moreover, the studies presented showed that the involvement of multimodal semantic properties during pragmatic processing might be intrusive in some atypical populations (e.g., schizophrenia) but not in others (e.g., motor conditions), thus indicating that multimodal effects are not only task-sensitive but also highly modulated by differences related to cognitive and clinical characteristics.

TABLE OF CONTENTS

LIST OF TABLES

STUDY 3:

STUDY 4:

STUDY 5:

LIST OF FIGURES

STUDY 1:

STUDY 2:

STUDY 3:

STUDY 4:

STUDY 5:

ACKNOWLEDGEMENTS

Many people deserve to be acknowledged for their contribution to the work presented in this dissertation. First, I would love to thank my PhD supervisor, Prof. Valentina Bambini, who inspired and supported my scientific growth, from being a student with a fascination for neurolinguistics to becoming a young researcher. I'm also thankful to friends and colleagues in the Neurolinguistics and Experimental Pragmatics (NEP) Lab at IUSS who contributed to the studies included in this thesis, in particular: Luca Bischetti, Maddalena Bressler, Paolo Canal, Veronica Mangiaterra, and Chiara Pompei.

I'd like also to acknowledge the contribution of all the people from other institutions (in Italy and abroad) who collaborated with us through vibrant and inspiring scientific exchanges. In particular, Prof. Penny M. Pexman and Emiko J. Muraki (University of Calgary, Canada), who contributed to Study 1, Prof. Mauro Ceroni and Dr. Luca Diamanti (IRCCS Mondino Foundation, Pavia, Italy), who contributed to Study 2, and Prof. Marta Bosia and Dr. Giulia Agostoni (Vita-Salute San Raffaele University, Milan, Italy), who contributed to Study 3 and Study 4.

Through the years I had the opportunity to present my work in several scientific contexts, from workshops to conferences: some of the most enriching discussions happened during my research period at the University College London (UK), where I had the honor to share my on-going studies with Prof. Robyn Carston, who always gave me inspiring and encouraging feedback. I could not feel any luckier for all those moments.

Finally, I'm grateful to my friends and loved ones, who helped me in recruiting participants and never said no when I had to test materials or pilot new experiments. I hope your investment in time, eventually, paid off!

My deepest gratitude goes to all the people who took part in the studies included in this dissertation, especially patients and their families. This work would not exist without your commitment.

INTRODUCTION

1. What is "meaning"?

Meaning is a central component of human language capacity, yet it is still an object of theoretical and empirical debate in several disciplines, from Linguistics to Philosophy, passing through Psychology and Neuroscience (Damasio, 1989; Grice, 1957; Putnam, 1975). But what is the content of meaning? And what do we *mean* when we diverge from literal uses of language, as in the case of metaphors?

In cooperative communication, whenever we utter a word or a sentence, be it used in literal or metaphoric ways, we *mean* something, that is we convey information on concepts (Putnam, 1975), mental instances carrying packages of knowledge on categories of the world (Murphy, 2002). The problem of the content of words meaning, namely their semantic representation (Vigliocco & Vinson, 2012), is then inherently linked to the issue of what concepts are made of and whether their format is "modality-specific" or "amodal", namely whether it is either linked or not to experience acquired via sensorial and motor activity (Michel, 2021).

The traditional view in Cognitive Science argues that conceptual knowledge is essentially symbolic and detached from a specific modality (Fodor, 1975). This view stems from a purely symbolic model of cognition and its main supporting arguments include: i) the existence of abstract concepts (e.g., *DEMOCRACY* and *JUSTICE*), which cannot be captured by representations acquired via sensory-motor experience; ii) the evidence of a systematic arbitrariness between the phonetic realization of words and their semantic representation; and iii) the idea that the abstract format of conceptual representations allows for a flexible and productive re-combination of symbols (as for syntax), which can lead to new connections among thoughts (for a discussion, see Dove, 2014). Elaborating on these arguments, purely amodal accounts of conceptual knowledge predict that the meaning of a word is understood when its symbolic lexical representation is retrieved from the mental lexicon

(Levelt, 1989) and via the activation of other words in the same semantic network (Landauer & Dumais, 1997).

Conversely, modality-specific accounts (or simply "modal" accounts) argue that concepts are grounded (i.e., embodied) in sensory-motor and emotional systems, which means that conceptual knowledge involves "re-enactments" or "simulations" of states in the sensory-motor and limbic areas of the brain (Barsalou, 2008; Glenberg & Gallese, 2012; Pulvermüller, 2012). Traditionally, the most relevant argument that distinguishes embodied accounts from the idea of a purely symbolic model of cognition concerns how mental representations are connected with external referents: while amodal accounts from the cognitivist tradition argue that referents and concepts are associated via a transformation of sensorial signals into symbolic representations (see Pylyshyn, 1986: 147–191), embodied views posits that mental representations are constructed by exploiting prior experience developed from the interaction with the environment, so that conceptual processing can be possible by re-enacting (or "simulating") such experience in modality-specific brain areas (Barsalou, 1999; Barsalou et al., 2003). Thus, current modal views support the hypothesis of an experience-based organization of semantic knowledge by relying on several converging streams of empirical evidence, in particular: i) behavioral data showing that sentences denoting arm motion in one direction (e.g., *Close the drawer*) are processed faster when coupled with a congruent movement executed with the same effector (Glenberg & Kaschak, 2002); ii) neurofunctional data reporting increased activations in sensory-motor areas of the brain during listening of words and sentences denoting motor activity (Hauk et al., 2004; Hauk & Pulvermüller, 2004; Rueschemeyer et al., 2010); and iii) neuropsychological evidence showing that motor or sensorial deprivation is associated with increased difficulties in processing concepts related to action or perception (Boulenger et al., 2008; Grossman et al., 2008; for a summary of empirical evidence, see Meteyard et al., 2012). It is also relevant to mention that more radical modal views argue that not only meaning but also the other linguistic levels, including syntax, can be thought to be "embodied" in perception-motor systems (for a discussion on the solidity of these proposals, see: Moro, 2014; Tettamanti & Moro, 2012). Within the modal framework, it is then quite relevant to distinguish between "strong" embodied accounts, supporting the hypothesis that cognition is entirely formatted in sensory-motor experience to the point that semantic processing is entirely based on simulations (Glenberg, 2015), and "weak" embodied accounts, which admit the coexistence of modal and amodal features of conceptual organization (for a comparison among different accounts, see: Meteyard et al., 2012; Muraki, Speed, et al., 2023). An example of such "weak" (or "hybrid") accounts is the highly influential "spokes-and-hub" model of semantic memory (Ralph et al., 2017): according to this model, semantic processing features the involvement of modality-specific "spokes", distributed in sensory-motor areas of the cortex, and a supra-modal "hub", which is anatomically served by modality-nonspecific brain regions (i.e., anterior temporal poles), responsible for controlled integration of modality-specific information into a coherent representation.

It goes without saying that modal accounts have received several criticisms over the years (Mahon, 2015), the most relevant ones arguing that sensory-motor activations in the brain are merely ancillary or epiphenomenal to an intrinsically abstract process of conceptual understanding (Mahon & Caramazza, 2008) or questioning the robustness of empirical evidence coming from neuropsychological studies (Mahon & Hickok, 2016). In recent years, experimental studies have tried to move beyond the simplistic opposition between modal and amodal theories. Accumulating experimental evidence has strengthened the idea that word meaning is organized in multiple dimensions (Dove, 2023), characterized not only by an abstract and symbolic format but also by representations coming from a plethora of "experiences", including sensory-motor simulations, bodily sensations, emotional processing, and social interactions (Borghi & Binkofski, 2014; Borghi & Cimatti, 2010; Connell, 2019; Pexman, 2020). In other words, rather than just *modal* or *amodal*, the general view on the content of semantic representations is now becoming *multimodal*.

2. When meaning is not "transparent": metaphors and pragmatic uses of language

So far, I have addressed only the case of meaning conveyed by single words or by utterances used in their literal sense. But what if we convey meaning that is not explicitly stated by the literal interpretation of an utterance, as when we say *Those dancers are butterflies* or *One swallow does not make a summer*?

This is the case of metaphors and other non-literal uses of language, where the speaker's communicative intention departs from the logical interpretation of the utterance and the actual conveyed meaning requires extra inferential operations to be grasped by the interlocutor (Bara, 2010; Domaneschi & Bambini, 2020). In the last 20 years, the understanding of the cognitive and neural mechanisms underlying pragmatic inferences has considerably expanded (Bischetti et al., 2024), yet the experimental attempts to understand the contribution of experience-based information are still limited (Cuccio, 2022). In looking at this literature, I will briefly focus on the notion of "metaphoric meaning" and the research that investigated its content and format from both theoretical and experimental perspectives.

Traditionally, metaphoric utterances are thought to convey a literal meaning that is closely related to the conventional sense of the words uttered, yet it is categorically false with respect to the speaker's intended meaning (Grice, 1975). To paraphrase Grice's view, metaphors are blatantly false utterances that require inference on the speaker's communicative intention (i.e., an "implicature") to go beyond the apparent logical incongruity and reach the true intended meaning. Elaborating on this intuition, post-Gricean models of pragmatics – the most prominent one being "Relevance Theory" (Sperber & Wilson, 1995; Wilson & Sperber, 2004) – have further expanded the description of the inferential machinery supporting our ability to understand non-literal uses of language. In particular, relevant theoretical accounts of metaphor processing assume that whenever an encoded proposition (e.g., the metaphor *Those dancers are butterflies*) fails to fully convey the proper intended meaning, progressive adjustments of the concepts denoted by the utterance take place, motivated and modulated by contextual information and expectations on the speaker's will to convey something communicatively "relevant" (Sperber & Wilson, 2008; Wilson & Carston, 2007). Conceptual adjustments occur in parallel with contextual integration in a top-down/bottom-up fashion until the expectation of relevance is met and the speaker's intended meaning is recovered. What both Gricean and post-Gricean pragmatic accounts of metaphor processing have in common is the assumption that the linguistic (and conceptual) system where pragmatic implicatures take place is essentially symbolic. Stated differently, inferential processes reflect logical operations and heuristic computations among symbols (Carston, 2010a), occurring within the boundaries of a purely amodal representation at lexical and conceptual levels, which does not account for any involvement of experience-based representations (e.g., information grounded in sensory-motor or emotional systems).

This fundamental abstract configuration of pragmatic processes is not shared by other non-Gricean accounts, such as the "Conceptual Metaphor Theory" that emerged in the field of Cognitive Linguistics (Lakoff & Johnson, 1980). Differently from inference-based views, metaphors are described by Cognitive Linguistics in terms of a conceptual mapping between a concrete source concept and an abstract target concept (e.g., the *UNDERSTANDING-IS-SEEING* conceptual metaphor). Crucial to this account is the idea that the mechanisms underlying conceptual mappings are essentially embodied, in the sense that they are derived through the sensation of the body in action, as well as perceptual experience and body-environment interactions (Gibbs, 2005).

Beyond the dichotomy between more abstract and experience-oriented theories of metaphor processing, other proposals positioned themselves in a more in-between territory, where the definition of metaphoric meaning (and, ultimately, pragmatic processing) can be assimilated to "hybrid" or "weak" modal accounts. It is the case of the "Dual Coding Theory" (Paivio, 1979), originally proposed to explain the difference between concrete and abstract concepts, then applied also to describe how metaphoric meaning is understood (Paivio & Walsh, 1993). The most relevant aspect of this theory is the idea that metaphor processing entails two parallel routes, both relevant to grasping metaphoric meaning: a linguistic and an "imagistic" route. In particular, the concepts lexically encoded in a metaphoric utterance are assumed to arouse mental images that operate in parallel with linguistic operations, facilitating the retrieval of conceptual information relevant to metaphor understanding from long-term memory. The idea that mental imagery – i.e., the ability to arouse perceptual and motor representations in the absence of direct sensory-motor stimulation (Kosslyn et al., 2001) – is central for metaphors is not new, yet stems from an earlier philosophical debate: in particular, seminal proposals had previously put forward the intuition that the meaning of a metaphor is not grasped via linguistic operations on the literal meaning of the utterance, but rather via "images" that allow to "see" similarities among entities (Davidson, 1978; Lepore & Stone, 2010). Besides philosophical positions, the involvement of mental imagery in metaphor processing is also compatible with modality-oriented accounts already mentioned (e.g., "Conceptual Metaphor Theory"), where imagination is seen to play a relevant role in the process of re-enacting the sense of the metaphor through embodied simulations (Gibbs & Matlock, 2008).

Up to now, there is a body of empirical evidence, both at the behavioral and neurofunction levels, that seems to support modality-oriented theories, namely the idea that also non-literal uses of language capitalize on sensory-motor and bodily experience, although the involvement of perceptual and motor simulations during pragmatic processing seems to vary depending on several variables, including the conventionality of figurative expressions (Cuccio, 2022). To quickly summarize the main findings, a number of studies documented the activation of brain areas involved with motor planning during the listening of metaphors denoting figurative actions, such as *The jury grasped the concept* (Desai et al., 2011; Romero Lauro et al., 2013). Moreover, at the behavioral level, studies reported that individuals suffering from motor conditions (e.g., Parkinson's disease) might exhibit difficulties in processing metaphoric action sentences compared to non-action-related metaphoric sentences (Fernandino et al., 2013). Such pieces of evidence, however, are often not easily replicated (see, for instance, Humphries et al., 2019 for neuropsychological evidence) and are still nowadays a subject of heated discussion on whether they reflect a proper functional role of sensory-motor simulations: as discussed by Casasanto and

Gijssels (2015), a significant number of studies investigating the involvement of sensory-motor experience during the processing of action meaning used in figurative sense reported brain activations occurring in modality-nonspecific cortical regions or failed in detecting a "somatotopic" topographic activation of sensory-motor cortex, as observed for literal action language (Hauk et al., 2004; Tettamanti et al., 2005).

In a nutshell, it is an obvious consequence that current debate – at both theoretical and experimental levels – demands a rethinking of the content (and format) of metaphoric meaning not just in the *amodal* vs. *modal* opposition, as most of the literature has implicitly or explicitly done so far, but rather within a broader, *multimodal* perspective of meaning.

3. Aims and thesis roadmap

The present dissertation summarizes the results of experimental work animated by the challenge of providing empirical support to multimodal accounts of language meaning, with a focus on the representation of utterances conveying non-literal meanings, such as metaphors.

Following this motivation, this work aimed to investigate the involvement of different dimensions contributing to semantic representations, by testing the extent to which not only motor and perceptual experience but also bodily sensations and mental imagery become available during language processing along the continuum from literal to figurative expressions. In pursuing this general objective, great emphasis has been placed on several aspects. First, this work draws attention to the fundamental role of data from atypical populations as a testing ground of multimodal theories, by exploiting prior evidence of disruption of modality-specific meaning representation in motor disorders (Birba et al., 2017), as well as extensively documented impairment in pragmatic uses of language in both neurological and psychiatric conditions (Bambini, Arcara, Bechi, et al., 2016; Bambini, Arcara, Martinelli, et al., 2016). Second, relevant space has been given to the importance of situational context throughout the entire work, to test how the weight of different modalities changed during language processing as a function of tasks (Connell

& Lynott, 2013). Third, the empirical investigation could not neglect the role of individual differences (Ibáñez et al., 2023; Muraki & Pexman, 2021; Yap, Balota, et al., 2012), encompassing neurocognitive and socio-cognitive abilities but also personal variability in lexical-semantic skills and the use of mental imagery.

This work is articulated in five studies, investigating multimodal accounts in single-word paradigms up to include literal and non-literal expressions:

- **Study 1** investigated the involvement of sensory-motor and bodily experience (operationalized as *relative embodiment*) along the language-memory interface, focusing on Italian verb processing tested using three different visual word recognition tasks. The study advances our knowledge of bodily experience effects in language not only by showing its flexible involvement across tasks but also by documenting language-specific factors influencing its role in verb processing.
- **Study 2** aimed to test the hypothesis of the motor grounding disruption (and its potential somatotopic distribution) in two groups of participants with Motor Neuron Diseases, namely Amyotrophic Lateral Sclerosis and Hereditary Spastic Paraplegia with SPG4 gene mutation, the latter characterized by selective lower-limb impairment. By testing the processing of lower-limb and upper-limb-related action verbs (e.g., *to run* vs. *to grasp*), presented either in isolation or in literal and metaphoric sentences, this study showed that somatotopically distributed action language deficits are present, yet characterized by great individual and cross-diagnostic variation and confined to literal uses of language.
- **Study 3** approached the case of *concretism* in schizophrenia, namely the tendency to provide literal and concrete interpretations of figurative expressions due to impaired abstract thinking. In particular, this study focused on the application of semi-automated approaches to measure the occurrence of words with high concreteness levels in participants' verbal explanations of idioms, metaphors, and proverbs, to test whether semantic content could underlie difficulties in figurative language understanding. The results showed that participants with schizophrenia

tended to produce verbal explanations of figurative meanings characterized by higher concrete content, which was also indicative of their pragmatic impairment.

- **Study 4** expanded the exploration of the relationship between multimodal experience and concretism in schizophrenia, by testing whether difficulties in understanding metaphoric meanings could be related to higher activation of visual mental images in patients. For this study, a metaphor priming experiment was used, to test the activation of multimodal properties of the literal meaning of the sentence, alongside amodal semantic properties. The results showed that individuals with schizophrenia were characterized by greater visual priming effects compared to controls, indicating higher activation of visual images of the literal, concrete meaning of metaphoric sentences. Importantly, such activations were also correlated with patients' performance in understanding metaphors, suggesting that visual mental images might interfere with pragmatic processing in schizophrenia.
- **Study 5** explored the neural correlates (via EEG recording) of the involvement of mental imagery during metaphor processing. In particular, this study employed a novel paradigm, aiming at testing whether mental images evoked by sentence processing (be them metaphoric or a simple literal description) influenced the perceptual processing of a subsequent matching or mismatching visual stimulus. Relevant to this experiment was also the comparison between linguistic-related imagery (evoked by literal and metaphoric sentences) and pure imagistic processes, namely mental images evoked in a controlled and deliberate way. Differences in the P300 ERP response for matching visual stimuli between metaphoric and literal sentences indicated that metaphor meaning, compared to literal counterparts, aroused a richer perceptual representation, which was also more malleable and open, as shown by reduced N200 response to mismatching pictures. Interestingly, greater N400 and Late Positive Potentials (LPP) for pure imagery compared to metaphors suggested greater integration costs when mental images are aroused in the absence of propositional representations, indicating that during metaphor

processing language too plays a great role, possibly in constraining the content of mental images.

Across these five studies, the experimental work incorporated data collected using not only behavioral but also neurophysiological and computational techniques. The methodological and theoretical background is anchored in Neuropragmatics (Bambini & Frau, 2022; Bischetti et al., 2024) and, more generally, in Experimental Pragmatics (Noveck, 2018), but was not limited to these fields: theories, views, and methodologies from other different, albeit complementary, disciplines – primarily Cognitive Psychology, but also Neuroscience and, ultimately, Medicine – inspired a significant part of this work. The cross-disciplinary contamination with other research fields is what allowed this dissertation to be fully *neuropragmatic* in its spirit.

STUDY ONE

UNDERSTANDING WITH THE BODY? TESTING THE ROLE OF VERB RELATIVE EMBODIMENT ACROSS TASKS AT THE INTERFACE OF LANGUAGE AND MEMORY[1]

Abstract

Multiple representation accounts of conceptual knowledge argue that information associated with sensory-motor experience contributes to word processing in addition to pure linguistic information. A number of issues, however, remain under-investigated, including the extent to which these dimensions affect verb processing (rather than nouns), especially in languages other than English, and their role across different tasks along the language-memory *continuum*. Here, we collected ratings for a verb-specific dimension linked to bodily experience (*relative embodiment*, RE) for 647 Italian verbs and we tested its effects in three tasks differently modulating semantic activation and memory processes (i.e., lexical decision, grammatical decision, and memory recognition). Our results showed reliable influences of RE during lexical decision and memory recognition, but not in grammatical decision, possibly due to Italian inflectional richness. The cross-task comparison showed that RE effects were substantially higher in memory recognition compared to lexical decision, indicating that semantic and episodic processes interact at the interface of language and memory. Overall, results support the flexible and context-dependent role of sensory-motor and bodily-related experience during verb processing, pointing also to language-specific factors and implications for the organization of declarative memory.

[1] This chapter is a manuscript currently submitted and under review as "Frau, F., Bischetti, L., Campidelli, L., Tonini, E., Muraki, E.J., Pexman, P.M., & Bambini, V., *Understanding with the body? Testing the role of verb relative embodiment across tasks at the interface of language and memory*" to the Journal of Memory and Language.

1.1. Introduction

Through decades of debate within cognitive science, several attempts have been made to describe the nature of conceptual knowledge along a spectrum of proposals (for an updated summary, see Dove, 2023; Muraki et al., 2023). These range from purely amodal accounts, arguing that conceptual knowledge is essentially symbolic and detached from any specific modality (Fodor, 1975), to modal (or modality-specific) accounts put forward in the Embodied Cognition Framework, claiming that concepts are grounded in sensory-motor systems and that semantic processing corresponds to "re-enactments" or "simulations" of states involving the sensory-motor areas of the brain (Barsalou et al., 2003; Glenberg & Gallese, 2012).

In recent years, multiple representation theories have tried to move beyond the traditional opposition between modal and amodal accounts, based on the idea that semantic knowledge is organized along different dimensions (Pexman, 2020), from pure linguistic information to experience-based dimensions (e.g., sensory-motor, emotional, and social experience). The multidimensional nature of conceptual knowledge is primarily supported by empirical evidence reporting that lexical access is generally facilitated when a target word is associated with a richer semantic representation, involving not only higher involvement of sensorial and motor-related experience (Siakaluk et al., 2008) but also higher imageability (i.e., how easily a word can arouse mental images, see Yap et al., 2012), higher emotional experience (Siakaluk et al., 2016), and greater social relevance (Diveica et al., 2022, 2023).

Going further, multiple representation theories of conceptual knowledge predict that meaning access works in a "trade-off" fashion, i.e., it involves specific modalities instead of others based on different task demands and contexts, as well as on individual differences (Connell, 2019). First, studies on noun processing showed that the activation of experience-based information varies as a function of task-specific demands (Newcombe et al., 2012; Tousignant & Pexman, 2012; Yap, Pexman, et al., 2012). In particular, a number of behavioral experiments reported significant, yet limited activation of sensory-motor information during tasks that require a shallower processing, such as lexical decisions (Yap et al., 2015; Yap, Pexman, et al., 2012), compared to tasks that involve deeper semantic processing, as in semantic decisions (Pexman et al., 2017). These task-related modulations of experience-based dimensions are generally explained by capitalizing on the interactive activation and competition (IAC) model (McClelland & Rumelhart, 1981), which is used to predict different patterns of top-down/bottom-up activations of orthographic, lexical, and semantic representations based on specific tasks (Balota, 1990). On the one hand, lexical decisions seem to rely more on lexical and orthographic properties of the stimuli, such as form familiarity, while semantic properties – including experience-based dimensions – affect performance only via feedback activations from semantic to orthographic representations (Pexman et al., 2002); on the other hand, semantic decisions require a complete semantic coding and are expected to involve higher activation of semantic-level proprieties compared to simple lexical decisions (Pexman et al., 2017). Second, studies on noun processing also documented a flexible interaction between experience-based dimensions and individual difference variables, such as vocabulary and mental imagery skills: in particular, individuals with higher vocabulary knowledge seem to capitalize more efficiently on the sensory-motor properties of the items in tasks that require deeper semantic processing (Pexman & Yap, 2018), while showing reduced sensitivity to word properties (including semantic ones) in tasks that require shallower processing (Yap, Balota, et al., 2012). Conversely, individual differences in motor imagery skills interact with sensory-motor properties of the items only in very specific ways (Muraki, Dahm, et al., 2023; Muraki & Pexman, 2021).

However, the majority of the studies documenting effects related to task demands and individual differences focused on noun processing, while similar effects are less documented for verb processing. There is initial evidence that the activation of semantic properties affects verb processing too, with task and context modulations. For instance, Sidhu et al. (2014) analyzed a set of 687 English verbs belonging to different semantic classes (e.g., action, perception, cognitive processes, communication, social interactions, physiological states, etc.) and observed that verb processing was influenced by a novel semantic dimension, which they defined as *relative embodiment*

(RE). RE expresses the experience related to "owning" and "sensing" the body (Borghi & Cimatti, 2010), including any kind of sensory-motor and bodily information deriving from voluntary actions, passive movements, proprioceptive experience, internal states, as well as from the interaction between the body and the environment. Interestingly, this study showed that during lexical decisions, RE played a greater role in responses than did verb imageability and was also a significant predictor of reaction times and accuracy in syntactic categorizations (i.e., decisions concerning grammatical categories of words), even when other lexical-semantic variables were controlled for, thus suggesting that bodily-related experience is particularly relevant for verb processing.

Another open issue concerns the extent to which the effect of experience-based dimensions spreads from lexical-semantic processing to other facets of language at the interface with memory skills (e.g., in word memory recognition). For instance, Sidhu and Pexman (2016) investigated the effect of the RE dimension on recognition memory for verbs, reporting that verbs associated with higher bodily experience were remembered more easily and more accurately compared to verbs less associated with bodily experience. Similarly, Muraki et al. (2022) reported that verbs characterized by higher RE ratings were easier to remember compared to various classes of abstract verbs (i.e., mental or emotional abstract verbs). This type of evidence seems to be in contrast with traditional theoretical views on declarative memory, which assume that semantic and episodic processes behave as distinct, segregated memory components (Tulving, 1972). In fact, these sub-domains interact (for a recent overview, see De Brigard et al., 2022) and, as sparse data from verb processing seem to suggest, the boundaries between semantic and episodic domains become more permeable in tasks at the interface between language and memory.

Overall, across the last ten years a significant number of studies provided support to multiple representation accounts of semantic processing, showing the role of several dimensions and additional situational factors in word processing. However, only a small handful of studies addressed the role of task-dependent semantic modulation in verb processing, including the interaction between semantic and episodic processes at the language-memory interface. Importantly, the majority of empirical studies documented task-related influences on English verb processing only, while little evidence is available on the generalizability of such effects to other languages. This is a significant limitation for multiple representation accounts, given existing evidence that documents cross-linguistic variation of sensory-motor effects on language processing (Gianelli et al., 2017; Kim et al., 2021).

1.1.1. The present study

The overarching aim of this study was to investigate a dimension related to bodily experience – namely RE – in a set of Italian verbs. In particular, we tested the effects of RE on verb processing across several tasks with different semantic loading along the language-memory interface.

To achieve our objective, we first acquired RE ratings for a large and heterogeneous set of Italian verbs, consistently with Sidhu et al. (2014), to test whether this experience-based dimension is relevant for Italian verbs. Then, we used the acquired RE ratings to test the role of this dimension on three visual word recognition paradigms with different levels of semantic demands, including lexical decision, grammatical decision (often referred to as *syntactic classification task*), and a memory recognition task involving both semantic and episodic processes. Overall, we expected to detect a flexible effect of RE during verb processing, consistent with multiple representation models; specifically, we expected RE effects to vary as semantic demands increase (up to involving episodic processes) and possibly due to language-specific factors.

To spell out the predictions in more detail, in the first place we relied on evidence collected for English. Specifically, we expected the effect of RE to be significant, yet moderate in a task with lower semantic demands, such as lexical decision. In particular, elaborating on evidence supporting the IAC model (McClelland & Rumelhart, 1981; see Yap et al., 2015), in this task we expected to observe larger effects of lexical-orthographic variables (i.e., lexical frequency, word length, and orthographic neighborhood, etc.) than for experience-based semantic dimensions, such as RE.

Going further, we predicted that, relative to shallower lexical decisions, tasks with higher semantic demands, such as grammatical decision, would trigger greater RE effects. Considering that participants are expected to perform a full semantic coding of verb meaning to make their decisions (i.e., if a presented word is a verb or a noun), thus resulting in higher activation of semantic-level proprieties (including those related to experience-based dimensions), verbs with higher RE should be processed faster than verbs with lower RE, in line with prior evidence in English (Sidhu et al., 2014; Yap & Pexman, 2016). Finally, the memory recognition task, in which participants have to decide whether a presented verb is either new or old (i.e., "old" items are those presented in a previous encoding task), is where we expected to observe the strongest effects of RE. Compared to previous tasks, the memory recognition task loads on different facets of declarative memory processes, including both lexical-semantic and episodic processing (Ballot et al., 2021; Cortese et al., 2010). Based on previous evidence from English verbs (see Muraki et al., 2022; Sidhu & Pexman, 2016), we then expected to see a significant generalizability of semantic-related facilitatory effects to episodic components of declarative memory (namely, better performance in the recognition of verbs with higher RE values compared to verbs with lower RE values). Importantly, as part of the multidimensional account, we also acknowledge possible language-specific effects (for a discussion, see Kemmerer, 2023) in determining the role of RE: while the above predictions are shaped on English data, the Italian material used here might disclose different patterns. The language-specific factor might become especially relevant when grammatical aspects are incorporated into the task (Bambini & Canal, 2021; Dillon & Keshev, to appear), such as in the case of a grammatical decision task, where the rich morphological profile of Italian might reduce the semantic demands of the task and therefore exhibit a different involvement of semantic properties of the stimuli (see Vigliocco et al., 2005).

Finally, we included in the study an analysis of the role of individual difference variables in vocabulary and motor mental skills in relation with RE, as they might play a role in modulating its effect in verb processing (see Muraki & Pexman, 2021; Pexman & Yap, 2018; Yap, Balota, et al., 2012). Given the scarcity of data in this domain, the analysis on this matter was conducted exploratorily.

1.2. Data availability

All data files, including novel RE rating values for Italian verbs, and the analysis codes for the present study are available through the Open Science Framework (OSF) at the following link: https://osf.io/axhn3/?view_only=b52fab19b9ea41a79b57be92597a28f2.

1.3. Rating task on Italian verbs

1.3.1. Method

1.3.1.1. Participants

One hundred Italian-speaking participants took part in the rating task. They were recruited via social media, public communications, and mailing lists. The sample was reduced to meet the distribution of participants' age reported in the English rating study reported in Sidhu et al. (2014), with a cut-off of |2| *SD* (i.e., 26 years of age). Twenty-six participants were excluded as they were outside the range of 18-26 years of age.

Seventy-four participants were included in the final sample (43 female; Age, M = 22.38, SD = 2.20; Education in years, M = 14.58, SD = 1.87). Informed consent was obtained from all participants. The study (including the rating task and the other three experiments) was approved by the Ethics Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (protocol number 76/2021).

1.3.1.2. Materials

The set of stimuli was constructed starting from the 687 English verbs used to collect RE ratings by Sidhu et al. (2014). We translated these verbs into Italian in their infinitive form following strict criteria: 1) we selected the most immediate Italian equivalent according to the definitions provided by bilingual and monolingual dictionaries; 2) in the case of synonyms in the Italian translation, a lexical cognate (if present) was preferred over the other possible translations (e.g., eng. *to emit* it. 'emettere' instead of 'emanare'); 3) in the case of verbs that function both in transitive and intransitive constructions in English and that in Italian can be translated either with or without the clitic pronoun *-si* (e.g., eng. *to comb* it. transitive 'pettinare', it. reflexive 'pettinarsi'; eng. *to amuse* it. transitive 'divertire', it. intransitive 'divertirsi'), we checked the most frequent use in English (based on the British National Corpus Online service, available on: http://www.natcorp.ox.ac.uk/) and we selected the Italian equivalent.

English items were excluded from the translation when: 1) the Italian translation would correspond to multi-word lexical units (e.g., eng. *to golf* it. 'giocare a golf'', eng. *to yodel* it. 'fare lo yodel'), or 2) a fitting translation was not available (e.g., *to braise*, *to cringe*, *to gargle*, *to scorch*). The English list also included near-synonymic pairs, that would lead to duplicate translations in Italian (e.g., eng. *to do/to make* > it. *fare*). In this case, only one English form was translated, choosing the closest in frequency with the Italian form. By these criteria, 30 verbs were excluded from the translation process.

The final translations were validated by comparing (log-transformed) lexical frequency values for English and Italian lemmas, extracted from Lund and Burgess (1996) and the "La Repubblica" corpus (Baroni et al., 2004), respectively. For four Italian verbs (.6% of the total set), log-transformed lexical frequency values were extracted from the itWaC corpus (Baroni et al., 2009), as they were not attested in the "La Repubblica" corpus. If a log-transformed value exceeded |2.5| *SD* from the mean difference between English and Italian frequency values, we considered the pairs a case of lexical frequency mismatch. Nine verbs were excluded after this process (e.g., *to elapse*, *to echo*, *to need*, etc.). The remaining 648 Italian verbs showed a strong association with English lexical frequency values ($r(646) = .72$, $p < .001$) and were considered suitable for the rating task.

1.3.1.3. Procedure

The set of 648 Italian verbs was randomly split into two lists, each of which was divided into four blocks of equal length. The eight blocks were matched for log-transformed lexical frequency ($F(7, 640) = 0.60$, $p = .754$) and RE values associated with the English counterparts ($F(7, 640) = 0.15$, $p = .994$).

Thirty-five participants rated one list of verbs ("Survey A"), while 39 participants rated the other list ("Survey B"). To limit participants' fatigue, blocks within surveys were separated by short pauses. Randomization between and within blocks was used to prevent order-effect bias.

Both surveys were designed on the web-based platform LimeSurvey® (https://limesurvey.org). After providing informed consent and demographics (i.e., age, gender, years of education), participants received the instructions for the rating task (translated and adapted from Sidhu et al., 2014) and were asked to judge (on a 1-7 scale) the degree to which the meaning of each verb involved the human body. Each participant completed the survey autonomously as an online task.

1.3.1.4. Statistical analysis

We first assessed response quality with the intra-individual response variability score (IRV, Dunn et al., 2018) available in the *careless* package (Yentes & Wilhelm, 2023), computed within the entire set of responses and each block. Then, we tested rater reliability by computing the Interclass Correlation Coefficient (ICC), selecting a two-way model testing agreement on the average score with the *irr* package (Gamer et al., 2022). Finally, we described the distribution of Italian RE rating values and we validated them by testing their agreement with English RE values via Pearson's correlation. The analysis was performed using RStudio (R Core Team, 2023), version 2023.09.0+463.

1.3.2. Results

We obtained relative embodiment ratings for a total of 647 Italian verbs. One verb (it. *asserire* eng. 'to assert') was excluded after the rating task due to technical issues related to the online survey platform. Seven raters were excluded from the analysis because they obtained an IRV score < 1 in one of the four blocks. The ICC showed good to excellent agreement among the raters ($ICC_{SurveyA}$ = .88, 95% CI [.86, .91], $p < .001$; $ICC_{SurveyB}$ = .93, 95% CI [.92, .95], $p < .001$). During the rating task, all the participants used a range of points ≥ 6 on the 7-point scale, suggesting that the rating scale was used properly as instructed by the whole sample.

Italian RE values were distributed as follows: *M* = 3.75, *SD* = 1.05, *Mdn* = 3.53, *min* = 2.08, *max* = 6.43, and *range* = 4.35. The distribution showed a moderate positive skewness (.64) and moderate negative kurtosis (-.54). The distribution of Italian RE ratings is shown in Figure 1.1A.

Italian and English RE ratings showed a strong positive correlation ($r(645) = .79$, $p < .001$), with a tight overlap between Italian and English RE rating distributions (see Figure 1.1B).

The database with the rating values for Italian verbs, alongside the original English rating values, is openly accessible for future uses at the following link: https://osf.io/axhn3/?view_only=b52fab19b9ea41a79b57be92597a28f2.

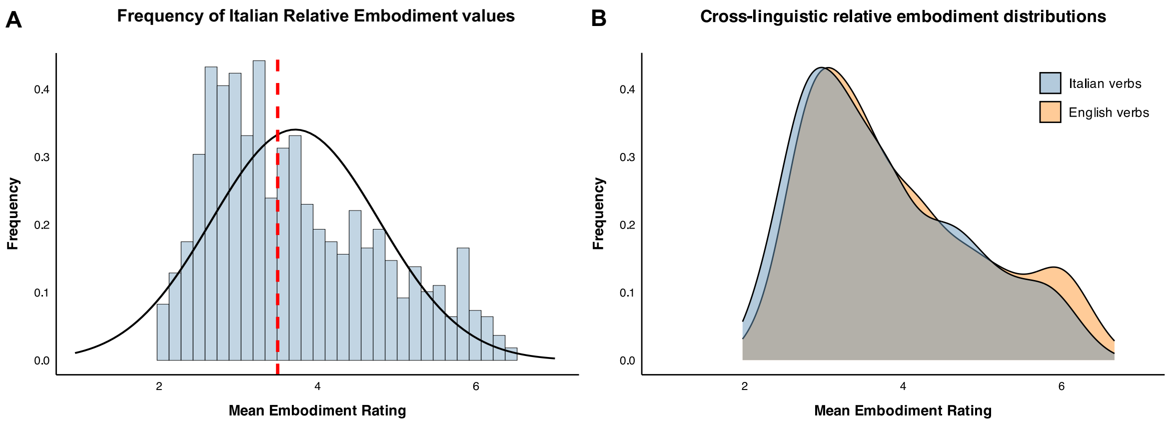


Figure 1.1. Relative Embodiment rating values for Italian verbs. *(A) The histogram displays the frequency of relative embodiment ratings for the 647 verbs used in the rating task (a black line representing a Gaussian distribution is plotted as reference). The red dotted line represents the median of the distribution. (B) The density plots show the distributions of Italian and English relative embodiment ratings in comparison (the grey area represents the overlap region between the two distributions).*

1.3.3. Discussion

Our data from a heterogenous set of Italian verbs showed that bodily experience and the sense of "owning a body" varies considerably across verbs, with psychological verbs (e.g., it. 'negare' eng. *to deny*, it. 'ponderare' eng. *to ponder*, it. 'presumere' eng. *to assume*, etc.) receiving lower RE scores (RE values ranging between 2.08 and 2.19 in the verbs mentioned) and verbs related to motor and sensorial experience (e.g., it. 'sedersi' eng. *to sit*, it. 'gustare' eng. *to taste*, it. 'calciare' eng. *to kick*, etc.), physiological processes (e.g., it. 'sanguinare' eng. *to bleed*, it. 'sudare' eng. *to sweat*, it. 'ingrassare' eng. *to fatten*, etc.), and body interaction (e.g., it. 'abbracciare' eng. *to hug*, it. 'coccolare' eng. *to cuddle*, it. 'lottare' eng. *to fight*, etc.) receiving relatively higher RE scores (RE values ranging between 5.86 and 6.43 for the verbs mentioned). Importantly, our results show remarkable stability across raters, supporting the idea that sensory-motor, bodily, and interoceptive experience contribute to characterizing the semantic representation of verb meaning.

Finally, the cross-language correlation suggests that the extent to which speakers subjectively perceive bodily experience to be involved in verb meaning is rather stable across languages. This finding expands previous evidence of cross-linguistic stability of other lexical-semantic variables, such as lexical frequency (Kuperman et al., 2024; Tjuka, 2020), imageability (Rofes et al., 2018), and emotional content (Marmolejo-Ramos et al., 2017).

Overall, the results of the rating task and the strong association with English values suggest that RE is a relevant semantic dimension for verbs across languages (Sidhu et al., 2014).

1.4. Experiment 1: Lexical decision task (LDT)

1.4.1. Method

1.4.1.1. Participants

We recruited 39 Italian-speaking young adult participants for the LDT. The sample size was decided following Sidhu et al. (2014), to have a comparable number of participants for this experiment. Exclusion criteria included being: 1) non-native speaker of Italian, 2) bilingual from birth, and 3) diagnosed with a learning disability. Four participants were excluded after checking for criteria 1) and 2). Thirty-five participants were included in the final sample (21 female; Age, $M = 22.85$, $SD = 2.85$; Education in years, $M = 14.49$, $SD = 2.01$). Informed consent was obtained from all participants (within ethic protocol 76/2021).

1.4.1.2. Materials

Following Sidhu et al. (2014), we selected 400 Italian verbs with relatively stable RE ratings, namely with low *SD* (the *SD* range in the resulting set was 1.04–1.88). The range of RE ratings in the Italian set was 2.08–6.43, spanning from low to high RE verbs. Focusing on our main hypothesis, namely that LDT item latencies would be affected by RE beyond other lexical variables, we ran a sensitivity power analysis with G*Power, version 3.1.9.6 (Faul et al., 2009), confirming that the set size of 400 items was adequate, with $a = .05$ and power ≥ 80%, to detect an effect size higher than $f^2 \geq .03$, in line with Sidhu et al. (2014). A total of 400 pseudo-verbs (i.e., nonwords with a correct inflectional infinitival morpheme of Italian verbs, such as *febire*, with *-ire* as the infinitival marker) were created as control stimuli, matched with the 400 verbs for length ($t(798) = 0.16$, $p = .876$). For the verbs in the final set, we extracted a number of item variables, including log-transformed lexical frequency values from the "La Repubblica" corpus (Baroni et al., 2004), Orthographic Levenshtein Distance (OLD20) values, indicating word orthographic neighborhood, from the PhonItalia database (Goslin et al., 2013), the number of morphemes determined using the DerIvaTario annotated lexicon (Talamo et al., 2016), and imageability values from the MEGAHR cross-linguistic repository (Ljubešić et al., 2018). The descriptive statistics of stimuli used for Experiment 1 are available in Table 1.1.

Table 1.1. Descriptive statistics (means and SD) of the verbs used for Experiment 1.

Variable	**Verbs (*N* = 400)**
Relative Embodiment	3.76 (1.23)
Lexical Frequency (log)	7.18 (1.81)
Length (characters)	8.77 (1.61)
Orthographic Levenshtein Distance	2.16 (0.47)
Number of Morphemes	2.33 (0.54)
Imageability	3.57 (0.52)

To exploratorily investigate the role of individual differences in motor mental imagery ability and vocabulary skills on LDT response, the Vividness of Movement Imagery Questionnaire-2 (VMIQ-2, Roberts et al., 2008) and the Verbal Meaning Task from the Primary Mental Abilities battery (Thurstone & Thurstone, 1963) were administered to participants as additional tasks.

The VMIQ-2 is a self-report questionnaire assessing the vividness of movement imagery. Participants were instructed to imagine 12 different actions from three different viewpoints (external visual imagery, internal visual imagery, and kinaesthetic imagery) on a 1-5 scale, with 1 corresponding to images that were perfectly clear and as vivid as normal vision or feeling of movement and 5 corresponding to no image at all. The questionnaire was adapted into Italian and implemented in a computerized format to be administered as an online task. To assess the reliability of the adapted version of the questionnaire, we tested its internal consistency with Cronbach's a computed for External, Internal, and Kinaesthetic subscales (Cronbach's a = .95, .96, and .94 respectively). To improve interpretability, the VMIQ-2 scores across subscales were reversed before inclusion in the analysis, so that high VMIQ-2 scores indicate high motor imagery vividness. The Verbal Meaning Task from the Primary Mental Abilities battery is a test measuring vocabulary skills, in particular vocabulary depth (i.e., how well words are known and integrated in the mental lexicon, see Schmitt, 2014). Participants were instructed to identify the correct synonym of 50 target words by choosing from four given alternatives within 8 minutes. The total score ranged from 0 to 50. The Italian version of the Verbal Meaning Task was validated for in-person administration on a sample of young and older Italian-speaking adults (Cronbach's a = .63, see Rosi et al., 2017). Here, we implemented the Italian version of this task (Rosi et al., 2017) in a web-based computerized format to be administered online while maintaining the time pressure. Our online adaptation showed good internal consistency (Cronbach's a = .80, computed with Kuder-Richardson formula 20).

Cronbach's a values for the VMIQ-2 and the Verbal Meaning Task were computed by pooling participants who completed Experiments 1, 2, and 3 (N = 162).

1.4.1.3. Procedure

Trials were divided into two blocks (each including 200 verbs and 200 pseudoverbs) matched for all the linguistic variables, separated by a short pause. Randomization between and within blocks was used to prevent order-effect bias.

Each trial began with a cross in the middle of the screen as a fixation point (lasting 1500 ms), followed by a target verb or a nonword that remained on the screen for 3000 ms. Participants were instructed to press the "L" key (or "A", counterbalanced) on their keyboard using their right hand to respond "word" and the "A" key (or "L", counterbalanced) using their left hand to respond "nonword". Participants were asked to respond as quickly and as accurately as possible. Each participant completed 20 practice trials (10 verbs and 10 nonwords) before the experiment. Participants received feedback during the practice trials. After completing the LDT, participants were administered the VMIQ-2 and the Verbal Meaning Task within the same online session.

The experiment was implemented on the web-based platform Gorilla Experiment Builder® (https://gorilla.sc).

1.4.1.4. Statistical analysis

We first investigated the effect of item variables on LDT latencies, by computing patterns of associations between dimensions via a correlogram computed using Pearson's correlations, adjusting *p*-values for False Discovery Rate (FDR, Benjamini & Hochberg, 1995). Then we assessed the effect of RE and other variables on LDT latencies by fitting a Linear Mixed-effects Model (LMM) on the log-transformed and *z*-scored reaction times of correct responses using *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017). The LMM included item variables and, exploratorily, individual difference variables as fixed predictors. Item variables included two sets of dimensions: control (log-transformed lexical frequency, word length, OLD, and number of morphemes) and semantic variables (RE and imageability values). Individual difference variables included the VMIQ-2 subscale scores (External, Internal, and Kinaesthetic) and the Verbal Meaning Task total score.

Likelihood-ratio tests on models of increasing complexity were used to evaluate whether the inclusion of continuous predictors improved the model's goodness of fit. We progressively added the control item variables to an intercept-only model, followed by semantic variables and individual differences variables, included in interaction with RE ratings. We kept in the model only those predictors that exerted a significant effect on the dependent variable. All continuous fixed predictors were standardized as *z*-scores before being included in the analysis. Inclusive R^2 (i.e., the total variance explained by a predictor) was computed for each fixed effect using the *PartR2* package (Stoffel et al., 2021).

The optimal random structure was determined in a stepwise fashion using the *buildmer* package (Voeten, 2023), starting from a maximal model including all possible fixed and random effects. The final random structure included by-subjects and by-items random intercepts only, as more complex structures did not lead to model convergence. For each specified random effect, we checked the Best Linear Unbiased Predictors (Baayen et al., 2008): participants and items with systematic variation from the group mean (i.e., > |2.5| *SD*) were treated as deviant points and removed from the analysis.

The analysis was performed using RStudio (R Core Team, 2023).

1.4.2. Results

LDT data were pre-processed by trimming participants (n = 0) and items (n = 8) with accuracy values < 70%. Incorrect trials and trials on which latencies were more than |2.5| *SD* from the participant's mean (2.91% of trials) were excluded from the analysis. Verb and nonword latencies and accuracy levels are reported in Table 1.2.

The correlogram showed that LDT log-transformed latencies for verbs were negatively correlated with log-transformed lexical frequency, RE, and Imageability, and positively correlated with word length, OLD, and number of morphemes (Figure 1.2). Since Imageability and RE showed a high association and a similar pattern of correlations, we decided to drop Imageability (i.e., the least relevant variable considering our hypotheses) from subsequent analyses.

Table 1.2. Descriptive statistics (means and SD) for latencies (in milliseconds) and accuracy associated with verbs and nonwords in Experiment 1.

Variable	**Mean**	**SD**
LDT Latencies		
Verbs	722	65.42
Nonwords	849	76.10
LDT Accuracy		
Verbs	.96	.04
Nonwords	.96	.04

Note. LDT = Lexical Decision Task.

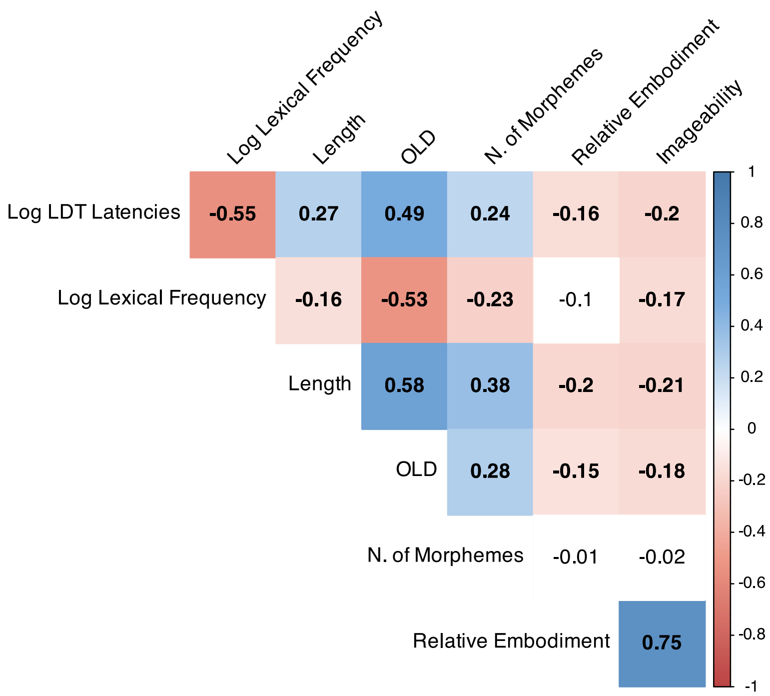


Figure 1.2. Correlogram between verb LDT latencies and item variables. *The magnitude of associations is depicted by colour (white cells indicate non-significant correlations after FDR correction, with significance level p < .05). Note: LDT = Lexical Decision Task; OLD = Orthographic Levenshtein Distance.*

The best model included control variables, RE, and Verbal Meaning Task total score (Table 1.3). The model showed a significant negative effect of log-transformed lexical frequency and RE, as well as a positive effect of OLD on LDT latencies: these effects indicate that LDT latencies were faster in verbs with higher frequency and RE values but slower in verbs with higher number of orthographic neighbours. Moreover, the model showed a significant interaction between RE and the Verbal Meaning Task total score, indicating that the effect of RE on latencies decreased as participants' vocabulary skills increased (see Table 1.4 and Figure 1.3).

Table 1.3. Likelihood-ratio tests to assess the goodness of fit of the LMMs on LDT log-transformed zRTs, based on the progressive inclusion of fixed effects.

Models' fixed effects	**AIC**	**BIC**	**logLik**	χ^2	**df**	***p*-value**
Control variables	25463	25522	-12724	172.34	4	**< .001**
Control variables + RE	25450	25515	-12716	15.51	1	**< .001**
Control variables + RE * (VMIQ-2 Ext + VMIQ-2 Int + VMIQ-2 Kin)	25457	25566	-12714	4.91	6	.555
Control variables + RE * Verbal Meaning Task	25443	25523	-12710	11.00	2	**.004**

Note. RE = Relative Embodiment; VMIQ-2 = Vividness of Movement Imagery Questionnaire 2; Ext = External; Int = Internal; Kin = Kinaesthetic. Control variables include log-transformed lexical frequency, word length, Orthographic Levenshtein Distance, and number of morphemes.

Table 1.4. Output of the best Linear Mixed-effects Model on LDT log-transformed reaction times (z-scores).

Fixed Effects	***B***	**95% *CI***	***p*-value**	***IR*^{2}**
(Intercept)	-0.00	[-0.19, 0.19]	.989	-
Lexical Frequency (log)	-0.11	[-0.14, -0.09]	**< .001**	.023
Word Length	0.00	[-0.03, 0.03]	.919	.005
Orthographic Levenshtein Distance	0.06	[0.03, 0.09]	**< .001**	.019
N. of morphemes	0.02	[-0.00, 0.04]	.101	.004
Relative Embodiment	-0.04	[-0.07, -0.02]	**< .001**	.002
Verbal Meaning Task	-0.14	[-0.33, 0.05]	.158	.009
Relative Embodiment * Verbal Meaning Task	0.02	[0.01, 0.04]	**.003**	< .001
Random Effects	**Variance**	**SD**		
$\text{Intercept}_{\text{Subject}}$	0.32	0.57		
$\text{Intercept}_{\text{Item}}$	0.02	0.13		
Residuals	0.61	0.78		
$\text{ICC}_{\text{SubjectItem}}$	0.36			
Model fit	Marginal	Conditional		
R^2	.047	.391		

Note. LDT = Lexical Decision Task; B = model estimates (standardized); CI = confidence intervals; IR^2 = Inclusive R^2.
*Only complete cases entered the model. One deviant subject was excluded from the analysis. Model formula: Log-transformed zRT ~ Log Lexical Frequency + Length + OLD + N. of morphemes + Relative Embodiment * Verbal Meaning Task + (1 | Subject) + (1 | Item).*

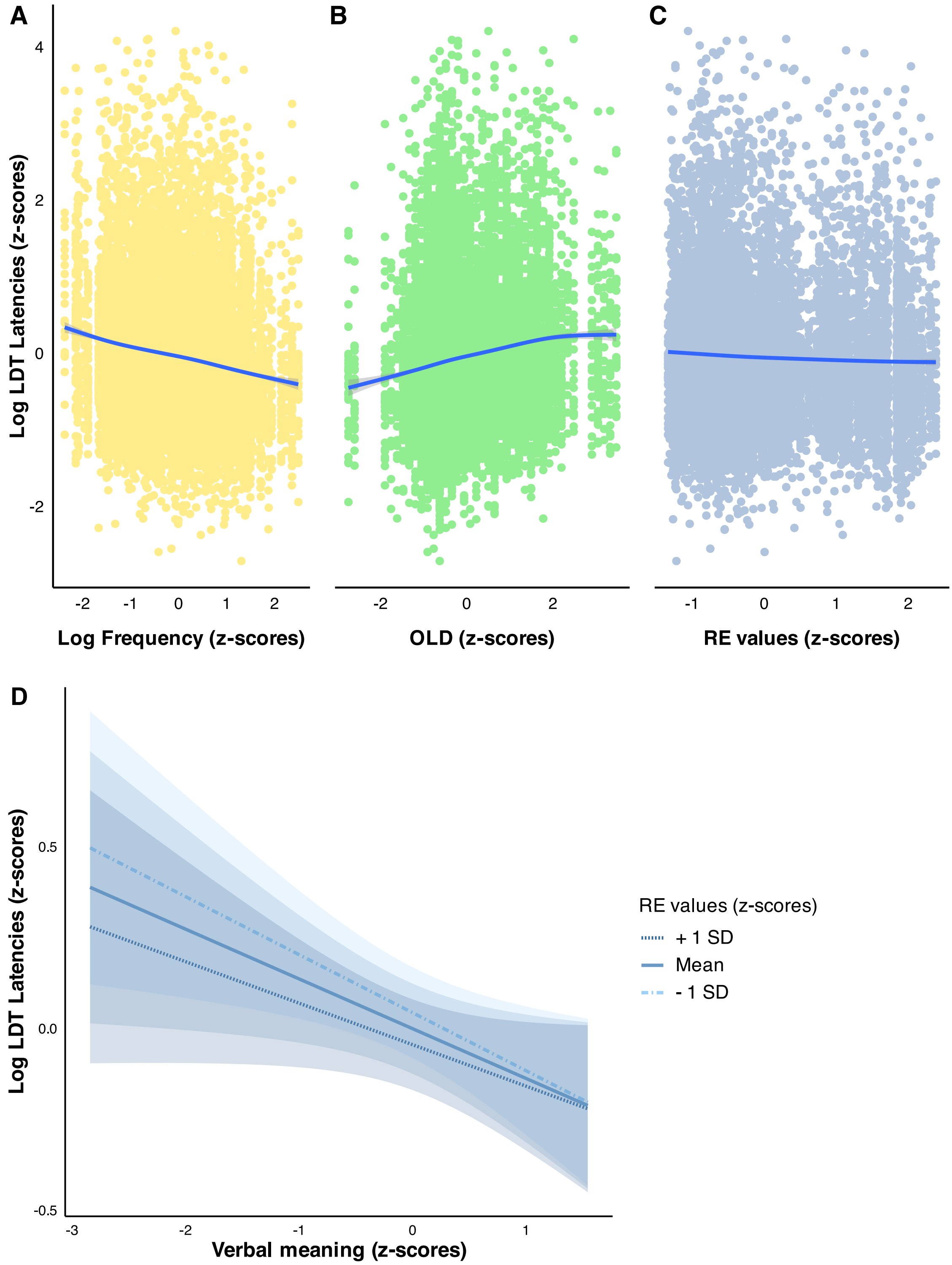


Figure 1.3. Significant effects of item and individual difference variables on lexical decision latencies. *(A-C) Fitted values for log-transformed z-scored Lexical Decision Task (LDT) latencies predicted by significant fixed-effects variables, namely log-transformed lexical frequency values, Orthographic Levenshtein Distance (OLD), and Relative Embodiment (RE) values (smoothed lines added to aid visualization). (D) Fitted values for log-transformed z-scored LDT latencies predicted by RE values in interaction with Verbal Meaning Task total score.*

1.4.3. Discussion

The findings from the LDT on Italian verbs align with our prediction based on English data (Sidhu et al., 2014) and confirm the strong effect of lexical variables, on the one hand, and the modest effect of RE, on the other hand, in this task. First, the coupled effects of lexical frequency and OLD support the idea that familiarity and orthographic processes occurring at the lexical level are the major drivers of performance in lexical decision. In particular, our data replicate the well-known word frequency (Gardner et al., 1987) and orthographic similarity (McClelland & Rumelhart, 1981) effects observed in LDTs, indicating that highly familiar words are processed faster, while words with a high number of orthographic neighbours are processed more slowly due to within-level lexical competition. Conversely, RE had only a modest effect on LDT latencies, as attested by the smaller portion of variance explained by this variable, confirming the idea that semantic properties facilitate word processing in lexical decisions via feedback activations to orthographic representations (Pexman et al., 2002).

Interestingly, the results also confirm our predictions concerning individual difference variables. In particular, we observed a significant interaction between verb RE values and participants' vocabulary skills (but not imagery): the magnitude of the facilitatory effect of RE on latencies decreased as participants' vocabulary skills increased. Conversely, we did not observe any effect of motor mental imagery skills. These findings are in line with the idea that higher vocabulary skills (but not imagery) facilitate lexical decisions (Muraki & Pexman, 2021), with individuals with higher vocabulary skills relying less on semantic dimensions of words (such as RE) to make lexical decisions, compared to less lexically skilled participants (see also Yap, Balota, et al., 2012).

1.5. Experiment 2: Grammatical Decision Task (GDT)

1.5.1. Method

1.5.1.1. Participants

For the GDT, we recruited 53 Italian-speaking young adult participants who did not take part in Experiment 1. Exclusion criteria included being: 1) non-native speaker of Italian, 2) bilingual from birth, and 3) diagnosed with a learning disability. Nine participants were excluded after checking for these criteria. Forty-four participants were included in the final sample (35 female; Age, M = 23.91, SD = 3.21; Education in years, M = 15.11, SD = 2.05). We ran a sensitivity power analysis based on our major research question, namely whether participants would process verbs with higher RE better compared to lower RE verbs: with a = .05 and power ≥ .80, our sample size was sufficient to detect effect sizes greater than Cohen's $d \geq 0.38$, in line with the effect sizes reported by Sidhu et al. (2014) for the same experiment. Informed consent was obtained from all participants (within the ethic protocol 76/2021).

1.5.1.2. Materials

Following Sidhu et al. (2014), we selected 80 Italian verbs from the set of stimuli used in the rating task and also 80 Italian nouns. Since in Italian infinitival forms are marked by specific morphological suffixes (i.e., *-are*, *-ere*, *-ire*) which are not present in nouns, all verbs were presented in the 3SG present form (i.e., with the inflectional marking *-a* or *-e*, as in it. *mangi-a*, eng. 's/he eats'), i.e., the most frequent form of most Italian verbs (Bertinetto et al., 2020). Italian verbs in the 3SG present form inflectional marking might result in homophones with nouns, so the set of items was adjusted in order to avoid verbs that could also be homophone nouns (e.g., it. *Cucina* 'kitchen' but also 's/he cooks'). We also substituted verbs with the clitic pronoun *-si* in the infinitival form resulting in multiword expressions in the 3SG present form (e.g., it. *allontanarsi* 'to depart' > *si allontana* 's/he departs'). Finally, to avoid orthographic confounds, we excluded nouns with endings not represented in the 3SG present form of verbs (i.e., *-o*, since 3SG verbs can end only with *-a* and *-e* in Italian). The final subset of verbs included 40 verbs associated with high (> 3.5) and 40 verbs associated with low (< 3.5) RE ratings. The two lists were matched for relevant lexical and semantic variables, such as log-transformed lexical frequency, word length, Orthographic Levenshtein Distance (OLD20), number of morphemes, and imageability (ts < .98, ps > .331), and they were significantly different only for RE values ($t(78) = 11.58$, p = < .001). The 80 verbs selected for the GDT were matched also with the set of nouns for lexical variables and imageability (ts < 1.41, ps > .160, see Table 1.5).

Table 1.5. Descriptive statistics (means and SD) of the verbs used for Experiment 2 (low and high Relative Embodiment verbs and the matched control nouns).

Variable	**Low RE verbs (*N* = 40)**	**High RE verbs (*N* = 40)**	**Nouns (*N* = 80)**
Relative Embodiment[a]	2.75 (0.40)	4.46 (0.84)	-
Lexical Frequency (log)[a]	7.94 (1.39)	7.93 (1.72)	8.11 (1.70)
Length (characters)[b]	6.93 (1.54)	7.10 (1.58)	7.10 (1.50)
Orthographic Levenshtein Distance[b]	1.91 (0.35)	2.00 (0.42)	2.04 (0.40)
Number of Morphemes[b]	2.23 (0.42)	2.25 (0.44)	2.21 (0.41)
Imageability[a]	3.54 (0.39)	3.63 (0.46)	3.65 (0.44)

Note. RE = Relative Embodiment.
[a]*value of the lemma (infinitive form) used for verbs*
[b]*value of the form (3SG present indicative) used for verbs*

1.5.1.3. Procedure

Trials (*N* = 160) were administered to participants in two separate blocks (each including 40 verbs and 40 nouns), matched for all the relevant dimensions. Randomization was applied within and between blocks. Blocks were separated by a short pause. Each trial began with a fixation cross for 1000 ms, followed by a blank screen for 1000 ms. Afterwards, a target verb or a noun was presented. The task was combined with a *go/no-go* paradigm, in which participants were instructed to decide whether each word was a verb or a noun. If they judged that the word was a verb, they had to press the space key on their keyboard. Otherwise, they were instructed to refrain from responding. The next trial began automatically after 3000 ms. Each participant completed 14 practice trials (7 verbs and 7 nouns) before starting the experiment. During the practice trials, participants received feedback. The experiment was administered on the web-based platform Gorilla Experiment Builder®. After completing the GDT, participants were administered the VMIQ-2 and the Verbal Meaning Task within the same online session.

1.5.1.4. Statistical analysis

Following a similar analytical strategy to the one adopted for Experiment 1 (see Section 1.4.1.4), we fit a Linear Mixed-effects Model (LMM) on the log-transformed *z*-scored reaction times of correct responses. The LMM included RE levels (low vs. high, sum contrast coded, baseline = low) and individual differences variables, namely VMIQ-2 reversed subscale scores (External, Internal, and Kinaesthetic) and the Verbal Meaning Task total score.

Likelihood-ratio tests were again used to evaluate the inclusion of standardized continuous predictors, starting from RE levels added to an intercept-only model, followed by individual differences variables in interaction with RE levels.

The final random structure included by-subjects and by-items random intercepts only, as the optimal structure (including by-items random slope for the Verbal Meaning task total score) led to convergence issues when individual differences variables were included in the model.

The analysis was performed using RStudio (R Core Team, 2023).

1.5.2. Results

GDT data were pre-processed by trimming participants (n = 1) with accuracy values < 70%. Incorrect trials and trials on which latencies exceeded |2.5| *SD* from the participant's mean were excluded from the analysis (3.24% of trials). Descriptive statistics for by-subject and by-item latencies and accuracy are reported in Table 1.6.

Table 1.6. Descriptive statistics (mean and standard deviations) of by-subject and by-item latencies (in milliseconds) and accuracy for low and high relative embodiment verbs.

Measure	Low RE	High RE
GDT Latencies		
By-subject	783 (221)	781 (201)
By-item	778 (48.60)	780 (64.90)
GDT Accuracy		
By-subject	.99 (.04)	.99 (.03)
By-item	.99 (.02)	.99 (.02)

Note. RE: Relative Embodiment; GDT: Grammatical Decision Task.

The hierarchical approach testing the effect of fixed predictions showed that the inclusion of RE levels and VMIQ-2 subscales did not improve model's fit. Only the inclusion of the Verbal Meaning Task total score showed a main effect, indicating that GDT latencies were generally faster for individuals with higher vocabulary skills. The output of the Likelihood-ratio tests and the best model are reported in Table 1.7 and 1.8.

Table 1.7. Likelihood-ratio tests to assess the goodness of fit of the LMMs on GDT log-transformed zRTs, based on the progressive inclusion of fixed effects.

Models' fixed effects	AIC	BIC	logLik	χ^2	df	*p*-value
RE levels	7317.2	7347.7	-3653.6	0.07	1	.798
VMIQ-2 Ext + VMIQ-2 Int + VMIQ-2 Kin	7321.2	7363.8	-3653.6	0.09	3	.993
Verbal Meaning Task	7295.1	7325.6	-3642.6	22.12	1	**< .001**

Note. RE = Relative Embodiment; VMIQ-2 = Vividness of Movement Imagery Questionnaire 2; Ext = External; Int = Internal; Kin = Kinaesthetic.

Table 1.8. Output of the best Linear Mixed-effects Model on GDT log-transformed and z-scored reaction times.

Fixed Effects	***B***	**95% *CI***	***p*-value**	***IR*²**
(Intercept)	-0.03	[-0.19, 0.12]	.696	-
Verbal Meaning Task	-0.31	[-0.48, -0.15]	**< .001**	.095
Random Effects	**Variance**	**SD**		
$\text{Intercept}_{\text{Subject}}$	0.24	0.49		
$\text{Intercept}_{\text{Item}}$	0.03	0.18		
Residuals	0.48	0.70		
$\text{ICC}_{\text{SubjectItem}}$	0.36			
Model fit	Marginal	Conditional		
R^2	.095	.418		

Note. GDT = Grammatical Decision Task; B = model estimates (standardized); CI = confidence intervals; IR² = Inclusive R². One deviant subject was excluded from the analysis. Model formula: Log-transformed zRT ~ Verbal Meaning Task + (1 | Subject) + (1 | Item).

1.5.3. Discussion

The results of the GDT did not align with the English-oriented prediction of strong activation of semantic properties when participants had to decide whether a word was a noun or a verb. English speakers were shown to exploit high activation of verb semantic properties in this task, resulting in faster and better responses for high RE verbs compared to low RE ones (Sidhu et al., 2014). Conversely, RE was not a significant predictor of GDT response in Italian. The pattern is in line with multiple representation models and shows that the activation of sensory-motor properties of words is influenced by language-specific factors, such as differences at the grammatical level (Gianelli et al., 2017). In particular, this pattern suggests that in a richly inflected language such as Italian, verb inflection integration interferes with other levels of processing (Bambini & Canal, 2021). In other words, in a task where judgments over grammatical categories are required, the performance of Italian-speaking individuals might be driven more by morphosyntactic information embedded in Italian verbs, rather than semantic properties.

As for individual differences, the only significant predictor for GDT latencies was vocabulary skills: this effect is in line with previous evidence showing that participants with deeper vocabulary knowledge and higher integration of words within the mental lexicon capitalize on more extended and robust word knowledge, resulting in faster grammatical decisions compared to less skilled participants (see Muraki & Pexman, 2021).

1.6. Experiment 3: Memory Recognition Task (MRT)

1.6.1. Method

1.6.1.1. Participants

Ninety-eight Italian-speaking young adult participants who did not take part in Experiments 1 and 2 were involved in the MRT. Exclusion criteria included being: 1) non-native speaker of Italian, 2) bilingual from birth, and 3) diagnosed with a learning disability. A total of 13 participants were excluded after checking for these criteria. Eighty-five participants were included in the final sample (48 female; Age, M = 25.56, SD = 4.14; Education, M = 16.34, SD = 2.46). A sensitivity power analysis based on the hypothesis that verbs with higher RE would be recognized better than lower RE verbs showed that our sample size was adequate to detect, with a = .05 and power ≥ 80%, effect sizes greater than Cohen's d ≥ .27 (comparable to the effect sizes reported by Sidhu & Pexman, 2016 for the same experiment administered online).

Informed consent was obtained from all participants (within the ethic protocol 76/2021).

1.6.1.2. Materials

Consistently with Sidhu and Pexman (2016), we selected 140 Italian verbs and 70 nouns. For this task, we expanded the set of verbs from Experiment 2, by adding 60 verbs from the list in the rating task. Verbs were again presented in their 3SG present indicative form, therefore the selection of the additional items followed the same criteria used in Experiment 2 (see Section 1.5.1.2). The set of stimuli was divided into two lists, each including 35 verbs with high RE values and 35 verbs with low RE values. The verb lists were matched for log-transformed lexical frequency, word length, Orthographic Levenshtein Distance (OLD), number of morphemes, imageability, and RE values (*t*s < .77, *p*s > .443). Verbs in each list were also matched for the same variables (excluding RE) with the set of nouns (*F*s < .99, *p*s > .370, see Table 1.9).

Table 1.9. Descriptive statistics (means and SD) of the verbs used for Experiment 3 (low and high Relative Embodiment verbs for list 1 and 2 with the matched control nouns).

Variable	List 1 (*N* = 70)		List 2 (*N* = 70)		Nouns (*N* = 70)
	Low RE verbs	**High RE verbs**	**Low RE verbs**	**High RE verbs**	
Relative Embodiment[a]	2.85 (0.37)	4.46 (0.76)	2.81 (0.36)	4.44 (0.80)	-
Lexical Frequency (log)[a]	7.85 (1.48)	7.89 (1.58)	8.08 (1.37)	7.95 (1.68)	8.00 (1.69)
Length (characters)[b]	6.89 (1.55)	6.71 (1.66)	7.14 (1.50)	6.80 (1.49)	6.99 (1.49)
OLD[b]	1.92 (0.34)	1.90 (0.33)	1.94 (0.41)	1.96 (0.44)	1.99 (0.36)
Number of Morphemes[b]	2.26 (0.44)	2.31 (0.47)	2.23 (0.43)	2.23 (0.43)	2.21 (0.41)
Imageability[a]	3.59 (0.44)	3.67 (0.44)	3.53 (0.31)	3.62 (0.44)	3.68 (0.43)

Note. RE = Relative Embodiment; OLD = Orthographic Levenshtein Distance.
[a]value of the lemma (infinitive form) used for verbs
[b]value of the form (3SG present indicative) used for verbs

1.6.1.3. Procedure

In this experiment, participants first took part in a *go/no-go* GDT (with the same procedure as in Experiment 2), during which they were presented with only one list of verbs and the matched nouns (140 trials). Participants were all unaware that their verb memory would be tested in a later task. List selection for the GDT was counterbalanced across participants. Afterwards, participants took part in the MRT including both lists of verbs (140 trials) presented in a randomized order, in which they had to decide whether a verb on the screen had already been presented to them in the previous task (i.e., an *old* verb) or not (i.e., a *new* verb). Each trial began with a cross in the middle of the screen as a fixation point (lasting 1000 ms), followed by a verb that remained on the screen for 3000 ms. Participants were instructed to press the "L" key (or "A", counterbalanced) on their keyboard using their right hand to respond "old" and the "A" key (or "L", counterbalanced) using their left hand to respond "new". The next trial began automatically after 3000 ms if no answer was provided. The MRT was preceded by a five-minute distraction task, in which participants had to answer as many addition questions (e.g., "364 + 133 = ?") as they could. The experiment was administered on the web-based platform Gorilla Experiment Builder®. After completing the MRT, participants were administered the VMIQ-2 and the Verbal Meaning Task within the same online session.

1.6.1.4. Statistical analysis

Building on the classic approaches to the analysis of memory recognition data, we first computed *d*-prime scores to compare participants' performance with low-RE verbs and high-RE verbs. *D*-prime scores were computed using the *psycho* package (Makowski, 2018) as: *z*-scored hit rate (i.e., percent of trials on which a previously seen word was correctly identified as an old word) *minus* *z*-scored false alarm rate (i.e., percent of trials on which a new word was incorrectly identified as an old one). Participants obtaining negative *d*-prime scores were flagged as having poor reliability and removed from the analysis, as they might reflect sampling error or response confusion (Stanislaw & Todorov, 1999). Then, we compared participants' *d*-prime scores, hit rates, and false alarms using paired *t*-tests.

Afterwards, we fitted a Linear Mixed-effects Model (LMM) on the log-transformed *z*-scored reaction times of the correct answers of the MRT, following the same criteria used in Experiments 1 and 2 (see Sections 1.4.1.4 and 1.5.1.4). The optimal random structure for model convergence included by-subjects and by-items random intercepts and a by-subjects random slope for RE levels. The analysis was performed using RStudio (R Core Team, 2023).

1.6.2. Results

We first inspected data quality to trim participants who were not appropriately committed during the entire experiment. The exclusion criteria were: a) < 70% correct answers in either critical or control trials of the encoding task ($n = 0$), b) < 10 additions in the distraction task ($n = 2$), c) > 20 minutes between the end of the encoding task and the end of the MRT (*n* = 7), d) > 30% of time-out trials in the MRT (*n* = 2), and e) obtaining negative *d*-prime scores (*n* = 3). To analyze latencies, we trimmed incorrect trials and trials on which latencies were more than |2.5| *SD* from the participant's mean (3.70% of trials).

Paired *t*-tests on participants' *d*-primes, hit rates, and false alarm rates showed that verbs with high RE ratings were more accurately recognized than verbs with low RE ratings (Table 1.10, Figure 1.4).

Table 1.10. Descriptive statistics (mean and standard deviations) and comparisons of participants' d*-prime scores, hits rates, and false alarm rates for low and high relative embodiment verbs.*

Variable	**Low RE**	**High RE**	***t*-statistics**	**Cohen's *d***	***p*-value**
D-prime	0.92 (0.46)	1.07 (0.48)	$t(73) = 2.90$	0.34	**.005**
Hit rate	0.67 (0.08)	0.70 (0.09)	$t(73) = 2.42$	0.28	**.018**
False alarm rate	0.33 (0.08)	0.39 (0.09)	$t(73) = -2.42$	-0.28	**.018**

Note. RE: Relative Embodiment.

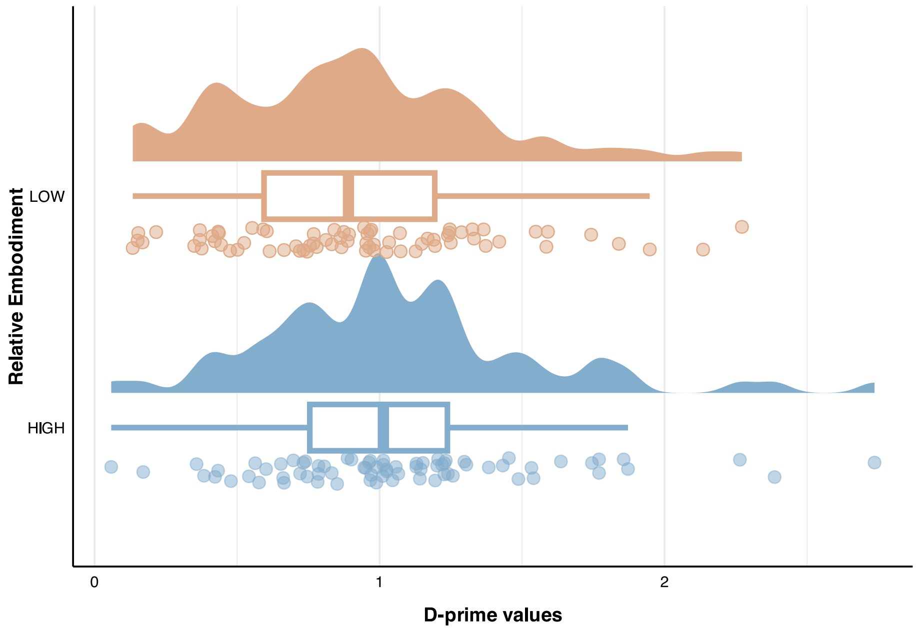


Figure 1.4. Accuracy values in the Memory Recognition task. *The raincloud plots show the distribution of* d*-prime scores obtained by participants for high-RE and low-RE verbs.*

As for the LMM on MRT latencies, the inclusion of RE levels and Verbal Meaning Task total scores to the baseline model explained most data variance: both variables showed a main effect, indicating that MRT latencies were faster for high RE verbs and in individuals with higher vocabulary skills. Conversely, no effects were observed for motor mental imagery (see Table 1.11 for the output of the Likelihood-ratio tests and Table 1.12 for the output of the best model).

Table 1.11. Likelihood-ratio tests to assess the goodness of fit of the LMMs on MRT log-transformed zRTs, based on the progressive inclusion of fixed effects.

Models' fixed effects	**AIC**	**BIC**	**logLik**	χ^2	**df**	***p*-value**
RE levels	17406	17454	-8696.1	4.20	1	**.040**
RE levels * (VMIQ-2 Ext + VMIQ-2 Int + VMIQ-2 Kin)	17413	17501	-8693.4	5.44	6	.488
RE levels * Verbal Meaning Task	17404	17465	-8693.0	6.23	2	**.044**

Note. RE = Relative Embodiment; VMIQ-2 = Vividness of Movement Imagery Questionnaire 2; Ext = External; Int = Internal; Kin = Kinaesthetic.

Table 1.12. Output of the best Linear Mixed-effects Model on MRT log-transformed and z-scored reaction times for all correct answers.

Fixed Effects	***B***	**95% *CI***	***p*-value**	***IR*²**
(Intercept)	0.01	[-0.09, 0.11]	.830	**-**
RE levels (Low vs. High)	-0.08	[-0.16, -0.01]	**.036**	**.002**
Verbal Meaning Task	-0.13	[-0.22, -0.03]	**.009**	**.016**
RE levels (Low vs. High) * Verbal Meaning Task	-0.03	[-0.09, 0.03]	.317	**< .001**
Random Effects	**Variance**	**SD**	**Corr**	
$Intercept_{Subject}$	0.17	0.41		
RE levels (Low vs. $High)_{Subject}$	0.04	0.21	-.35	
$Intercept_{Item}$	0.02	0.13		
Residuals	0.76	0.87		
$ICC_{SubjectItem}$	0.20			
Model fit	Marginal	Conditional		
R^2	.018	.218		

*Note. MRT = Memory Recognition Task; B = model estimates (standardized); CI = confidence intervals; IR² = Inclusive R². One deviant subject was excluded from the analysis. Model formula: Log-transformed zRT ~ RE Levels * Verbal Meaning Task + (1 + RE Levels | Subject) + (1 | Item).*

1.6.3. Post-hoc cross-task comparisons

After investigating the role of RE in different tasks, we aimed to exploratorily compare its involvement across tasks, by comparing the results obtained in the two tasks where it exerted a significant effect, namely LDT (Experiment 1) and MRT (Experiment 3).

To achieve this objective, we compared trimmed log-transformed *z*-scored latencies and binomial Accuracy scores (0 = incorrect, 1 = correct) obtained in the LDT and the MRT by fitting a LMM and a Generalized Linear Mixed-effects Model (GLMM). For this analysis, we included latencies and accuracy obtained by participants in a subset of verbs present in both experiments (*n* = 83 verbs, 21% of the LDT set and 59% of the MRT set). We included Task (LDT vs. MRT, with sum contrast coding and MRT as baseline) in interaction with RE ratings as fixed predictors. Based on previous analysis, showing a significant effect of specific item dimensions and vocabulary skills, we also included control item variables (i.e., log-transformed lexical frequency, word length, OLD, and the number of morphemes) and Verbal Meaning Task total score as covariates. The LMM random structure included by-subjects and by-items random intercepts and by-subjects random slope for log-transformed lexical frequency, while the GLMM random structure included by-subjects intercepts only. The analysis was performed using RStudio (R Core Team, 2023) with packages already mentioned in Section 1.4.1.4.

The comparison between LDT and MRT log-transformed *z*-scored latencies (Table 1.13, Figure 1.5A) showed a main effect of the Verbal Meaning Task total score and RE values as well as a significant interaction between RE and Task, indicating that latencies were faster for verbs with higher RE and for individuals with higher vocabulary skills, but the effect of RE was significantly higher in the MRT.

Table 1.13. Output of the Linear Mixed-effects Model on LDT and MRT log-transformed z-scored reaction times for all correct answers.

Fixed Effects	***B***	**95% *CI***	***p*-value**	***IR²***
(Intercept)	-0.02	[-0.12 – 0.08]	.702	**-**
Lexical Frequency (log)	0.01	[-0.03 – 0.04]	.699	**.000**
Word Length	0.04	[-0.00 – 0.08]	.074	**.004**
Orthographic Levenshtein Distance	0.02	[-0.01 – 0.06]	.256	**.003**
N. of morphemes	0.02	[-0.01 – 0.04]	.252	**.001**
Verbal Meaning Task	-0.10	[-0.19 – -0.01]	**.033**	**.011**
Relative Embodiment	-0.04	[-0.06 – -0.01]	**.006**	**.002**
Task (LDT vs. MRT)	-0.02	[-0.22 – 0.18]	.839	**.022**
Relative Embodiment * Task (LDT vs. MRT)	0.08	[0.04 – 0.12]	**< .001**	**.005**
Random Effects	**Variance**	**SD**	**Corr**	
$Intercept_{Subject}$	0.22	0.47		
Lexical Frequency $(log)_{Subject}$	0.01	0.11	.27	
$Intercept_{Item}$	0.00	0.07		
Residuals	0.71	0.84		
$ICC_{SubjectItem}$	0.25			
Model fit	Marginal	Conditional		
R^2	.017	.263		

*Note. LDT = Lexical Decision Task; MRT = Memory Recognition Task; B = model estimates (standardized); CI = confidence intervals; IR² = Inclusive R². One deviant subject was excluded from the analysis. Model formula: Log-transformed zRT ~ Log Lexical Frequency + Length + OLD + N. of morphemes + Verbal Meaning Task + Relative Embodiment * Task + (1 + Log Lexical Frequency | Subject) + (1 | Item).*

The comparison between LDT and MRT accuracy scores showed no main effect of RE values, yet a significant main negative effect of lexical frequency and word length, indicating that accuracy decreased as word length and lexical frequency increased, as well as a main effect of Task, showing that accuracy levels were significantly lower in MRT (Table 1.14, Figure 1.5B).

Table 1.14. Output of the Generalized Linear Mixed-effects Model on LDT and MRT binomial Accuracy scores.

Fixed Effects	***OR***	**95% *CI***	***p*-value**	***IR*²**
(Intercept)	11.04	[9.46, 12.88]	**< .001**	-
Lexical Frequency (log)	0.86	[0.81, 0.91]	**< .001**	.002
Word Length	0.83	[0.75, 0.92]	**< .001**	.056
Orthographic Levenshtein Distance	1.05	[0.96, 1.15]	.267	.010
N. of morphemes	0.96	[0.91, 1.02]	.219	< .001
Verbal Meaning Task	1.01	[0.94, 1.09]	.770	.017
Relative Embodiment	0.99	[0.86, 1.14]	.889	< .001
Task (LDT vs. MRT)	36.11	[26.22, 49.71]	**< .001**	.310
Relative Embodiment * Task (LDT vs. MRT)	0.85	[0.64, 1.12]	.249	.001
Random Effects	**Variance**	**SD**		
$\text{Intercept}_{\text{Subject}}$	0.07	3.29		
$\text{ICC}_{\text{SubjectItem}}$	0.02			
Model fit	Marginal	Conditional		
R^2	.391	.435		

Note. LDT = Lexical Decision Task; MRT = Memory Recognition Task; OR = odds ratios; CI = confidence intervals; IR² = Inclusive R². Model diagnostics were inspected using DHARMa *R package* (Hartig, 2022). *One deviant subject was excluded from the analysis. Model formula: Accuracy ~ Log Lexical Frequency + Length + OLD + N. of morphemes + Verbal Meaning Task + Relative Embodiment * Task + (1 | Subject).*

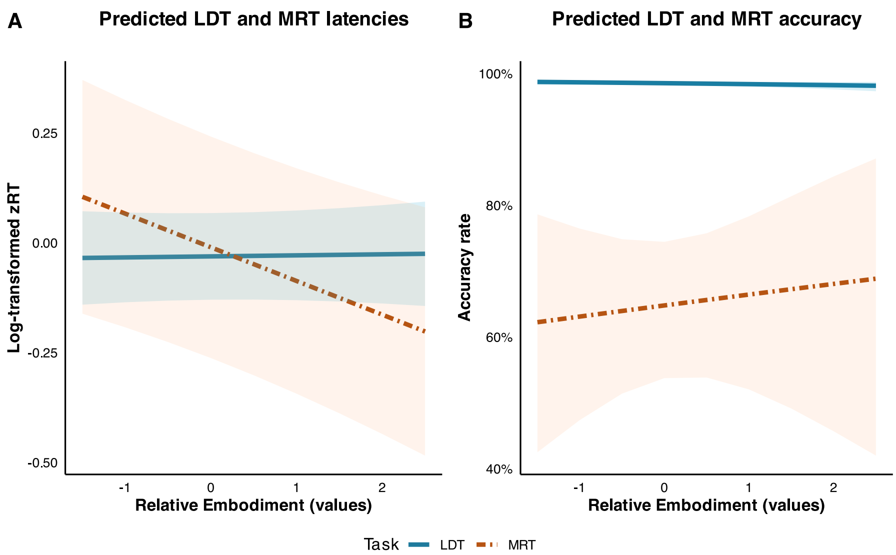


Figure 1.5. Results of the cross-task comparison between LDT and MRT. *(A) Predicted values of log-transformed z-scored latencies obtained in the Lexical Decision Task (LDT) and the Memory Recognition Task (MRT. (B) Predicted values of accuracy scores obtained in the LDT and the MRT.*

1.6.4. Discussion

The results of the MRT showed that, for Italian verbs, RE effects generalize from word processing to memory processes and facilitate word memorization and recognition, thus confirming the prediction based on English data (Sidhu & Pexman, 2016). The across-task comparison is also in line with prediction and confirms the idea, put forward by multiple representation models, that the activation of sensory-motor properties of words varies as a function of task demands. In particular, our results support the greater role of experience-based semantic information in tasks with multiple memory processes involved (e.g., both lexical-semantic and episodic processes).

The results of the MRT also showed a main effect of vocabulary skills (furtherly confirmed by the cross-task comparison) without any interaction with RE levels. This finding suggests that individuals with deeper word knowledge can more efficiently capitalize on their vocabulary experience, but not on sensory-motor properties of words.

1.7. General discussion

In this study, we tested the effect of a semantic dimension known as *relative embodiment* (RE), on verb processing in Italian. In particular, we collected RE ratings for a heterogeneous set of Italian verbs – which are offered as a novel, openly accessible resource to be reused for further studies – and to test the modulation of RE effects during verb processing across tasks and at the interface of language and memory. To achieve this latter aim, we considered specifically a task with moderate activation of semantic properties (i.e., lexical decision), a task that in English has shown strong involvement of semantic processing (i.e., grammatical decision), and a task with high semantic demands and additional involvement of other memory processes, such as episodic processes (i.e., memory recognition task). Elaborating on multiple representation theories of word meaning (see Muraki et al., 2023), we made predictions based on data collected from English (Sidhu et al., 2014; Sidhu & Pexman, 2016): in particular, we expected semantic effects to increase across tasks in parallel with semantic demands and episodic memory involvement. Additionally, we also acknowledged potential modulations of RE effects due to Italian-specific characteristics, which might manifest especially in tasks where semantic processing is coupled with grammatical processes (i.e., in grammatical decision).

Overall, our results showed that RE is a relevant dimension of Italian verbs, with RE rating values being strongly associated with ratings acquired for English verbs. Beyond that, the results of the three experiments showed clear variation of RE effects on verb processing and confirmed that the extent to which RE became available and relevant during verb processing varied across contexts. On the one hand, we observed a significant effect of RE in lexical decision and memory recognition tasks (stronger in the latter compared to the former), indicating that verbs with higher RE values are more easily retrieved from the mental lexicon and more accurately remembered. On the other hand, we did not find similar effects in the grammatical decision task, where verbs with higher RE values were not processed faster than verbs with lower RE values. The pattern that we observed across tasks is in line with multiple representation theories, which predict that both experience-based and amodal information contribute to shaping the representation of word meaning, in a trade-off that varies as a function of task demands to capitalize on the most salient source of information needed to provide optimal responses (see Connell, 2019; Connell & Lynott, 2013). Notably, the pattern that we observed aligns with predictions from English for two out of three tasks, differing only in the grammatical task. This seems to indicate that task demands interact with language-specific characteristics – such as inflectional richness of grammatical categories – and this interaction drives the activation of either linguistic or experience-based representations.

Focusing on task modulations, lexical decision and memory recognition emerged as the two contexts where RE played a role. These two tasks equally require the activation of experience-based properties of the stimuli yet at a different level of processing. In lexical decisions, we observed that lexical and orthographic properties of the stimuli were more relevant to solving the task, while experience-based properties of verbs were ancillary, thus only indirectly recruited (Yap et al., 2015). This pattern can be interpreted along the lines of models assuming that during shallow lexical operations, semantic representations are not fully activated, but only partially involved via indirect feedback from lexical-orthographic levels (Balota, 1990). Conversely, in tasks requiring full semantic decoding and capitalizing on additional mnestic processing (as in memory recognition, Ballot et al., 2021; Cortese et al., 2010), we observed that experience-based properties of verbs were highly salient and relevant, even when the effect of lexical variables and imageability controlled. Contrary to previous data on noun processing, where imageability was a better predictor of memory recognition performance than motor-related dimensions (Khanna & Cortese, 2021), our results suggest that the activation of sensory-motor and bodily-related experience is especially relevant for encoding verbs in memory, even beyond amodal properties of words and perceptual-based semantic dimensions.

Results from the memory recognition task offer relevant theoretical implications for the understanding of the language-memory interface and the organization of declarative memory. Originating from Tulving's influential model of declarative memory (Tulving, 1972), the debate in the memory field revolved around whether semantic and episodic processes can be described as distinct submodules, with little interaction (De Brigard et al., 2022). The strong dissociability between episodic and semantic processes has been challenged by empirical evidence accumulating over the years, showing that these two facets of declarative memory, in fact, partially overlap and are intertwined at both behavioral and neural levels (Greenberg & Verfaellie, 2010; Renoult et al., 2019). Far from being decisive in this long-standing debate, our data provide additional support to the idea that semantic processes influence specific episodic processes involved in word recognition: in particular, the encoding of stimuli associated with richer semantic representations promotes stronger storage of those items in episodic memory and better performance in identifying them as familiar in the retrieval phase (Stern & Hasselmo, 2009). This type of semantic-to-episodic prompting effect has also been observed in non-word memory recognition studies: for instance, Hovhannisyan et al. (2021) tested visual object and word recognition, showing that semantic features associated with the conceptual representation of a given object influenced memory performance and contributed to better performance in both word and object retrieval skills. Overall, these data contribute to elucidating which factors – including semantic experience-based dimensions, such as RE – facilitate the process of memorizing certain words and not others. Further studies should explore whether the prompt of bodily-related semantic information to episodic processes also extends to more explicit operations, such as recollecting processes (Stern & Hasselmo, 2009), namely if high-RE verbs are not only easier to recognize as old or new stimuli but also if RE facilitates the process of explicitly recalling items from memory, as in free recall tasks (see Makri & Jarrold, 2021). This would extend to language and word memory pieces of evidence already observed in autobiographical memory recall tasks, where specific facets of bodily experience – such as interoception – were shown to influence episodic processes (Messina et al., 2022).

Grammatical decision stands out as the task where we found that bodily information plays no role and is where Italian and English results diverged the most. On the one hand, this finding might cast doubts on the role of bodily information in understanding verbs, on the other hand it is fully compatible with the multiple representation account, as it indicates that the context-modulations of experience-based semantic properties of the stimuli might change from language to language, especially at the syntactic-semantic interface. Our findings are consistent with some previous – albeit still scarce – evidence of cross-linguistic variation of sensory-motor effects in language processing. For instance, Gianelli et al. (2017) compared the effect of verbs combined with pronouns on the activation of motion-related representations in Italian and German. The presence of the 2SG personal pronoun (i.e., "you") as the subject of either action or interaction verbs, triggered a motor effect in Italian but not in German. This effect was interpreted in light of structural differences between Italian and German languages: the non-mandatory status of pronouns in Italian might have promoted perspective taking and the re-activation of motor experience during word processing. Our data align with this type of pattern, showing that greater inflection richness may reduce semantic processes in favour of morphosyntactically-driven operations in a grammatical task (see also Vigliocco et al., 2005, 2011). These data bring additional implications at the methodological level: when it comes to studying the role of verb semantic representation in grammatical decisions in richly inflected languages, such as those within the Romance family, these data suggest that it may be important to consider the interaction between the morphosyntax-semantics interface and structural language-specific differences (Bambini & Canal, 2021).

An additional point worth discussing is the role of individual differences in verb processing and their interaction with RE. In this study, we tested the role of vocabulary and mental imagery skills in an explorative fashion, capitalizing on initial evidence that individual differences interact with semantic processes alongside task and contextual differences (Ibáñez et al., 2023; Pexman, 2020). Our results confirm that individual differences matter in Italian verb processing, highlighting especially the role of between-subject variability at the level of vocabulary skills: participants with deeper vocabulary knowledge and higher lexical quality could capitalize on more extended and robust word knowledge, showing better performance compared to less skilled participants in all three tasks. However, this vocabulary advantage showed remarkable task-related differences, as it directly interacted with bodily properties of the stimuli only in one out of three tasks. In particular, high vocabulary skill resulted in less sensitivity to semantic properties in the lexical decision task but overall faster grammatical decisions and memory recognitions, suggesting a quite broad effect of this skill. Our data confirm patterns already observed in previous studies on English language, showing that individual differences in vocabulary skills interact with experience-based semantic dimensions in lexical decisions (Pexman & Yap, 2018; Yap, Pexman, et al., 2012) and facilitate grammatical decisions (Muraki & Pexman, 2021), but our findings also provide new evidence on the role of vocabulary in word memory recognition.

Conversely, we did not find an effect of mental imagery in any of the three tasks. Previous studies in English showed that, while mental imagery does not seem to affect lexical decisions, it interacts at least with body-object interaction properties during grammatical decisions (Muraki & Pexman, 2021). Here we did not find such effects, which might be due to the fact that – especially in single-word paradigms – the processes underlying mental imagery might diverge from embodied simulations involved in semantic processing (Meteyard et al., 2012; Willems et al., 2010). It is also important to underline that our analysis only addressed the vividness of first-person, third-person, and kinaesthetic motor imagery via a self-administered questionnaire: as pointed out by recent viewpoints (see, for instance, Schwarzkopf, 2023), such questionnaires might not capture imagery skills with high levels of precision as a combination of implicit tasks would do. Nevertheless, as observed by Muraki et al. (2023), the investigation of mental imagery is a promising avenue to unravel the embodied simulation processes underlying conceptual knowledge.

Finally, there are certainly some limitations in our study, which might inspire future investigations on the effects of RE in language and memory tasks. The most relevant limitation concerns the fact that we could compare the effects of RE between the lexical decision and the memory recognition tasks only in a subset of items used in both tasks and using responses from two unbalanced samples of participants. This limitation affects the power of the post-hoc analysis, which nevertheless can inform us concerning the different weights of RE in these two tasks. Interestingly, most of the studies comparing semantic effects across tasks limited their observations to qualitative inspections of differences in effect sizes (see, for instance, Khanna & Cortese, 2021): our study, albeit less robust in terms of power, may inspire more direct comparisons of semantic effects across tasks in future studies. Similarly, we approached the role of individual differences in vocabulary and motor mental imagery skills only in an exploratory fashion, as sample size across experiments was determined to be powered enough to detect effects involving item-level predictors, but not individual difference variables. This limitation reflects the fact that individual differences are still marginally addressed in the literature investigating semantic richness effects, despite their potential relevance to test the predictions of multiple representation accounts (Muraki, Speed, et al., 2023; Muraki & Pexman, 2021): future studies should exploit more robustly their role in modulating semantic effects in word processing to deepen our understanding of the processes underlying their interaction with task properties, contexts, and languages.

Overall, our work documents consistent effects of RE across languages and contributes to enriching the available evidence supporting the idea that richer semantic representations facilitate word processing and memorizing. Moreover, our data show that these effects are flexible, as they not only seem to vary as a function of semantic demands along the language-memory continuum but they become negligible when task demands interact with language-specific grammatical features. Beyond these findings, the semantic richness effects described in our work emphasized also that different memory components (e.g., semantic and episodic processes) likely interact at the language-memory interface: this piece of evidence is expected to significantly contribute to the current theoretical debate on the organization of declarative memory and its relationship with language ability.

STUDY TWO

TESTING MOTOR GROUNDING AND SOMATOTOPIC EFFECTS FOR LITERAL AND FIGURATIVE ACTION-LANGUAGE IN MOTOR NEURON DISEASES[2]

Abstract

Motor Neuron Diseases (MND) may represent a test-ground to assess motor cortex involvement in action-language processing, from literal to figurative uses, in a stringent fashion. Here, a sample of patients with Amyotrophic Lateral Sclerosis (ALS) and another sample with the *SPG4* variant of Hereditary Spastic Paraplegia (HSP-SPG4), a rare condition affecting specifically lower limbs, were tested with two experimental tasks employing upper-limb and lower-limb motion verbs occurring either as isolated words or in literal and metaphorical sentences. Action-language impairment occurred in 24% of ALS patients and approximately 60% of HSP-SPG4 patients (with the expected effector-specific asymmetries) when considering isolated verbs and literal sentences, but was negligible in metaphorical sentences. These findings support the motor simulation account at the literal level, albeit with large individual variation, but cast doubts on the motor grounding of metaphors. Results also encourage a thorough consideration of the neuropsychological profile of MND in accounting for figurative meaning impairment.

[2] This chapter is a manuscript in preparation for submission to a peer-review journal as "Frau, F., Losi, G., Coppa, G., Tonini, E., Bischetti, L., Canal, P., Diamanti, L., Ceroni, M., & Bambini, V., *Testing motor grounding and somatotopic effects for literal and figurative action-language in Motor Neuron Diseases*".

2.1. Introduction

Motor Neuron Diseases (MND) are a group of rare, adult-onset neurodegenerative diseases characterized by progressive loss of upper and lower motor neurons (Bäumer et al., 2014; McDermott & Shaw, 2008). The most common form of MND is Amyotrophic Lateral Sclerosis (ALS), a neurological condition with an incidence between 1.11 and 5.55 cases per 100,000 individuals-years (Wolfson et al., 2023): in ALS, degeneration can affect bulbar and spinal motor neurons, resulting in gradual disruption of motor functioning in different body regions leading to distinctive motor symptomatology, which includes weakness, limb muscular atrophy, and spasticity, as well as dysarthria and dysphagia (especially in bulbar-onset patients) up to respiratory failure (Hardiman, 2011; van Es et al., 2017). The taxonomy of MND also includes a group of even rarer genetic conditions called Hereditary Spastic Paraplegias (HSP), a cluster of mostly autosomal dominant neurological diseases whose most frequent phenotype is associated with the mutation of the *SPG4* gene (Meyyazhagan & Orlacchio, 2022; Solowska & Baas, 2015), occurring with a reported global prevalence of 0.90 cases per 100,000 individuals (Vander Stichele et al., 2022). In the "pure" forms, as in the case of *SPG4* mutation, the clinical profile of HSP includes progressive muscular weakness and spasticity affecting the lower limbs selectively (Parodi et al., 2017). Traditionally, neurological conditions such as ALS or HSP-SPG4 have been ascribed to purely neuromuscular diseases, due to their predominant motor manifestations. Nevertheless, it's nowadays widely accepted that their clinical profile might include a number of non-motor features, with behavioral and neurocognitive changes becoming more frequent in patients as the disease progresses (Faber et al., 2016; Shojaie et al., 2023). Non-motor impairments have been largely documented in the case of ALS, especially in the domains of executive functions, social cognition, and language (Pinto-Grau et al., 2018; Woolley & Rush, 2017), with linguistic difficulties extending also to higher-level language uses, such as the ability to produce and understand discourse and figurative language expressions such as metaphor and humor (Bambini, Arcara, Martinelli, et al., 2016; Bambini, Bischetti, et al., 2020; Bambini & Ceroni, 2021). Less research has been conducted on HSP-SPG4, although initial evidence suggests that a subtle decline of some cognitive domains, such as executive functions and social cognition, might be detected in this population as well (Chamard et al., 2016; Tallaksen et al., 2003). Given their neuropsychological profile in combination with motor impairment, MND have been considered a relevant case to test theories of embodied cognition and motor simulation, in particular with respect to language processing (Bak & Chandran, 2012; Maggio et al., 2022).

In their strongest formulation, Embodied Cognition Theories (ECT; for a summary of different declinations of the theory, see Meteyard et al., 2012) argue that word meaning is represented in modal formats, which capitalize on information derived from perceptual and motor simulations in sensory-motor areas of the brain (Barsalou et al., 2003; Gallese & Lakoff, 2005). Along this vein, so-called 'embodied effects', namely the involvement of motor simulations during action meaning processing, have been investigated in several neuroimaging studies, showing the activation of motor and premotor cortex during the processing of action verbs presented as single words (e.g., *kick*, see Hauk et al., 2004) or embedded in declarative sentences (e.g., *I kick the ball*, see Tettamanti et al., 2005). Interestingly, these studies evidenced not only motor system activations during action-related meaning processing but also the recruitment of motor and premotor neural activations following (at least partially) a somatotopic distribution, that is an effector-specific mapping into the motor cortex (Hauk et al., 2004; Tettamanti et al., 2005). In this context, data from neurological conditions involving motor functioning, became relevant for investigating to what extent an intact motor system is necessary for language processing (Birba et al., 2017). In other words, the question is whether neuropathological alterations in neural activity and connectivity within the motor system lead also to disruption of action meaning representations partially, which would support the idea that action meaning is indeed grounded in motor regions (Gallese & Cuccio, 2018).

Such "motor grounding disruption" has been investigated mostly in patients with Parkinson's disease, reporting evidence that they are, compared to matched healthy individuals, less accurate in action-verb naming and processing (Bocanegra et al., 2017; York et al., 2014), and slower in decision tasks involving action verbs or action sentences (Fernandino et al., 2013). Also MND have been considered in this context, with studies documenting difficulties in processing nouns and verbs related to action in individuals with ALS (Bak & Hodges, 2004; Cousins et al., 2018; Grossman et al., 2008). However, such results are not easily replicated, as recent studies reported no evidence of motor grounding disruption linked to action language in ALS (Aiello et al., 2023; Papeo et al., 2015) and in other neurological conditions, such as Parkinson's disease (Aiello et al., 2022; Argiris et al., 2020) or focal lesions in the motor cortex (Weiss et al., 2016), which raises questions about the ease with which effects of motor impairment on action meaning can be reliably captured at the behavioral level (Meteyard et al., 2012).

A further issue in Embodied Cognition is the extent to which motor grounding effects reported for the processing of action language extend to abstract uses of such expressions, as in the case of figurative meanings. The motor grounding of figurative language has been particularly emphasized in approaches such as the Conceptual Metaphor Theory (Lakoff & Johnson, 1980), which assumes that the comprehension of metaphorical expressions is rooted in patterns of bodily experience (Gallese & Lakoff, 2005; Gibbs, 2005). For instance, a conceptual metaphor such as *AFFECTION IS WARMTH* is understood via the simulation of the experience of being hugged since childhood (Lakoff & Johnson, 1999). Beyond this, some authors showed that metaphors based on action verbs expressing figurative motion (e.g., *The jury grasped the concept*) were linked to activations in areas involved in motor planning and action control, like the inferior and superior parietal lobule (Desai et al., 2011). Similarly, Romero Lauro et al. (2013) reported the activation of the left precentral gyrus during action metaphor comprehension, although such activations did not show any robust somatotopic distribution as observed for action meaning used in literal senses (Hauk et al., 2004; Tettamanti et al., 2005). The activation of the left precentral gyrus and inferior/superior parietal lobule in action metaphor comprehension was also confirmed by a meta-analysis by Yang & Shu (2016), yet again without evidence of a somatotopic distribution of such activations. This latter point is particularly important, as it suggests that current neuroimaging evidence of embodied effects for metaphors is too weak to robustly support the hypothesis that also figurative meaning is grounded in modality-specific systems (Casasanto & Gijssels, 2015). Or, alternatively, it implies that the involvement of sensory-motor simulations might not be crucial for all cases of metaphoric uses but might exhibit a more relevant role in particularly creative uses (Cuccio, 2022). As for clinical populations, albeit more limited, the available evidence seems to reveal a puzzling scenario, with some studies reporting poor understanding in patients with neurodegenerative disorders (Fernandino et al., 2013) but others not replicating such effects (Humphries et al., 2019). In all, the literature on Embodied Cognition is now being challenged with more stringent somatotopic criteria: along this vein, we could expect MND to allow us to shed new light on the frequency of the disruption of motor embodied effects, if present, and to explore more in depth the somatotopic distribution of such effects including also figurative action meanings. This latter point is deemed to be crucial to test the extent to which motor grounding and its disruption in neurological conditions generalize to figurative meaning.

2.1.1. The present work

In this work, we aimed to investigate the involvement of the motor cortex in literal and figurative language processing as predicted by ECT by considering the case of MND. By including a sample with widespread motor impairment (ALS) and a sample with limb-specific impairment (SPG4-HSP), we were able to test not only motor grounding disruption in processing literal and figurative meanings, but also the somatotopicity of these effects in a stringent fashion, i.e., by testing differences in action language processing related to effector-specific motor dysfunction.

To pursue these aims, we designed two experimental tasks, one for verbs in isolation (Experiment 1) and the other for literal and figurative sentences (Experiment 2), that were administered to a sample of patients with ALS (Study 1) and with HSP-SPG4 (Study 2). Both experimental tasks included lower-limb and upper-limb-related action verbs and were used to collect accuracy values and latencies from participants. In particular, Exp 1 compared upper-limb and lower-limb-related verb processing to psychological verbs, while Exp 2 compared literal sentences with upper-limb and lower-limb-related action verbs (e.g., *The boy draws a portrait of his mother* vs. *The soccer player kicks the ball with strength*) to metaphorical ones (e.g., *Alice draws her future with Alberto* vs. *Carlo kicks the criticisms of envious people*).

Following ECT, we predicted that both populations would show a generalized impairment in processing action verbs compared to psychological verbs (i.e., Motor effect) in Experiment 1, consistently with previous findings reporting impaired action verbs processing in individuals with ALS (Grossman et al., 2008; York et al., 2014). Moreover, we expected patients with HSP-SPG4 to show limb asymmetries in these deficits, with lower-limb-related action verbs more impaired than upper-limb-related action verbs (i.e., Limb effect). For Experiment 2, we expected to observe a general Metaphor effect in both populations, following the evidence of impaired figurative language processing in MND (Bambini & Ceroni, 2021). Furthermore, for HSP-SPG4 we expected to observe Limb effects, i.e., worse performance in literal sentences with lower limb verbs, extending to metaphorical sentences. Given the rare incidence of MND and hence the small size of our sample, we based our analysis on a single-case approach, to test the presence of effects in patients at the individual level with respect to a control group. This approach has been already applied to inspect individual deficits in neurodegenerative conditions (Baumard et al., 2018; La Corte et al., 2021) and is considered a powerful tool to test theories through neuropsychological evidence, as it allows to detect patterns of impairment in a finer-grained fashion beyond group differences (Nickels et al., 2022).

Overall, through the investigation of Motor, Limb, and Metaphor effects we expect this study to have a double impact: on the one hand, we expect to provide a more stringent assessment of ECT and the involvement of sensory-motor systems in figurative language processing; on the other hand, we expected to deepen our understanding of language and cognitive impairment in MND.

2.2. Study 1: Amyotrophic Lateral Sclerosis (ALS)

2.2.1. Methods

2.2.1.1. Participants

Fourteen Italian-speaking patients with ALS and 13 healthy matched controls were enrolled to take part in Study 1. All patients were recruited in the General Neurology Department of the National Neurological Institute "Casimiro Mondino" (Pavia, Italy) and had a diagnosis of probable or definite ALS based on Revised El Escorial Criteria and electrodiagnostic criteria (Brooks et al., 2000; de Carvalho et al., 2008). At enrollment, none of the patients exhibited severe neuropsychological impairment (e.g., dementia) or met consensus criteria for diagnosis of frontotemporal dementia (Neary et al., 1998). Patients showing major comorbid medical, neurological, or psychiatric history were also excluded from the study. The local Ethics Committee approved the study. Informed consent was obtained from all participants following the principles of the Declaration of Helsinki.

2.2.1.2. Assessment

All participants in the ALS group underwent a comprehensive evaluation of motor and neurocognitive impairment. Motor impairment was assessed using the Italian version of the Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised (ALS-FRS-R, Cedarbaum et al., 1999). The ALS-FRS-R includes 12 items targeting patient's residual functionality in five domains (i.e., bulbar functions, upper and lower limbs fine motor tasks, gross motor tasks, and respiratory function) on a five-point scale (0 = complete disability, 4 = normal function). The maximum score on this scale is 48.

Neurocognitive impairment was assessed using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS, Poletti et al., 2016), which targets cognitive and behavioral alterations using two sub-scales: 1) the ALS-Specific Functions subscale (max. score: 100), covering cognitive domains specifically impaired in ALS (i.e., executive and social functioning, language, and verbal fluency);

2) the Non-ALS-Specific Functions subscale (max. score: 36), including tasks targeting domains not specifically impaired in ALS (i.e., memory and visuospatial ability).

2.2.1.3. Experimental task 1

Experimental task 1 consisted in a semantic decision task, in which participants had to decide whether a word reflected an action (either physical or mental) or not. In this experiment, participants were presented with verbs presented in isolation, denoting either motion (involving lower and upper limbs) or mental processes.

As material, we selected 90 Italian verbs presented in their 3SG indicative present form, including: i) 30 action verbs related to upper-limb motion (e.g., It. 'disegna', Eng. *draws*), ii) 30 lower-limbs-related action verbs (e.g., It. 'calcia', Eng. *kicks*), and iii) 30 psychological verbs expressing mental and cognitive processes (e.g., It. 'nega', Eng. *denies*). Additionally, 90 filler items, including 45 nouns (e.g., It. 'autista', Eng. *driver*) and 45 adjectives (e.g., It. 'costoso', Eng. *expensive*). The different sets of verbs and the fillers were matched for log-transformed lexical frequency values, extracted from the "Corpus e Lessico di Frequenza dell'Italiano Scritto" (CoLFIS, Bertinetto et al., 2005), and word length (frequency: $F(3,176) = 1.18$, $p = .320$; length: $F(3,176) = 0.55$, $p = .647$).

Prior to running the experiment, we performed a rating task to evaluate limb relatedness for the set of upper-limb and lower-limb action verbs included in the experiment. Thirty-one Italian-speaking young adults (16 females, Age: $M = 25.48$, $SD = 2.72$; Education: $M = 15.90$, $SD = 2.50$) were recruited, each rating half of verbs: all items were presented in the 3SG present form and participants were asked to judge on a 7-point scale the extent to which each verb referred to a) an action performed using hands or arms (1 = no upper-limb involvement, 7 = significant upper-limb involvement); b) an action performed using feet or legs (1 = no lower-limb involvement, 7 = significant lower-limb involvement). The results of the rating task showed that upper-limb relatedness was significantly higher for upper-limb verbs compared to lower-limb ($t(58) = 25.28$, $p < .001$) and psychological verbs ($t(58) = 27.65$, $p < .001$). Conversely, lower-limb relatedness was significantly higher for lower-limb verbs compared to higher-limb ($t(58) = 15.64$, $p < .001$) and psychological verbs ($t(58) = 37.61$, $p < .001$). Item properties for Experiment 1 are summarized in Table 2.1.

Table 2.1. Means and standard deviations for item measures across conditions (upper-limb verbs, lower-limb verbs, psychological verbs, and fillers) in Experiment 1.

Measure	**Upper-limb verbs**	**Lower-limb verbs**	**Psychological verbs**	**Filler**
Lexical frequency (log)	6.27 (1.42)	6.77 (1.63)	6.70 (1.51)	6.66 (1.82)
Length (in character)	2.07 (0.64)	1.98 (0.91)	2.24 (0.40)	2.00 (0.58)
UL relatedness	6.45 (1.42)	2.23 (0.81)	2.59 (0.63)	-
LL relatedness	2.48 (1.27)	6.39 (0.51)	1.83 (0.44)	-

Note. UL = upper-limb; LL = lower-limb.

Selected verbs were embedded in a semantic decision task combined with a go/no-go paradigm, in which participants had to decide whether a word reflected an action (either physical or mental) or not. To overcome the motor slowdown in patients, the go/no-go paradigm was implemented in a vocal modality: participants were instructed to utter the word «VAI[3]» (Eng., *go*) as clearly and fast as possible on a microphone when the word referred to an action, while they had to refrain from responding if the word referred to a noun or an adjective. A total of 180 trials were administered to participants, divided into 6 different blocks to ensure a short break every 30 trials. Items and blocks were randomized. In each trial, a fixation cross appeared on the screen for 500 ms, followed by a target word lasting on the screen for 2750 ms and an inter-trial interval lasting 1250 ms (Figure 2.1A).

[3] The word «VAI» was chosen as the beginning phoneme [v] could be easily captured by the microphone in case of dysarthria.

2.2.1.4. Experimental task 2

The same set of verbs from Experiment 1 was also used to derive the set of sentences for Experiment 2, where participants were presented with action verbs embedded in literal and metaphoric sentences.

In particular, we created 30 plausible literal sentences with upper-limb-related action meaning (e.g., *The boy draws a portrait of his mother*) and 30 plausible literal sentences with lower-limb-related action meaning (e.g., *The soccer player kicks the ball with strength*). From the same verbs, we derived also 30 metaphoric sentences with upper-limb-related figurative meaning (e.g., *Alice draws her future with Alberto*) and 30 metaphoric sentences with lower-limb-related figurative meaning (e.g., *Carlo kicks the criticisms of envious people*). A set of 120 implausible sentences, included as fillers, were created using the psychological verbs (e.g., *The carrot denies the involvement in the robbery*) and the adjectives (e.g., *The philosopher is expensive despite the sales*) used in Experiment 1. For plausible sentences, the use of proper and common nouns as subjects was counterbalanced. Literal, metaphoric, and implausible sentences were matched for the number of words ($F(2,177) = 0.09$, $p = .918$) and the number of characters ($F(2,177) = 0.56$, $p = .572$). Furtherly, upper-limb-related and lower-limb-related sentences across literal and metaphoric conditions did not differ for the number of words (Limb relatedness × Figurativity: $F(1,116) = 0.00$, $p = 1.00$) and letters (Limb relatedness × Figurativity: $F(1,116) = 0.39$, $p = .533$).

Before the experiment, all items were rated for sensicality and familiarity by a group of 60 Italian-speaking young adults (35 females, Age: $M = 26.65$, $SD = 3.85$; Education: $M = 15.80$, $SD = 2.15$). Participants were asked to judge on a 7-point scale the extent to which each sentence was familiar (1 = not familiar, 7 = definitely familiar) and meaningful (1 = not meaningful, 7 = definitely meaningful) to them. All items were randomly divided into four lists and each participant rated only one list. The results of the rating task showed that implausible sentences were rated as less familiar and meaningful than both literal (Familiarity: $t(118) = 45.64$, $p < .001$; Sensicality: $t(118) = 72.88$, $p < .001$) and metaphoric (Familiarity: $t(118) = 25.78$, $p < .001$; Sensicality: $t(118) = 36.32$, p

< .001) sentences. Moreover, familiarity and sensicality scores were higher for literal sentences compared to metaphoric sentences (Familiarity: $t(59) = 7.21$, $p < .001$; Sensicality: $t(59) = 5.74$, $p < .001$). Among literal and metaphoric sentences, upper-limb-related and lower-limb-related sentences were not significantly different in familiarity (Limb relatedness × Figurativity: $F(1,116) = 2.38$, $p = .126$) and sensicality (Limb relatedness × Figurativity: $F(1,116) = 0.38$, $p = .541$). Item properties for Experiment 2 are summarized in Table 2.2.

Table 2.2. Means and standard deviations for item measures across sentences type (literal, metaphoric, and implausible sentences) and verb limb-relatedness (upper-limb verbs and lower-limb verbs) in Experiment 2.

Measure	**Literal sentences**		**Metaphoric sentences**		**Implausible sentences**
	UL-related	**LL-related**	**UL-related**	**LL-related**	
Number of characters	41.27 (2.57)	41.60 (3.04)	41.47 (3.06)	42.47 (2.92)	41.43 (3.66)
Number of words	6.27 (1.41)	6.77 (1.63)	6.27 (1.41)	6.77 (1.63)	6.52 (1.52)
Sensicality	6.02 (0.27)	5.97 (0.28)	5.38 (0.81)	5.46 (0.71)	1.41 (0.40)
Familiarity	5.96 (0.67)	5.71 (0.72)	4.54 (0.93)	4.75 (0.96)	1.36 (0.29)

Note. UL = upper-limb; LL = lower-limb.

Stimuli were embedded in a sensicality judgment task combined with a vocal go/no-go paradigm, in which participants had to decide whether a sentence was meaningful or not. Participants were instructed to utter the word «VAI» as clearly and faster as possible on a microphone when the sentence was plausible, while they had to refrain from responding if the sentence was implausible. To prevent participants' fatigue, the set of items was randomly split into two lists (List A and List B), each including 120 sentences divided into 3 blocks, separated by a short break. Each participant was assigned to either List A or List B. Items and blocks were randomized. Each trial started with a fixation cross lasting on the screen for 500 ms, followed by a target sentence presented on the screen for 4250 ms and an inter-trial interval lasting 1250 ms (Figure 2.1B).

Both experiments were implemented on DMDX Display© software (Forster & Forster, 2003) combined with CheckVocal© software (Protopapas, 2007), which allowed to record reaction times and accuracy values from vocal responses.

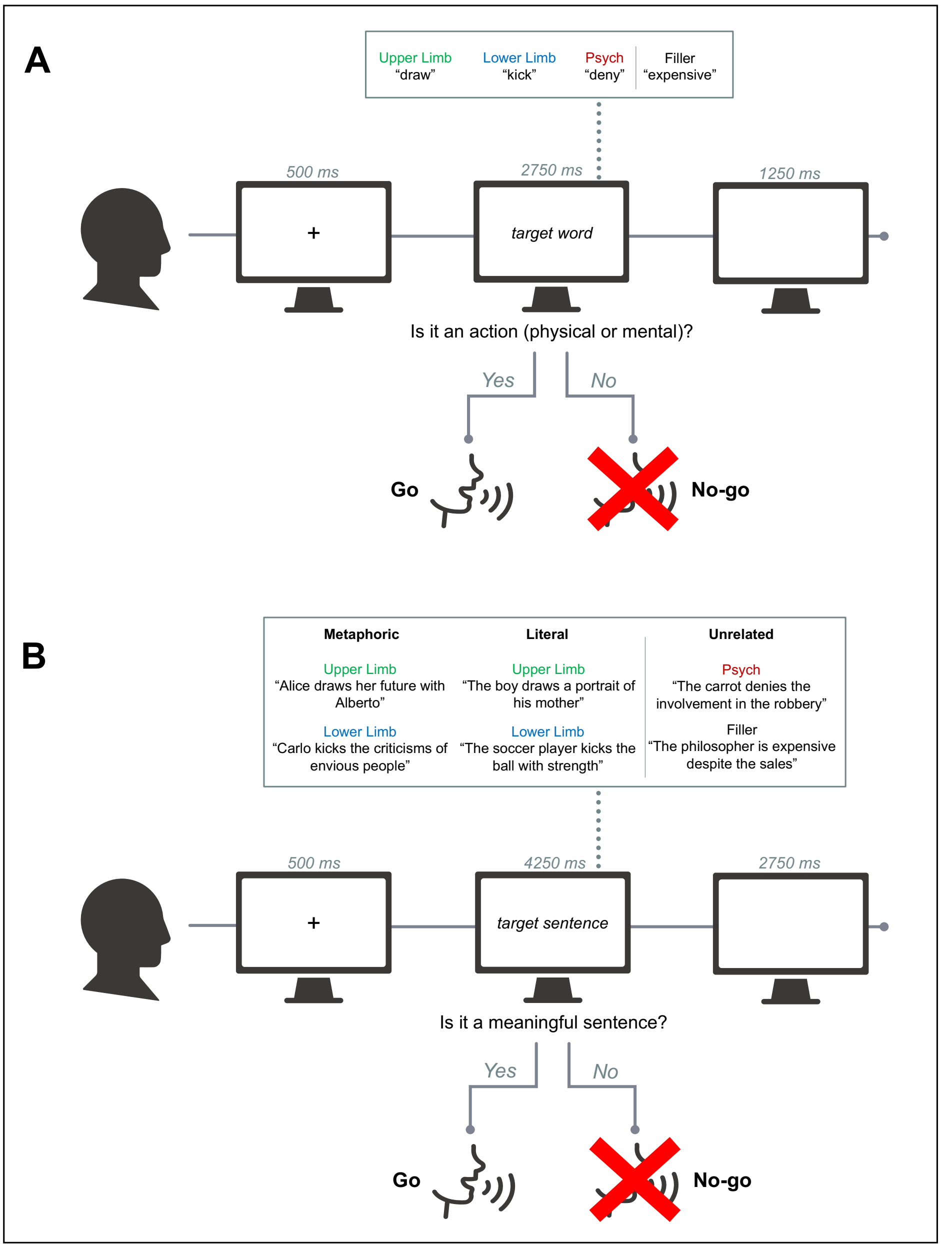


Figure 2.1. Task structure and procedure of Experiment 1 and 2. *(A) Structure of the semantic decision task (with vocal go/no-go) used in Experiment 1. Participant were presented with a single target word and had to decide whether it reflected an action (either physical or mental) or not by verbalizing «VAI» or holding their response. (B) Structure of the sensicality judgment task (with vocal go/no-go) used in Experiment 2. Participants were presented with a single sentence and had to decide whether it was meaningful or not by verbalizing «VAI» or holding their response.*

2.2.1.5. Statistical analysis

In this study, given the relatively small sample size, we adopted a single-case analysis to test for the presence of effects at the individual level with respect to a small control group.

We first computed *z*-scores for mean Accuracy and Reaction Time measures obtained by ALS patients and healthy controls in Experiment 1 and 2. For each hypothesis, we computed a differential effect from *z*-scored Accuracy (Acc) and Reaction Times (RT), operationalized as follows:

1. Motor effect: the difference in accuracy and latencies between psychological and action verbs

$\Rightarrow \textit{Motor effect}_{Acc} = Psych_{Acc} - \frac{Lower_{Acc} + Upper_{Acc}}{2}$ | $\textit{Motor effect}_{RT} = \frac{Lower_{RT} + Upper_{RT}}{2} - Psych_{RT}$

2. Limb effect (computed for action verbs in Experiment 1 and literal/metaphoric sentences in Experiment 2): the difference in accuracy and latencies between upper-limb related and lower-limb related action verbs

$\Rightarrow \textit{Limb effect}_{Acc} = Upper_{Acc} - Lower_{Acc}$ | $\textit{Limb effect}_{RT} = Lower_{RT} - Upper_{RT}$

3. Metaphor effect (computed for Experiment 2 only): the difference in accuracy and latencies between literal and metaphoric sentences

$\Rightarrow \textit{Metaphor effect}_{Acc} = Literal_{Acc} - Metaphoric_{Acc}$ | $\textit{Metaphor effect}_{RT} = Metaphoric_{RT} - Literal_{RT}$

To test differences between patients and controls in Motor, Limb, and Metaphor effects we used Crawford & Howell's (1998) one-tailed Test for a Deficit (TD), a modified *t*-test comparing one case against the mean of the control group. The TD returns modified *t*-statistics alongside a one-tailed *p*-value (alpha set at 0.05) and a standardized effect size indicating Case-Control difference (z_{CC}). In the case of patients with suspected cognitive impairment – namely scoring below the cut-off scores (i.e., obtaining an equivalent score ≤ 1 based on Italian normative data, see Siciliano et al., 2017) at the ECAS test in either one or both subscales – we used the Bayesian Test for a Deficit allowing for Covariates (BTD-Cov), developed by Crawford et al. (2011), to control for the role of cognitive variables (either one or both ECAS Non-ALS-Specific/ALS-Specific Functions *z*-

transformed scores were included in the analysis). The test returns a measure of significance (one-tailed *p*-value) and a standardized effect size indicating Case-Control difference with Covariates (z_{CCC}).

2.2.2. ***Results***

2.2.2.1. Sample description

Three participants (one patient and two controls) could not complete one of the two experiments and were therefore excluded from the analysis. The final sample consisted of 13 patients (5 females, Age: M = 65.54, SD = 10.41; Education: M = 9.31, SD = 2.36) and 11 controls (6 females, Age: M = 66.91, SD = 9.04; Education: M = 8.82, SD = 1.89). Twelve patients had a predominant spinal onset type, while one patient had a both spinal and bulbar type of onset. The mean illness duration in months for patients with ALS was 41.69 (SD = 37.37). Patients' mean ALS-FRS-R total score was 29.77 (SD = 9.24), with a mean ECAS total score of 107.75 (SD = 16.46). Two patients (SL005 and SL010) scored below the cut-off in one or both ECAS subscales (ECAS Non-ALS-Specific and ALS-Specific Functions scores). One participant from the control group scored below the cut-off in the ECAS total score and was therefore excluded from the analysis. The remaining control participants had a mean ECAS total score of 121.05 (SD = 5.95). Demographic, motor, and cognitive measures of the ALS group are reported in Table 2.3.

Table 2.3. Demographic, clinical, and cognitive measures of the participants in the ALS group (cognitive measures were corrected for age and education).

Patient	Age	Education	Onset type	Limb onset	Duration (months)	ALS-FRS-R (Tot)	ECAS-Sp	ECAS-NSp	ECAS-Tot
SL001	69	13	Spinal	LL/UL	98	24	87.56	31.25	118.82
SL002	62	8	Spinal	LL	34	13	93.79	35.79	123.71
SL003	43	8	Spinal	LL	34	25	78.36	23.58	106.23
SL004	57	8	Spinal	LL/UL	18	21	80.45	28.45	104.66
SL005	70	13	Spinal	UL	17	28	54.70*	15.70*	67.89*
SL006	81	9	Spinal	LL	25	21	91.70	30.70	119.89
SL007	64	8	Spinal	UL	85	35	92.79	32.79	119.71
SL008	69	8	Spinal	LL	8	41	72.79	19.79*	86.71
SL009	60	8	Spinal	LL	12	39	82.79	21.79	98.71
SL010	80	6	Spinal/bulbar	LL/UL	17	41	77.92	34.92	104.78
SL011	75	13	Spinal	LL	46	40	94.70	32.70	124.89
SL012	56	8	Spinal	LL	126	35	93.45	26.20	119.66
SL013	66	11	Spinal	LL	22	24	78.56	25.56	103.82

Note. LL = lower-limb onset; UP = upper-limb onset; ALS-FRS-R = Amyotrophic Lateral Sclerosis Functional Rating Scale Revised; ECAS-Sp/-NSp/-Tot = Edinburgh Cognitive and Behavioural ALS Screen Specific Functions/Non-specific Functions/Total scores.
** indicates a score below the cut-off relative to Italian normative data (equivalent score ≤ 1) as reported in* Siciliano et al. (2017).

2.2.2.2. Experimental Task 1

Table 2.4 and Figure 2.2 report the descriptive statistics for Accuracy and Reaction Time values obtained by patients with ALS and healthy controls in Experiment 1. The analysis of Reaction Times was conducted on correct answers only, after trimming latencies faster than 150 msec and exceeding |2.5| standard deviations from participants' means across conditions (1.44% of correct answers).

Table 2.4. Descriptive statistics (mean and standard deviation) of Experiment 1 (Accuracy and Reaction Times).

	Condition	ALS group	Control group
Accuracy	*Psychological*	0.81 (0.40)	0.80 (0.40)
	Upper Limb	0.88 (0.32)	0.84 (0.37)
	Lower Limb	0.84 (0.37)	0.87 (0.34)
Reaction Times	*Psychological*	1147.84 (292.21)	1058.73 (256.10)
	Upper Limb	1106.45 (285.92)	993.07 (215.01)
	Lower Limb	1149.06 (294.04)	1050.13 (275.34)

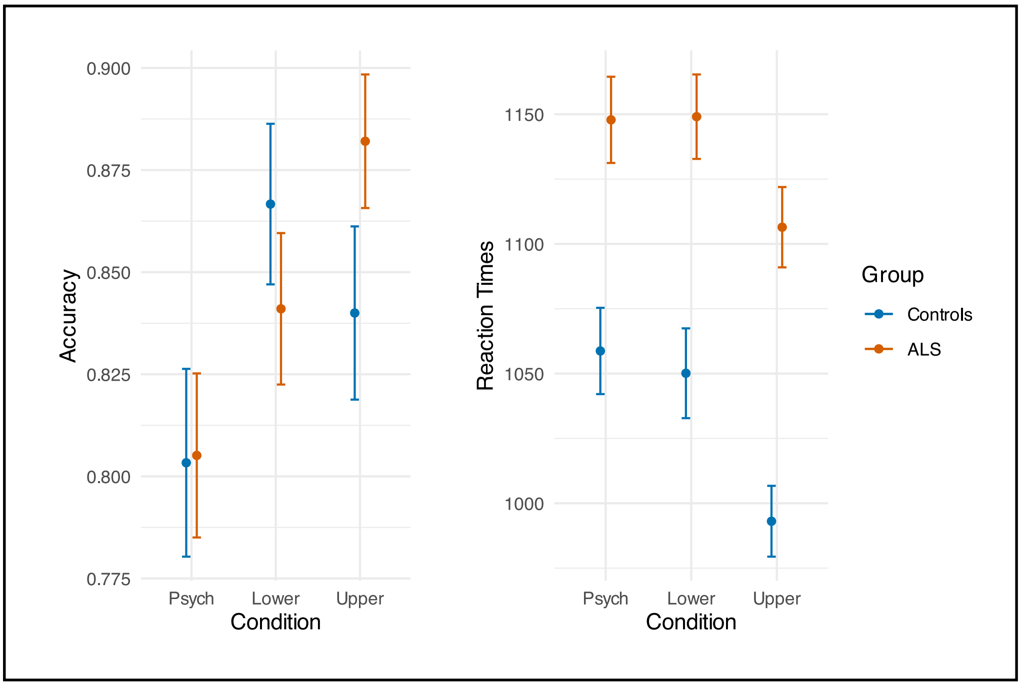


Figure 2.2. Group-level Accuracy and Reaction Time values for patients with ALS from Experiment 1. *Mean values of Accuracy and Reaction Times obtained by patients with ALS and healthy controls in each condition of Experiment 1, namely psychological verbs (Psych), lower-limb-related (Lower), and upper-limb-related (Upper) action verbs (error bars indicate the standard error).*

The single-case analysis on Accuracy measures of Experiment 1 showed that no patient from the ALS group exhibited a Motor effect (i.e., lower accuracy values for action verbs compared to psychological verbs in the ALS group relative to the control group). Two patients (15.4% of the sample) exhibited a significant (SL010: $t_{modified}(9) = 2.90$, $p = .009$, $z_{CC} = 3.04$) or marginally significant (SL009: $t_{modified}(9) = 1.58$, $p = .074$, $z_{CC} = 1.66$) Limb effect (i.e., lower accuracy values for lower-limb action verbs compared to upper-limb action verbs in the ALS group relative to the control group). Since these two patients scored above the cut-off in the ECAS Non-ALS-Specific/ALS-Specific Functions subscales, the role of cognitive variables was not controlled for. Individual means and standard errors for Accuracy values obtained by patients with ALS across conditions are reported in Figure 2.3A.

The single-case analysis on Reaction Time measures of Experiment 1 showed that one patient (SL010, i.e., 7.7% of the sample) exhibited a significant Motor effect ($t_{modified}(9) = 3.09$, $p = .006$,

z_{CC} = 3.24). The same patients exhibited also a significant Limb effect ($t_{modified}(9)$ = 3.51, p = .003, z_{CC} = 3.69). No other patients exhibited Motor or Limb effects. Individual means and standard errors for Reaction Time values obtained by patients with ALS across conditions are reported in Figure 2.3B.

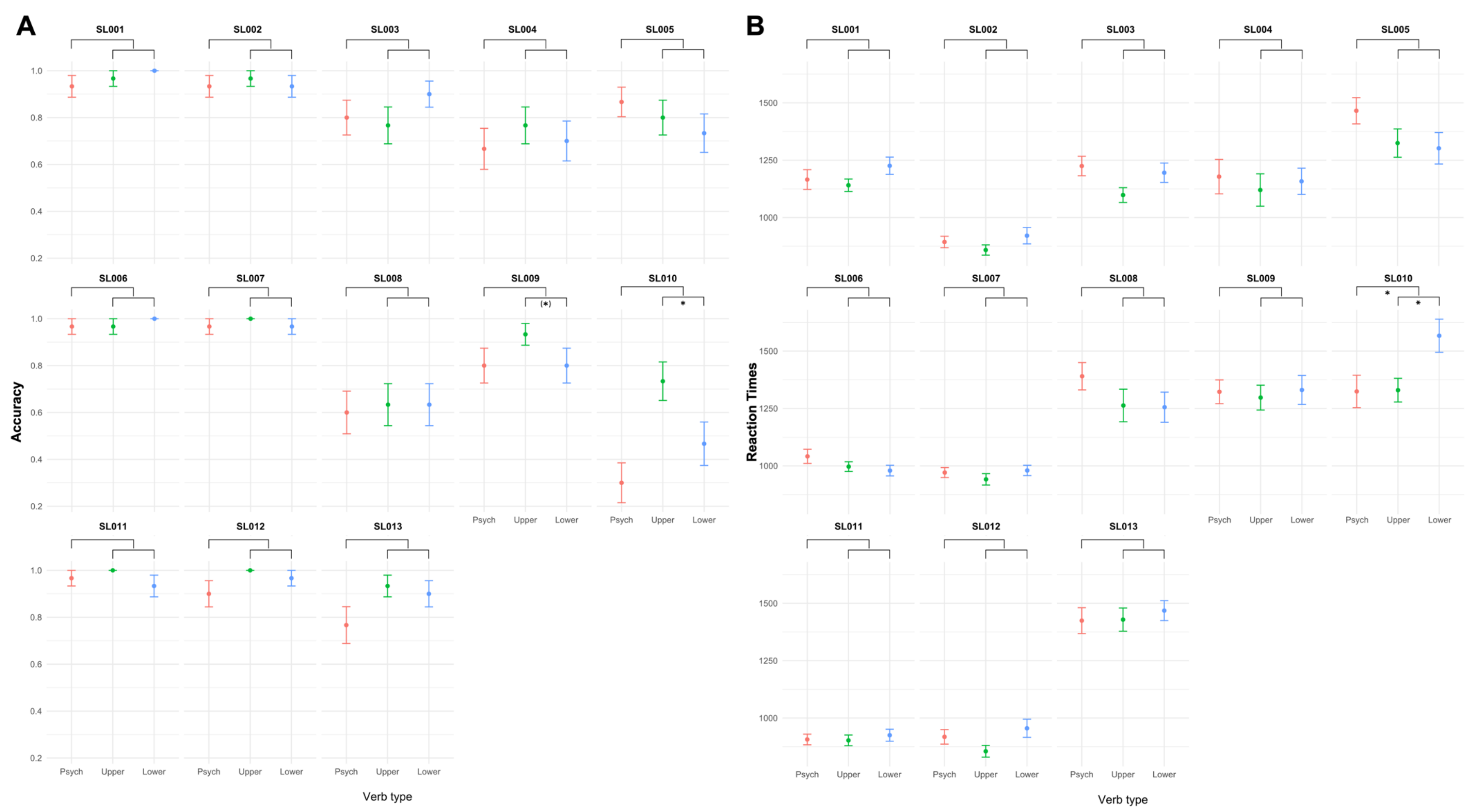


Figure 2.3. Individual Accuracy (A) and Reaction Time (B) values for patients with ALS from Experiment 1 (means and standard errors). *Squared brackets indicate the comparison between psychological (Psych) and motor verbs and the comparison between upper-limb-related (Upper) and lower-limb-related (Lower) verbs relative to the mean performance of the control group. Significant (p < .05) and marginal (p < .10) effects are reported before controlling for cognitive measures and are indicated with the symbols * and (*), respectively.*

2.2.2.3. Experimental Task 2

Table 2.5 and Figure 2.4 report the descriptive statistics for Accuracy and Reaction Time values obtained by patients with ALS and healthy controls in Experiment 2. As in Experiment 1, the analysis of Reaction Times was conducted on correct answers only (trimming latencies faster than 150 msec and exceeding |2.5| standard deviations from participants' means across conditions, which accounted for the 1.19% of correct answers).

Table 2.5. Descriptive statistics (mean and standard deviation) of Experiment 2 (Accuracy and Reaction Times).

	Condition	**ALS group**	**Control group**
Accuracy	*Literal Upper Limb*	0.91 (0.28)	0.95 (0.21)
	Literal Lower Limb	0.91 (0.29)	0.97 (0.16)
	Metaphor Upper Limb	0.73 (0.44)	0.82 (0.39)
	Metaphor Lower Limb	0.74 (0.44)	0.83 (0.38)
Reaction Times	*Literal Upper Limb*	2317.57 (531.72)	2094.60 (455.07)
	Literal Lower Limb	2453.52 (544.56)	2207.20 (424.24)
	Metaphor Upper Limb	2601.03 (498.48)	2379.92 (462.40)
	Metaphor Lower Limb	2584.02 (528.87)	2384.99 (506.16)

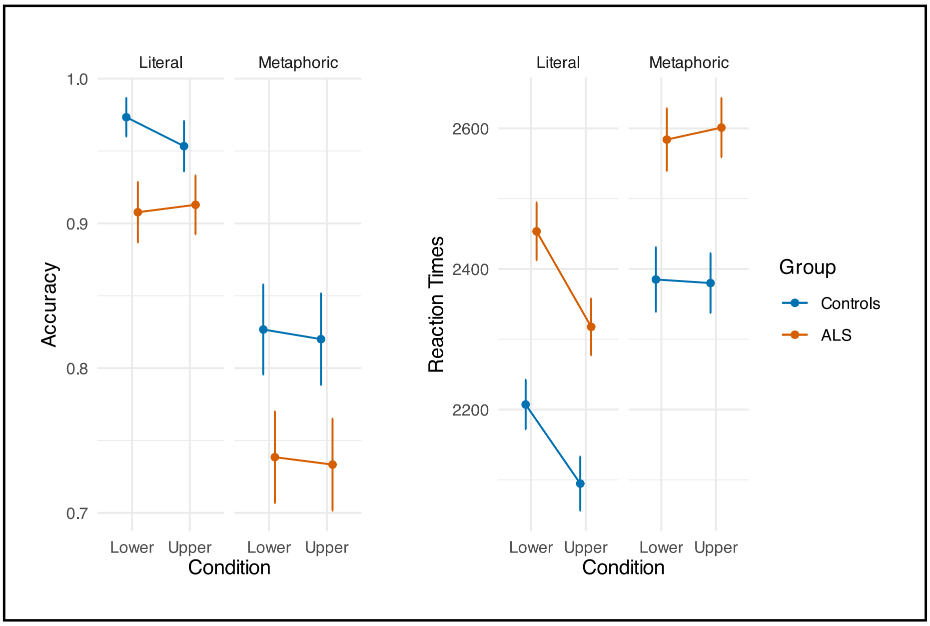


Figure 2.4. Group-level Accuracy and Reaction Time values for patients with ALS from Experiment 2. *Mean values of Accuracy and Reaction Times obtained by patients with ALS and healthy controls in each condition of Experiment 2, namely lower-limb-related (Lower) and upper-limb-related (Upper) action verbs used in Literal and Metaphoric sentences (error bars indicate the standard error).*

The single-case analysis on Accuracy measures of Experiment 2 showed that the Metaphor effect (i.e., lower accuracy values for metaphoric sentences compared to literal sentences in the ALS group relative to the control group) was marginally significant in one patient, namely 7.7% of the sample (SL008: $t_{modified}(9) = 1.61$, $p = .071$, $z_{CC} = 1.69$). This patient obtained a score below the cut-off in both ECAS subscales (i.e., ECAS Specific and Non-specific Functions), and the Metaphor effect became significant after controlling for ECAS subscores ($z_{CCC} = 2.88$, $p = .046$). The Limb effect for literal sentences was significant in two patients (15.4% of the sample), namely SL005 ($t_{modified}(9) = 2.71$, $p = .012$, $z_{CC} = 2.85$) and SL008 ($t_{modified}(9) = 1.89$ $p = .046$, $z_{CC} = 1.98$). Both patients scored below the cut-off in one or both ECAS subscales, and the Limb effects remained significant after controlling for the role of cognitive covariates (SL005: $z_{CCC} = 5.86$, $p = .009$; SL008: $z_{CCC} = 3.72$, $p = .009$). The Limb effect for metaphoric sentences was marginally significant in one patient (7.7% of the sample; SL008: $t_{modified}(9) = 1.46$, $p = .089$, $z_{CC} = 1.53$), but it became not significant after controlling for ECAS subscores ($z_{CCC} = 1.75$, $p = .119$). Individual means and standard errors for Accuracy values obtained by patients with ALS across conditions figurative types are reported in Figure 2.5A.

The single-case analysis on Reaction Time measures of Experiment 2 showed that two patients (15.4% of the sample) exhibited a marginal Metaphor effect (SL009: $t_{modified}(9) = 1.76$, $p = .056$, $z_{CC} = 1.85$; SL013: $t_{modified}(9) = 1.50$, $p = .084$, $z_{CC} = 1.57$). The Limb effect in literal sentences was observed in one patient (7.7% of the sample; SL013: $t_{modified}(9) = 2.01$, $p = .038$, $z_{CC} = 2.11$). The Limb effect in metaphor sentences was significant in one other patient (7.7% of the sample; SL010: $t_{modified}(9) = 1.93$, $p = .043$, $z_{CC} = 2.02$). Individual means and standard errors for Reaction Time values obtained by patients with ALS across conditions figurative types are reported in Figure 2.5B.

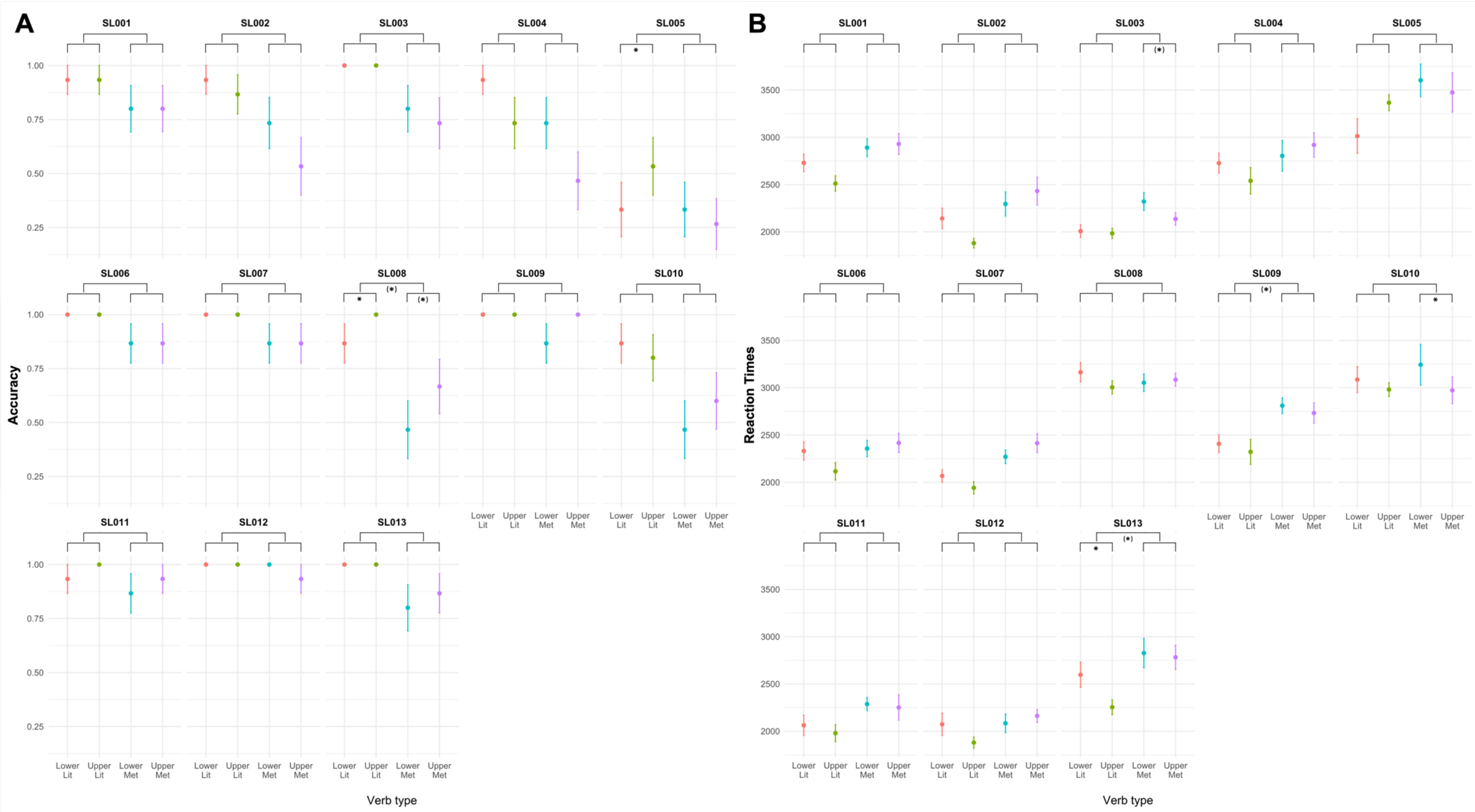


Figure 2.5. Individual Accuracy (A) and Reaction Time (B) values for patients with ALS from Experiment 2 (means and standard errors). *Squared brackets indicate the comparison between literal (Lit) and metaphoric (Met) sentences, as well as the comparison between upper-limb-related (Upper) and lower-limb-related (Lower) verbs in either literal or metaphoric sentences relative to the mean performance of the control group. Significant ($p < .05$) and marginal ($p < .10$) effects are reported before controlling for cognitive measures and are indicated with the symbols * and (*), respectively.*

The summary of Experiment 1, with the percentage of ALS patients showing Motor and Limb effects for single verbs, and Experiment 2, with the percentage of patients showing Metaphor and Limb effects in literal and metaphoric sentences, is reported in Figure 2.6A.

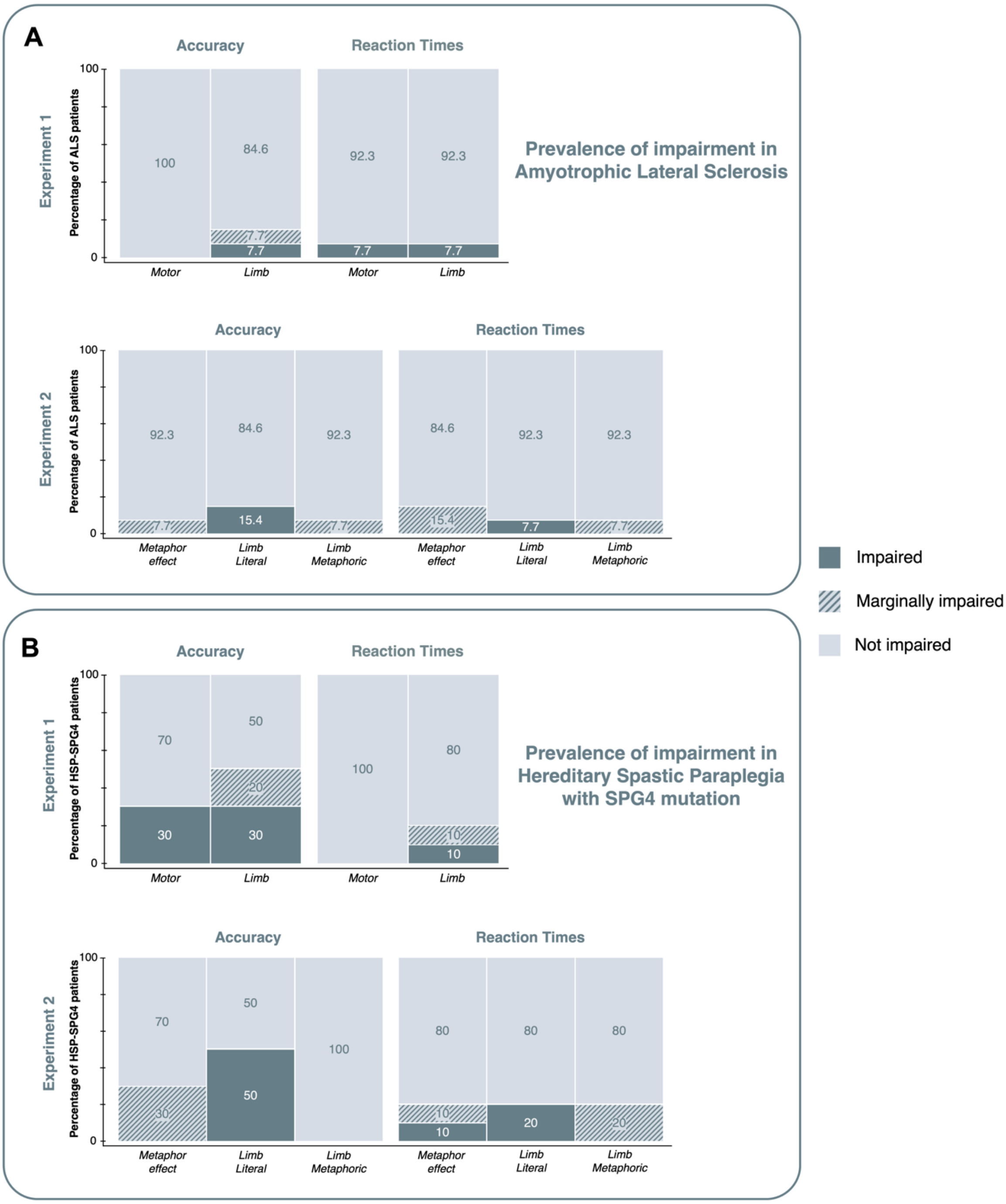


Figure 2.6. Frequency of impairment in ALS (A) and HSP-SPG4 (B) across experiments. *The panels report the percentage of participants with ALS and HSP-SPG4 showing Motor or Limb effects in Experiment 1, as well as Metaphor and Limb (in literal or metaphoric sentences) effects in Experiment 2, before controlling for the role of cognitive measures.*

2.2.3. *Discussion*

In Study 1, we recruited a group of participants with ALS and we tested the presence of Motor effects for action verbs presented as single words (Experiment 1), namely higher difficulty in processing action verbs (as a whole) compared to psychological verbs. We then tested whether patients would exhibit any difficulty in processing action verbs denoting motion of lower-limbs compared to upper-limb motion verbs (Limb Effect), presented in isolation (Experiment 1) or embedded in literal and metaphoric sentences (Experiment 2). By using both literal and metaphoric sentences in Experiment 2, we also investigated whether patients would show a general difficulty in processing figurative meanings compared to literal ones (Metaphor effect). We expected patients with ALS to show a marked Motor effect in Experiment 1 and a Metaphor effect in Experiment 2, with no specific Limb effects.

The results from Experiment 1 showed that only a small percentage of participants exhibited a Motor effect (only 7.7% of participants) or a Limb effect (15.4% of participants, cumulating unique cases exhibiting marginal and significant effects in accuracy or reaction time measures). In Experiment 2, 23.1% of participants presented a marginal Metaphor effect only in either latencies or accuracy scores (cumulating unique cases for accuracy and reaction time values), while the Limb effect was significantly present only in 23.1% of participants for literal sentences and marginally present in 15.4% of cases for metaphoric sentences (considered again unique cases cumulatively for accuracy and reaction times values).

Overall, the results from Experiment 1 showed that difficulties in processing action-related meanings were not widespread in ALS and were detectable more in accuracy values than in reaction times. Our findings suggest that effects of motor grounding disruption are limited in ALS and might emerge when assessed with finer-grained case-level observations, while might not be captured at the group level (Aiello et al., 2023; Papeo et al., 2015). Going further, the results from Experiment 2 confirm that individuals with ALS might exhibit difficulties in understanding figurative meanings, even if the frequency of such deficits in our sample is relatively lower compared to previous studies (see Bambini et al., 2016, 2020, reporting percentages of receptive pragmatic deficits ranging between 27% and 36% of cases, respectively).

2.3. Study 2: Hereditary Spastic Paraplegia positive to SPG4 mutation (HSP-SPG4)

2.3.1. Methods

2.3.1.1. Participants

Ten Italian-speaking patients with HSP positive to SPG4 gene mutation (5 females, Age: M = 54.90, SD = 7.82; Education: M = 10.5, SD = 5.95) and 10 healthy matched controls (5 females, Age: M = 57.50, SD = 7.89; Education: M = 13.20, SD = 4.13) were enrolled to take part in Study 2. The diagnosis was genetically confirmed at the General Neurology Department of the National Neurological Institute "Casimiro Mondino" (Pavia, Italy). Patients were included in the study only if they were not showing severe neurocognitive impairment (e.g., dementia) or major comorbid medical, neurological, or psychiatric history. The local Ethics Committee approved the study. Informed consent was obtained from all participants following the principles of the Declaration of Helsinki.

2.3.1.2. Assessment

All participants in the HSP-SPG4 group were evaluated for motor functionality and neurocognitive impairment. Motor impairment was assessed using the Italian version of the Spastic Paraplegia Rating Scale (SPRS, Schule et al., 2006). The SPRS includes 13 items assessing different aspects of motor functionality (i.e., walking, gait quality, limb spasticity, weakness, etc.) on a five-point scale (0 = no affection, 4 = most severe affection). The maximum score on the SPRS is 52.

Considering the similarity in the cognitive profile between ALS and HSP-SPG4 patients (Murphy et al., 2009), neurocognitive impairment was assessed again using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS, Poletti et al., 2016), as described in Section 2.2.1.1.

2.3.1.3. Experimental Tasks 1 and 2

The participants taking part in Study 2 were administered Experiments 1 and 2 following the same materials and procedure used in Study 1 (see Sections 2.2.1.2. and 2.2.1.3).

2.3.1.4. Statistical analysis

To detect effects at the individual level, we adopted the same single-case approach used in Study 1 (see Section 2.2.1.4).

2.3.2. Results

2.3.2.1. Sample description

All participants completed both Experiments 1 and 2. For patients with HSP-SPG4, the mean illness duration in months was 41.69 (*SD* = 37.37), with a mean SPRS total score of 20.33 (*SD* = 8.31) and a mean ECAS total score of 94.25 (*SD* = 21.18). Only one patient could not perform the motor evaluation due to medical difficulties (i.e., recent right knee arthroscopy). Three patients (SPG_007, SPG_008 and SPG_009) scored below the cut-off in either one or both ECAS subscales (ECAS Non-ALS-Specific and ALS-Specific Functions scores). All participants from the control group obtained an ECAS total score above the cut-off (*M* = 118.41, *SD* = 8.24). Demographic, motor, and cognitive measures of the HSP-SPG4 group are reported in Table 2.6.

Table 2.6. Demographic, clinical, and cognitive measures of the HSP-SPG4 group (cognitive measures were corrected for age and education).

Patient	Age	Education	SPRS	ECAS-Sp	ECAS-NSp	ECAS-Tot
SPG_001	59	15	20	60.47	25.00	87.47
SPG_002	51	9	15	81.78	24.55	106.78
SPG_003	53	24	-	73.47	28.00	103.48
SPG_004	53	13	15	73.78	29.55	103.78
SPG_005	66	8	22	90.71	26.92	115.71
SPG_006	59	8	10	93.66	23.20	115.66
SPG_007	57	3	22	45.50*	14.60*	56.50*
SPG_008	64	5	38	55.22*	19.31*	74.54*
SPG_009	40	12	21	45.98*	24.98	68.35*
SPG_010	47	8	15	83.58	26.63	110.23

Note. SPRS = Spastic Paraplegia Rating Scale; ECAS-Sp/-NSp/-Tot = Edinburgh Cognitive and Behavioural ALS Screen Specific Functions/Non-specific Functions/Total scores.

** indicates a score below the cut-off relative to Italian normative data (equivalent score ≤ 1), as reported in* Siciliano et al. (2017).

2.3.2.2. Experimental Task 1

Table 2.7 and Figure 2.7 report the descriptive statistics for Accuracy and Reaction Time values obtained by patients with HSP-SPG4 and healthy controls in Experiment 1. We analyzed Reaction Times on correct answers only, after trimming latencies faster than 150 msec and exceeding |2.5| standard deviations from participants' means across conditions as in Study 1 (1.44% of correct answers were removed).

Table 2.7. Descriptive statistics (mean and standard deviation) of Experiment 1 (Accuracy and Reaction Times).

	Condition	HSP-SPG4 group	Control group
Accuracy	*Psychological*	0.75 (0.43)	0.96 (0.20)
	Upper Limb	0.83 (0.38)	0.98 (0.15)
	Lower Limb	0.81 (0.39)	0.97 (0.16)
Reaction Times	*Psychological*	1111.87 (260.20)	1021.64 (488.97)
	Upper Limb	1058.36 (212.50)	927.68 (223.92)
	Lower Limb	1049.25 (230.67)	925.40 (259.21)

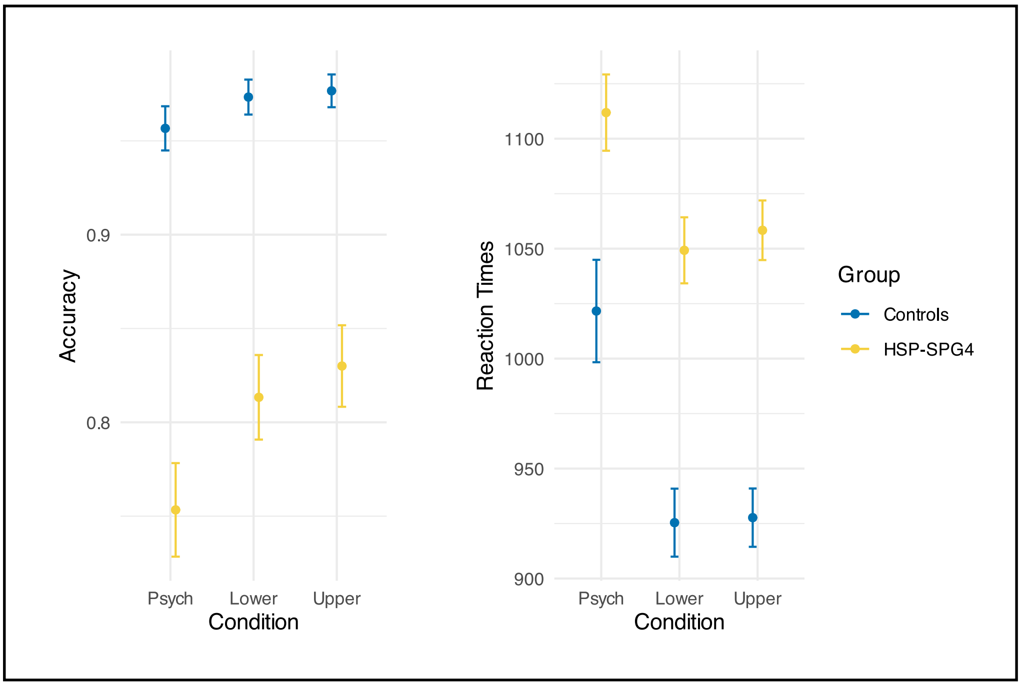


Figure 2.7. Group-level Accuracy and Reaction Time values for patients with HSP-SPG4 from Experiment 1. *Mean values of Accuracy and Reaction Times obtained by patients with HSP-SPG4 and healthy controls in each condition of Experiment 1, namely psychological verbs (Psych), lower-limb-related (Lower), and upper-limb-related (Upper) action verbs (error bars indicate the standard error).*

The single-case analysis on Accuracy measures of Experiment 1 showed that three patients from the HSP-SPG4 group (30% of the sample) exhibited a significant Motor effect (SPG_005: $t_{modified}(9) = 1.91$, $p = .045$, $z_{CC} = 1.99$; SPG_007: $t_{modified}(9) = 4.14$, $p = .001$, $z_{CC} = 4.34$; SPG_008: $t_{modified}(9) = 3.39$, $p = .004$, $z_{CC} = 3.57$). We controlled for the role of cognitive variables for patients SPG_007 and SPG_008, as both patients scored below the cut-off in ECAS subscales (i.e., ECAS Specific and Non-specific Functions), yet Motor effect remained significant (SPG_007: $z_{CCC} = 6.93$, $p = .031$; SPG_008: $z_{CCC} = 5.58$, $p = .034$). The Limb effect was significant in three patients (30% of the sample; SPG_004: $t_{modified}(9) = 3.19$, $p = .005$, $z_{CC} = 3.35$; SPG_005: $t_{modified}(9) = 3.19$, $p = .005$, $z_{CC} = 3.35$; SPG_007: $t_{modified}(9) = 3.19$, $p = .005$, $z_{CC} = 3.35$) and marginally significant in two more patients (20% of the sample; SPG_001: $t_{modified}(9) = 1.51$, $p = .082$, $z_{CC} = 1.56$; SPG_003: $t_{modified}(9) = 1.51$, $p = .082$, $z_{CC} = 1.59$). The Limb effect in patient SPG_007 became non-significant after controlling for ECAS subscores ($z_{CCC} = 1.70$, $p = .311$). Individual means and standard errors for

Accuracy values obtained by patients with HSP-SPG4 across conditions are reported in Figure 2.8A.

The single-case analysis on Reaction Time measures of Experiment 1 showed that no patients exhibited a Motor effect, while a Limb effect was significant in one patient (10 % of the sample; SPG_007: $t_{modified}(9) = 2.74$, $p = .012$, $z_{CC} = 2.87$) and marginally significant in one other patient (10% of the sample; SPG_002: $t_{modified}(9) = 1.44$, $p = .091$, $z_{CC} = 1.51$). The effect observed in SPG_007 became non-significant after controlling for ECAS subscores ($z_{CCC} = 3.10$, $p = .180$). Individual means and standard errors for Reaction Time values obtained by patients with HSP-SPG4 across conditions are reported in Figure 2.8B.

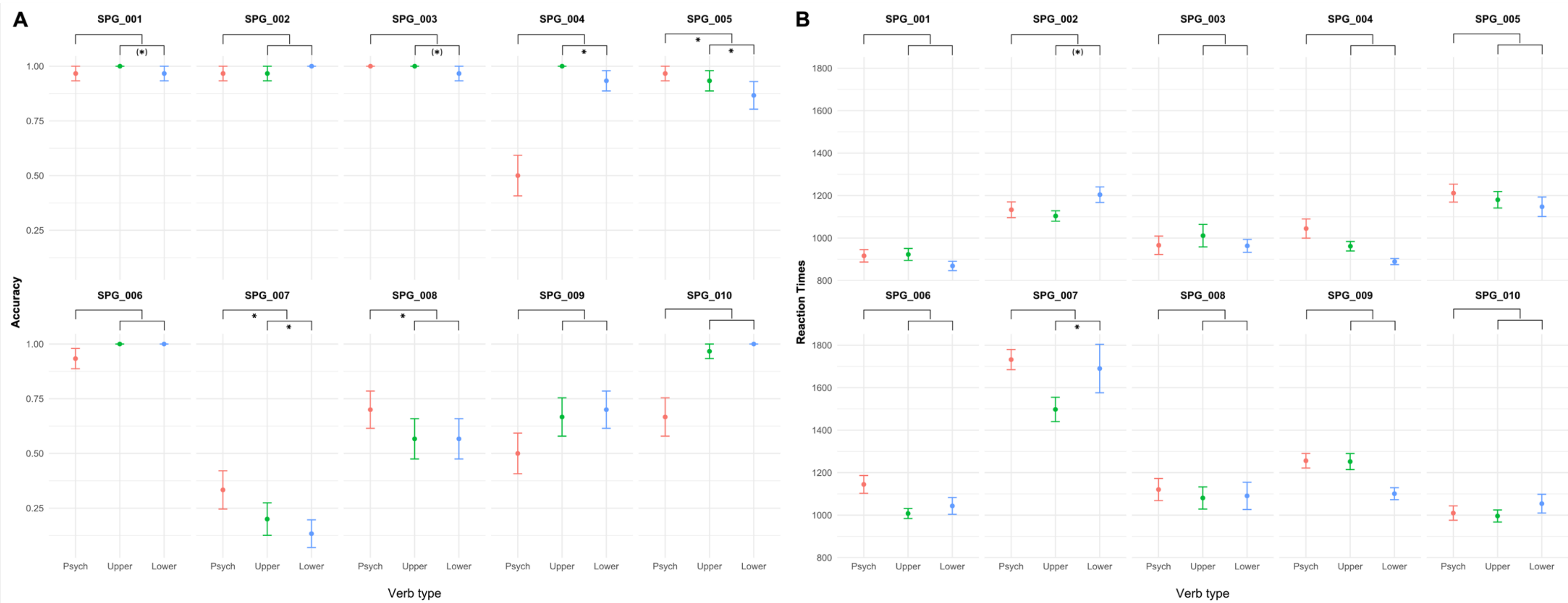


Figure 2.8. Individual Accuracy (A) and Reaction Time (B) values for patients with HSP-SPG4 from Experiment 1 (means and standard errors). *Squared brackets indicate the comparison between psychological (Psych) and motor verbs and the comparison between upper-limb-related (Upper) and lower-limb-related (Lower) verbs relative to the mean performance of the control group. Significant ($p < .05$) and marginal ($p < .10$) effects are reported before controlling for cognitive measures and are indicated with the symbols * and (*), respectively.*

2.3.2.3. Experimental Task 2

Table 2.8 and Figure 2.9 report the descriptive statistics for Accuracy and Reaction Time values obtained by patients with ALS and healthy controls in Experiment 2. Reaction Times were analyzed after removing incorrect answers. Latencies of correct answers were trimmed to remove responses faster than 150 msec and exceeding |2.5| standard deviations from participants' means across conditions (1.10% of correct answers were removed).

Table 2.8. Descriptive statistics (mean and standard deviation) of Experiment 2 (Accuracy and Reaction Times).

	Condition	**HSP-SPG4 group**	**Control group**
Accuracy	*Literal Upper Limb*	0.87 (0.33)	0.99 (0.12)
	Literal Lower Limb	0.83 (0.37)	1.00 (0.00)
	Metaphor Upper Limb	0.65 (0.35)	0.85 (0.35)
	Metaphor Lower Limb	0.67 (0.47)	0.82 (0.39)
Reaction Times	*Literal Upper Limb*	2311.29 (555.68)	1969.20 (392.71)
	Literal Lower Limb	2407.43 (576.80)	2048.83 (412.63)
	Metaphor Upper Limb	2512.07 (512.19)	2253.39 (459.30)
	Metaphor Lower Limb	2505.27 (476.25)	2210.32 (379.45)

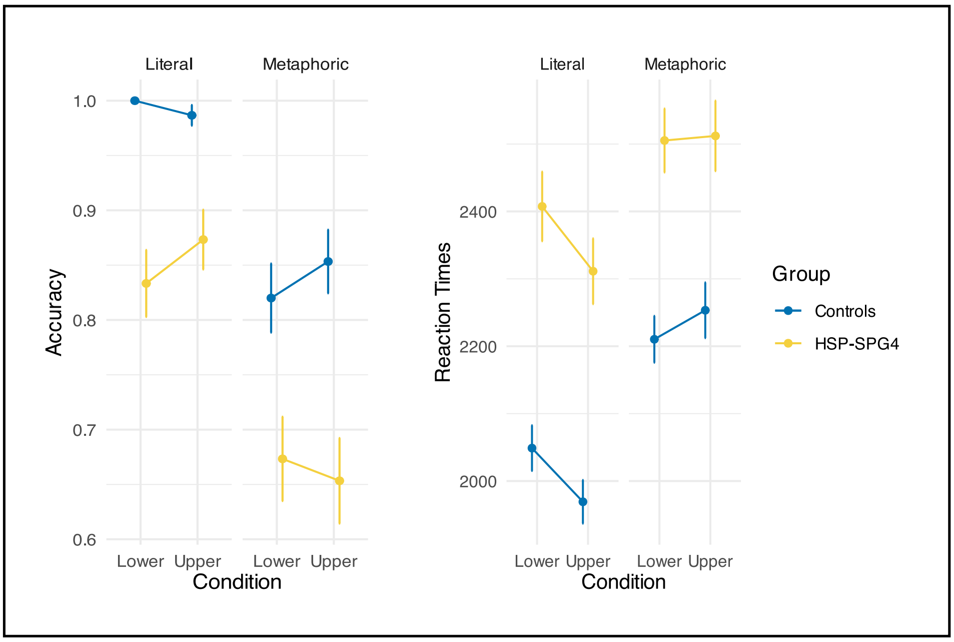


Figure 2.9. Group-level Accuracy and Reaction Time values for patients with HSP-SPG4 from Experiment 2. *Mean values of Accuracy and Reaction Time values obtained by HSP-SPG4 and Control groups in Experiment 2 (error bars indicate standard error).*

The single-case analysis on Accuracy measures of Experiment 2 showed that the Metaphor effect was marginally significant in three patients (30% of the sample; SPG_002: $t_{modified}(9) = 1.45$, $p = .091$, $z_{CC} = 1.52$; SPG_003: $t_{modified}(9) = 1.79$, $p = .054$, $z_{CC} = 1.87$; SPG_004: $t_{modified}(9) = 1.45$, $p = .091$, $z_{CC} = 1.52$). The Limb effect for literal sentences was significant in five patients (50% of the sample; SPG_001: $t_{modified}(9) = 2.71$, $p = .012$, $z_{CC} = 2.85$; SPG_002: $t_{modified}(9) = 4.97$, $p < .001$, $z_{CC} = 5.22$; SPG_006: $t_{modified}(9) = 4.97$, $p < .001$, $z_{CC} = 5.22$; SPG_008: $t_{modified}(9) = 2.71$, $p = .012$, $z_{CC} = 2.85$; SPG_009: $t_{modified}(9) = 4.97$, $p < .001$, $z_{CC} = 5.22$). The Limb effect became non-significant for patient SPG_008 ($z_{CCC} = 2.68$, $p = .162$) and marginally significant for patient SPG_009 ($z_{CCC} = 4.99$, $p = .073$) after controlling for ECAS subscores. No patients showed a Limb effect for metaphoric sentences. Individual means and standard errors for Accuracy values obtained by patients with HSP-SPG4 across conditions figurative types are reported in Figure 2.10A.

Single-case analysis on Reaction Time measures of Experiment 2 showed that only one patient exhibited a significant Metaphor effect (10% of the sample; SPG_006: $t_{modified}(9) = 3.03$, $p = .007$, $z_{CC} = 3.18$), which was marginally significant in one other patient (10% of the sample; SPG_003: $t_{modified}(9) = 1.73$, $p = .059$, $z_{CC} = 1.81$). The Limb effect in literal sentences was significant in two patients (20% of the sample; SPG_002: $t_{modified}(9) = 2.11$, $p = .032$, $z_{CC} = 2.22$; SPG_005: $t_{modified}(9) = 2.82$, $p = .010$, $z_{CC} = 2.96$). The Limb effect in metaphor sentences was marginally significant in two patients (20% of the sample; SPG_005: $t_{modified}(9) = 1.48$, $p = .086$, $z_{CC} = 1.55$; SPG_006: $t_{modified}(9) = 1.62$, $p = .070$, $z_{CC} = 1.70$). Individual means and standard errors for Reaction Time values obtained by patients with HSP-SPG4 across conditions figurative types are reported in Figure 2.10B.

The summary of Experiment 1, with the percentage of HSP-SPG4 patients showing Motor and Limb effects for single verbs, and Experiment 2, with the percentage of patients showing Metaphor and Limb effects in literal and metaphoric sentences, is reported in Figure 2.6B.

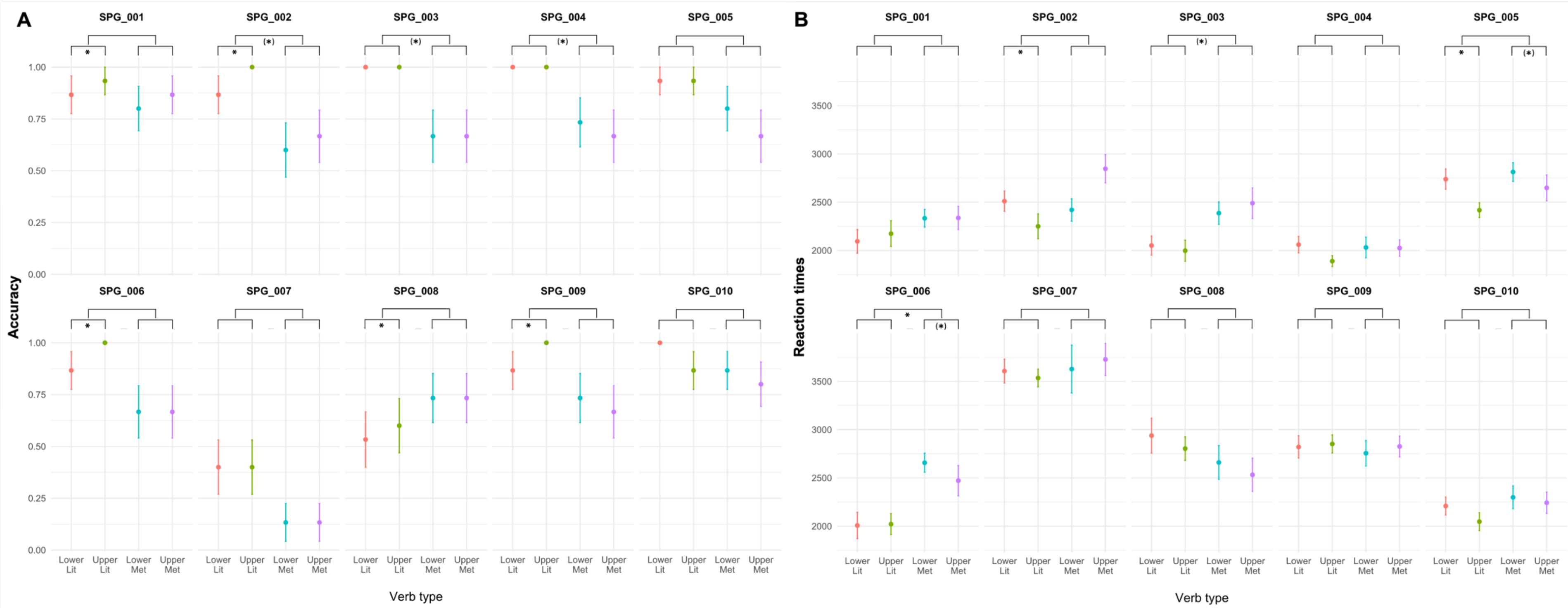


Figure 2.10. Individual Accuracy (A) and Reaction Time (B) values for patients with HSP-SPG4 from Experiment 2 (means and standard errors). *Squared brackets indicate the comparison between literal (Lit) and metaphoric (Met) sentences, as well as the comparison between upper-limb-related (Upper) and lower-limb-related (Lower) verbs in either literal or metaphoric sentences relative to the mean performance of the control group. Significant (p < .05) and marginal (p < .10) effects are reported before controlling for cognitive measures and are indicated with the symbols * and (*), respectively.*

2.3.3. Discussion

In Study 2, we aimed to test the presence of somatotopic, limb-specific distributions of deficits in action-language processing in a cohort of patients with HSP-SPG4, characterized by asymmetrical motor impairment affecting selectively the lower limbs. Importantly, we were interested in testing not just whether patients would exhibit difficulties in processing action verbs in general (Motor effect) but rather whether they would show more specific deficits with verbs denoting movements of the lower limbs (Limb effect) when presented as single words (Experiment 1) and when embedded in literal or metaphoric sentences (Experiment 2).

Patients with HSP-SPG4 showed an expected somatotopic distribution of action-language difficulties, observed in 60% of the sample (cumulating unique cases exhibiting marginal or significant effects in accuracy or reaction time measures) for lower-limb-related action verbs presented as isolated words and in literal sentences, even in the absence of cognitive impairment. Motor effects were observed in 30% of patients only. These results suggest that the Motor effect was not associated with a widespread difference between patients and controls, while a disease-specific pattern emerged – albeit with great inter-individual variation – when we compared upper-limb and lower-limb-related action verbs presented in isolation or embedded in literal declarative sentences.

Finally, Metaphor effects, namely an impaired processing of metaphoric sentences compared to literal ones, were present in 40% of HSP-SPG4 patients (cumulating unique cases exhibiting marginal or significant effects in accuracy or reaction time measures). However, a Limb effect for metaphoric sentences was marginally observed in 20% of patients only. Overall, our findings suggest that action language difficulties seem to mirror – at least partially – the lower-limb-specific motor symptomatology characterizing the clinical profile of HSP-SPG4 (Solowska & Baas, 2015), yet did not massively generalize to figurative language processing, opening relevant theoretical implications worthy of being addressed in the general discussion.

2.4. General discussion

In this work, we presented two studies designed to test the predictions of ECT concerning the involvement of "motor grounding disruption" effects in MND through two tasks including lower-limb and upper-limb-related action verbs presented either in isolation (Experiment 1) or in literal and metaphoric sentences (Experiment 2). The novelty of the study is in the test of rare neurodegenerative diseases within the spectrum of MND, namely ALS and HSP-SPG4. HSP-SPG4, in particular, is a genetic condition specifically affecting lower-limb motor functioning, which allowed a stringent assessment of somatotopic effects. Somatotopicity is indeed a key aspect for assessing the robustness of neuroscientific evidence of ECT, especially for the case of more abstract uses of language, such as metaphors (Casasanto & Gijssels, 2015). We tested specific predictions based on ECT, which would expect to find a generalized impairment in processing action verbs relative to psychological verbs (i.e., Motor effect) in ALS. As for HSP-SPG4, ECT would predict worse performance in processing lower-limb-related action verbs compared to upper-limb-related action verbs (i.e., Limb effect), thus supporting the idea that word meaning grounding in the motor system reflects a somatotopic organization. A central question of both studies was whether disruption of motor grounding effects would generalize from action verbs used with their literal meaning to more abstract language uses, as in the case of action metaphoric sentences. Theories supporting an embodied account of metaphors (Gibbs, 2005; Lakoff & Johnson, 1999) would expect to find similar patterns also in non-literal uses of language, along with a general difficulty in processing metaphoric sentences compared to literal ones (i.e., Metaphor effect). The results were surprising in several respects, the main one being that the metaphorical level did not show widespread motor or somatotopic effects in both populations. Motor effects (general or limb-specific) were observed in both ALS and HSP-SPG4 – albeit with different frequency rates – but were confined only to the literal level with great individual and cross-diagnostic variation.

Starting with a deeper inspection of the results of Experiment 1, for the sample with ALS, we observed that the Motor effect occurred in 7.7% of cases, while the Limb effect was detectable in 15.4% of patients. The frequency of occurrence of these effects showed that generalized difficulties in processing action meaning are rare in ALS, while its more specific variant involving specifically lower-limb-related action verbs is slightly more frequent. Overall, these results suggest that motor grounding effects are not widespread in the population of patients with ALS, as they might occur in a small subsample of cases. Importantly, with such limited occurrence of action language impairment, is relevant to emphasize the importance of adopting finer-grained approaches investigating the presence of deficits at the individual level: while previous works on ALS (Aiello et al., 2023; Papeo et al., 2015) and other motor conditions, such as Parkinson's disease (Aiello et al., 2022; Møller et al., 2023), did not observed significant action language difficulties at the group level, we might still expect to detect some case-specific patterns of impairment. It is still relevant to inspect the potential explanations behind the lack of generalizability of action language impairment in our sample. In this regard, it is interesting to observe that a number of previous studies brought the dissociation between object name and action verb processing as evidence to support the disruption of action-meaning representation in ALS (Bak & Hodges, 2004; Grossman et al., 2008; Taylor et al., 2013). However, as pointed out by Papeo et al. (2015), most studies did not consider the higher executive involvement required during verb compared to noun processing (Vigliocco et al., 2011), neglecting that action-verb impairment in ALS might rather reflect executive dysfunctions, which are a central feature of patients' cognitive decline during disease progression (Woolley & Rush, 2017). Capitalizing on these arguments, namely that difficulties in action-verb processing documented in ALS could be more generally ascribed to cognitive deficits affecting the executive functioning domain, we hypothesize that the lack of widespread Motor and Limb effects in our sample might be related to participants' cognitive profile, which was unimpaired in most of our patients. In other words, better preserved cognitive skills in our sample might have resulted in more marginal difficulties in action verb processing, compared to previous literature reporting effects in the opposite direction (Cousins et al., 2018; Grossman et al., 2008; York et al., 2014).

As for the group of participants with HSP-SPG4, we observed that difficulties in processing action verbs related to lower-limb movements occurred in 60% of patients, even without cognitive impairment. This pattern confirms that action language might exhibit – at least in part – a somatotopic organization of motor grounding (Hauk et al., 2004; Tettamanti et al., 2005), which is mirrored also in its disruption in disorders characterized by focal, limb-specific motor impairment not complicated by other neurological or extra-motor symptoms (Panza et al., 2022; Rossi et al., 2022). Although similar effector-specific effects have been reported for Parkinson's disease in relation to different motor impairment onset sites (Roberts et al., 2017), this is the first study to document a somatotopic distribution of action-language deficits in individuals with a rare congenital condition, such as HSP-SPG4. However, despite the evidence of overlapping language and motor impairment, it is relevant to underline that the cohort of patients in our study showed remarkable variability at the individual level, indicating that effector-specific action-language deficits are highly susceptible to variation even in an apparently homogeneous population. This inter-individual variability in Limb effects might be interpreted by considering the complex interaction of clinical and genetic factors related to HSP-SPG4. First, despite the relatively consistent motor impairment, patterns of clinical and genetic heterogeneity can still be detected in this population: as observed by Rossi et al. (2022), patients with HSP-SPG4 might considerably vary in terms of clinical factors, such as disease onset and duration, motor impairment severity, prognosis, and long-term disability. Such factors interact also with the genetic level, as patients might also exhibit variants of *SPAST* gene mutation that affect the course of the disease and motor functioning (Rossi et al., 2022). Beyond clinical and genetic determinants of variability, it is also relevant to consider the role of environmental factors modulating the relationship between genotype and phenotype in neurodegenerative disease (Bradley et al., 2018; Mo et al., 2015; Zou et al., 2017), as well as compensatory neural or cognitive mechanisms that might be developed by individuals with brain damage to cope with cognitive dysfunction (Stern, 2009; for brief discussion related to 'embodied' effects, see Aiello et al., 2022; Ibáñez et al., 2023). Taken together, all these factors might thus result not only in clinical heterogeneity but also in individual differences related to how (and how frequently) motor grounding disruption manifests in HSP-SPG4 and related disorders.

The results from Experiment 1 – in particular those emerging from HSP-SPG4 – have relevant implications for ECT, as seem to support the involvement of motor simulations during motion verbs processing, reflecting a somatotopically organized engagement of cortical motor areas in response to action language (Hauk et al., 2004; Hauk & Pulvermüller, 2004). In particular, our study supports the hypothesis that action meaning grounding might be disrupted in neurological disorders involving motor system not only at the level of degeneration of frontostriatal circuits (Birba et al., 2017) but also at the level of spinal motor neuron degeneration, as in MND (Bak & Chandran, 2012). A relevant implication for ECT is that motor grounding disruption effects, and in particular their somatotopic distribution in HSP-SPG4, were not ubiquitous and emerged in an individual-level analysis, inspecting case-by-case patterns of deficits. This is a significant aspect, as it confirms that the involvement of motor simulations in language processing are highly modulated by individual factors (Muraki & Pexman, 2021; Pexman & Yap, 2018) and that embodied effects should be inspected using finer-grained tools, as might be dampened when investigated at the group-level only. This might be crucial in the case of neurological conditions, where sources of variability might include cross-diagnostic differences (York et al., 2014), as well as possibly environmental factors and neural mechanisms of compensation (Gregory et al., 2017; Papoutsi et al., 2014; Stern, 2009), which might mitigate the emergence of action-language deficits in some individuals (Ibáñez et al., 2023).

Moving to the results of Experiment 2, involving action literal and figurative sentences, we observed a pattern that partially echoed Experiment 1 for what concerns literal uses of language: while participants with ALS showed no widespread motor impairment (detectable in less than ¼ of the sample), 60% of participants with HSP-SPG4 exhibited effector-specific difficulties involving predominantly lower-limb-related action sentences. The novel aspects that emerged from this experiment concern the processing of action metaphoric sentences, in particular: i) the occurrence of deficits in figurative language processing, present in 23.1% and 40% of individuals with ALS and HSP-SPG4, respectively; ii) the limited occurrence of Limb effect for metaphoric sentences in ALS and HSP-SPG4 patients, detected in 15.4% and 20% of cases, respectively. The first point can be explained by considering that pragmatic difficulties are strongly associated with cognitive dysfunctions in MND (Bambini, Arcara, Martinelli, et al., 2016; Bambini, Bischetti, et al., 2020; Bambini & Ceroni, 2021) and might then not occur in samples of participants with relatively spared cognitive profile.

The second point, however, was rather unexpected and has strong implications for ECT, as it suggests that action language disruption in motor disorders does not significantly affect figurative uses of verbs denoting motions. This finding is actually consistent with similar studies focusing on other motor conditions, such as Parkinson's disease (Humphries et al., 2019), but also with studies on sensory-modality-related metaphors (e.g., metaphors expressing figurative tactile meaning), which reported unimpaired tactile metaphor comprehension in congenital conditions affecting somatosensation (Phillips et al., 2023). If we combine all this converging neuropsychological evidence, we are induced to acknowledge that 'embodied effects' for modality-related metaphors might be weaker in figurative compared to literal uses of language and that information activated by motor or perceptual simulations might become less available in more abstract linguistic operations. Insights supporting this claim come also from neurofunctional data acquired from neurotypical individuals, showing that embodied effects in action metaphor processing – albeit reported by several neurofunctional studies (Desai et al., 2011; Romero Lauro et al., 2013; Yang & Shu, 2016) – are not as strong as those observed for literal sentences and lack substantial somatotopic distribution (Casasanto & Gijssels, 2015). Beyond this interpretation, the absence of action language deficits involving metaphor processing might also be explained by considering that the involvement of sensory-motor simulations during action metaphor processing might be influenced by the level of conventionality of the stimuli (Desai et al., 2011; Yang & Shu, 2016). In our study, we employed metaphors that were rated as moderately familiar during a pre-experimental validation study, although not perceived as fully conventionalized: compared to novel and creative metaphors, relatively familiar ones might not elicit bodily simulations, as their processing relies more on the activation of a pre-stored meaning (Cuccio, 2022). This alternative interpretation is in line with other studies investigating embodied effects along the continuum from literal to more abstract uses of language, documenting the lack of sensory-motor activations in the case of highly conventionalized idiomatic expressions (e.g., *kick the bucket*) compared to less conventionalized figurative language uses (see Cacciari et al., 2011; Romero Lauro et al., 2013; Yang & Shu, 2016).

This pattern has significant implications for ECT, as it shows that somatotopic, limb-specific involvement of the motor cortex is more likely to be observed during the processing of literal uses of language, while it fades in more figurative ones. In particular, these findings confirm recent elaborations of ECT, which not only account for inter-individual variations in the involvement of sensory-motor information during language processing (Pexman & Yap, 2018), but also emphasize that the involvement of embodied simulations is not ever-present across tasks (Tousignant & Pexman, 2012) and might become marginal especially when contextual and situational requests do not require it (Connell, 2019; Connell & Lynott, 2013; see also Frau, Bischetti et al., *under review*, described in §Study 1). In the specific case of metaphor comprehension, our results show that higher-level pragmatic uses of language, while partially supported by modality-specific information, still ground their processing machinery on abstract, inferential operations required to grasp the interlocutor's intended meaning (Sperber & Wilson, 2008). Such predominantly abstract representation might spare pragmatic processing at the behavioral level in motor impairment, which might still affect more literal language uses even in individuals with relatively intact cognitive functioning.

This study is expected not only to have implications related to the theoretical debate within ECT but also to elucidate the cross-diagnostic description of the linguistic profile within the spectrum of MND. In particular, our results contribute to the description of cognitive deficits in MND (Chamard et al., 2016; Strong et al., 1999; Tallaksen et al., 2003), by adding relevant information for the characterization of the linguistic profile of individuals with ALS and HSP-SPG4 (Bak & Hodges, 2004), also extending to difficulties involving higher-level pragmatic abilities (Bambini & Ceroni, 2021). Along these lines, our findings can inform current and future research on how selective linguistic deficits are shaped by differences in the profile of motor impairment associated with MNDs, as well as by cross-population differences determined by etiological factors. In our study, we tested action-language processing in two conditions characterized by different etiologies, i.e., predominantly non-familial for the case of ALS, with genetic mutations accounting for less than 10% of sporadic cases (Sabatelli et al., 2013; Zou et al., 2017), and congenital for the case of HSP-SPG4, with cases characterized by a mostly homogeneous genetic mutation involving the *SPG4* gene. Future studies might then investigate how strongly genetic factors shape the connection between motor impairment and action-language disruption in neurological conditions with such a direct link between genotype and motor phenotype (for similar considerations on language impairment in hereditary conditions, see Bak et al., 2006; García et al., 2017).

To conclude, in our study we tested the solidity of ECT, by addressing the hypothesis that neural changes in the motor cortex translate into disruption of action-meaning representations not only in literal but also in figurative uses of language. We investigated this hypothesis in a stringent fashion, involving two groups of participants within the spectrum of MND, namely ALS and HSP-SPG4, the latter being particularly relevant to detecting a somatotopic distribution of action-language deficits. By applying an individual-level analysis, we showed that neurodegenerative processes involving the motor system might disrupt action meaning in literal sentences while leaving more abstract language uses relatively more intact. And even in the presence of marked action-language deficits, with somatotopic distribution mirroring the profile of motor impairment (as in HSP-SPG4), individual variation showed that motor grounding is flexible and its disruption is likely to be influenced by environmental, clinical, and – possibly – genetic factors.

STUDY THREE

FROM SEMANTIC CONCRETENESS TO CONCRETISM IN PSYCHOPATHOLOGY: AN AUTOMATED LINGUISTIC ANALYSIS OF FIGURATIVE LANGUAGE INTERPRETATION IN SCHIZOPHRENIA[4]

Abstract

Lack of abstract thinking, often referred to as concretism, is a well-known psychopathological feature of schizophrenia, which includes the tendency to adhere to concrete aspects of stimuli and difficulties in understanding figurative language. Inspired by the similarity between “concretism” as defined in psychopathology and “concreteness” as defined in linguistics, namely the semantic properties of words referring to perceptual experience, we tested the idea that impairment in deriving figurative meanings could depend on impairment at the lexical/semantic level, involving concreteness. We analyzed the speech samples produced by 63 individuals with schizophrenia and 47 controls, who were asked to explain the meaning of a series of figurative expressions (idioms, metaphors, and proverbs). By automatically extracting linguistic values from participants’ speech samples, we observed that patients’ answers exhibited higher concreteness at the word level, especially in the verbal explanation of proverbs, while not different in measures of lexical richness and in the pause-to-word ratio. Word concreteness in patients’ explanations was also predictive of their ability to understand proverbs and was indicative of their global pragmatic and cognitive abilities. This supports the idea that concretism is rooted in the lexical/semantic level, with an association between concrete figurative interpretations and the use of concrete words. The study also discloses new areas of interest in the automated analysis of speech in psychosis for a better

[4] This chapter is a manuscript in preparation for submission to a peer-review journal as “Bambini, V.*, Frau, F.* (*co-first authors), Bischetti, L., Agostoni, G., Mevio, C., Bechi, M., Buonocore, M., Sapienza, J., Spangaro, M., Guglielmino, C., Cocchi, F., Cavallaro, R., & Bosia, M., *From semantic concreteness to concretism in psychopathology: An automated linguistic analysis of figurative language interpretation in schizophrenia*”.

characterization of the linguistic profiles as well as further applications to method development and the identification of potentially relevant linguistic dimensions to be used in clinical research.

3.1. Introduction

Altered thought processes have always undeniably dominated the clinical characterization of schizophrenia since Eugen Bleuler's early descriptions (Bleuler, 1911). Lately labeled in psychiatric literature as "formal thought disorder", impaired thinking in this condition includes a plethora of psychopathological manifestations, leading to disorganized streams of thoughts, unusual and irrelevant conceptual associations, including biases toward more "concrete" modes of thinking (Spitzer, 1997). Instances of concrete thinking in schizophrenia have been extensively documented since Bleuler's works (Spitzer, 1993): whether it manifests as an inability to grasp similarities among objects, to change strategies during a task, or to generalize relevant aspects from a situation, *concretism* can be briefly summarized as a generalized difficulty in going beyond immediate experience whenever it is required to adopt an abstract mode of thinking (Goldstein, 1944, 1959; Wright, 1975).

Among the various manifestations of concretism in schizophrenia, the literature traditionally includes also difficulty in understanding figurative expressions, such as metaphors and proverbs, as documented in early and more recent studies (Goldstein, 1944; Mossaheb et al., 2014; Spitzer, 1993). In particular, individuals with schizophrenia might fail to go beyond the literal meaning of such expressions, being rather more likely to remain anchored to the most immediate, concrete, and literal sense of words (Harrow, 1974; Kircher et al., 2007). In this view, concretism is tightly connected with pragmatic disorder, namely a more general communicative impairment hampering the ability to manage articulated discourse and use language appropriately to context, as in the case of figurative language understanding (Bambini, Arcara, Bechi, et al., 2016; Colle et al., 2013). Along this vein, a recent study by Bambini et al. (2020) addressed specifically the pragmatic manifestation of concretism in schizophrenia, investigating the range of impairment in understanding figurative meanings across different types of expressions (e.g., idioms, metaphors, and proverbs) and different types of response formats used to evaluate comprehension skills (e.g., multiple-choice vs. verbal explanations). The results confirmed a generalized difficulty in individuals with schizophrenia in understanding different figurative expressions and showed that proverbs were perceived as particularly challenging, likely due to the particularly abstract moral content conveyed by such expressions (Sperber & Wilson, 1995).

From Bleuler's initial, more comprehensive idea of "disordered conceptual associations", concretism in schizophrenia has received several explanations, linking biases toward concrete thinking to more general difficulties in abstract conceptualization (Wright, 1975) or altered organization of semantic networks in associative memory (Spitzer, 1993), or more recently impaired cognitive and socio-cognitive mechanisms (see Frau, Cerami et al., 2024). Little consideration has been paid to the idea that concretism – at least for what concerns its pragmatic manifestation – could be linked to difficulties in the building blocks of language. In particular, this work starts from the idea that the bias towards more concrete interpretations of a linguistic stimulus might also reflect a tendency to prefer more concrete aspects of the semantic representation of words.

Following theoretical models in pragmatics arguing that the process of understanding non-literal language relies on the elaboration of a concept and its properties (Carston, 2010a; Sperber & Wilson, 2002, 2008), we hypothesize that individuals with schizophrenia might have difficulties in integrating concrete properties of concepts encoded in figurative expressions, resulting in concrete, literal interpretations. If we take the example of a metaphor such as *That lawyer is a shark*: concrete properties of the concept *shark* (e.g., 'swims' or 'has a fin') might be preferred to more abstract ones (e.g., 'being ruthless' or 'being aggressive'), resulting in a concrete explanation of the meaning of the sentence (e.g., 'that lawyer swims very fast').

Several pieces of evidence support this idea, including data showing that language impairment is multi-dimensional in schizophrenia (Bambini, Frau, et al., 2022; Covington et al., 2005), encompassing both lexical (Maher et al., 1983; Manschreck et al., 1984) and semantic (Barattieri di San Pietro et al., 2023; Marini et al., 2008; Pomarol-Clotet et al., 2008) aspects of language ability, mirroring also patients' symptomatologic profiles (Bambini, Frau, et al., 2022). Going further, additional support comes from studies addressing concretism from the perspective of associative memory, such as semantic priming studies showing that concrete properties of figurative expressions remained significantly active only in patients but not in healthy controls (Spitzer, 1993).

3.1.1. The present study

This study aimed to test the hypothesis that the pragmatic manifestation of concretism in schizophrenia, i.e., impairment in figurative language understanding, reflects a bias towards concrete properties of the semantic representation of concepts (e.g., concreteness). If this hypothesis is confirmed, then verbal explanations of the meaning of non-literal expressions would result in higher use of concrete words, namely words denoting concepts more easily perceived using sensorial experience (Paivio, 1979).

We tested this hypothesis by using a verbal explanation task, where participants were asked to explain the meaning of different figurative expressions. We applied an automated pipeline on participants' speech samples, to extract semantic properties of words, alongside control linguistic measures (e.g., fluency and lexical variables), capitalizing on the highly documented effectiveness of computational methods in capturing linguistic correlates of psychopathological processes in schizophrenia (Corcoran et al., 2020; Elvevåg et al., 2007; Hitczenko et al., 2021; Oomen et al., 2022).

We expected to observe a higher occurrence of words denoting concrete concepts in patients' explanations compared to controls, with variables measuring general lexical or fluency aspects not differing between groups. We also expected that word concreteness would reflect patients' accuracy in providing a correct interpretation of figurative expressions, as well as global pragmatic performance. Finally, we expected that word concreteness would be associated with clinical, psychopathological, cognitive, and sociocognitive measures in the schizophrenia group.

As a follow-up testing ground of the task-specific effect of word concreteness, we applied the same pipeline to a different speech elicitation task, namely a semi-structured interview. We expected to find no differences between groups in the use of concrete words in the semi-structured interview, consistently with the idea that preference towards more concrete aspects of the semantic representation of words is task-dependent in schizophrenia, namely reflects difficulties that arise specifically when figurative interpretations are at stake.

3.2. Methods

3.2.1. Participants

Sixty-three individuals with a diagnosis of schizophrenia based on DSM-5 criteria (American Psychiatric Association, 2013) were recruited from the Department of Clinical Neurosciences, IRCCS San Raffaele Scientific Institute, Milan, Italy. All participants were native speakers of Italian and were part of a sample involved in a larger study (Bambini, Frau, et al., 2022). Inclusion criteria were: age 18–65 years; being clinically stabilized and treated with a stable dose of the same antipsychotic therapy for at least 6 months. Exclusion criteria were: severe traumatic brain injury or neurological disorders, intellectual disability, alcohol or substance abuse in the preceding 6 months, and severe psychotic exacerbation in the preceding 3 months. Additionally, 47 Italian-speaking healthy controls were recruited from a larger sample involved in the validation of assessment tools of pragmatic ability (Arcara & Bambini, 2016; Bischetti et al., 2023), balanced for age and education with the sample of patients.

All participants provided informed consent. The study was approved by the local ethical committee, following the principles of the Declaration of Helsinki.

3.2.2. Assessment

Both groups were assessed for global pragmatic skills with the Assessment of Pragmatic Abilities and Cognitive Substrates test (APACS; Arcara & Bambini, 2016), a validated battery developed to assess pragmatic skills (both in production and comprehension) in Italian-speaking individuals.

Patients were further evaluated with a battery of tests assessing: psychopathology, with the Positive and Negative Syndrome Scale for Schizophrenia (PANSS; Kay et al., 1987), including the disorganization dimension (van der Gaag et al., 2006); cognitive skills, with the Italian version of the Brief Assessment of Cognition in Schizophrenia (BACS; Anselmetti et al., 2008; Keefe, 2004), using the average equivalent score as overall measure of the cognitive profile (Bambini, Arcara, Bechi, et al., 2016); social cognition, with Theory of Mind Picture Sequencing Task (ToM-PST; Brüne, 2003).

3.2.3. Speech samples and automated analysis

3.2.3.1. Elicitation task

The APACS Figurative Language 2 task was used to elicit speech samples from participants. In particular, the prompting items included five highly familiar idioms (e.g., *My brother is always in the red*) extracted from existing norms (Tabossi et al., 2011), five novel metaphors (e.g., *Some voices are trumpets*) from a previous study (Bambini et al., 2013), and five common proverbs (e.g., *A swallow does not make a summer*) from a dictionary of Italian proverbs (Guazzotti & Oddera, 2006). For the post-hoc task-specific assessment, we used speech samples elicited using the APACS Interview task (focusing on autobiographical topics, i.e., family, home, work, and organization of the day), which were already partially available from a previous study (Bambini, Frau, et al., 2022).

Verbal explanations were recorded using a one-channel audio recorder oriented towards the participant. The recordings were acquired in a quiet room within a controlled laboratorial setting, then converted to .wav files to be imported into the PRAAT software (Boersma & Weenink, 2021), with a standard quality of 44.10 kHz (capturing 44100 samples per second).

3.2.3.2. Automated analysis

The audio recordings of participants' verbal explanations were pre-processed and transcribed verbatim before undergoing the automated pipeline for the extraction of linguistic measures (the pre-processing and transcription phases are described in the Supplementary materials, see Appendix A).

Linguistic measures included semantic variables (e.g., word concreteness and imageability), alongside a set of control variables (i.e., response length, number of pauses, pause-to-word ratio, lexical frequency, and type-token ratio) extracted to investigate fluency and lexical richness differences between patients and controls. Table 3.1 includes a description of all linguistic measures, while Figure 3.1 depicts a representation of the pipeline used to extract them from transcripts.

Concreteness and imageability values were extracted for lemmas using the multilingual MEGAHR repository (Ljubešić et al., 2018), covering 94.42% of lemmas in our transcripts. Lexical frequency values (log-transformed) were extracted for tokens from the Corpus and Frequency Lexicon of Written Italian (CoLFIS; Bertinetto et al., 2005), covering 94.47% of tokens in our dataset. Type-token ratio was computed after removing stop words from the transcripts using Python's Natural Language Toolkit (NLTK; Bird et al., 2009) list of Italian stop words.

All variables (excluding measures of pauses) were also extracted from the transcripts of the APACS Interview, with concreteness and imageability ratings covering 92.81% of the lemmas.

The automated pipeline was developed on R Studio (R Core Team, 2023).

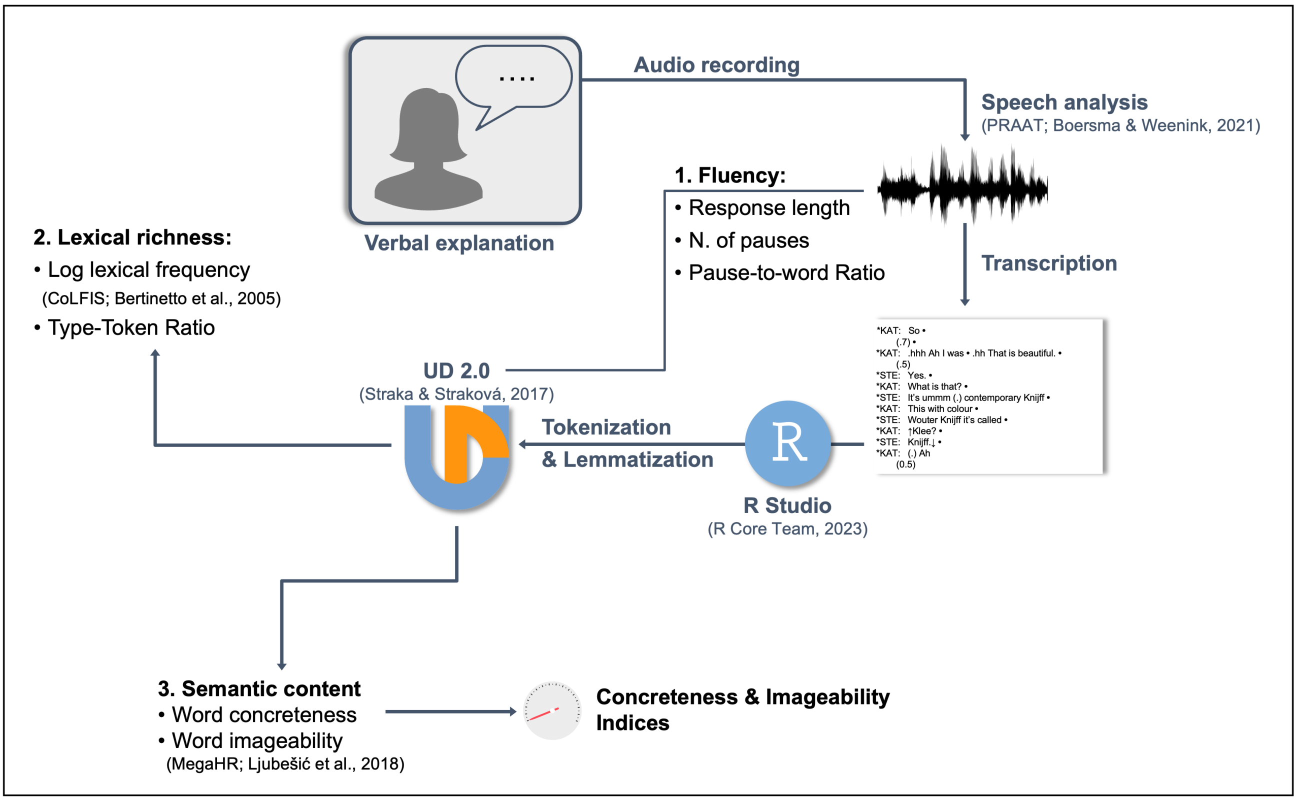


***Figure 3.1**. The automated pipeline applied to participants' verbal explanations for the APACS Figurative Language 2 task.* *Verbal explanations were first audio-recorded and analyzed with PRAAT software to identify silent pauses. Audio samples were then transcribed, imported on RStudio software, and tagged using Universal Dependencies (UD 2.0) treebank. From tokenized and lemmatized transcripts, we extracted fluency and lexical richness measures, as well as semantic variables (word concreteness and imageability), then normalized and standardized as concreteness and imageability indices.*

Table 3.1. Description of linguistic features.

Linguistic dimensions	**Measures**	**Description**
Fluency	Response length	Total number of words uttered by the participants in their verbal explanations; this value might indicate a tendency to either verbosity or poverty of speech.
	Number of pauses	Total number of long silent pauses (defined as silences ≥ 1 second) and filled pauses (e.g., "uhm", "ehm", etc.) for each verbal explanation; this measure reflects planning and self-monitoring processes (de Boer, van Hoogdalem, et al., 2020).
	Pause-to-word ratio	Total number of pauses divided by the total number of words for each verbal explanation; this value can be considered as an indicator of processing speed (de Boer, van Hoogdalem, et al., 2020).
Lexical Richness	Lexical frequency (log-transformed)	Log-transformed frequency value of words uttered by the participants, extracted from the Corpus and Frequency Lexicon of Written Italian (CoLFIS); it indicates whether participants used more low- or high-frequency words.
	Type-token ratio	The number of unique words (types) divided by the total number of words (tokens) in each verbal explanation, ; this measure is considered an indicator of lexical variety and might depict language deviance or thought disorder (Manschreck et al., 1981).
Semantic content	Concreteness index	The percentage of concreteness content for each verbal response, obtained by averaging and standardizing concreteness ratings for single lemmas, extracted from the MEGAHR repository; it indicates whether participants produced verbal explanations using words rated as more or less concrete.
	Imageability index	The percentage of imageability content for each verbal response, obtained by averaging and standardizing imageability ratings extracted from the MEGAHR repository; it indicates whether participants produced verbal explanations using words rated as more or less imageable, i.e., that can more or less easily arouse mental images.

3.2.4. Statistical Analysis

The rationale of the analysis included three steps: i) we first investigated group differences between individuals with schizophrenia and healthy controls in semantic (concreteness and imageability) and control variables, ii) we tested whether semantic variables predicted participants' accuracy in the APACS Figurative Language 2 task scores, as well as global pragmatic performance, and iii) we explored the association between semantic variables and psychopathological, cognitive, and sociocognitive measures in the schizophrenia group.

In step (i), we compared the two groups using independent sample *t*-tests (after checking for the homoskedasticity and normality assumptions) across the control and semantic variables. All *p*-

values were adjusted for False Discovery Rate (FDR; Benjamini & Hochberg, 1995). We further inspected group differences for semantic variables in the different item types (i.e., idioms, metaphors, and proverbs) using Linear Mixed-effects Models (LMMs), with Group in interaction with Item Type (i.e., idioms, metaphors, and proverbs) as fixed predictors. Random structure was determined upon convergence following a parsimonious approach. Post-hoc pairwise comparisons were performed on the estimated means, with Tukey *p*-values correction.

In step (ii), we fitted a Generalized Linear Mixed-effects Model (GLMM) on APACS Figurative Language 2 score (binomial: 0 = incorrect, 1 = correct), with semantic variables included as fixed effects in interaction with Group and Item Type, alongside control linguistic variables included as covariates. The inclusion of fixed effects was determined via likelihood-ratio tasks, while random structure was determined upon model convergence. We fitted a series of linear models with APACS Figurative Language 1, Composite Comprehension, and Total scores as dependent variables, including semantic variables in interaction with Group as predictors.

For step (iii), we explored correlation patterns in the schizophrenia group using Pearson's correlations. For the post-hoc task-specific assessment, we applied the same rationale as in step (i) to detect group differences across linguistic measures.

All statistical analyses were run in R, v. 4.3.1 (R Core Team, 2023), with the R Studio editor, v. 2023.09.1+494.

3.3. Results

3.3.1. Sample description and assessment

Table 3.2 shows sample characteristics and the results of the global assessment. The correlations among assessment variables in the schizophrenia group are reported in Appendix B, Supplementary Figures 3.1 and 3.2. Overall, patients with schizophrenia obtained lower scores in all APACS measures (reflecting lower pragmatic performance), especially in the Figurative Language 2 task.

Table 3.2. Demographic and assessment measures (mean and standard deviations) of patients and controls, alongside group comparisons for global pragmatic variables.

Measures	**Patients Mean (SD)**	**Controls Mean (SD)**	**Test Statistics**	***p*-value**
Age	39.37 (10.93)	42.15 (13.12)	$t(108) = 1.21$	$p = .228$
Education	11.89 (2.76)	12.79 (3.18)	$t(108) = 1.58$	$p = .117$
Sex (F/M)	24/39	28/19	-	-
APACS Figurative Language 2	18.89 (5.25)	27.43 (2.79)	$t(97.14) = 10.92$	$p < .001$
Idioms	8.61 (1.75)	9.89 (0.34)	$t(67.15) = 5.52$	$p < .001$
Metaphors	7.24 (2.39)	9.52 (1.01)	$t(87.05) = 6.77$	$p < .001$
Proverbs	3.03 (2.35)	8.04 (2.11)	$t(106) = 11.45$	$p < .001$
APACS Production	0.93 (0.06)	0.99 (0.02)	$t(75.74) = 6.42$	$p < .001$
APACS Comprehension	0.74 (0.15)	0.95 (0.03)	$t(69.10) = 10.37$	$p < .001$
APACS Total	0.84 (0.10)	0.97 (0.02)	$t(68.62) = 10.50$	$p < .001$
Illness onset	24.43 (6.51)	-	-	-
Illness duration	15.08 (10.69)	-	-	-
Chlorpromazine-equivalent dose (mg/d)	450.33 (202.04)	-	-	-
PANSS Positive	17.11 (4.11)	-	-	-
PANSS Negative	20.87 (4.93)	-	-	-
N 5 Item	3.44 (1.04)	-	-	-
PANSS General	38.59 (6.55)	-	-	-
PANSS Disorganization	21.17 (5.20)	-	-	-
BACS[a]	1.54 (0.90)	-	-	-
ToM-PST (Total score)	44.63 (11.97)	-	-	-

Notes: APACS = Assessment of Pragmatic Abilities and Cognitive Substrates, PANSS = Positive and Negative Syndrome Scale, BACS = Brief Assessment of Cognition in Schizophrenia, ToM-PST = Theory of Mind Picture Sequencing Task.
Non-integer degrees of freedom are due to Welch independent t*-test computation, used in the case of heteroskedasticity.*
[a] *Mean of the equivalent scores from each subtask of the BACS*

3.3.2. Group comparisons across linguistic variables

Compared to controls (Table 3.3), patients produced longer verbal explanations (measured as number of words) with more pauses. Patients' explanations were also characterized by higher Concreteness and Imageability indices, especially in the case of proverbs.

Table 3.3. Descriptive statistics (mean and standard deviations) and group comparisons related to linguistic variables (control and semantic content variables) extracted from verbal explanations provided by participants to the APACS Figurative Language 2 task.

Measures	**Patients Mean (SD)**	**Controls Mean (SD)**	**Test Statistics**	***p*-value**
Response length	14.30 (8.83)	10.90 (5.01)	$t(101.53) = 2.53$	$p = .042$
N. of pauses	1.25 (0.83)	0.88 (0.59)	$t(107.73) = 2.73$	$p = .032$
Pause-to-word ratio	0.09 (0.05)	0.08 (0.07)	$t(80.53) = 0.51$	$p = .720$
Lexical frequency (log)	5.30 (0.30)	5.33 (0.27)	$t(108) = -0.58$	$p = .720$
Type-token ratio	0.99 (0.01)	0.99 (0.01)	$t(108) = -0.54$	$p = .720$
Concreteness index (total)	63.90 (2.30)	62.50 (1.79)	$t(108) = 3.44$	$p = .005$
Idioms	65.20 (2.70)	64.60 (3.30)	$t(108) = 1.20$	$p = .602$
Metaphors	62.80 (4.04)	62.20 (2.82)	$t(107.56) = 0.92$	$p = .668$
Proverbs	63.60 (3.60)	60.70 (2.80)	$t(108) = 4.70$	$p < .001$
Imageability index (total)	71.06 (1.58)	70.10 (1.23)	$t(108) = 3.45$	$p = .002$
Idioms	71.37 (1.98)	71.28 (2.15)	$t(108) = 0.21$	$p = .838$
Metaphors	70.76 (2.66)	70.22 (1.92)	$t(107.91) = 1.24$	$p = .293$
Proverbs	71.08 (2.45)	68.82 (1.96)	$t(108) = 5.22$	$p < .001$

Notes: all p-values are FDR adjusted. Degrees of freedom in the t-tests vary due to missing values on some tests. Non-integer degrees of freedom are due to Welch independent t*-test computation, used in the case of heteroskedasticity.*

We inspected possible collinearity between Concreteness and Imageability indices: given their high correlation across responses ($r(1614) = .89$, $p < .001$), we kept only the Concreteness Index in the further steps of the analysis.

In the LMM with Concreteness index as the dependent variable, we found a significant main effect of Group and a significant Group × Item Type interaction, indicating that patients with schizophrenia used more concrete words in explaining Proverbs compared to Metaphors (Table 3.4 and Figure 3.2).

Table 3.4. Output of the linear mixed-effects model with the z-centered Concreteness index as the dependent variable.

Fixed Effects	**B**	**SE**	**95% CI**	***t*-statistics**	***p*-value**
(Intercept)	-0.03	0.13	[-0.28, 0.22]	-0.25	.801
Group: CON vs SCZ	0.15	0.05	[0.05, 0.25]	2.81	**.005**
Item type: Idi vs Met	-0.37	0.31	[-0.97, 0.23]	-1.20	.230
Item type: Met vs Prov	-0.01	0.31	[-0.62, 0.59]	-0.05	.962
Group: CON vs SCZ × Item type: Idi vs Met	0.04	0.09	[-0.14, 0.23]	0.46	.646
Group: CON vs SCZ × Item type: Met vs Prov	0.31	0.09	[0.12, 0.49]	3.27	**.001**
Random Effects	**Variance**	**SD**			
$\text{Intercept}_{\text{Subject}}$	0.04	0.19			
$\text{Intercept}_{\text{Item}}$	0.23	0.48			
Residuals	0.57	0.77			
$\text{ICC}_{\text{SubjectItem}}$	0.32				
Model fit	Marginal	Conditional			
R^2	.046	.350			

*Notes: B = model estimates; SE = standard error; CI = confidence intervals; CON = control group; SCZ = schizophrenia group. Concreteness Index trimmed for missing values and values exceeding |2.5| standard deviations (4% of observations). Model formula: Concreteness Index (scaled) ~ Group * Item type + (1 | Subject) + (1 | Item).*
Group was included with sum contrast coding (Controls = reference level); Item Type was included with forward difference contrast coding (Idioms vs. Metaphors, Metaphors vs. Proverbs).

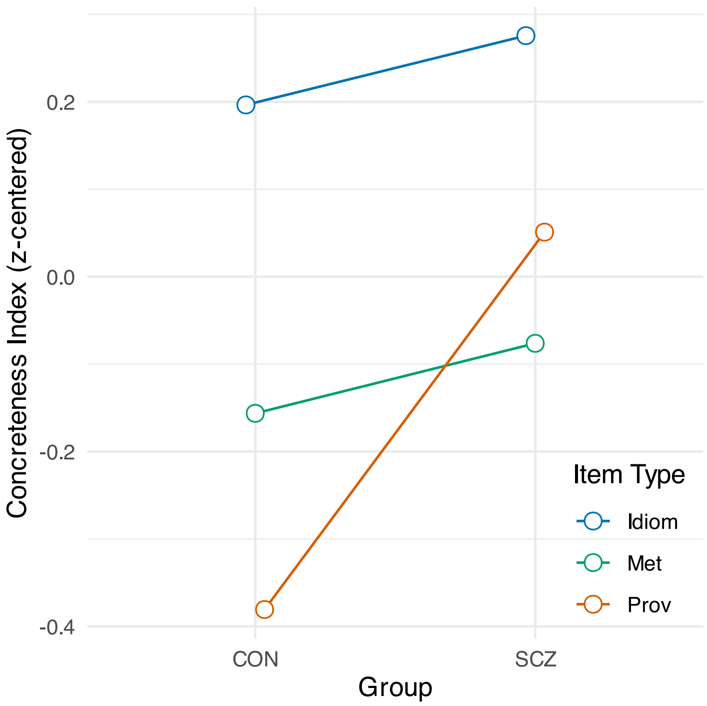


Figure 3.2. Output of the linear mixed-effects model on Concreteness Index across groups and item types. *The plot shows the estimated standardized means of Concreteness Index across groups (CON = control group; SCZ = schizophrenia group) and Item types (i.e., idioms, metaphors, and proverbs).*

The post-hoc pairwise comparisons on the estimated marginal means confirmed that, compared to controls, patients with schizophrenia produced significantly more concrete words in Proverbs only (Table 3.5).

Table 3.5. Post-hoc pairwise comparisons on the estimated means between groups across item types.

Item type	**B**	**SE**	***t*-ratio**	***p*-value**
Idioms	-0.02	0.08	-0.24	.811
Metaphors	-0.06	0.08	-0.82	.414
Proverbs	-0.37	0.08	-4.89	< .001

Notes: B = estimates; SE = standard error. The control group was coded as the baseline in the reported contrasts.

3.3.3. Effect on accuracy in figurative interpretations and pragmatic abilities

The GLMM on APACS Figurative Language 2 score showed a main effect and interaction of Group and Item Type, with patients exhibiting a higher probability of providing an incorrect explanation compared to controls, especially in the difference between metaphors and proverbs. The model also showed a significant interaction between the Concreteness index and Item type in the difference between metaphors and proverbs, indicating that in both groups as concreteness increases, accuracy decreases in the difference between metaphors to proverbs (Table 3.6 and Figure 3.3).

Table 3.6. Output of the generalized linear mixed-effects model with APACS Figurative Language Accuracy score as the dependent variable with Likelihood-Ratio Tests (LRT).

Fixed Effects	**OR**	**SE**	**95% CI**	***z*-value**	***p*-value**
(Intercept)	4.36	0.98	[2.80, 6.77]	6.53	**< .001**
Response length	0.99	0.13	[0.76, 1.29]	-0.07	.943
N. of pauses	0.78	0.10	[0.60, 1.01]	-1.86	.064
Pause-to-word ratio	0.95	0.09	[0.78, 1.15]	-0.53	.598
Lexical Frequency (log)	1.02	0.10	[0.84, 1.23]	0.17	.867
Type-token ratio	1.12	0.08	[0.96, 1.29]	1.43	.152
Group: CON vs SCZ	0.08	0.03	[0.04, 0.15]	-7.66	**< .001**
Item type: Idi vs Met	0.40	0.20	[0.15, 1.08]	-1.81	.070
Item type: Met vs Prov	0.12	0.06	[0.04, 0.30]	-4.48	**< .001**
Concreteness Index	0.85	0.10	[0.68, 1.07]	-1.40	.162
Group: CON vs SCZ × Item type: Idi vs Met	1.70	1.08	[0.49, 5.90]	0.83	.406
Group: CON vs SCZ × Item type: Met vs Prov	0.27	0.14	[0.10, 0.75]	-2.50	**.012**
Group: CON vs SCZ × Concreteness Index	1.36	0.31	[0.88, 2.12]	1.37	.170
Item type: Idi vs Met × Concreteness Index	0.80	0.23	[0.45, 1.41]	-0.78	.437
Item type: Met vs Prov × Concreteness Index	0.43	0.10	[0.26, 0.69]	-3.47	**.001**
(Group: CON vs SCZ × Item type: Idi vs Met) × Concreteness Index	1.20	0.67	[0.40, 3.57]	0.33	.743
(Group: CON vs SCZ × Item type: Met vs Prov) × Concreteness Index	1.21	0.57	[0.48, 3.06]	0.40	.687
Random Effects	**Variance**	**SD**			
$Intercept_{Subject}$	1.07	1.04			
$Intercept_{Item}$	0.44	0.66			
Group: CON vs. SCZ_{Item}	0.25	0.50	-.44		
$ICC_{SubjectItem}$	0.32				
Model fit	Marginal	Conditional			
R^2	.467	.638			
Likelihood-Ratio Tests (LRT)					
Fixed factors	**AIC**	**BIC**	**Loglik**	**Chi Test**	***p*-value**
Control variables	1342.1	1395.7	-661.03	$\chi^2(5) = 16.99$	.005
Control variables + (Group × Item type)	1295.4	1375.8	-632.71	$\chi^2(5) = 56.64$	< .001
Control variables + (Group × Item type × Concreteness Index)	1282.4	1394.9	-620.20	$\chi^2(6) = 25.02$	< .001

Notes: OR = odds ratio; SE = standard error; CI = confidence intervals; CON = control group; SCZ = schizophrenia group. Missing values and Concreteness Index values exceeding |2.5| standard deviations were removed (4.85% of observations).
*Model formula: Accuracy ~ Control variables (z-scaled) + Group * Item Type * Concreteness Index (z-scaled) + (1 | Subject) + (1 + Group | Item).*
Group was included with sum contrast coding (Controls = reference level); Item Type was included with forward difference contrast coding (Idioms vs. Metaphors, Metaphors vs. Proverbs).

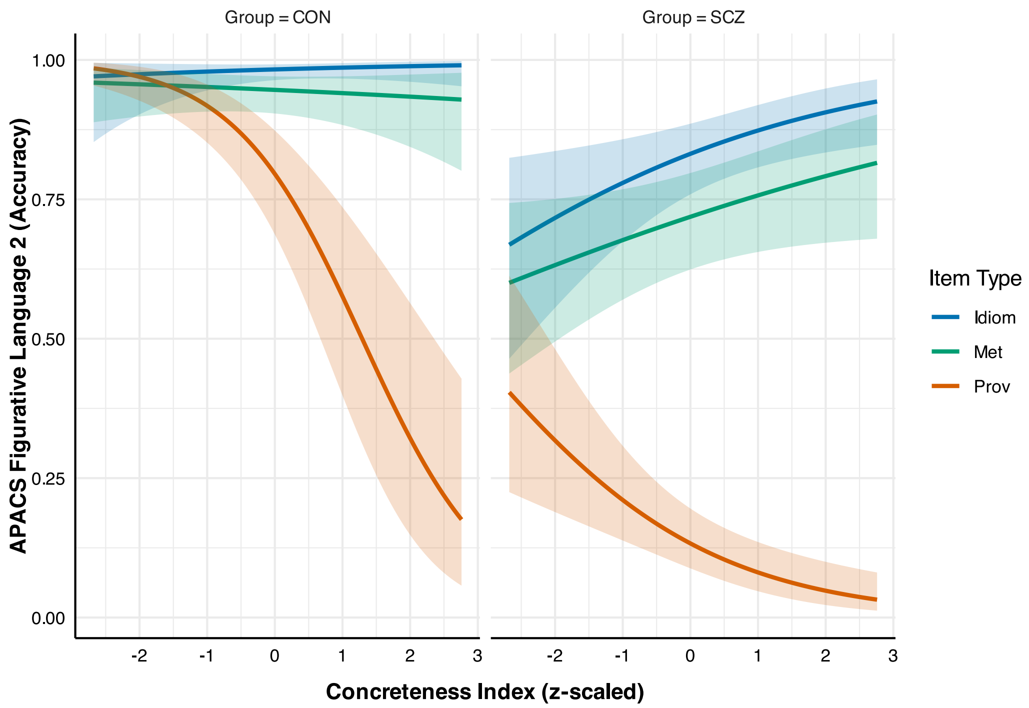


Figure 3.3. APACS Figurative Language 2 Accuracy probability as predicted by Concreteness Index. *Estimated probabilities for the Accuracy score of the APACS Figurative Language 2 task as predicted by Item types (i.e., idioms, metaphors, and proverbs) in interaction with the Concreteness Index (z-scaled) across groups (CON = control group; SCZ = schizophrenia group).*

The linear model on APACS Figurative Language 1 score showed a main effect of Group and Concreteness Index, as well as a significant interaction between both predictors, indicating that patients with schizophrenia performed worse than controls, especially if they tend to use more concrete words. The models with APACS Composite Comprehension and Total scores showed a main effect of Group and Concreteness Index, without significant interactions, indicating that both APACS scores were significantly lower in the schizophrenia group and were associated with the use of word concreteness in both groups (Table 3.7 and Figure 3.4).

Table 3.7. Output of linear models on pragmatic performance as measured with the APACS test.

Dependent variable	Predictors	B	SE	*t*-value	*p*-value
APACS Figurative Language 1	Group	-0.09	0.03	-3.45	< .001
	Concreteness Index (mean)	-0.04	0.01	-2.83	.006
	Group × Concreteness Index (mean)	-0.08	0.03	-2.70	.008
APACS Composite Comprehension	Group	-0.19	0.02	-8.24	< .001
	Concreteness Index (mean)	-0.03	0.01	-2.30	.024
	Group × Concreteness Index (mean)	-0.03	0.02	-1.26	.210
APACS Total	Group	-0.12	0.01	-8.30	< .001
	Concreteness Index (mean)	-0.02	0.01	-2.28	.025
	Group × Concreteness Index (mean)	-0.02	0.02	-1.07	.286

Notes: B = estimates; SE = standard error; APACS = Assessment of Pragmatic Abilities and Cognitive Substrates. The control group was coded as the baseline in the reported contrasts.

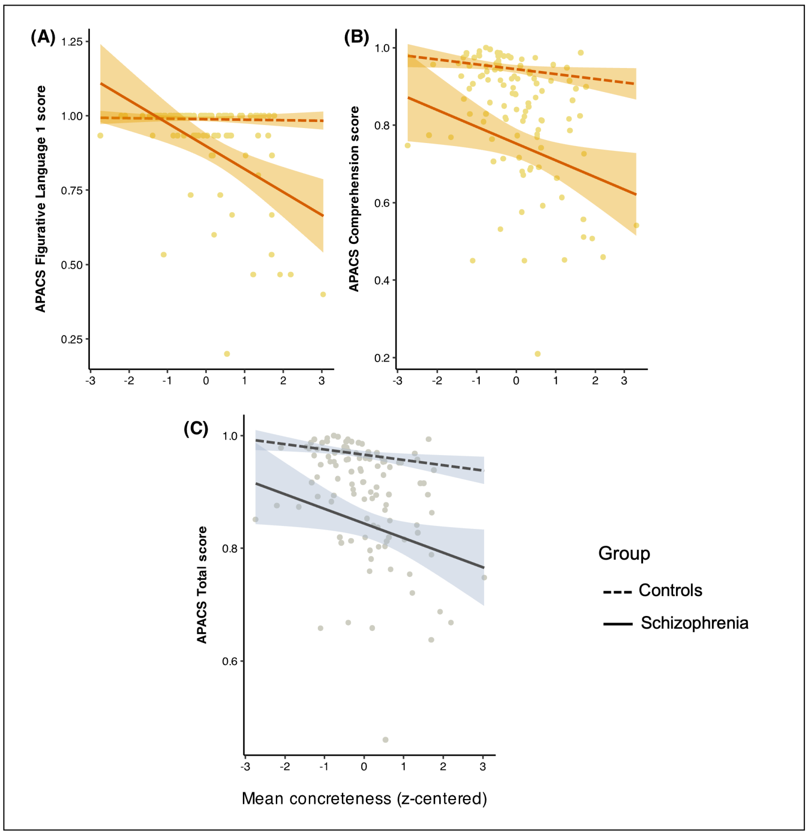


Figure 3.4. Association between mean Concreteness Index and pragmatic measures. *The plot shows the correlation between the mean Concreteness Index and APACS Figurative Language 1 proportionate score (A), APACS Comprehension composite score (B), and APACS Total score (C) in the schizophrenia and the control groups. Note: APACS = Assessment of Pragmatic Abilities and Cognitive Substrates.*

3.3.4. Relation to demographic, clinical, psychopathological, and cognitive variables

The patterns of correlations showed that the Concreteness Index was associated with Education and the Chlorpromazine-equivalent dose (Figure 3.5A), indicating that individuals with lower education and higher chlorpromazine intake exhibited higher concrete content in their verbal explanations (especially in proverbs). Concreteness Index was also associated with cognitive and sociocognitive skills, indicating that individuals with better preserved cognitive and sociocognitive functioning used less concrete vocabulary in their verbal explanations (Figure 3.5B).

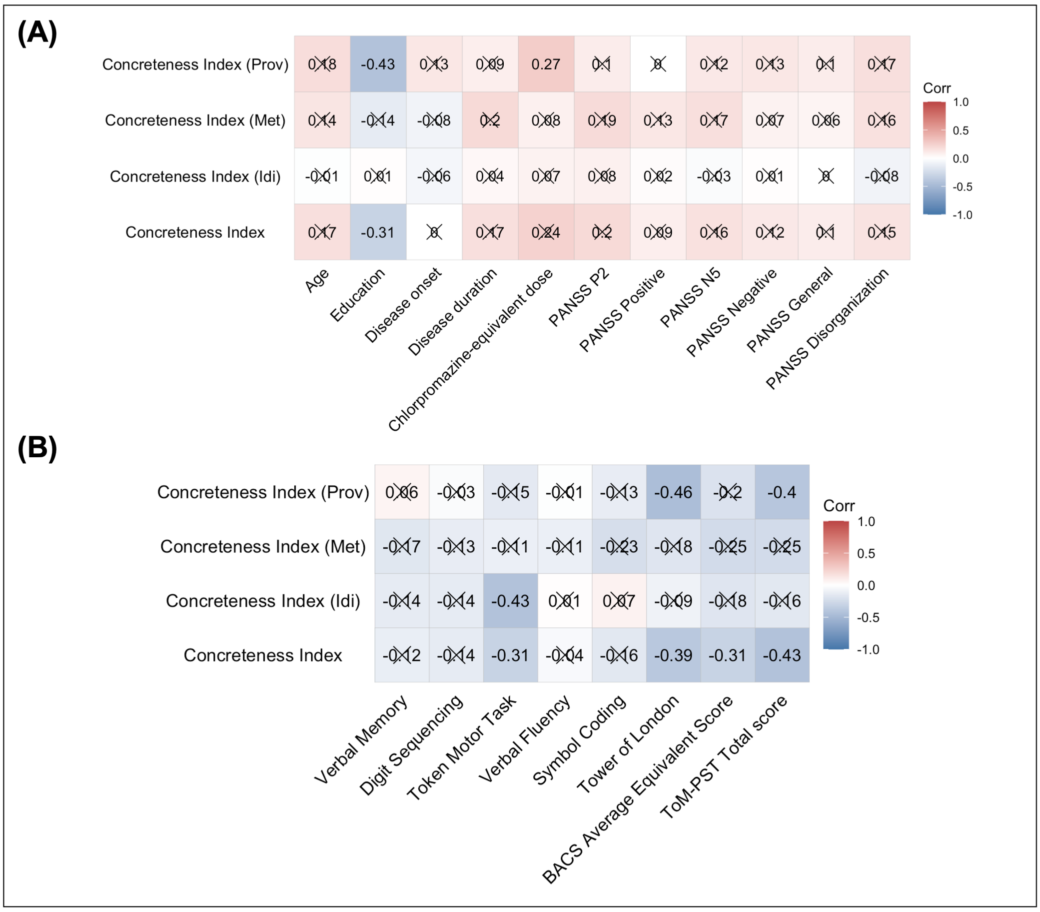


Figure 3.5. Correlogram between Concreteness Index and individual difference variables in the schizophrenia group. *The plots show correlations between patients' Concreteness Index (global mean value and sub-values for idioms, metaphors, and proverbs) and demographic, clinical, and psychopathological measures (A), as well as cognitive and sociocognitive measures (B). The magnitude of associations is depicted by color (crossed cells indicate non-significant correlations, with significance level $p < .05$). Note: PANSS = Positive and Negative Syndrome Scale; BACS = Brief Assessment of Cognition in Schizophrenia; ToM-PST = Theory of Mind Picture Sequencing Task.*

3.3.5. Post-hoc task-specific assessment

The descriptive statistics for the linguistic measures extracted from participants' APACS Interviews are reported in the Supplementary Materials (Appendix B, Supplementary Table 3.1). A linear model on Concreteness Index (Group as sum contrast-coded predictor, with Control as the baseline) showed that the two groups did not significantly vary in the concrete content of their interviews (B = -0.71, SE = 0.43, t-value = -1.64, p-value = .104) and the difference became even more negligible when we included interviews length (in words) as a covariate (Group: B = 0.03, SE = 0.40, t-value = 0.07, p-value = .948; Length: B = -0.01, SE = 0.01, t-value = -5.47, p-value < .001).

3.4. Discussion

This study explores the novel idea that the pragmatic side of concretism, manifesting as an impairment in understanding figurative expressions, reflects difficulties in concrete aspects of the semantic representation of words. Here, we analyzed verbal explanations of non-literal sentences (i.e., idioms, metaphors, and proverbs) provided by patients and healthy controls using an automated pipeline, focusing on concreteness values of the words used in such explanations. Our results showed that verbal explanations provided by patients with schizophrenia were characterized by a higher occurrence of concrete words, especially for proverbs, and that more concrete explanations were also more likely to be incorrect. Interestingly, the tendency to rely more on a concrete vocabulary in explaining figurative expressions was also indicative of participants' pragmatic abilities, especially in the schizophrenia group. Finally, our results showed a pattern in the use of concrete vocabulary that was specific for a task involving figurative meanings, while it did not generalize to a semi-structured interview on autobiographical topics, thus confirming the link between word concreteness and pragmatic concretism.

Going more in depth, this study discloses the linguistic roots of concretism, by showing that the tendency to adhere to concrete aspects of the stimuli might underlie a propensity to remain anchored to concrete properties of the semantic representation of words. The evidence of higher use of concrete vocabulary in individuals with schizophrenia, as well as the link to pragmatic performance, strengthens the idea that communicative difficulties in patients might be linked to – and likely influenced by – deficits in the building blocks of language (Moro et al., 2015; Salavera et al., 2013; Tavano et al., 2008). This view reinforces the hypothesis that language impairment is multidimensional in schizophrenia (Bambini, Frau, et al., 2022), manifesting as specific deficits at different levels of processing and contributing to impairment in higher-level uses of language (Covington et al., 2005). Here, we focused on the semantic level, where individuals with schizophrenia are already known to exhibit difficulties linked to the relation among words within their semantic networks (Barattieri di San Pietro et al., 2023; Pomarol-Clotet et al., 2008; Spitzer et al., 1993) or related to the use of words from specific semantic classes (Buck et al., 2015; Minor et al., 2015). To existing literature, we added novel evidence that also the integration of semantic features related to conceptual representations and linked to perceptual experience, such as concreteness, might be impaired in schizophrenia. We can better elaborate on this finding by looking at the correlation between concreteness and imageability in our data, which suggests that individuals with schizophrenia might not only remain anchored to sensorial aspects of concrete concepts but possibly to the mental images aroused by them (Oertel et al., 2009). These considerations are particularly relevant, as implicate that difficulties in processing and integrating perceptual-based semantic features might directly interfere with figurative language understanding in schizophrenia (Tian & Poeppel, 2012).

Interestingly, pragmatic performance was more strongly affected by semantic features in the case of proverbs, which confirms that these expressions are particularly challenging for individuals with schizophrenia (Bambini et al., 2020; see also Felsenheimer & Rapp, 2023). We hypothesize that the extra difficulty in proverbs might be linked to the linguistic characteristics of these expressions. In particular, proverbs are an exceptional case of figurative language uses, where a highly concrete and plausible state of affairs allows for an abstract interpretation, conveying additional moral and social wisdom content (Sperber & Wilson, 1995). Crucially, the literal meaning plays a significant role in prompting the figurative interpretation of proverbs, serving as a "bridge" toward a more abstract level of representation (Unger, 2019). In such cases, where the literal interpretation remains active in the interlocutor's mind until the figurative one is reached (for other similar cases, see Carston, 2010b), it is plausible that individuals with schizophrenia might struggle in moving beyond the concrete interpretation due to concretism, with literal representation becoming an obstacle to the elaboration of the figurative one.

We did not find similar patterns (as the one observed for proverbs) in idioms and metaphors, although patients exhibited greater difficulty in understanding them, as reported in previous studies (Bambini, Arcara, Bechi, et al., 2016; Bambini, Arcara, et al., 2020; Langdon, Coltheart, et al., 2002; Langdon, Davies, et al., 2002; Pesciarelli et al., 2014; Sela et al., 2015). Especially in the case of metaphors, we assume that these findings might be explained by considering that the items included in the verbal explanation task (e.g., *Some voices are trumpets* or *Some handbags are boulders*) all convey information based on sensorial experience and require inferences on physical properties (Lecce et al., 2019). In this kind of expressions, lexical and semantic operations are applied to concrete properties (e.g., 'being shrieky' for the case of trumpets or 'being heavy' for the case of boulders), so word concreteness alone might not be a sensitive marker of concretism, as even incorrect interpretations might be as concrete as the correct ones. Conversely, we believe that metaphors conveying psychological information (e.g., *Some lawyers are sharks*) would more strongly detect the relationship between word concreteness and concretism, as such expressions would require conceptual operations on abstract properties to grasp a figurative interpretation about mental states (see Canal et al., 2022).

Another relevant point is that the correlation analysis reflected associations already observed for pragmatic impairment, generalizable also at the level of building blocks of language. In particular, patients' bias towards concrete properties of words patterned with lower neurocognitive skills, especially executive functioning, and lower mentalizing ability, mirroring the link between the pragmatic impairment and cognitive/sociocognitive deficits extensively documented in schizophrenia (Bambini et al., 2016; Binz & Brüne, 2010; Parola et al., 2018; for a comprehensive view, see also Frau, Cerami et al., 2024). These correlations also confirmed that more general difficulties in abstract thinking are highly related to deficits in other cognitive domains (Oh et al., 2015), including social-related skills (Berg et al., 2017). Interestingly, we also found a link between the higher use of concrete vocabulary and higher chlorpromazine daily intake: this finding is quite relevant, as it might either reflect iatrogenic effects of pharmacologic treatment – in line with other studies linking language deficits to specific antipsychotic (e.g., high dopamine D_2 receptor (D2R) occupancy antipsychotics were associated with more severe speech disturbances, see de Boer, Voppel, et al., 2020) – or might index a global measure of clinical severity impacting also on language ability. Notably, we did not observe a significant association between word concreteness and other psychopathological measures, including the N5 item of the PANSS assessing lack of abstract thinking. This finding, albeit surprising, suggests that concretism reflects a wider spectrum of psychopathological features, as already emphasized by early studies (Goldstein, 1944, 1959; Wright, 1975), while word concreteness seems to specifically capture its pragmatic manifestation. We should also notice that this study involved highly stabilized patients with a long-term course of disease, which might explain the lack of association between word concreteness and standardized measures of psychopathological symptoms.

A relevant point to understand the scope of the implications of our findings was to test how specific the effect of word concreteness was with respect to the task used. Our post-hoc task-specific assessment showed that patients' overuse of concrete words, albeit linked to their pragmatic impairment, did not generalize to a semi-structured interview on autobiographic topics. This task-dependent effect has strong methodological implications for the application of automated methods to the study language in schizophrenia (Corcoran et al., 2020; Hitczenko et al., 2021), as it shows that task properties matter in determining the sensitivity of linguistic measures acquired via computational approaches (see also Elvevåg et al., 2007).

Along this vein, the application of the investigation of word concreteness in schizophrenia has a great implication for natural language processing applied to psychiatric conditions. Now that the application of automated approaches to language in psychiatric conditions is being not only consolidated (Bora et al., 2021) but also expanded to natural language generation (Palaniyappan et al., 2023), our study opens perspectives on novel linguistic dimensions to be tested in method development and clinical research: in particular, we showed that word concreteness is a potentially relevant linguistic measure that should receive more attention in future studies, especially to test its reliability and cross-cultural generalizability (Parola et al., 2022). Through an adequate speech elicitation task, word concreteness might exhibit useful applications to assess higher-level aspects of communication and cognitive functioning in chronic individuals, extending also to progress monitoring in integrated treatment programs addressing pragmatic, cognitive, and sociocognitive skills (Bambini, Agostoni, et al., 2022; Bechi et al., 2020; Buonocore et al., 2018; Lindenmayer et al., 2013).

To conclude, in this study we unveiled novel aspects of the linguistic profile of schizophrenia, by showing that the pragmatic manifestation of concretism is rooted in lexical and semantic processes. These findings emphasize the centrality of the research on language for advancing the understanding of the cognitive phenotype of schizophrenia and other psychopathological disorders (Hinzen & Palaniyappan, 2024). But it also shows novel potential routes for the applications of natural language processing methods to the study of language in psychiatric conditions.

Appendix A. Technical details on audio and text pre-processing.

Audio file pre-processing and transcription.

The audio recordings from the Figurative Language 2 task were first converted to .wav files to be imported into the PRAAT software (Boersma & Weenink, 2021), with a standard quality of 44.10 kHz (capturing 44100 samples per second). Audio files were segmented using the PRAAT software, to obtain a single audio file for each item. Afterwards, audio files were manually transcribed verbatim by a trained transcriber. Interviewer turns, non-verbal vocalizations (e.g., laugh, yawn, cough, etc.), and false starts (e.g., it. "non ha dat- fatto la multa" eng. "he didn't giv- wrote a fine") were removed from the transcripts.

Once the manual transcribing process was completed, each audio recording underwent a speech analysis on the PRAAT software, to identify silent and filled pauses. Silent pauses were identified using PRAAT's automated silence detection algorithm, which extracted the number and the duration of silences from the audio files. The process of automated silence detection was followed by a manual check to prevent false positives and false negatives in pause identification. Only long pauses (i.e., 1-second or longer intra-turn silences) were considered and meticulously marked in participants' transcripts using an unambiguous diacritic mark (i.e., &&-).

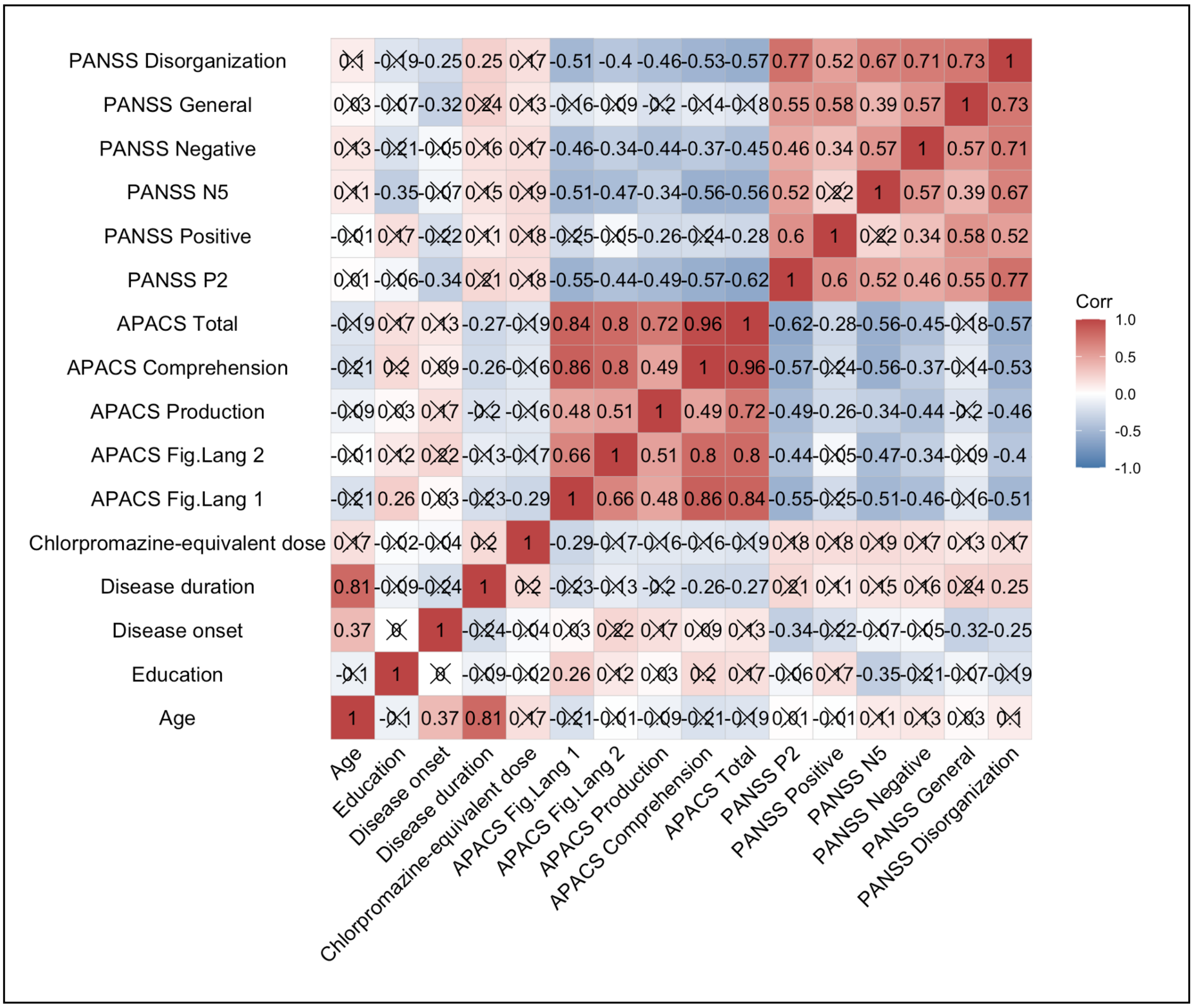
The PRAAT software was also used to manually check the occurrence of filled pauses in participants' speech, operationalized as non-lexicalized vocalizations within a turn representing a hesitation in speech (e.g., "ehm", "mh", etc.). Filled pauses were marked in the transcripts using the diacritic mark "&-" followed by the type of vocalization produced (i.e., "&-ehm", "&-mh", etc.).

Tokenization and lemmatization on transcripts.

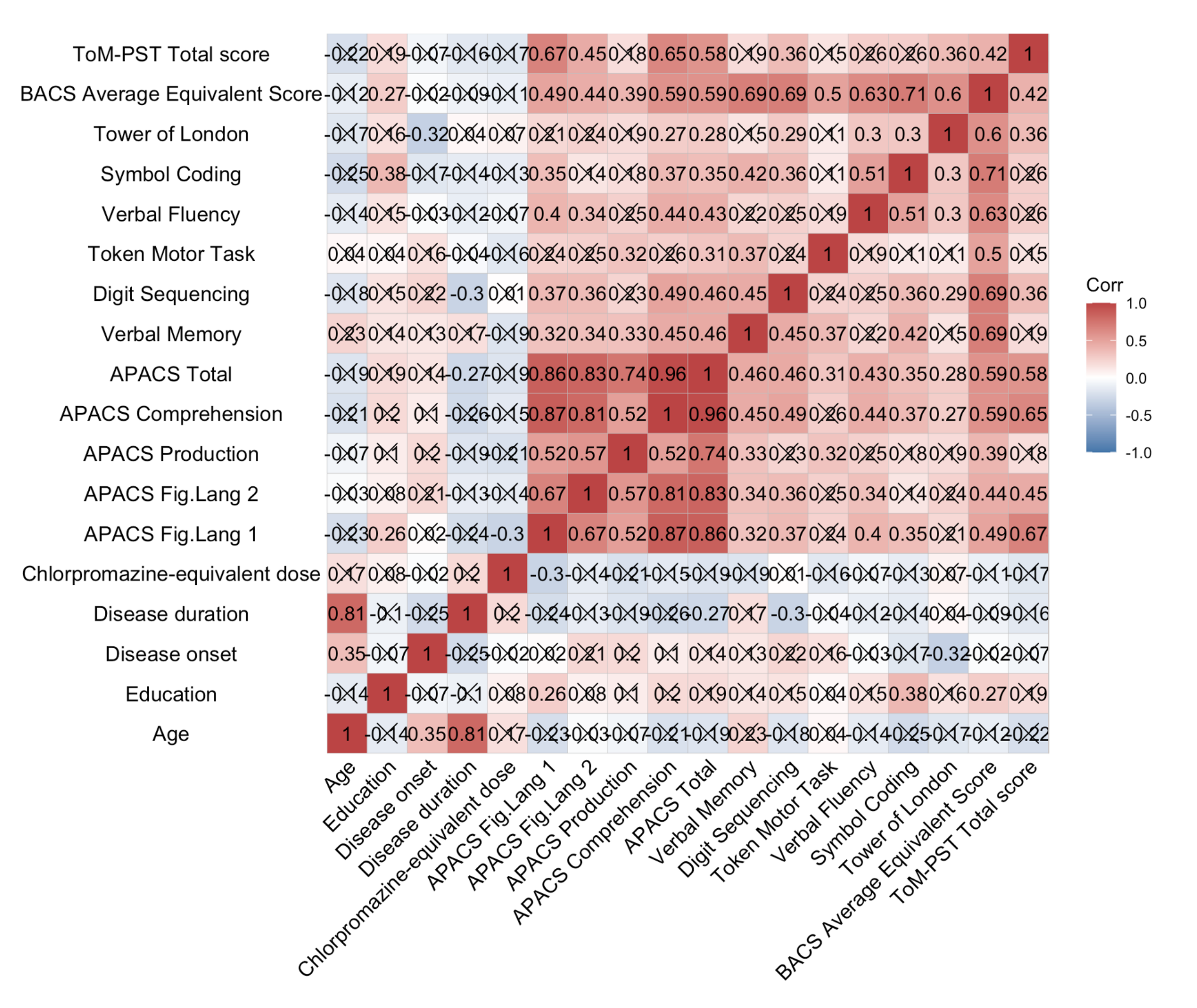
Participants' transcripts underwent a process of automated tokenization and lemmatization in R Studio software (R Core Team, 2023) to allow the extraction of the linguistic features.

First, text samples were automatically segmented into word units (i.e., tokens) using the *unnest_token()* function of *tidytext* package (Silge & Robinson, 2017). Then, tokens were automatically tagged into parts of speech (POS) and lemmatized using *udpipe* package, a text annotation pipeline that provides pre-trained linguistic models for more than 65 languages, including Italian. In particular, *udpipe_annotate()* function was used to perform POS tagging and lemmatization, applying a pre-trained model on Italian based on Universal Dependencies 2.0 treebanks (http://universaldependencies.org/): this model showed great performance in validation studies, with 97.2% accuracy in POS tagging and 97.4% accuracy in lemmatization compared to human performance (Straka & Straková, 2017).

Appendix B. Additional data to the correlations among assessment variables in the schizophrenia group and the post-hoc task-specific assessment.



Supplementary Figure 1. Association between assessment measures in the schizophrenia group (i). *The correlogram shows the correlations among patients' demographic, clinical, psychopathological, and pragmatic measures. The magnitude of associations is depicted by color (crossed cells indicate non-significant correlations, with significance level p < .05). Note: PANSS = Positive and Negative Syndrome Scale; APACS = Assessment of Pragmatic Abilities and Cognitive Substrates.*



Supplementary Figure 2. Association between assessment measures in the schizophrenia group (ii). *The correlogram shows the correlations among patients' demographic, clinical, pragmatic, cognitive, and sociocognitive measures. The magnitude of associations is depicted by color (crossed cells indicate non-significant correlations, with significance level $p < .05$). Note: APACS = Assessment of Pragmatic Abilities and Cognitive Substrates; BACS = Brief Assessment of Cognition in Schizophrenia; ToM-PST = Theory of Mind Picture Sequencing Test.*

Supplementary Table 1. Descriptive statistics (mean and standard deviations) and group comparisons related to linguistic variables extracted from the transcripts of APACS Interviews.

Measures	**Patients Mean (SD)**	**Controls Mean (SD)**	**Test Statistics**	***p*-value**
Interview length (in words)	160.37 (77.30)	107.00 (58.50)	$t(92) = 3.43$	$p = .002$
Lexical frequency (log)	5.66 (0.19)	5.66 (0.24)	$t(92) = 0.00$	$p = .999$
Type-token ratio	0.93 (0.04)	0.96 (0.03)	$t(92) = -4.49$	$p < .001$
Concreteness index	62.94 (1.96)	63.74 (2.52)	$t(92) = -1.68$	$p = .160$
Imageability index	70.94 (1.49)	71.29 (1.85)	$t(92) = -1.00$	$p = .399$

Notes: Due to limited availability of audio recordings, speech samples were available for 62 participants in the schizophrenia group and 32 participants from the control group. All p-values are FDR adjusted. Degrees of freedom in the t-tests vary due to missing values on some tests. Non-integer degrees of freedom are due to Welch independent t*-test computation.*

STUDY FOUR

HIGHER VISUAL MENTAL IMAGERY ACTIVATION DURING METAPHOR PROCESSING IN SCHIZOPHRENIA: A NEW WINDOW INTO CONCRETISM?[5]

Abstract

It is well-known that individuals with schizophrenia show impairment in metaphor comprehension, with a tendency to provide concrete interpretations of figurative expressions, often sticking to their literal meaning. What causes this concrete bias is, however, largely unexplained. In this work, we tested the hypothesis that concrete interpretations in schizophrenia might be linked to a greater and longer activation of visual images generated by metaphorical expressions. A sample of 66 individuals with schizophrenia and 70 healthy controls were administered a novel paradigm where metaphors (e.g., *Wisdom is a flashlight*) were used as primes for target words associated based on visual (e.g., *microphone*) vs. semantic features (e.g., *lamp*). While in healthy participants metaphors activated semantically associated words, in individuals with schizophrenia metaphor primed visually related words, with an effect lasting up to 1400 ms. Correlation analysis showed that the greater the visual priming effect the lower the metaphor comprehension skills in patients. These results suggest not only that visual images of the stimuli are activated during metaphor processing but also that sticking to these images might be a cause of figurative language impairment in schizophrenia. These results contribute to elucidating the mechanisms underlying concretism in schizophrenia and – at the theoretical level – they support a multimodal account of metaphor processing by spelling out its implications in case of impairment.

---

[5] This chapter is a manuscript in preparation for submission to a peer-review journal, with F. Frau and V. Bambini as co-first authors and M. Bosia as last author. This work received support from the European Research Council under the EU's Horizon Europe programme, ERC Consolidator Grant "PROcessing MEtaphors: Neurochronometry, Acquisition and Decay, PROMENADE" (G.A. number: 101045733). The content of this chapter is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

4.1. Introduction

Figurative language expressions, such as metaphors, are ubiquitous in everyday spontaneous speech (Glucksberg, 1989) and most of the speakers, when they run into them during a conversation, can intuitively and correctly process their meaning. However, non-literal expressions might not be as easily understood by individuals with schizophrenia, who tend to be anchored to a more literal, concrete interpretation of such expressions (Bambini, Arcara, et al., 2020; Rosen et al., 2021). Difficulty in grasping the figurative meaning of an utterance is often referred to using the clinical label of *concretism*, which reflects a more generalized impairment in the use of an abstract-symbolic mode of thinking (Harrow, 1972).

Concretism is currently considered one of the most relevant and pervasive features of the pragmatic profile of schizophrenia (Bambini, Arcara, Bechi, et al., 2016), as it disrupts not only metaphor (Deamer et al., 2019; Kircher et al., 2007; Mashal et al., 2013) but also idiom (Schettino et al., 2010; Sela et al., 2015; Titone et al., 2002) and proverb (Brüne & Bodenstein, 2005; Felsenheimer & Rapp, 2023) comprehension. Several studies pointed to the role of co-occurrent cognitive deficits – affecting in particular executive and sociocognitive skills (Brüne & Bodenstein, 2005; Langdon, Coltheart, et al., 2002; Li et al., 2017; Parola et al., 2020) – in explaining the mechanisms underlying concretism as manifested in pragmatic behavior, yet the variability among individuals and specific patterns of impairment cannot be fully explained by considering only these cognitive substrates (Frau, Cerami, et al., 2024). For instance, a recent study by Bambini et al. (*in preparation*, see §Study 4) focused on the interpretations given to figurative meanings and the link between concretism and the lexical-semantic properties, by testing whether patients with schizophrenia would rely more on the use of concrete vocabulary (i.e., words with higher concreteness, which are also associated with greater perceptual experience) when providing verbal explanations of non-literal expressions. The results confirmed the hypothesis and showed that a higher use of 'concrete' words was indicative of patients' difficulty in explaining figurative meanings. Capitalizing on such evidence and on the assumption that words denoting highly concrete concepts can also more easily arouse mental images (Paivio, 1979), here we will explore a more in-depth hypothesis, namely that the tendency to adhere to the literal meaning is linked to a lower-level difficulty in integrating multimodal information during pragmatic processing, particularly in moving beyond the concrete properties of words, and possibly the mental images aroused by the linguistic stimuli.

According to recent theoretical proposals in pragmatics, the process of grasping the figurative meaning of a metaphor entails not only abstract linguistic operations but also the activation of 'mental images' with multimodal content (e.g., visual, motor, auditory, etc.). While some accounts (e.g., within the Relevance Theory framework, Sperber & Wilson, 1995; Wilson & Sperber, 2004) argue that such images might be supportive of the pragmatic inferential procedure in most cases (Carston, 2010b, 2018), other positions claim that mental images activated during metaphor comprehension involve automated perceptual and motor simulations, which are supposed to be central to the pragmatic comprehension (Gibbs & Matlock, 2008). While the role of such mental images during pragmatic inferencing remains to be tested, experimental work seems to support the hypothesis that metaphor processing involves, at least in part, the activation of certain properties of the literal meaning of metaphoric sentences (Rubio Fernández, 2007), including multimodal features linked to sensory-motor experience (Al-Azary & Katz, 2021). Interestingly, mental images featuring sensory-motor properties of the literal meaning seem to become available in a flexible way, as they are more strongly activated in novel and less conventionalized metaphors (Al-Azary & Katz, 2021; Desai et al., 2011; Jamrozik et al., 2016).

The integration of representations coming from different modalities, however, might be difficult for individuals with schizophrenia. In particular, a number of studies documented generalized impairment in multi-sensory integration in this population (Tseng et al., 2015), reflecting a reduced ability to combine congruent streams of information from different modalities (e.g., visual or auditory signals) and integrate them into a coherent and unitary representation (de Gelder et al., 2003; Williams et al., 2010). Moreover, individuals with schizophrenia might also exhibit difficulties in managing mental imagery *per se* (Oertel et al., 2009; Stephan-Otto et al., 2017), which has also been linked to language dysfunction (Hoffman, 1986). This might also apply to the case of metaphors, where patients might struggle to integrate multimodal images spontaneously generated by metaphors with a more abstract verbal meaning.

4.1.1. The present study

The aim of this study was to test whether the activation of multimodal images during metaphor processing – in particular, visual properties aroused by the literal meaning of the sentence – might be related to difficulties in figurative language comprehension in schizophrenia.

To this purpose, we developed a novel metaphor priming experiment by capitalizing on evidence of priming effects in single-word and metaphor paradigms in both typical and atypical populations. First, we exploited previous evidence from single-word priming studies (Lam et al., 2015), showing that lexical access is facilitated when the target word is primed by another word that shares visual features such as shape (e.g., the recognition of "screwdriver" is facilitated when primed with "soldering iron") or motor features such as manipulation characteristics (e.g., the recognition of "screwdriver" is facilitated also when primed with "housekey"). Second, we also capitalized on evidence collected from typical adult individuals, showing that metaphors can be used to trigger priming effects on the lexical access of semantically related target words (Rubio Fernández, 2007) and target words related to the metaphor vehicle based on sensory-motor properties (Al-Azary & Katz, 2021). Finally, figurative expression priming effects can be impaired in psychiatric conditions, including schizophrenia (Titone et al., 2002), where concretism might induce higher priming effects on concrete-associated target words compared to abstract target words (Spitzer, 1993).

In our experiment, metaphors (e.g., *Wisdom is a flashlight*) were used as primes in a lexical decision task, followed by four different target conditions: i) semantic, ii) multimodal visual, iii) multimodal action, iv) and unrelated target words. In the semantic condition, target words were related to the metaphor vehicle (i.e., *flashlight*) based on amodal semantic properties (i.e., meaning associations in the same semantic network, as for *lamp*) and were used to compute semantic priming effects with respect to unrelated target words (Sperber et al., 1979), namely words with no association with the metaphor vehicle (e.g., *scissors*). In the multimodal visual and action conditions, target words were also related to the metaphor vehicle based on multimodal features, namely either visual (i.e., visual shape, as for *microphone*) or action (i.e., manipulability, as for *remote*) properties, to test for the activation of visual mental images beyond semantic priming effects, also potentially extending to the motor domain. Target words were presented at different inter-stimulus intervals (ISI), to test the temporal resolution of priming effects associated with different target types. Finally, we correlated priming effects with an offline measure of metaphor comprehension, alongside a more extended assessment addressing individual differences in psychopathology, neurocognition, lexical-semantic skills, and mental imagery.

We hypothesized that metaphors would trigger the activation of multimodal images – especially images linked to the visual properties of the metaphor vehicle – and such images, rather than helping, would become intrusive in schizophrenia. Specifically, we expected greater and longer metaphor priming effects on visual target words in individuals with schizophrenia compared to controls, as well as a significant association between visual activations and metaphor comprehension skills in schizophrenia.

4.2. Methods

4.2.1. Participants

Sixty-six individuals with a diagnosis of schizophrenia based on DSM-5 criteria (American Psychiatric Association, 2013) were recruited from the Department of Clinical Neurosciences, IRCCS San Raffaele Scientific Institute, Milan, Italy. All participants were Italian native speakers. Inclusion criteria were: age 18–65 years; being clinically stabilized and treated with a stable dose of the same antipsychotic therapy for at least 6 months. Exclusion criteria were: severe traumatic brain injury or neurological disorders, intellectual disability, alcohol or substance abuse in the preceding 6 months, and severe psychotic exacerbation in the preceding 3 months. Additionally, 73 Italian-speaking healthy controls were recruited, matched for age and education with the sample of patients. Inclusion criteria for the control group were: a) being a native speaker of Italian; b) not being bilingual from birth; c) absence of a diagnosis of psychiatric or neurological disorders. Two control subjects were excluded from the sample as they did not meet the inclusion criteria, while another control subject dropped out of the study and was not included in the sample.

An *a priori* power analysis was performed using G*Power software (Faul et al., 2009) to estimate the optimal sample size for this study: in line with moderate-to-large effect sizes reported by Lam et al. (2015) in a similar experiment ($0.20 \leq f^2$s ≤ 0.33, determined following Lakens, 2013), the power analysis indicated the required sample size to achieve 80% power for detecting a medium effect ($f^2 \geq 0.15$, Cohen, 1988), at a significance criterion of $a = .05$, was $N = 114$, accounting for the potential exclusion of participants during data cleaning (~15% of subjects).

All participants provided informed consent. The study was approved by the local ethical committee, following the principles of the Declaration of Helsinki.

4.2.2. Assessment

Participants from the schizophrenia group were assessed for psychopathology by a trained psychiatrist. Additionally, all participants (patients and controls) underwent a comprehensive assessment, administered by trained psychologists, including general lexical and semantic skills, pragmatic abilities, neurocognition, psychosocial well-being, and mental imagery. To contain patients' fatigue, the assessment in the schizophrenia group was conducted in three sessions, each lasting approximately one hour. In the control group, the assessment was performed in a single session (with a short break in between), lasting approximately 80 minutes.

4.2.2.1. Psychopathology

Psychopathology was assessed using the Positive and Negative Syndrome Scale for Schizophrenia (PANSS; Kay et al., 1987), a standardized measure for the clinical evaluation of typological and dimensional symptoms along the positive (e.g., delusions and disorganized thinking, score range: 7–49) and negative (e.g., poverty of speech and difficulty in abstract thinking, score range: 7–49) scales, as well as general psychopathology (e.g., depression and disorientation, score range: 16–112). Following previous works (Bambini, Frau, et al., 2022; Lucarini et al., 2022), we also computed a composite score assessing the disorganization dimension, obtained by summing the items of conceptual disorganization (P2), difficulty in abstraction (N5), stereotyped thinking (N7), mannerism (G5), disorientation (G10), poor attention (G11), lack of judgment and insight (G12), and disturbance of volition (G13).

4.2.2.2. Lexical and semantic skills

Lexical and semantic skills were assessed using the Italian version (Orsini & Laicardi, 1997) of the vocabulary subtest of the Weschler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981). The task includes a list of 35 words of increasing difficulty, presented to participants on a laptop screen one item at the time. Participants are asked to explain the meaning of each term. Scores are assigned as follows: wrong explanation (0 points), partial explanation (1 point), correct and complete explanation (2 points). The range of possible scores is 0–70.

4.2.2.3. Pragmatic skills

Pragmatic abilities were evaluated using the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS; Arcara & Bambini, 2016) Brief version, a novel tool currently under validation for its use in person, but already validated to be administered remotely (see Bischetti et al., 2023). The APACS Brief was designed for rapid evaluation of pragmatic skills in both production and comprehension along five short tasks, addressing different facets of pragmatic ability:

- *Interview*: participants are involved in a short structured autobiographic interview about childhood. Over-informativity, under-informativity, missing references, and abrupt topic shifts are evaluated as communicative difficulties (score range: 0–4).
- *Narratives*: participants are asked to carefully listen to a short news-like story and to answer different questions about story content and the meaning of two figurative expressions included in the text (score range: 0–6).
- *Figurative Language 1*: participants are presented with a multiple-choice task, including 3 different figurative expressions (one idiom, one metaphor, and one proverb), and are asked to select the correct interpretation of each expression among three alternatives (score range: 0–3).
- *Humor*: participants are presented with a multiple-choice story completion task, where are asked to select the most fitting (i.e., funniest) ending for short passages with humorous content (score range: 0–2).
- *Figurative Language 2*: participants are presented with different figurative expressions (two metaphors and one proverb) and are asked to verbally explain their meaning (score range: 0–3).

The APACS Brief total score is obtained by averaging task scores transformed in proportions (score range: 0–1).

4.2.2.4. Neurocognition

Neurocognitive abilities were assessed through the Italian version (Anselmetti et al., 2008b) of the Brief Assessment of Cognition in Schizophrenia (BACS; Keefe, 2004), a validated tool aiming at assessing cognitive functions that are consistently impaired and related to schizophrenia spectrum disorders. The BACS includes seven short tasks, assessing six different cognitive domains:

- Verbal Memory (*List learning task*): participants are presented with a list of 15 words and are asked to recall as many as possible in five trials. The total number of words recalled per trial (in any order) is measured (score range: 0–75).

- Working memory (*Digit sequencing task*): participants are presented with arrays of numbers of increasing length and are asked to repeat the numbers in ascending order. The number of correct responses is measured (score range: 0–28).
- Motor speed (*Token motor task*): participants are given 100 plastic tokens and are asked to simultaneously place them two at a time in a container using both hands within one minute. The number of tokens correctly placed into the container in the first half-minute, in the second half-minute, and in the whole minute are measured (score range: 0–100).
- Semantic and Phonetic Fluency (*Category instances* and *Controlled oral word association tasks*): participants are given 60 seconds to produce as many words as possible within a certain category (semantic fluency) or starting with a given letter (phonetic fluency). The number of unique and valid words produced is used as score.
- Attention and Speed of Information Processing (*Symbol coding task*): participants are given 90 seconds to write as many numerals (1-9) as possible to match specific symbols on a response sheet. The number of correct digits is measured (score range: 0–110).
- Executive Functions (*Tower of London task*): Participants are presented with two pictures (A and B), showing a vertical abacus with three rods and balls placed in a unique pattern, and are asked to come up with the total number of movements required for the balls in picture B to represent picture A. The number of correct responses is measured (score range: 0–22).

Raw scores were corrected for age, gender, and education. Equivalent scores were calculated for each task, indicating whether a participant's performance fell within the neurotypical range (score range: 0–4). An average equivalent score across the six cognitive domains was also computed.

4.2.2.5. Mental imagery

The assessment of mental imagery focused on two specific facets known for being potentially impaired in the schizophrenia spectrum (Pearson et al., 2013), namely vividness as well as maintenance and inspection of mental images.

The vividness of mental images was assessed using the visual and kinesthetic subscales of the shortened form of Betts' Questionnaire upon Mental Imagery (QMI; Sheehan, 1967), translated into Italian and adapted to be administered to individuals with schizophrenia. In our adapted version, participants are asked to evoke specific visual (e.g., the sun as it is sinking below the horizon, score range: 5–35) or motor (e.g., thinking of performing certain actions such as running upstairs, score range: 5–35) mental images and rate the vividness of such images on a 7-point scale ranging from 1 (i.e., "No image present at all") and 7 (i.e., "Perfectly clear and as vivid as the actual experience"). A composite score of mental imagery vividness was obtained by summing all item scores from both subscales (score range: 10–70).

The ability to maintain and inspect mental images was assessed using the Mental Imagery Test (MIT; Di Nuovo et al., 2014), a validated battery developed to evaluate several facets of imagistic processes using standardized methods. In particular, for this study we selected two tasks that could be easily administered to individuals with schizophrenia, including:

- *The Brooks' "F" test*: participants are presented with a printed upper-case letter "F" and then are asked to walk along its contour using their imagination, saying whether the edges of the contour were internal or external (score range: 0–10).
- *Mental exploration of a map*: participants are presented with a printed map and then are asked to answer four questions about the comparative distance between elements represented in the map (score range: 0–6).

A composite score of mental imagery maintenance and inspection was obtained by summing the subscores from both tasks (score range: 0–16).

4.2.2.6. Psychosocial well-being

Psychosocial well-being was evaluated using the 42-item version of Ryff's Psychological Well-Being Scales (PWB; Ryff, 1989), validated to be used in Italian (Ruini et al., 2003). The PWB is a self-administered questionnaire, evaluating psychosocial well-being along six dimensions (each including 7 items): Autonomy, Environmental mastery, Personal Growth, Positive Relations, Purpose in life, and Self-acceptance. For each item, the participant is asked to rate the degree of agreement with the content of the sentence using a 6-point scale ranging from 1 (i.e., "Strongly disagree") and 6 (i.e., "Strongly agree"). Negatively phrased items are reversed. The total score of the PWB is obtained by summing all scores across items (score range: 1–252).

4.2.3. The metaphor priming experiment

4.2.3.1. Materials

4.2.3.1.1. Prime metaphors

A set of prime metaphors (all with an *X is Y* structure) was constructed to be employed in the priming experiment in critical and filler trials. Critical trials included 20 novel prime metaphors each involving a vehicle denoting familiar tools or manipulable objects and a topic denoting either concrete (e.g., *That cake is a sponge*) or abstract concepts (e.g., *Wisdom is a flashlight*). All critical metaphors were rated for familiarity and difficulty by a group of 16 Italian-speaking adults (9 females, Age: M = 28.75, SD = 7.52; Education: M = 18.13, SD = 1.31). Participants were asked to judge on a 7-point scale the extent to which each sentence was familiar to them (1 = not familiar, 7 = definitely familiar) and how difficult it was to understand the meaning of the sentence (1 = not difficult, 7 = definitely difficult). On average, critical prime metaphors were perceived as mostly unfamiliar (M = 2.39, SD = 1.40, range: 1.25–6.44), yet not too difficult to understand (M = 3.08, SD = 1.06, range: 1–4.69).

Finally, filler trials included 20 additional prime metaphors to be used in filler trials. Filler prime metaphors were extracted from available databases and were already normed for familiarity and difficulty in previous studies (Bambini et al., 2013).

4.2.3.1.2. Target words

Critical prime metaphors were paired with four different target words, all denoting manipulable objects, corresponding to the four priming conditions: i) semantic, ii) multimodal visual, ii) multimodal action, or iv) unrelated. The semantic condition was created to test for semantic priming effects and included target words related to the metaphor vehicle by amodal semantic properties (e.g., meaning association), without any visual or action relatedness (e.g., *lamp* for the prime metaphor *Wisdom is a flashlight*). The visual and action conditions were created to test for multimodal priming effects and included: a) visual target words, namely targets related to the metaphor vehicle by visual appearance or shape, without strong action or meaning associations (e.g., *microphone* for the prime metaphor *Wisdom is a flashlight*); b) action target words, namely targets related to the metaphor vehicle by manipulability (i.e., the hand-related motor schema required to use the object), with no overt visual or meaning associations (e.g., *remote* for the prime metaphor *Wisdom is a flashlight*). Finally, the unrelated condition indicated target words without any association with the metaphor vehicle for either multimodal or amodal properties (e.g., *scissors* for the prime metaphor *Wisdom is a flashlight*). The prime metaphor vehicle was matched with target words across conditions for log-transformed lexical frequency ($F(4,95) = 0.85$, $p = .499$), extracted from the Corpus and Frequency Lexicon of Written Italian (CoLFIS; Bertinetto et al., 2005), and length as measured in number of syllables ($F(4,95) = 0.28$, $p = .892$). See Table 4.1.

Table 4.1. Means and standard deviations for log-transformed lexical frequency and length of prime metaphor vehicles and target words across conditions.

Measure	Prime vehicles	Target words			
		Semantic	Visual	Action	Unrelated
Lexical frequency (log)	3.85 (1.59)	3.38 (1.64)	3.76 (1.65)	3.02 (1.67)	3.83 (2.14)
Length (in syllables)	3.05 (0.76)	3.25 (1.07)	3.20 (1.11)	3.35 (0.88)	3.25 (0.79)

To validate our prime-target associations, we performed a rating task, involving 17 Italian-speaking adults (8 females, Age: $M = 31.18$, $SD = 12.38$; Education: $M = 18.12$, $SD = 1.27$), who had not participated in the validation of the prime metaphors. Participants were presented with word pairs (the vehicle of a prime metaphor presented with a target word, e.g. *flashlight – lamp*) and were asked to rate the extent to which the two words were associated by meaning associations (i.e., semantic relatedness), visual properties (i.e., visual relatedness), and manipulability (i.e., action relatedness) using a 7-point scale (1 = not related, 7 = definitely related). Metaphor vehicles were presented in association with all target words in random order and the resulting pairs were rated for all relatedness dimensions. Except for the unrelated condition, a vehicle-target pair to be considered acceptable had to obtain an average score ≥ 3.5 rating points on the target relatedness dimension, namely the dimension for which the target word was selected. Starting from semantic relatedness, word pairs including metaphor vehicles and semantic target words were more strongly associated by semantic relatedness compared to word pairs matching metaphor vehicles and either visual, action, or unrelated target words (prime-semantic vs. prime-visual: $t(38) = 10.80$, $p < .001$; prime-semantic vs. prime-action: $t(38) = 8.09$, $p < .001$, prime-semantic vs. prime-unrelated: $t(38) = 22.44$, $p < .001$). Moving to visual relatedness, the word pairs including metaphor vehicles and visual target words obtained higher ratings for this dimension, compared to word pairs matching metaphor vehicles and either action, semantic, or unrelated target words (prime-visual vs. prime-action: $t(38) = 4.93$, $p < .001$; prime-visual vs. prime-semantic: $t(38) = 6.80$, $p < .001$, prime-visual vs. prime-unrelated: $t(38) = 11.82$, $p < .001$). As for action relatedness, word pairs including metaphor vehicles and action target words obtained higher ratings in this dimension compared to word pairs matching metaphor vehicles and either visual, semantic, or unrelated target words (prime-action vs. prime-visual: $t(38) = 2.42$, $p = .020$; prime-action vs. prime-semantic: $t(38) = 5.44$, $p < .001$, prime-action vs. prime-unrelated: $t(38) = 10.80$, $p < .001$). Rating values for all word pairs across the three relatedness dimensions are reported in Table 4.2.

Table 4.2. Mean and standard deviations for rating values, measuring the association between metaphor vehicles (prime) and each target word (semantic, visual, action, and unrelated) for meaning association (semantic relatedness), visual properties (visual relatedness), and manipulability (action relatedness).

Measure	**Prime-Semantic**	**Prime-Visual**	**Prime-Action**	**Prime-Unrelated**
Semantic relatedness	5.32 (0.81)	2.19 (1.02)	2.57 (1.29)	1.22 (0.13)
Visual relatedness	2.17 (1.24)	4.78 (1.18)	2.90 (1.23)	1.38 (0.51)
Action relatedness	2.67 (1.14)	3.66 (1.35)	4.59 (1.09)	1.57 (0.61)

Exploratorily, we also checked the semantic similarity between prime metaphor vehicles and the associated target words, computed using *LSAfun* R package (Günther et al., 2015) with a semantic space built on ItWac corpus (Baroni et al., 2009): the cosine similarity was significantly lower for Prime-Unrelated pairs compared to Prime-Semantic ($t(36) = 7.32$, $p < .001$), Prime-Visual ($t(36) = 2.36$, $p = .024$), and Prime-Action ($t(33) = 4.76$, $p < .001$) pairs, while it was significantly higher for Prime-Semantic pairs compared to Prime-Visual ($t(36) = 4.47$, $p < .001$) and Prime-Action ($t(33) = 2.83$, $p = .007$) pairs.

4.2.3.2. Procedure

The priming experiment was administered as a lexical decision task, preceded by the presentation of a prime metaphor. Each participant was randomly assigned to a single ISI group (400, 1000, and 1400 ms) and was presented with 120 trials (80 critical and 40 filler trials), divided into four blocks with a short break between blocks. Items and blocks were randomized. After a fixation cross (500 ms), the prime metaphor appeared on the screen for 1500 ms followed by a variable ISI.

Afterwards, the target word appeared and lasted on the screen for 2000 ms, during which the participants had to express whether the target word was an existing Italian word or not by pressing two different buttons on a Cedrus® RB-540 response pad. In critical trials, a (real) target word was presented, while in filler trials a pseudo-word (e.g., *matolca*) was presented. The button order was counterbalanced across participants to control for handedness confounds. After 2000 ms, the next trial automatically began, preceded by a blank screen lasting 500 ms. Before starting with the actual experiment, each participant completed 10 practice trials. The trial structure and the rationale of the procedure is exemplified in Figure 4.1. The experiment was implemented on PsychoPy© software (Peirce et al., 2019).

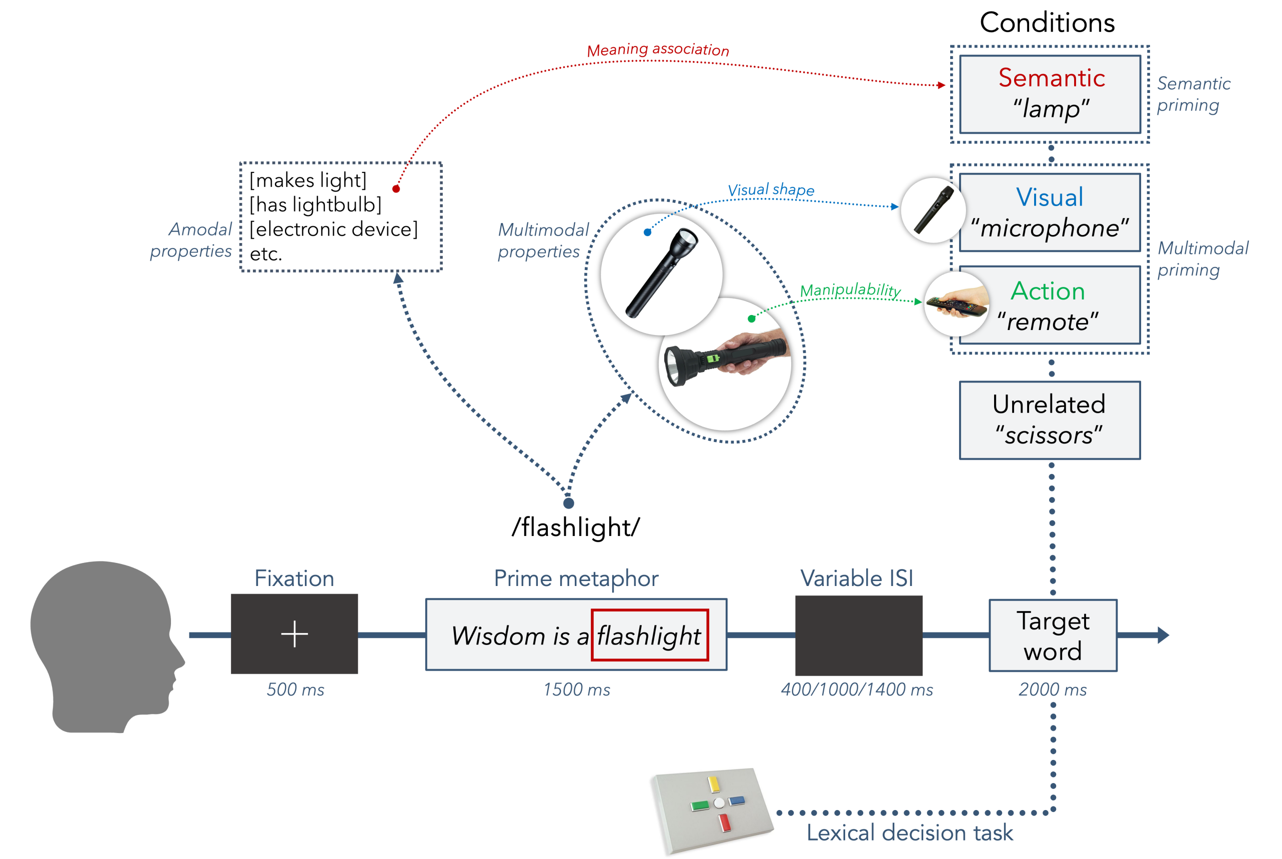


Figure 4.1. The trial structure of the metaphor priming experiment. *Participants were asked to read a prime metaphor (e.g., "Wisdom is a flashlight", with "flashlight" as the metaphor vehicle) and then were presented with a target word after a variable inter-stimulus interval (ISI). The target word could be completely unrelated with respect to the metaphor vehicle (i.e., "scissors") or could be associated with it by either amodal semantic properties (i.e., meaning association, as with "lamp") or multimodal properties, such as visual (i.e., visual shape, as with "microphone") or action (i.e., manipulability, as with "remote") characteristics. Participants were asked to perform a lexical decision task, namely to say whether the target word was an existing Italian word or not by pressing two buttons on the button box.*

After the metaphor priming experiment, a multiple-choice offline metaphor comprehension task was administered to ensure that participants were able to understand the prime metaphors. Participants were presented with prime metaphors (one item at the time) on a laptop monitor and were presented with three possible interpretations (correct, incorrect concrete, and incorrect unrelated interpretations). They were instructed to choose the correct interpretation among the three alternatives. A binomial score (1 = correct, 0 = incorrect) was assigned for each metaphor (score range: 0–20).

4.2.4. Statistical Analysis

The strategy for statistical analysis included three stages: i) we first compared the schizophrenia group with the control group for all the measures acquired during the assessment, ii) then we analyzed the output of the metaphor priming experiment by comparing responses in the two groups across conditions and inter-stimulus intervals, and iii) finally we tested whether the activation effects in the metaphor priming experiment could be related to participants' ability to understand metaphors – as measured using the multiple-choice offline metaphor comprehension task – as well as individual difference variables.

To accomplish stage i), we run a series of independent sample *t*-tests along the measures acquired in both groups (i.e., lexical and semantic skills, pragmatic abilities, neurocognition, psychosocial well-being, and mental imagery), to test whether schizophrenia patients exhibited a different performance compared to control participants. The *t*-tests' normality assumption was checked by visual inspection of the distribution of the dependent variables, while homogeneity of variance assumption (homoskedasticity) was checked with tests of variance (using *var.test()* function in R). Welch's independent sample *t*-test was used in the case of heteroskedasticity. All *p*-values were adjusted for False Discovery Rate (FDR; Benjamini & Hochberg, 1995).

As for stage ii), we first aimed to test whether the activation of semantic, meaning-related properties of the metaphor vehicle would elicit semantic priming effects, namely faster response times in the recognition of semantically related target words. Then, we aimed to test whether the activation of multimodal properties of the vehicle (i.e., Visual and Action targets) would trigger stronger priming effects on Visual and Action target words beyond the activation of amodal semantic priming effects. To do so, we fitted a Linear Mixed-effects Model on participants' *z*-transformed latencies of correct responses using *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017), including Condition (within subjects, four levels: Visual, Action, Semantic, and Unrelated) in interaction with Group (between subjects, two levels: Patients and Controls) as fixed factors nested within ISI (between subjects, three levels: 400, 1000, and 1400 ms). Following Al-Azary and Katz (2021), we used two different baseline conditions to compute our priming effects to maximize the chance of intercepting robust activations: semantic priming effects were computed as the difference between the Unrelated and the Semantic target words, while the multimodal priming effects were computed as the difference between Semantic and either Visual or Action target words. Therefore, we constructed the contrast coding matrix for Condition in a way that would allow us to test these differences in the same model (Semantic Priming: Unrelated – Semantic; Additional Visual Priming: Semantic – Visual; Additional Action Priming: Semantic – Action). Exploratorily, we also fitted another model including also Metaphor Familiarity as a factor (within subjects, two levels: Low and High) in interaction with Group and Condition, to explore whether multimodal priming effects relative to semantic activations could be modulated by differences in item familiarity. In all models, the Group factor was sum-coded (Controls = 0.5, Schizophrenia = –0.5).

The optimal random structure was always determined in a stepwise fashion using the *buildmer* package (Voeten, 2023), starting from a maximal model including all possible fixed and random effects. The final random structure included by-subjects and by-items random intercepts only, as more complex structures did not lead to model convergence.

As for stage iii), we first inspected patterns of correlations across relevant inter-stimulus intervals using correlograms, computed for the schizophrenia and control groups separately using Spearman's correlations. Before exploring the correlations among relevant variables, we run a series of one-way ANOVA tests to check that participants assigned to different ISI groups were not significantly different for demographic, clinical, and cognitive variables. We then tested which variables predicted metaphor comprehension scores in both groups: the variables showing a significant association with the metaphor comprehension total score were used as predictors in a multiple regression model, with the total score of the metaphor comprehension task as the dependent variable nested across relevant inter-stimulus intervals and groups.

The analysis was performed using RStudio (R Core Team, 2023), version 2023.12.0+369.

4.3. Results

4.3.1. Demographic, clinical, and cognitive characteristics

The sample included 66 patients with schizophrenia, who were all stabilized and treated with antipsychotic therapy for at least 3 months. The control group included 70 participants, matched for age and education with the schizophrenia group. Descriptive statistics for demographic, clinical, and cognitive measures as well as between-group comparisons are reported in Table 4.3.

The *t*-tests comparing the two groups showed that participants with schizophrenia performed significantly worse than control participants in the WAIS-R Vocabulary subtask (assessing lexical-semantic skills), the APACS Brief test (assessing pragmatics skills), and the BACS test (assessing neurocognitive skills). Participants with schizophrenia reported lower vividness of mental images, as evaluated using the QMI subscales, but did not score worse than controls in tasks assessing manipulability and maintenance of mental images (as evaluated using the MIT subtasks). Participants with schizophrenia reported reduced psychosocial well-being compared to control subjects, as resulted from the PWB questionnaire, and showed a slightly worse performance in the multiple-choice offline metaphor comprehension task compared to control participants.

Table 4.3. Descriptive statistics (mean and standard deviations) of patients and controls with group comparisons across descriptive, clinical, and cognitive measures.

Measures	**Patients Mean (SD)**	**Controls Mean (SD)**	**Test Statistics**	***p*-value**
Age	40.03 (13.01)	42.13 (15.42)	$t(134) = 0.86$	$p = .394$
Education (years)	11.62 (2.69)	11.99 (2.61)	$t(134) = 0.80$	$p = .424$
Gender (F/M)	17/49	50/20	$\chi^2(1) = 28.35$	$p < .001$
Illness onset (years)	23.08 (6.20)	-	-	-
Illness duration (years)	16.95 (10.53)	-	-	-
Mean chlorpromazine-equivalent dose (mg/d)	493.55 (227.80)	-	-	-
PANSS Positive	16.98 (4.75)	-	-	-
PANSS Negative	21.61 (6.38)	-	-	-
PANSS General	40.37 (9.20)	-	-	-
PANSS Disorganization	21.17 (6.44)	-	-	-
WAIS-R Vocabulary	44.86 (11.6)	52.22 (9.72)	$t(131) = 3.97$	$p < .001$
APACS Brief Total	0.61 (0.11)	0.87 (0.08)	$t(59.61) = 12.81^b$	$p < .001$
BACS[a]	1.70 (0.92)	2.63 (0.71)	$t(112.07) = 6.46$	$p < .001$
QMI Composite	53.48 (8.21)	61.00 (6.97)	$t(132) = 5.73$	$p < .001$
MIT Composite	14.70 (8.21)	15.01 (6.97)	$t(132) = 0.94$	$p = .351$
PWB Total	159.29 (22.20)	193.87 (16.20)	$t(124.43) = 10.34$	$p < .001$
Metaphor Comprehension Total	16.30 (2.41)	17.34 (3.42)	$t(115.97) = 2.04$	$p = .044$

Notes: PANSS = Positive and Negative Syndrome Scale, BACS = Brief Assessment of Cognition in Schizophrenia, QMI = Questionnaire upon Mental Images, MIT = Mental Imagery Test, PWB = Psychological Well-Being Scales. All p*-values were adjusted for False Discovery Rate (FDR).*

[a] *Mean of the equivalent scores from each subtask of the BACS*

[b] *The comparison refers to 108 participants only, as the test could not be administered to other participants*

4.3.2. Results of the metaphor priming experiment

The metaphor priming data were pre-processed and cleaned by excluding all participants with < 70% of correct answers in the lexical decision task ($n = 2$ participants from the schizophrenia group and $n = 2$ participants from the control group were removed). We also trimmed trials that were: i) incorrect, ii) faster than 250 ms, and iii) exceeding |2.5| standard deviations from participants' means across conditions (4.34% and 6.81% deleted trials for controls and patients, respectively). Finally, only responses for which the full condition quadruplet (i.e., Unrelated, Semantic, Action, and Visual) was available were included in the analysis. The descriptive measures of participants' latencies are reported in Table 4.4.

Table 4. Latencies in milliseconds (mean and standard deviations) for patients and controls across conditions and inter-stimulus intervals (ISI), alongside semantic and multimodal (visual and action) priming effects.

Condition	ISI 400 ms		ISI 1000 ms		ISI 1400 ms	
	Controls	Patients	Controls	Patients	Controls	Patients
Unrelated	728.86 (190.71)	859.14 (306.96)	781.08 (241.61)	811.84 (260.79)	735.85 (186.48)	811.65 (208.42)
Semantic	726.12 (228.88)	854.36 (297.30)	765.32 (214.89)	833.87 (264.40)	737.10 (201.91)	839.97 (238.43)
Visual	734.09 (255.55)	854.30 (300.09)	788.03 (230.42)	821.25 (266.32)	769.41 (227.77)	832.36 (234.93)
Action	753.17 (213.56)	894.95 (334.48)	811.31 (223.64)	863.33 (266.89)	788.50 (231.10)	897.40 (251.44)
Semantic Priming	2.73	4.77	15.76	-22.03	-1.25	-85.75
Multimodal Visual priming	-7.96	0.06	-22.71	12.63	-32.31	7.60
Multimodal Action priming	-27.04	-40.58	-45.99	-29.45	-51.39	-57.43

Notes: Semantic priming effects are computed as the difference between the latencies of the Unrelated minus the Semantic target words. Multimodal priming effects are computed as the difference between the latencies of the Semantic minus the Visual/Action target words.

The Linear Mixed-Effects model aiming at testing the presence of Semantic priming effects, on the one hand, and Multimodal (Visual and Action) priming effects beyond amodal semantic activations, on the other hand, showed a main effect of Group at 400 ms and – marginally – at 1400 ms, indicating that in these time windows patients with schizophrenia were globally slower than controls. Across ISI, the model also showed a main negative effect of Condition for Multimodal Action priming effects, indicating that in both groups Action target words elicited slower responses compared to the Semantic condition. There was also a significant Group ×

Condition interaction for the Semantic Priming at 1000 ms: in this time window, control subjects showed faster latencies for Semantic target words (relative to the Unrelated) compared to patients. Moreover, the model also detected a significant Group × Condition interaction for the Multimodal Visual priming at 1000 and 1400 ms: this interaction indicates that in these time windows the Visual target words were processed faster than the Semantic ones in the schizophrenia group compared to the control group. The output of the model is reported in Table 4.5 and Figure 4.2.

Table 4.5. Output of the Linear Mixed-Effects Model on participants' z-transformed latencies, testing the presence of Semantic and Multimodal (Visual and Action) priming effects across Inter-Stimulus Intervals (ISI) in the schizophrenia and control groups (significant effects are highlighted in bold).

Fixed Effects	**Inter-Stimulus Interval (ISI)**											
	ISI 400				**ISI 1000**				**ISI 1400**			
	B	***SE***	***t***	***p***	***B***	***SE***	***t***	***p***	***B***	***SE***	***t***	***p***
Group	**-0.16**	**0.07**	**-2.40**	**.017**	-0.07	0.07	-1.13	.259	**-0.12**	**0.07**	**-1.83**	**.067•**
Condition: Semantic	0.01	0.01	0.98	.328	-0.01	0.01	-0.66	.511	-0.01	0.01	-1.26	.206
Condition: Multimodal Visual	-0.00	0.01	-0.22	.824	-0.00	0.01	-0.39	.699	-0.01	0.01	-1.36	.174
Condition: Multimodal Action	**-0.04**	**0.01**	**-4.13**	**<.001**	**-0.05**	**0.01**	**-4.75**	**<.001**	**-0.06**	**0.01**	**-6.31**	**<.001**
Group × Condition: Semantic	0.01	0.02	0.40	.687	**0.04**	**0.02**	**2.08**	**.038**	0.03	0.02	1.36	.175
Group × Condition: Multimodal Visual	-0.00	0.02	-0.24	.810	**-0.04**	**0.02**	**-2.27**	**.023**	**-0.04**	**0.02**	**-2.19**	**.028**
Group × Condition: Multimodal Action	-0.00	0.02	-0.12	.903	-0.02	0.02	-1.16	.245	0.01	0.02	0.26	.794
Random Effects	**Variance**	**SD**										
$Intercept_{Subject}$	0.05	0.22										
$Intercept_{Item}$	0.00	0.06										
Residuals	0.03	0.18										
$ICC_{SubjectItem}$	0.60											
Model fit	Marginal	Conditional										
R^2	.049	.615										

*Notes: B = model estimates; SE = standard error; t = model statistics; p = p-value; Semantic = Semantic priming; Multimodal Visual/Action = Multimodal Visual/Action priming. The model included only the complete cases. Model formula: Log-transformed RT ~ ISI / (Group * Condition) + (1 | Subject) + (1 | Item). The fixed effects for ISI are omitted. Contrasts of Group: Controls – Schizophrenia. Contrasts of Condition: Semantic = Unrelated – Semantic; Multimodal Visual = Semantic – Visual; Multimodal Action = Semantic – Action.*

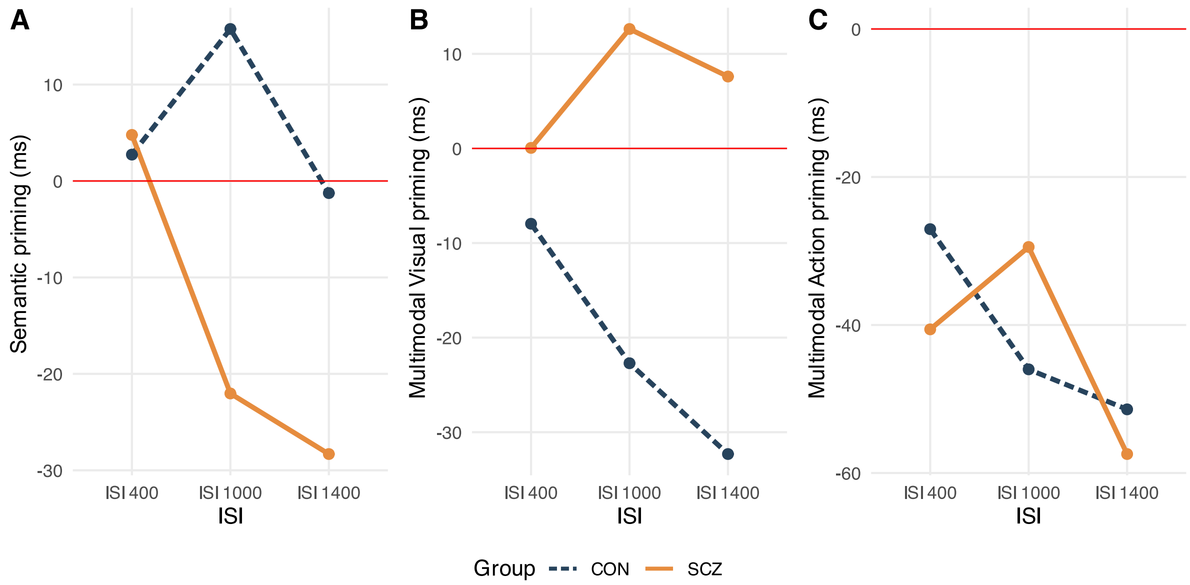


Figure 4.2. Semantic and Multimodal (Visual and Action) priming effects across groups and inter-stimulus intervals (ISI). *The plot in (A) depicts the mean Semantic priming effects (in milliseconds) for schizophrenia (SCZ) and control (CON) groups, alongside Multimodal Visual and Action priming effects (plots in B and C), across 400, 1000, and 1400 millisecond inter-stimulus intervals. Positive values indicate facilitatory effects of the condition relative to the baseline, negative values indicate inhibitory effects. The red line indicates the absence of activations.*

The Linear Mixed-Effects model including also Familiarity as a fixed factor in interaction with Group and Condition confirmed the effects of the previous model, but showed also a significant Condition × Familiarity interaction across all ISI. In particular, the model showed that in both groups Semantic priming effects were higher in highly familiar metaphors relative to more unfamiliar ones, while Multimodal priming effects for both Visual and Action target words were higher in unfamiliar metaphors relative to more familiar ones. The output of this model is reported in Table 4.6 and Figure 4.3.

Table 4.6. Output of the Linear Mixed-Effects Model on participants' z-transformed latencies, testing the presence of differential activations for Visual and Action targets relative to Semantic targets in interaction with metaphor Familiarity across Inter-Stimulus Intervals (ISI) in the schizophrenia and control groups (significant effects are highlighted in bold).

Fixed Effects	Inter-Stimulus Interval (ISI)											
	ISI 400				ISI 1000				ISI 1400			
	B	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Group	**-0.16**	**0.07**	**-2.40**	**.016**	-0.07	0.07	-1.14	.255	**-0.12**	**0.07**	**-1.84**	**.066•**
Condition: Semantic	0.01	0.01	1.30	.194	-0.00	0.01	-0.28	.776	-0.01	0.01	-0.65	.515
Condition: Multimodal Visual	-0.01	0.01	-0.65	.514	-0.01	0.01	-1.09	.275	**-0.02**	**0.01**	**-2.10**	**.036**
Condition: Multimodal Action	**-0.05**	**0.01**	**-4.72**	**<.001**	**-0.05**	**0.01**	**-5.32**	**<.001**	**-0.07**	**0.01**	**-7.04**	**<.001**
Familiarity	0.02	0.03	0.91	.364	0.02	0.03	0.59	.553	0.03	0.03	1.26	.208
Group × Condition: Semantic	0.01	0.02	0.41	.678	**0.04**	**0.02**	**1.97**	**.049**	0.03	0.02	1.41	.158
Group × Condition: Multimodal Visual	-0.01	0.02	-0.33	.742	**-0.04**	**0.02**	**-2.19**	**.029**	**-0.04**	**0.02**	**-2.07**	**.038**
Group × Condition: Multimodal Action	-0.01	0.02	-0.32	.746	-0.02	0.02	-1.05	.296	0.00	0.02	0.21	.834
Group × Familiarity	-0.01	0.01	-0.57	.572	-0.01	0.01	-0.80	.426	-0.01	0.01	-0.45	.656
Condition: Semantic × Familiarity	**0.06**	**0.02**	**2.74**	**.006**	**0.05**	**0.02**	**2.51**	**.012**	**0.08**	**0.02**	**3.92**	**<.001**
Condition: Multimodal Visual × Familiarity	**-0.07**	**0.02**	**-3.53**	**<.001**	**-0.10**	**0.02**	**-4.93**	**<.001**	**-0.10**	**0.02**	**-4.80**	**<.001**
Condition: Multimodal Action × Familiarity	**-0.10**	**0.02**	**-4.82**	**<.001**	**-0.08**	**0.02**	**-4.11**	**<.001**	**-0.10**	**0.02**	**-5.04**	**<.001**
Group × Condition: Semantic × Familiarity	-0.02	0.04	-0.45	.653	-0.01	0.04	-0.35	.727	0.03	0.04	0.64	.524
Group × Condition: Multimodal Visual × Familiarity	-0.00	0.04	-0.01	.991	-0.01	0.04	-0.15	.877	0.02	0.04	0.41	.682
Group × Condition: Multimodal Action × Familiarity	-0.03	0.04	-0.68	.498	0.01	0.04	0.23	.821	-0.02	0.04	-0.59	.556
Random Effects	**Variance**	**SD**										
$Intercept_{Subject}$	0.05	0.22										
$Intercept_{Item}$	0.00	0.06										
Residuals	0.03	0.18										
$ICC_{SubjectItem}$	0.60											
Model fit	Marginal	Conditional										
R^2	.049	.620										

*Notes: B = model estimates; SE = standard error; t = model statistics; p = p-value; Semantic = Semantic priming; Multimodal Visual/Action = Multimodal Visual/Action priming. The model included only the complete cases. Model formula: Log-transformed RT ~ ISI / (Group * Condition * Familiarity) + (1 | Subject) + (1 | Item). The fixed effects for ISI are omitted. Contrasts of Group: Controls – Schizophrenia. Contrasts of Condition: Semantic = Unrelated – Semantic; Multimodal Visual = Semantic – Visual; Multimodal Action = Semantic – Action. Contrasts of Familiarity: High – Low.*

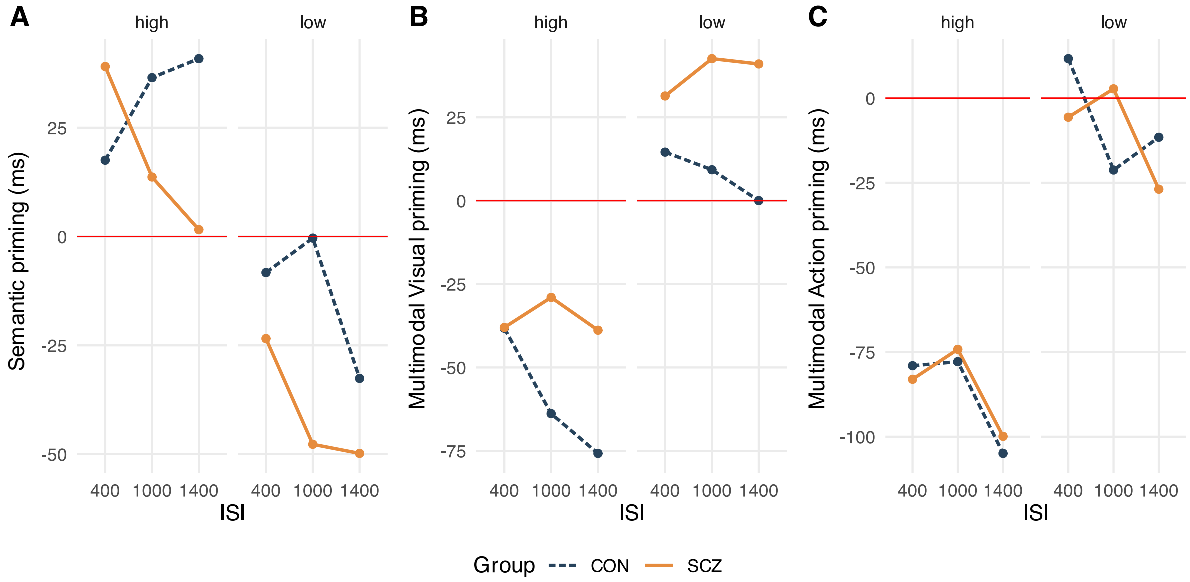


Figure 4.3. Semantic and Multimodal (Visual and Action) priming effects in interaction with Familiarity. *The plots depict the mean Semantic priming effects (in milliseconds) and the Multimodal (Visual and Action) priming effects for schizophrenia (SCZ) and control (CON) groups across Familiarity levels (high vs. low) and inter-stimulus intervals (ISI). Positive values indicate facilitatory effects of the condition relative to the baseline, negative values indicate inhibitory effects. The red line indicates the absence of activation.*

4.3.3. Relation with metaphor comprehension and individual differences

Before exploring the relationship between metaphor comprehension and other relevant variables (i.e., priming effects and individual difference variables), we checked that the different ISI subgroups in the schizophrenia and control samples were not significantly different for relevant variables. The One-way ANOVA tests showed that the participants from the schizophrenia group assigned to the different ISI subgroups did not differ in any individual difference variable (Fs < 0.64, ps > .529), namely demographic (i.e., age and education), clinical (i.e., illness onset and duration, psychopathology severity, and mean chlorpromazine equivalent dose), and cognitive measure (i.e., lexical-semantic, neurocognitive, and imagistic skills), except for the PANSS General score, which was lower for participants assigned to the ISI 1000 subgroup ($F(2,59) = 4.21$, $p = .020$; pairwise post-hoc comparisons: ISI 400 > ISI 1000; ISI 1000 = ISI 1400; ISI 400 = ISI 1400).

Similarly, the participants from the control group did not differ across ISI subgroups for any demographic and cognitive measure (*F*s < 2.29, *p*s $> .109$).

We then tested whether significant priming effects observed in the schizophrenia group, namely Multimodal Visual priming effects at 1000 and 1400 ms time-windows, could be indicative of patients' ability to understand the metaphor used in the priming experiment, as well as other relevant individual difference variables (i.e., psychopathology, neurocognition, mental imagery, lexical-semantic skills, and psychological well-being). To do so, we first inspected patterns of correlations in both the schizophrenia and control groups, focusing in particular on the associations between the output variables of the metaphor priming experiment (i.e., priming effects – averaged by subjects – and the total score of the multiple-choice metaphor comprehension task) and individual difference variables. The correlogram on patients' measures (Figure 4.4) showed that the mean Multimodal Visual priming effects were negatively associated with metaphor comprehension score at 1000 ms (but not at 1400 ms), indicating that higher activations of visual properties of the metaphor vehicle in that time window were indicative of worse metaphor comprehension performance. Additionally, metaphor comprehension was positively correlated with WAIS-R Vocabulary total score in both time windows and with BACS Verbal Memory, Verbal Fluency, and Average Equivalent Score at 1400 ms only, indicating that the ability to understand metaphors was higher in patients with better lexical-semantic and neurocognitive skills.

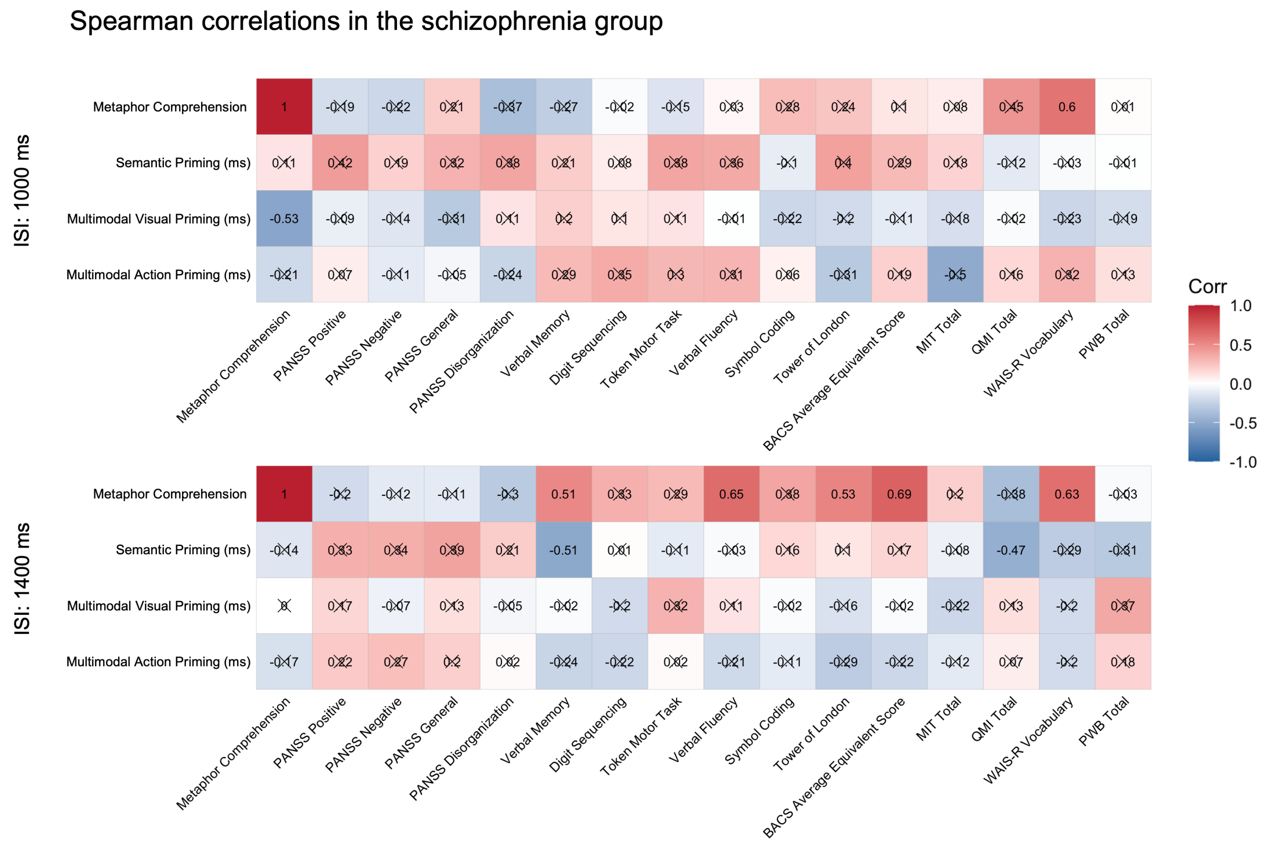


Figure 4.4. Patterns of associations in the schizophrenia group for relevant inter-stimulus intervals (ISI). *The correlogram shows Spearman's correlations between patients' mean priming activation effects (Semantic, Multimodal Visual, and Multimodal Action priming effects) in milliseconds (ms) and behavioral measures, namely participants' metaphor comprehension skills, psychopathological symptom severity, neurocognitive abilities, mental imagery skills, and lexical-semantic skills, as well as psychosocial wellbeing. The correlogram refers to patients' belonging to the ISI 1000 and 1400 ms subgroups. The magnitude of associations is depicted by color (crossed cells indicate non-significant correlations, with significance level p < .05). Note: Metaphor Comprehension = total score of the multiple-choice metaphor comprehension task; PANSS = Positive and Negative Syndrome Scale; BACS = Brief Assessment of Cognition in Schizophrenia; MIT = Mental Imagery Test; QMI = Questionnaire upon Mental Images; PWB = Psychological Well-Being Scales.*

The correlogram on controls' measures (Figure 4.5) showed that metaphor comprehension score was positively associated with the mean semantic priming effects at 1000 ms (but not at 1400 ms) time window, indicating that higher activations of amodal semantic properties of the metaphor vehicles in this group were indicative of better comprehension of metaphors. Metaphor comprehension total score was also positively correlated with lexical semantic skills (WAIS-R Vocabulary total score) at 1000 and 1400 ms, but also with sustained attention (Symbol Coding) at 1400 ms only.

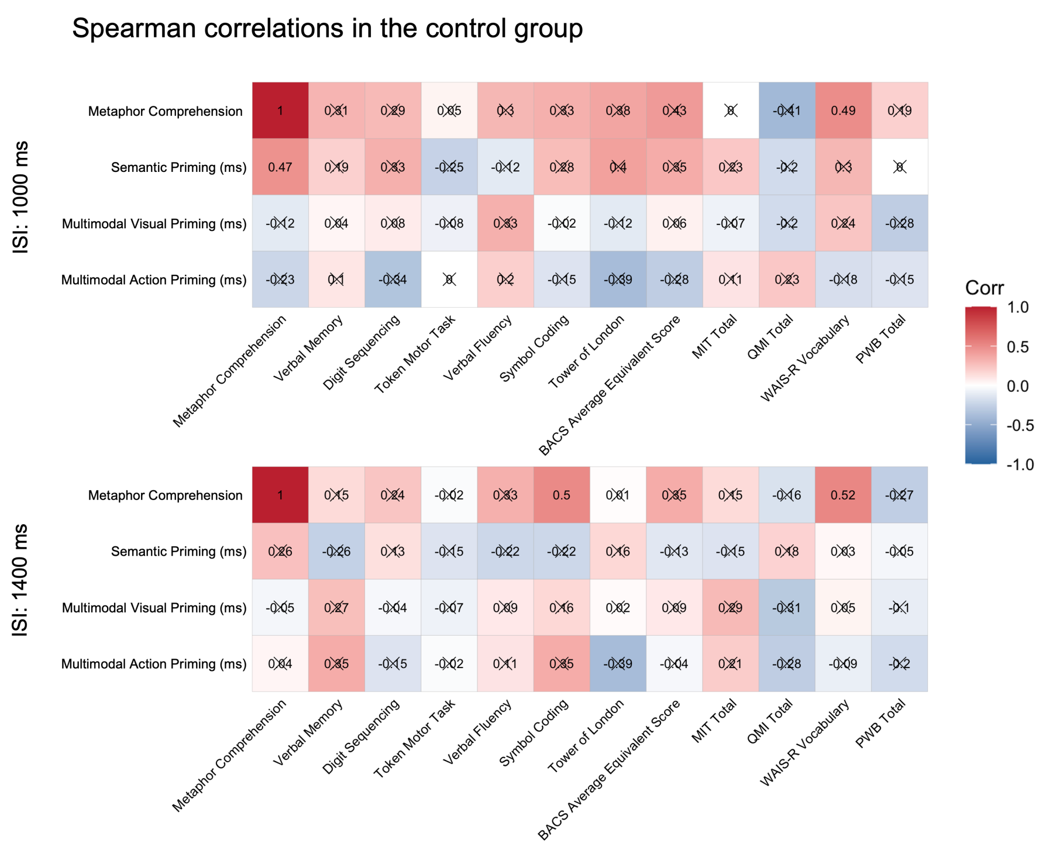


Figure 4.5. Patterns of associations in the control group for relevant inter-stimulus intervals (ISI). *The correlogram shows Spearman's correlations between controls' mean priming activation effects (Semantic, Multimodal Visual, and Multimodal Action priming effects) in milliseconds (ms) and behavioral measures, namely participants' performance in metaphor comprehension, neurocognitive, imagistic, and lexical-semantic tasks, as well as psychosocial wellbeing. The correlogram refers to controls belonging to the ISI 1000 and 1400 ms subgroups. The magnitude of associations is depicted by color (crossed cells indicate non-significant correlations, with significance level* $p < .05$*). Note: Metaphor Comprehension = total score of the multiple-choice metaphor comprehension task; BACS = Brief Assessment of Cognition in Schizophrenia; MIT = Mental Imagery Test; QMI = Questionnaire upon Mental Images; PWB = Psychological Well-Being Scales.*

Finally, we tested which variable – among the mean Multimodal Visual priming effects, lexical-semantic skills, and neurocognitive abilities – better predicted participants' ability to understand metaphors, distinguishing in particular the performance of participants from both groups at 1000 and 1400 ms inter-stimulus intervals. The multiple regression model (see Table 4.7 and Figure 4.6) confirmed that in the schizophrenia group the differential visual activations predicted metaphor comprehension at 1000 ms even when lexical-semantic skills were controlled for, while the neurocognitive skills were the sole predictor of metaphor comprehension at 1400 ms. No variable significantly predicted controls' metaphor comprehension performance at 1000 and 1400 ms time windows.

Table 4.7. Output of the multiple regression on participants' z-transformed total score obtained in the Metaphor Comprehension Task, with the mean Multimodal Visual Priming effects, lexical-semantic skills, and neurocognitive abilities as the independent predictors nested across groups and relevant inter-stimulus-intervals (ISI). Significant effects are highlighted in bold.

Predictors	**Group**	**Inter-Stimulus Interval (ISI)**							
		ISI 1000				**ISI 1400**			
		B	***SE***	***t***	***p***	***B***	***SE***	***t***	***p***
Mean Multimodal Visual Priming	**Patients**	-0.35	0.18	-1.96	.054•	0.32	0.21	1.54	.129
WAIS-R Vocabulary (Total score)		**0.57**	**0.21**	**2.78**	**.007**	0.45	0.27	1.63	.108
BACS Average Equivalent Score		-0.29	0.30	-0.95	.345	**0.64**	**0.23**	**2.74**	**.008**
Mean Multimodal Visual Priming	**Controls**	-0.19	0.28	-0.67	.506	0.02	0.18	0.12	.904
WAIS-R Vocabulary (Total score)		0.15	0.30	0.49	.622	0.21	0.34	0.62	.535
BACS Average Equivalent Score		0.12	0.31	0.38	.709	0.19	0.38	0.51	.613
Model fit	**R^2**	**R^2 adjusted**							
	.481	.358							

Notes: B = model estimates; SE = standard error; t = model statistics; p = p-value.
The model included only the complete cases. Two subjects from the control group were removed from the model as they showed highly deviant values in the dependent variable. Model formula: Metaphor Comprehension (Total) ~ ISI / Group / Mean Multimodal Visual Priming + WAIS-R Vocabulary (Total) + BACS Average Equivalent Score. All continuous variables were z-transformed before entering the model. The main effects for ISI and Group are omitted.

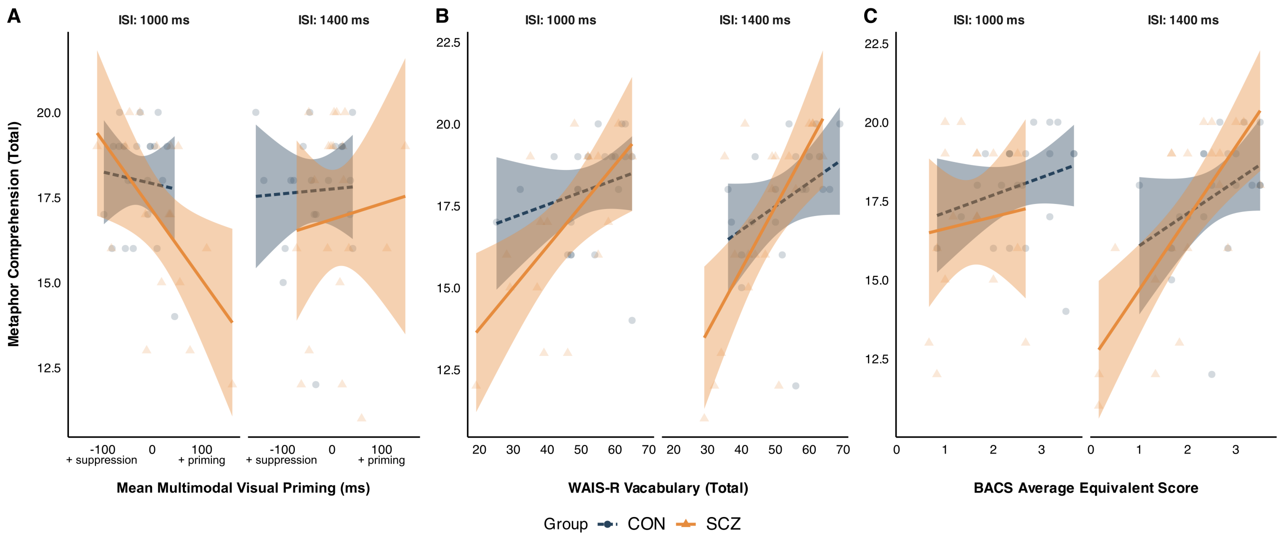


Figure 4.6. Association between metaphor comprehension and mean Multimodal Visual priming effects, lexical semantic skills, and neurocognitive abilities. *The scatterplots depict the association among variables in patients with schizophrenia and control participants, distinguishing between participants belonging to 1000 and 1400 ms inter-stimulus intervals (ISI) subgroups. BACS = Brief Assessment of Cognition in Schizophrenia.*

4.4. Discussion

In this study, we developed a metaphor priming paradigm with a group of individuals with schizophrenia, pursuing the main objective of testing the novel hypothesis that visual images evoked by metaphors might interfere with metaphor comprehension in individuals with schizophrenia. Capitalizing on largely documented evidence that patients tend to linger on concrete aspects of linguistic stimuli due to concretism (Bambini, Frau, et al., *in preparation*, presented in §Study 3; Goldstein, 1959; Harrow, 1972), we developed a metaphor priming paradigm to test whether metaphors would prime target words related to the metaphor vehicle based on multimodal properties, including visual features (e.g., shape) but also motor characteristics (e.g., manipulability). We expected to detect higher multimodal priming effects in schizophrenia, especially for visually-related target words, and to observe a relationship between such activations and patients' ability to understand metaphors.

The results confirmed our hypotheses, by showing that, compared to controls, patients with schizophrenia exhibited a stronger priming effect on visual-related target words 1000 ms after the presentation of the metaphor, lasting up to 1400 ms. Importantly, the visual priming effect observed in patients was also significantly associated with their performance in the offline metaphor comprehension task, indicating that the participants with higher activation of visual properties of the metaphor vehicle were also those with lower metaphor comprehension skills. Conversely, relative to patients, participants in the control group were characterized by stronger semantic priming effects at 1000 ms, which were correlated with control subjects' ability to understand metaphors.

To better elaborate on the multimodal visual priming effects, we observed that patients with schizophrenia, compared to controls, showed higher activations of visual properties of the metaphor vehicle, linked to the literal concrete meaning of the sentence. These findings seem to support the hypothesis proposed here, namely that individuals with schizophrenia, when reading a sentence such as *Wisdom is a flashlight*, activate visual images of the metaphoric vehicle (i.e., a visual representation linked to the concept *flashlight*) and remain attached to them in a way that differs from controls and hinders comprehension. In particular, visual priming effects persisted even after 1400 ms from the offset of the sentence, confirming that visual images tend to linger in individuals with schizophrenia. Our results are consistent with previous pioneering studies employing priming paradigms in individuals with schizophrenia: in particular, Spitzer (1993) showed that metaphors tend to elicit abnormal priming effects on concretely (but not abstractly) related target words in patients and that such effects were detectable after 1200 ms from the prime offset, suggesting that concrete properties of metaphoric stimuli were indeed still active. Our study expands this piece of evidence in two different directions: first, our results showed that priming activations elicited by metaphoric sentences involved not just concrete properties of the stimuli but actually multimodal properties – especially visual perceptual features – linked to the lingering of mental images; second, our results highlighted that such persistence of mental images during metaphor processing had an impact on patients' pragmatic difficulties.

While participants from the schizophrenia group showed higher activations of multimodal visual properties of metaphoric vehicles beyond amodal semantic properties, the control subjects exhibited an opposite pattern, namely their visual priming effects were reduced, indicating that during metaphor processing perceptual-based information related to the metaphoric vehicle were not activated to a greater extent with respect to more amodal semantic features. In a similar study conducted on typical adults, Lam et al. (2015) showed that single words denoting manipulable objects can facilitate the recognition of target words related to the prime by visual properties, although such priming effects were detectable only 1000 ms after the presentation of the priming word. The results from our control subjects diverge from those reported by Lam et al. (2015) and seem to suggest that, when presented in metaphoric sentences, highly concrete words do not necessarily require the activation of perceptual information embedded in their semantic representation. This pattern is consistent with other studies documenting cross-task differences and contextual effects in the activation of sensory-motor information during language comprehension (Frau, Bischetti, et al., *under review*, see §Study 1; Tousignant & Pexman, 2012), showing also that the involvement of experience-based information is higher in single-word or literal sentence processing, while it becomes less relevant in more abstract operations, such as metaphor comprehension (Frau, Losi, et al., *in preparation*, see §Study 2).

Going further with multimodal priming effects, we also tested whether mental imagery activations would involve images including motor experience as well. Interestingly, both schizophrenia and control participants showed no priming effects for action-related target words at all latencies, indicating that motor properties of the concept denoted by the metaphor vehicle were not activated during metaphor processing. This finding – and in particular the results in the control group – needs to be discussed in relation to two other pieces of evidence in literature: first, motor properties of concept denoted by words, such as the manipulability, are known to activate and prime actionally-associated target words in single-word paradigms (Lam et al., 2015); second, similar priming effects were also documented for metaphor priming paradigms (Al-Azary & Katz, 2021), where metaphors (e.g., *That lawyer is a shark*) were shown to prime target words related to metaphoric vehicles based on body-object interaction features (e.g., *bite*). With respect to the findings in single-word paradigms, our results confirm that motor-related features of semantic representation might not become available when words are used in non-literal senses, as in the case of metaphors. Motor activations in metaphor processing seem to be less relevant especially when action-related target words are associated with the metaphoric vehicle only based on the similarity in manipulability with the concept denoted by the metaphor vehicle (e.g., considering again the metaphor *Wisdom is a flashlight*, the concepts denoted by *flashlight* and the action-related target word *remote* share similar motor patterns). Conversely, sensory-motor dimensions might become more relevant in figurative language processing when target words refer not only to bodily actions associated with the metaphor vehicle but are also consistent with the metaphoric meaning (as in the case of *bite* for the metaphor *That lawyer is a shark*, which is associated with *shark* but is also consistent with the figurative interpretation of the sentence, i.e., *being aggressive*; see Al-Azary & Katz, 2021). Based on our data, we can therefore hypothesize that in more abstract linguistic tasks, where metaphors (and not single words) are used as prime stimuli, motor representations might need to be closely related to metaphoric meaning to be more easily activated.

Moving to the results concerning the semantic priming effects, we observed that only the participants from the control group showed faster reaction times in processing semantically related target words compared to unrelated target words (at 1000 ms only), while patients from the schizophrenia group exhibited reduced priming effects for semantically-related target words compared to the unrelated ones. In other words, while control subjects showed greater activation of abstract semantic properties during metaphor reading, this was not the case for patients with schizophrenia, as they processed semantically related target words more slowly than unrelated ones. The results in the control group align with previous studies showing that metaphors activate can prime target words related to the vehicle based on amodal semantic properties, such as meaning associations within the same semantic network (Rubio Fernández, 2007; Weiland et al., 2014). It should be also noted that typical adults in our study showed semantic priming effects only at 1000 ms, while such effects have been reported also in earlier time windows (see Rubio Fernández, 2007). Differences in time resolution of such effects are not surprising if we consider that, compared to other studies that tested metaphor semantic priming in highly educated individuals (e.g., university students), we employed a sample with a lower educational level and semantic priming effects are reported to be slower in individuals with lower literacy levels (Assink et al., 2004). Conversely, patients with schizophrenia showed reduced activation of semantic properties of the metaphoric stimuli at all latencies, displaying a pattern reported also in other studies with figurative expressions used as primes (e.g., in literal plausible idioms, see Titone et al., 2002). Early works on single-word priming paradigms documented increased (rather than reduced) semantic priming effects in individuals with schizophrenia, generally explained as an enhanced automatic spread of activation through semantic memory linked to formal thought disorder (Manschreck et al., 1988; Spitzer et al., 1993). However, previous studies showed that semantic priming effects can be also reduced in schizophrenia, especially at longer latencies, due to impairment in the cognitive skills required to strategically use semantic knowledge (Barch et al., 1996; Kuperberg, 2010). More recently, other works argued that such reduced semantic priming effects in schizophrenia can be explained in terms of "semantic interference": in particular, the semantic similarity between words might trigger the activation of multiple competing alternatives in patients, reflecting higher cognitive costs to select one single item (Almeida & Radanovic, 2021). In other words, a single word seems to prime a higher number of nodes in patients' semantic network and the suppression of competing alternatives for the selection of the desired response results in slower latencies. Such semantic interference effects in patients are also attributed to difficulties in inhibitory control (Lecardeur et al., 2007), which might limit the ability to suppress competing semantically related nodes activated after the presentation of a priming word.

What does not change across groups is the role of familiarity on semantic and multimodal priming effects. In particular, when comparing low against highly familiar metaphors, we observed that semantic priming effects were more pronounced in more conventionalized metaphors, while multimodal priming activations were stronger in more novel expressions. This dissociation confirms previous studies showing that the comprehension of novel and conventionalized metaphors capitalizes on different processes: while more familiar metaphors are more lexicalized and are processed via more direct access to abstract information, less familiar metaphors seem to go through an initial processing phase that encompasses sensory-motor simulations (Al-Azary & Katz, 2021; Cuccio, 2022; Desai et al., 2011; Romero Lauro et al., 2013).

The last point worthy of being discussed concerns the association between multimodal visual priming effects in schizophrenia and other behavioral variables, in particular clinical and psychopathological measures, as well as cognitive, lexical-semantic, and mental imagery skills. Our correlation analysis did not show any significant pattern of association between visual priming effects and other individual difference variables besides metaphor comprehension skills: this suggests that increased activation of perceptual simulations triggered by metaphoric stimuli, which we hypothesize reflects the permanence of visual images due to a concrete mode of thinking, cannot be directly put in relation with other behavioral measures and is not easily captured by other standardized measures, such as those used to assess psychopathology or neurocognition in this population. Higher visual priming effects were not associated with mental imagery variables either, namely vividness ratings and implicit tasks assessing the ability to maintain and inspect mental images. Whether the automated activation of mental images should be dissociated from more controlled and conscious imagistic processes is still debated, yet they are generally thought to reflect partially different neural mechanisms (Muraki, Speed, et al., 2023). What our results seem to suggest with respect to this debate is that greater visual priming activations detected in our experiment do not reflect altered controlled imagery processes or increased imagery vividness, but rather alterations in uncontrolled, automated mental imagery activations that might echo more lower-level sensory-motor simulation processes (Gibbs & Matlock, 2008). It is also worth noticing that several studies questioned the actual sensitivity of self-administered questionnaires rating subjective imagery vividness (for a discussion, see Schwarzkopf, 2023), including their reliability with schizophrenia patients (see Bell & Halligan, 2010; Pearson et al., 2013), which stresses the need for the development of more robust paradigms to explore imagery skills in both typical and atypical populations.

To summarize the results of the metaphor priming experiment, our major finding concerns the evidence of greater activation of mental images in individuals with schizophrenia, manifesting as a sustained greater activation of visuo-perceptual properties of the metaphoric vehicle, which linger to the point of becoming intrusive for patients, making it hard for them to go beyond the literal meaning of figurative expressions. Different underlying cognitive and neural mechanisms might explain the increased visual priming effects in schizophrenia. As argued above, multi-sensory integration is certainly a good candidate: in particular, impairment in integrating signals from different modalities, especially when information is conveyed by language (Williams et al., 2010), has been largely documented in schizophrenia (Gröhn et al., 2022; Tseng et al., 2015), often attributed to attentional deficits hampering the ability to exploit different streams of information (de Jong et al., 2010). Capitalizing on this evidence, we might interpret our results in terms of a more generalized difficulty in combining multimodal information, interfering with the process of integrating visual mental images with more abstract propositional representations. Alternative explanations might deal with the alteration of associative memory, as suggested by pioneering works on concretism (Spitzer, 1993), as associative memory is known to be involved in visual imagery generation, in particular in retrieving the structural representation of objects (Thompson & Kosslyn, 2000). Along this vein, alterations in the process of activating and retrieving information stored in associative memory might lead to a tendency to stick to visual representations during imagery generation. Interestingly, biases towards the activation of visual representations with respect to purely linguistic ones were hypothesized to be linked to hallucinatory and delusional processes, thought to arise from dysfunctioning top-down and bottom-up inferences matching external sensorial stimuli with *a priori* beliefs and expectations (Hugdal, 2009; Stephan-Otto et al., 2017; Tschacher et al., 2017). Finally, the hypothesis that increased and intrusive perceptual experience might be linked to altered inhibitory mechanisms, which might also be responsible for the arousal of visual hallucinatory processes (Silverstein & Lai, 2021), deserves to be mentioned as well. This idea stems from the more general hypothesis that altered cognitive inhibition might be a hallmark of thought and consciousness disorders in this population (Frith, 1979), yet might open relevant keys of interpretation for the behavior described in our study. Future research is then deemed to explore the cognitive and neurobiological bases of increased activations of visual images during metaphor processing in schizophrenia and their relationship with multi-sensory integration, hallucinatory experience, and inhibitory processes.

To conclude, this study has several potentially impactful implications for both clinical and theoretical studies. First, the major implication for clinical applications concerns the characterization of the cognitive substrates of concretism in schizophrenia, in particular its relationship with pragmatic impairment. The cognitive correlates of pragmatic impairment have been extensively studied, with a specific focus on the role of executive functioning and social cognition (Frau, Cerami, et al., 2024). Poorly studied is the role of other cognitive dysfunction in explaining certain aspects of pragmatic deficits in this population, including alterations in the ability to process and integrate sensory-motor information with abstract linguistic processes (Tseng et al., 2015; Williams et al., 2010). Deeper investigations on this topic might indeed induce a beneficial effect on the remediation of cognitive deficits (Bechi et al., 2015; Kluwe-Schiavon et al., 2013), including those directly addressing pragmatic impairment (Bambini, Agostoni, et al., 2022; Bosco et al., 2016), which currently lack proper consideration of the potential role of alterations in multimodal processes linked to mental imagery, as well as sensory-motor and bodily experience.

Finally, the second implication of this study involves the theoretical debate concerning the role of multimodal experience-based processes during metaphor comprehension. Currently, there is no universal agreement among scholars on whether pragmatic operations rely, by any means, on mechanisms beyond the operations required to adjust the propositional representation of metaphoric sentences with the integration of context (see, for instance, Sperber & Wilson, 2008). Among theoretical models that acknowledge some kind of involvement of non-propositional operations during metaphor processing, it is still disputed whether such operations may play any central role in pragmatic inferencing, with positions arguing that sensory-motor simulations might be not always functionally relevant (Carston, 2010b, 2018), whereas others argue the opposite (Gibbs & Matlock, 2008). Far from putting an end to this debate, our results support models that combine both propositional and non-propositional processes (Paivio & Walsh, 1993) and demonstrate that non-propositional effects might hamper the pragmatic inferential machinery in some fragile populations, as in the case of schizophrenia.

To conclude, this study provides novel data supporting a multimodal account of figurative meaning processing and innovatively brings relevant pieces of evidence from individuals with schizophrenia to show the influence of multimodal experience on pragmatic impairment. Most importantly, by showing how non-propositional representations might become relevant to characterize figurative language difficulties in fragile populations, this study opens up new, innovative research frontiers towards a profound understanding of the role of multimodal experience during pragmatic inferencing.

STUDY FIVE

BRIDGING THE GAP BETWEEN PROPOSITIONS AND IMAGES: THE NEURAL DYNAMICS OF MENTAL IMAGERY INVOLVEMENT DURING METAPHOR PROCESSING[6]

Abstract

Several theoretical accounts of metaphor comprehension argue that our ability to understand figurative meaning involves mental images. However, while there is evidence of images related to literal aspects of the metaphorically used words, no studies address whether complex representations of the global metaphorical meaning are formed in the brain. In this study, we explored the neural signature of imagery activation during metaphor processing via EEG. We developed a novel experimental paradigm, where participants: i) read metaphoric sentences or ii) literal description or iii) deliberately evoked a mental image based on an adjective prompt, or iv) were presented with a visual picture. Across conditions, they were asked to indicate whether a subsequent picture was matching or mismatching with respect to the preceding linguistic, imagistic, or perceptual stimulus.

The results showed that metaphor-evoked imagery elicited greater P300-like and LPP responses compared to literal paraphrases, indicating that such images were more compatible with visual stimuli representing a figurative meaning and required a greater integrative effort. Conversely, in response to mismatching stimuli, metaphor-related images showed reduced N200-like responses, indicating less constraining content compared to literal-sentence-evoked ones. Both linguistic-

[6] This chapter is a manuscript in preparation for submission to a peer-review journal as "Frau, F., Canal, P., Bressler, M., Pompei, C., & Bambini, V., *Bridging the gap between propositions and images: the neural dynamics of mental imagery involvement during metaphor processing*". This work received support from the European Research Council under the EU's Horizon Europe programme, ERC Consolidator Grant "PROcessing MEtaphors: Neurochronometry, Acquisition and Decay, PROMENADE" (G.A. number: 101045733). The content of this chapter is the sole responsibility of the authors. The European Commission or its services cannot be held responsible for any use that may be made of the information it contains.

related imagistic processes differed compared to pure imagery ones, which triggered greater later positivity and negativity responses, indexing greater integration effort.

Our findings support a multimodal model of metaphor comprehension by showing that not only propositional but also non-propositional representations are involved in figurative language understanding, with non-propositional operations being also more relevant and perceptually richer compared to literal sentence processing.

5.1. Introduction

Mental imagery refers to the access to perceptual or motor representations stored in long-term memory without direct sensorial stimuli (Kosslyn et al., 2001). The experience of "seeing with the mind's eye" varies in the degree of voluntariness (as when we willingly evoke the mental image of a 'red apple' vs. when we experience flashbacks, see Pearson & Westbrook, 2015) and can be either conscious or unconscious (Nanay, 2021). It is still debated whether the ability to mentally evoke "images" relies on either perceptual or symbolic representations (Sterelny, 1986), as it involves not only perceptual simulations in related sensory-motor brain regions (Palmiero et al., 2009; Pearson et al., 2015) but also more complex cognitive operations, such as episodic memory processes (Byrne et al., 2007; Laeng et al., 2014) or mental time travel (Rasmussen & Berntsen, 2014), allowing flexible combination and manipulation of perceptual and motor experience into novel representations (Hassabis et al., 2007). There is also evidence that mental imagery accompanies many other cognitive activities, including language processing (e.g., reading a sentence activates mental images related to its meaning, see Bergen et al., 2007; Just et al., 2004), yet it is still a matter of theoretical discussion whether mental images are activated also in more abstract uses of language, as in the case of figurative meaning understanding.

The idea that mental images might be central to metaphor understanding was first proposed in the philosophical debate concerning the nature of metaphorical meaning. Within this debate, early scholars first accounted for metaphoric meaning in a purely propositional fashion, seeing it as abstract symbolic content that reflects the output of linguistic operations applied to the literal meaning of the utterance to derive its figurative interpretation (Black, 1955; Grice, 1957). In opposition to purely propositional accounts, other scholars argued in favor of the "non-propositional" nature of metaphoric meaning. In particular, Davidson (1978) claimed that metaphors convey no other meaning beyond the literal interpretation of the utterance, while making the interlocutor see "one thing as another" and notice aspects or similarities between things that were never grasped before (Lepore & Stone, 2010). In-between positions were initially proposed within the field of psychology, capitalizing on the "dual coding theory", according to which conceptual representations consist of both symbolic/verbal format and mental images (Paivio, 1979). Applied to the case of metaphor processing, the "dual-coding theory" predicts that both verbal/symbolic and imagistic processes are required to understand figurative meanings (Paivio & Walsh, 1993): by reading a metaphor such as *Some hairstyles are bushes*, linguistic cues (i.e., *hairstyles* and *bushes*, representing respectively the metaphoric topic and vehicle) serve as conceptual "pegs" to retrieve mental images and verbal information from long-term memory; mental images are then particularly crucial as they allow fast and efficient memory search that supports the conceptual connection between topic and vehicle, required to reach an appropriate figurative interpretation. Other scholars within the Conceptual Metaphor Theory framework also argued that a relevant part of metaphor processing is served by mental imagery, specifically defined in terms of automated perceptual and motor simulation that might occasionally even reach the level of awareness (Gibbs & Matlock, 2008). It is also worthy of being mentioned that recent proposals within post-Gricean theoretical frameworks of pragmatics, specifically Relevance Theory, acknowledged that mental images might be activated during metaphor processing as a "by-product" of propositional operations, even though such images are supposed to accompany propositional processes and might have a role in inferences only in the case of highly demanding and syntactically extended figurative expressions, such as those embedded in literary texts (Bardzokas, 2023; Carston, 2010b, 2018).

Theoretical positions that admit that metaphor processing activates not only symbolic but also imagistic representations are actually supported by a number of experimental works. For instance, early studies showed that imagery might be involved during figurative language understanding via indirect evidence: for instance, Marschark and Hunt (1985) showed that metaphors that can easily evoke a mental image in the speakers' mind are also more easily memorized, while Bucci (1984) reported that individuals with higher imagery skills tend to use more metaphors in spontaneous production. Recent psycholinguistic studies address this topic using more direct paradigms, such as metaphor priming, showing that sensory-motor properties of the metaphoric vehicle, linked to the literal meaning of the expression, are activated during metaphor processing (Al-Azary & Katz, 2021) and might also interfere with figurative interpretations in some individuals, such as people with schizophrenia (Bambini, Frau, et al., *in prep.*; see above §Study 4). Both studies indicate that multimodal features of the literal representation of the metaphor are activated and seem to suggest that especially the mental image of the vehicle might serve as a crucial conceptual peg to constrain information on the topic and guide figurative interpretations (Paivio & Clark, 1986; Paivio & Walsh, 1993).

Besides behavioral evidence, a few studies addressed the potential neural mechanisms involved in mental imagery processes related to metaphor understanding. For instance, Yang et al. (2009) administered a set of literal and metaphoric sentences to a group of participants and asked them whether they could evoke a mental image related to the sentence meaning: the results showed that mental image generation for metaphors was associated with greater activation of visual-imagery-related areas, such as the right fusiform gyrus and the left precuneus, the latter being also relevantly involved in memory retrieval processes (Fletcher et al., 1995). Along this vein, Mashal et al. (2014) also underlined the role of the precuneus in mental image generation during metaphor processing. A more recent study by Bambini et al. (2023) investigated the ERP components of verbal and pictorial metaphors, also addressing the role of individual differences in mental imagery skills: the results showed that for both types of metaphors participants with higher skills in maintaining and manipulating mental images showed a pattern of more frontal distribution of the N400, indicating a greater tendency to process metaphoric stimuli in a picture-like fashion.

Overall, despite the abundance of work on the relationship between mental imagery and metaphors, two relevant points have never been addressed by previous literature: first, behavioral studies documented the possible activation of images of the metaphor vehicle, yet it is still poorly investigated whether metaphor processing entails the activation of more complex mental images, such as a more elaborated representation of the metaphoric meaning (Katz et al., 1988). Second, it remains to be investigated whether the application of neurophysiological techniques, such as EEG recording, can help us to capture the brain signature of the involvement of such complex images.

5.1.1. The present study

In this study, we aimed to investigate the neural correlates of the involvement of mental images during the reading of metaphoric sentences, with a particular focus on complex elaborated images of the figurative meaning.

By capitalizing on the evidence that both deliberate non-linguistic imagery and imagery evoked by the processing of literal sentences influence (in a similar way) the ERP response associated with the perceptual elaboration of subsequent auditory stimuli (Dudschig et al., 2016), we developed a novel experimental paradigm aiming at comparing the effects of pure non-linguistic imagery and imagery evoked by reading of literal and metaphoric sentences on subsequent visual stimuli processing. In this paradigm, we assess the effects on the perceptual elaboration of picture stimuli in a matching–mismatching task preceded by different tasks, which allows us to compare images generated by metaphoric vs. literal stimuli with respect to pure imagery processes, also including a more physical condition as a control baseline. Specifically, the four tasks included in the experimental design were: i) a Physical Task, namely a picture-picture matching task, included to assess the time course as well as the visual perceptual processes underlying the match–mismatch effects of pictures; ii) an Imagery Task, in which participants were asked to evoke a mental image of a person with a specific property denoted by an adjective (e.g., *uncombed*); iii) a Literal Task, where participants were presented with sentences conveying a literal statement (e.g., *Some hairstyles are uncombed*), either matching or mismatching with a subsequent picture; and iv) a Metaphoric Task, where participants were presented with metaphoric sentences (e.g., *Some hairstyles are bushes*).

We based our predictions on previous studies employing picture-picture verification tasks, showing that: i) two subsequent matching pictures elicit a centroparietally distributed P300 component, which reflects that the expectation of a pre-activated perceptual representation is confirmed by a visual stimulus (Friedman et al., 1988; Polich, 2007); ii) two subsequent mismatching pictures elicit a N200 component, indexing a deviation from a pre-activated perceptual representation (Folstein & Van Petten, 2008; Wang et al., 2004). The modulation of the N200 response, in particular, was also observed when the first stimulus was an imagined picture (Wu et al., 2012), indicating that an imagined visual picture can form an internal perceptual template that provides a recent context for a subsequent matching visual target (Tian & Poeppel, 2012).

Therefore, we made the general prediction that perceptual representations evoked by pure non-linguistic and linguistic-related imagery processes would elicit a P300/N200-like pattern during the elaboration of a matching or mismatching visual stimulus, similar to the ERP components associated with picture-picture verification tasks. We would then expect that the amplitude of the ERP components would reflect the extent to which such pre-activated representations either match or mismatch with the content of the subsequent pictures. In particular, if metaphor understanding capitalizes on processes involving the activation of more complex and enriched images compared to literal sentences, we would predict that: i) the P300/N200-like response after the presentation of a matching/mismatching picture to be greater in the Metaphor tasks compared to the ERP response in the Literal Task; ii) the ERP response for metaphors would not differ from the ERP response observed in the Imagery Task. Conversely, if imagery processes activated during metaphor understanding involve images whose content is not more distinctive than the images activated by literal sentences, we would predict that: i) the P300/N200-like response would not vary between the Metaphoric and Literal tasks and ii) both conditions would differ in the terms of ERP response from the Imagery Task.

We also inspected the ERP response in later time windows across conditions, to investigate whether metaphor-evoked mental images would differ from literal-sentence-evoked and pure imagistic processes for what concerns the integration of perceptual and conceptual representations (Dudschig et al., 2016).

Overall, we expect our results to have a significant impact on the theoretical debate opposing fully propositional and fully non-propositional accounts of metaphor meaning, by expanding the empirical evidence around the involvement of mental images during figurative language processing.

5.2. Methods

5.2.1. Participants

A group of 41 Italian-speaking right-handed participants were recruited for the EEG study. Exclusion criteria included being: 1) non-native speaker of Italian, 2) bilingual from birth, 3) diagnosed with a learning disability, and 4) left-handedness, as evaluated using the Edinburgh Handedness Inventory (Oldfield, 1971). One participant was excluded after checking for criteria 1) and 2), while another participant was excluded due to technical issues that occurred during the EEG recording. Thirty-nine participants were included in the final sample (25 females; Age, M = 22, SD = 2.08; Education in years, M = 13.92, SD = 1.88). All participants had normal or corrected to normal vision and reported not to have any neurological or psychiatric disorders nor to be under medication at the time of the experiment. Participants received monetary compensation for carrying out the EEG experiment. Informed consent was obtained from all participants. The study (including the rating tasks carried out to evaluate stimuli properties) was approved by the Ethics Committee of the Department of Brain and Behavioral Sciences of the University of Pavia (protocol number 123/2023).

5.2.2. Assessment

All participants were assessed for mental imagery and lexical-semantic skills (Table 5.1). The assessment of mental imagery included self-administered questionnaires and behavioral implicit tasks. The self-administered questionnaires included: a) the Italian adaptation of the Spontaneous Use of Imagery Scale (SUIS; Nelis et al., 2014), in which the participant is asked to evaluate the tendency to use visual mental imagery in daily life across 12 items using a 5-point scale (score range: 12–60); b) the Italian version of the Vividness Of Visual Imagery Questionnaire (VVIQ; Antonietti & Crespi, 1995; Marks, 1973), in which participants are asked to rate the vividness of visual mental images across 16 items using a 5-point scale (score range: 16–80). The behavioral implicit tasks included four sub-tests of the Mental Imagery Test (MIT; Di Nuovo et al., 2014) evaluating the ability to evoke, maintain, inspect, and manipulate mental images, namely: a) the *Clock Test*, requiring participants to imagine a clock reflected in a mirror and indicating 10 minutes past 10:00, and then to say what time the clock shows and what time it will show after 10 minutes (score range: 0–4); b) the *Cube Test*, in which participants are shown a cube composed of nine small cubes per face (3 × 3), with the external faces colored, and then are asked to state how many small cubes have three external (colored) faces, how many have two, how many one or none (score range: 0–8); c) the *Subtraction of Parts Test*, in which participants are shown the picture of a digital display with the number 88 (composed of small segments) and then are asked to imagine what two-digit number will remain after subtracting the parts of a new figure from the figure with the number 88 (score range: 0–12); d) the *Mental Exploration of a Map Test*, where participants are presented with a printed map and then are asked to answer four questions about the comparative distance between elements represented in the map (score range: 0–6). We then computed a composite score for implicit mental imagery skills by summing the subscores from all MIT tasks (score range: 0–30).

Lexical-semantic skills were assessed using the Italian version of the vocabulary subtest of the Weschler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981). Participants are asked to explain the meaning of 35 words of increasing difficulty, presented on a laptop screen one item at the time. Scores are assigned as follows: wrong explanation (0 points), partial explanation (1 point), correct and complete explanation (2 points). The range of possible scores is 0–70.

Table 5.1. Characteristics of the sample of participants that took part in the EEG experiment.

Sample characteristics	**Mean (SD; range)**
Age	22 (2.08; 19–28)
Education	13.92 (1.88; 13–18)
Gender (F/M)	25/14
EHI	72.70 (20.67; 15–100)
SUIS	44.28 (5.07; 34–54)
VVIQ	61.79 (9.05; 46–80)
MIT Total	22.59 (3.44; 13–28)
WAIS-R Vocabulary	54.26 (6.61; 40–67)

Notes: EHI = Edinburgh Handedness Inventory; SUIS = Spontaneous Use of Imagery Scale; VVIQ = Vividness Of Visual Imagery Questionnaire; MIT = Mental Imagery Test.

5.2.3. Materials

5.2.3.1. Linguistic stimuli

The set of linguistic stimuli used in the EEG experiment included 42 metaphoric sentences and 42 literal sentences to be used in the Literal and Metaphoric tasks, alongside 42 adjectives to be used in the Imagery task. The properties of the linguistic stimuli are reported in Table 5.2.

The set of metaphoric sentences was constructed by selecting items from already normed datasets, used in previous behavioral and EEG experiments (Bambini, Bertini, et al., 2016; Bambini et al., 2013; Canal et al., 2022). All metaphors in the set had the *Spec Xs are Ys*, with topics (*X*s) denoting human beings (plural nouns referring to social roles) or human body parts and vehicles (*Y*s) denoting concrete non-human entities associated with *X*s on the basis of either physical (e.g., *Some hairstyles are bushes*) or mental (e.g., *Some politicians are peacocks*) characteristics. The mean metaphor familiarity was 3.72 (*SD* = 1.26). For each metaphor, we derived a literal equivalent to be included in the set of literal sentences. All literal sentences had the *Spec Xs are Zs* structure, with Zs including adjectives denoting a literal description of the physical (e.g., *Some hairstyles are uncombed*) or mental (e.g., *Some politicians are vain*) characteristic of *X*s used in the metaphoric equivalents. All the adjectives in *Z*s denoted either physical or mental human characteristics and were used to derive the set of adjectives for the Imagery task. We extracted log-transformed lexical frequency values from the CoLFIS Database (Bertinetto et al., 2005) for the adjectives included in the task: within the entire set, the mean lexical frequency was 7.28 (SD = 2.28).

We conducted a rating study to collect relevant measures to characterize the linguistic stimuli, including imageability and physicality values for sentences (both metaphoric and literal ones), as well as imageability measures for the adjectives (presented as isolated words). We recruited 64 young Italian-speaking participants who had not participated in the EEG experiment (41 females; Age, M = 24.13, SD = 2.47; Education in years, M = 15.77, SD = 2.22) and completed the rating tasks as an online experiment on the web-based platform LimeSurvey® (https://limesurvey.org). Forty participants rated literal and metaphoric sentences for imageability and physicality. Metaphoric and literal sentences were randomly divided into two lists, each including a metaphoric sentence but not its literal equivalent (and vice versa). Each participant was randomly assigned to a list and was asked to rate: i) the ease with which each sentence could arouse a mental image (imageability) using a 7-point scale (1 = not easy at all; 7 = extremely easy); ii) whether the meaning of the sentence concerned physical properties of the people the sentence was referring to (physicality) using a 7-point scale (1 = not at all; 7 = completely). Raters showed excellent agreement for both imageability (ICC = .97, 95% CI [.96, .98], $p < .001$) and physicality (ICC = .97, 95% CI [.95, .98], $p < .001$) ratings. Literal sentences were rated as more imageable than metaphoric sentences ($t(82) = 2.51$, $p = .014$), while the two sets were not significantly different for physicality values ($t(82) = 0.67$, $p = .508$). Overall, the two dimensions showed a strong correlation across the entire set ($r(82) = .85$, $p < .001$).

Twenty-four participants rated the set of adjectives for imageability, namely were asked to rate the ease with which each adjective (presented in isolation) could arouse a mental image using a 7-point scale (1 = not easy at all; 1 = extremely easy). Raters showed good agreement for imageability ratings (ICC = .79, 95% CI [.68, .87], p < .001). Adjectives were rated as more imageable than metaphoric (t(82) = 2.75, p = .007) but not literal (t(82) = -0.18, p = .856) sentences.

Table 5.2. Characteristics of the linguistic stimuli used in the EEG experiment.

Measures	**Linguistic stimuli**		
	Literal sentences	**Metaphors**	**Adjectives**
Length (in characters)	29.21 (2.86)	27.57 (3.10)	8.69 (2.15)
Familiarity	-	3.72 (1.26)	7.28 (2.28)[a]
Imageability	4.86 (1.10)	4.28 (1.03)	4.82 (0.78)
Physicality	3.55 (2.04)	3.28 (1.65)	-

Notes: [a] for the set of adjectives, we indicated mean and standard deviation of log-transformed lexical frequency.

5.2.3.2. Picture stimuli

The set of picture stimuli included 84 photographs depicting human referents (e.g., a woman with a messy haircut), to be used as prompt stimuli in the Physical task and as target stimuli in all tasks (Physical, Imagery, Literal, and Metaphoric tasks). All pictures were collected from online databases (e.g., www.unsplash.com, www.freepik.com, etc.) and were compressed to be stored in .PNG format at 96 dots per inch (DPI). Half of the set was presented in the "match" condition and was selected to be a visual representation of the meaning of the linguistic stimuli across all tasks (i.e., a visual representation of metaphoric sentences and literal equivalents, as well as adjectives presented in isolation), while the other half was presented in the "mismatch" condition and was therefore not compatible with the meaning of any of the linguistic stimuli.

To avoid potential confounds caused by differences in the physical properties of the stimuli presented in match and mismatch conditions, we checked that the two sets of pictures were perfectly paired for perceptual characteristics, such as relative luminance (i.e., the brightness of a visual stimulus as perceived by the human eyes), contrast (i.e., the range of brightness, from lightest to darkest, in an image), self-similarity (i.e., how similar a specific part of an image is to the entire image), complexity (i.e., whether an image is dense and contains few redundancies), and symmetry (i.e., whether some parts of a picture mirror other parts in either vertical or horizontal directions). Relative luminance was extracted using *imhist* R package in R (https://rdrr.io/github/mokazuma/imhistR/), while the other measures were all extracted using *imagefluency* R package (Mayer, 2021). The sets of match and mismatch stimuli were paired across all perceptual measures (*t*s < 1.42, *p*s > .159). Visual stimuli properties are reported in Table 5.3.

Table 5.3. Characteristics of the picture stimuli used in the EEG experiment.

Measures [range values]	**Conditions of visual stimuli**		
	Match	**Mismatch**	**Difference**
Relative luminance [0–1]	.50 (.15)	.51 (.15)	$t(82) = -0.24, p = .809$
Contrast [0–1]	.24 (.06)	.23 (.05)	$t(82) = 0.85, p = .397$
Self-similarity [0– -∞]	-0.96 (0.38)	-1.06 (0.25)	$t(82) = 1.42, p = .159$
Complexity [0–1]	.50 (.15)	.48 (.15)	$t(82) = 0.72, p = .472$
Symmetry [0–1]	.30 (.14)	.30 (.19)	$t(82) = -0.02, p = .985$

Notes: relative luminance = higher values indicate higher luminance; contrast = higher values indicate higher contrast; self-similarity = values closer to 0 indicate lower self-similarity; complexity = higher values indicate higher complexity; symmetry = higher values indicate higher symmetry.

5.2.4. Procedure

The experimental procedure was implemented using Presentation® software (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). All stimuli were visually presented on a 24-inch monitor screen (resolution 1920 × 1080, 60 Hz), connected to a Cedrus® RB-540 button box. The linguistic stimuli were displayed in black on a grey background, while picture stimuli were presented on a grey background using a 400 × 333 aspect ratio. Participants sat in a comfortable chair approximately 90 cm from the display in a dimly lit room.

The EEG experiment included four different tasks: i) a Physical Task, in which participants had to indicate whether two subsequent pictures matched or mismatched; ii) an Imagery Task, where participants had to indicate whether a picture matched a previously imagined image, evoked by a previously presented word (an adjective denoting human characteristics); iii) a Literal Task, where participants were asked to indicate whether a picture matched the meaning of a previously presented literal sentence; and iv) a Metaphor Task, in which participants had to indicate whether a picture matched the meaning of a previously presented metaphoric sentence. Each task included 84 trials, with short breaks every ten trials to reduce participants' fatigue.

In each task, trials began with a fixation cross, lasting on the screen for 300 ms, followed by a blank screen for 200 ms. In the Physical Task, a picture was presented for 1000 ms, followed by a short inter-stimulus interval (ISI) lasting 700 ms, and a second picture (either matching or mismatching), presented for 1000 ms. After 2000 ms from the picture offset, a response screen was displayed, indicating a question (i.e., *Is it matched?*) and which key participants had to press on the button box to indicate whether the second picture matched or mismatched with the first one. In the Imagery Task, a single word (an adjective denoting human physical or mental characteristics, e.g., *uncombed*) was presented in isolation for 500 ms, followed by an inter-stimulus interval lasting 3000 ms, during which participants were instructed to evoke a mental image of a person with the characteristic denoted by the previously presented word. Then, a picture (either matching or mismatching) was presented for 1000 ms, followed by a 2000 ms blank screen and the response screen. In the Literal Task, a literal sentence (e.g., *Some hairstyles are uncombed*) was presented word-by-word, with each word presented for 300 ms and separated by a 200 ms blank screen. After a 700 ms inter-stimulus interval, either a matching or mismatching picture was presented for 1000 ms, followed by a 2000 ms blank screen and the response screen. The Metaphoric Task had the same structure as the Literal Task, with the difference being that the linguistic stimuli were all metaphoric sentences. Before starting each task, participants completed eight practice trials and received feedback. The trial structure for all tasks is summarized in Figure 5.1, with an example of matching and mismatching picture stimuli.

The order of tasks varied across participants: half of the participants performed the linguistic tasks (Literal and Metaphoric Tasks) at the beginning of the experiment and the Imagery Task as the last one, while the other half started the experiment with the Imagery Task and performed the linguistic ones at the end of the experiment. The order of the Literal and Metaphoric tasks was also counterbalanced across participants, so four versions of the experiment were created: i) Literal-Metaphoric-Physical-Imagery, ii) Metaphoric-Literal-Physical-Imagery, iii) Imagery-Physical-Literal-Metaphoric, and iv) Imagery-Physical-Metaphoric-Literal. The order of items was pseudorandomized to counterbalance the presentation order of matching and mismatching stimuli for each item, to ensure that a picture stimulus presented first in the matching condition in one task was then presented first in the mismatching condition in the following task (and vice versa). The pseudorandomized order also ensured that matching and mismatching stimuli associated with the same item were separated by at least ¼ of the trials (i.e., 21 trials). Finally, the key assignments for match and mismatch responses varied randomly from trial to trial to prevent response planning before the onset of the response screen.

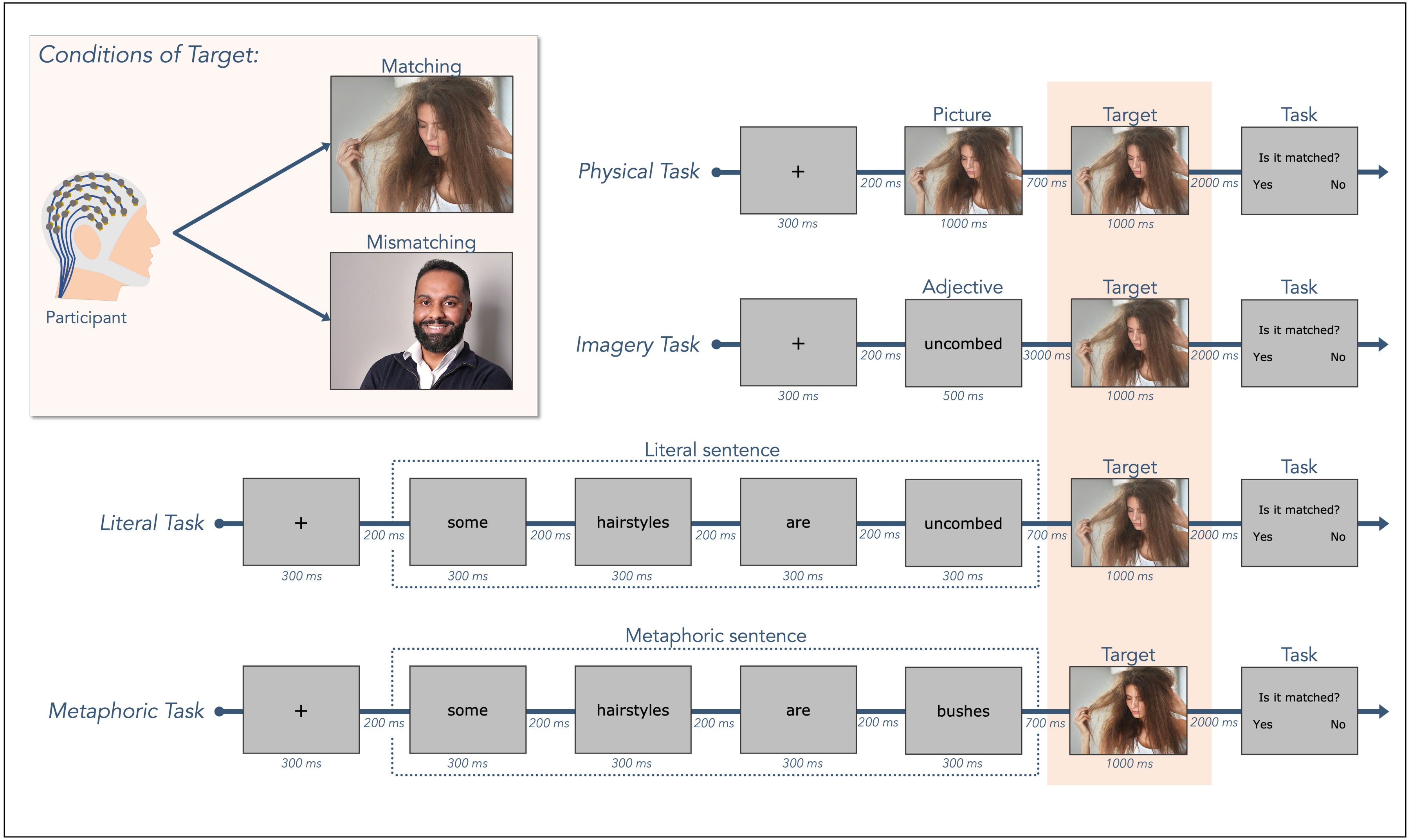


Figure 5.1. Trial structure of the four tasks included in the EEG experiment and example of matching and mismatching picture stimuli.

5.2.5. EEG recording and analysis

The electroencephalogram (EEG) was acquired at a 1000 Hz sampling rate in alternating current with a hardware low cut-off filter (10 s time constant) using a Brainamp® 64 channel system (Brain Products GmbH, Gilching, Germany). Fifty-eight electrodes were placed on the EEG cap according to the 10–20 International System: Fp1, Fpz, Fp2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, CP5, CP3, CP1, Cpz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz. AFz electrode was used as the online reference electrode, two electrodes (TP9 and TP10) were placed on the mastoids, while four electrodes (TP7 and TP8, placed above and below the left eye, and FT9 and FT10 at the lateral canthi) were used to monitor eye movements.

Pre-processing was carried out in MATLAB® (The MathWorks Inc., 2022) with *EEGLAB* (Delorme & Makeig, 2004) and *FieldTrip* (Oostenveld et al., 2011) toolboxes. Offline, the EEG was high-pass (0.10 Hz) and low-pass filtered (45 Hz) and re-referenced to the average activity of the two mastoids. Independent component analysis (ICA) decomposition was used to identify and remove eye-related activity only. The EEG was segmented into epochs around the presentation of the target picture stimulus (from −1870 to 3170 ms). The EEG activity in the selected epochs was averaged to derive ERPs of matching and mismatching trials for all tasks. ERP waveforms were inspected for artifact rejection in a time window from –500 to 2000 ms. Rejection of artifacts was carried out using an amplitude threshold of ±80 μV, keeping all participants who had a rejection rate lower than 40% of total epochs (no participant was excluded). The resulting average rejection rate was 10.60%, with an average of 75.40, 74.20, 75.50, and 75.30 epochs per participant for Physical, Imagery, Literal, and Metaphoric Tasks, respectively.

ERP grand averages were then analyzed with a data-drive approach using Montecarlo cluster-based permutation tests with N = 1000 permutations (see Maris & Oostenveld, 2007), namely a non-parametric statistical test specifically designed to deal with multiple comparison problems resulting from the spatiotemporal structure of the EEG signal (i.e., the signal is sampled at multiple channels and multiple time points). Permutation tests were performed on a matrix of 1201 time points (from 0 to 1200 ms) and 60 channels, with the constraint that at least two adjacent channels had to show a significant effect. For each significant cluster, we reported the value of t (maxsum) and the associated p-value ($\alpha = 0.025$, two-tailed). Permutation tests were applied to test the difference between matching and mismatching waveforms across all tasks (Physical, Imagery, Literal, and Metaphoric). Then, further permutation tests were applied on matching and mismatching waveforms separately, comparing: i) Metaphoric with Literal, ii) Imagery with Metaphoric, and iii) Imagery with Literal tasks.

5.3. Results

5.3.1. Behavioral data

We first inspected behavioral data acquired during the EEG experiment, showing the accuracy rate of responses provided by participants across all tasks. Overall, the average accuracy rate was high ($M = .96$, $SD = .01$), with the Physical and the Literal tasks with the highest accuracy rates (.97 both), followed by Imagery and Metaphoric tasks (.95 both). We inspected task and condition differences in accuracy rates by running a two-way ANOVA test, with Task (Physical, Imagery, Literal, and Metaphoric) and Condition (Match, Mismatch) as within-subject facts (4 × 2). The ANOVA test showed a main effect of Condition ($F(1,38) = 12.30$, $p = .001$), with no significant main effect of Task ($F(3,114) = 2.30$, $p = .081$) nor a significant Task × Condition interaction ($F(3,114) = 0.99$, $p = .402$). Post-hoc comparisons for Condition were carried out using paired-sample t-tests (adjusted for False Discovery Rate, Benjamini & Hochberg, 1995), showing that accuracy rates were higher in the mismatch condition for the Physical and Metaphoric tasks, but not for the Imagery and Literal tasks. Accuracy rates across tasks and conditions, alongside post-hoc t-tests, are reported in Table 5.4.

Table 5.4. Accuracy rates of the EEG experiment across tasks and conditions with the results of paired-sample t*-tests.*

Task	**Accuracy rates across conditions**		
	Match	**Mismatch**	**Difference**
Physical Task	.96 (.05)	.98 (.04)	$t(38) = -3.21, p = .011$
Imagery Task	.95 (.04)	.96 (.05)	$t(38) = -1.03, p = .308$
Literal Task	.97 (.04)	.98 (.03)	$t(38) = -2.07, p = .061$
Metaphoric Task	.94 (.10)	.96 (.10)	$t(38) = -2.71, p = .020$

Notes: p*-values are adjusted for False Discovery Rate (FDR).*

Exploratorily, we also inspected patterns of correlations among the behavioral measures acquired with the EEG and the variables acquired from the assessment, measuring individual differences in mental imagery and lexical-semantic skills. The correlogram (Figure 5.2) showed that accuracy rates in the Physical task were positively associated with accuracy rates in the Metaphoric task, while accuracy rates in the Imagery task were positively associated with accuracy rates in both the Literal and Metaphoric tasks. Interestingly, no behavioral measure from the EEG experiment was related to any variable acquired during the assessment. Among individual difference variables, the total scores of the questionnaires measuring mental imagery vividness (SUIS and VVIQ) were significantly correlated but were not associated with implicit mental imagery tasks (as measured using the MIT). As for the MIT measures, the *Cube Test* was strongly associated with the MIT total score, indicating that participants' performance in this test was highly indicative of their global ability to maintain, inspect, and manipulate mental images. Both measures were also associated with lexical-semantic skills, as measured using the WAIS-R Vocabulary subtest.

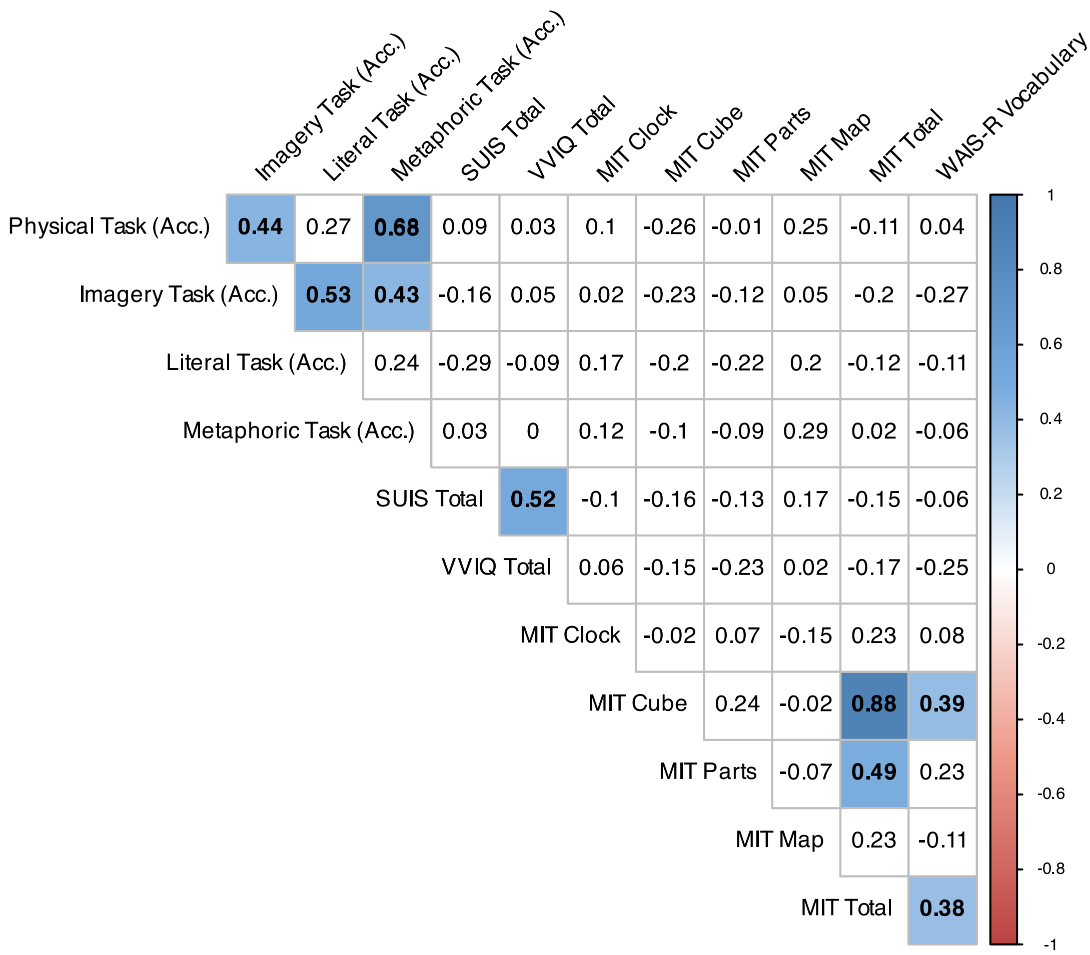


Figure 5.2. Patterns of associations among behavioral variables. *The correlogram shows the correlation between variables acquired during the EEG experiment (accuracy rates for all tasks) and behavioral measures acquired during the assessment, addressing mental imagery and lexical-semantic skills. The magnitude of associations is depicted by color (white cells indicate non-significant correlations, with significance level $p < .05$). SUIS = Spontaneous Use of Imagery Scale; VVIQ = Vividness Of Visual Imagery Questionnaire; MIT = Mental Imagery Test.*

5.3.2. Permutation tests on EEG data

5.3.2.1. Matching vs. mismatching waveforms

The grand means of ERP waveforms in matching and mismatching conditions across all tasks for fronto-central and parieto-occipital sites are reported in Figure 5.3.

We performed cluster-based permutation tests to explore the difference between matching and mismatching waveforms in all tasks (Figure 5.4). Starting from the Physical task, the permutation test revealed one significant negative cluster (t maxsum = -156540, $p < .001$) from 118 to 817 ms, with a prevalent fronto-central topographical distribution, indicating that the mismatching condition was associated with a waveform characterized by a significantly more negative amplitude compared to the matching condition. Similar patterns were also observed in the other tasks: in the Imagery task, the permutation test detected one significant negative cluster (t maxsum = -340290, $p < .001$) from 177 to 1200 ms, with a topographical distribution starting from fronto-central sites and then spreading up to occipital locations after 400 ms. A significant negative cluster was also identified by permutation tests for the Literal (t maxsum = -214570, $p < .001$), from 174 to 1200 ms, and Metaphoric (t maxsum = -274200, $p < .001$) tasks from 173 to 1200 ms; in both tasks, the topographical distribution spread from frontro-central to parietal sites, with the maximum extension in scalp distribution between 300 and 550 ms.

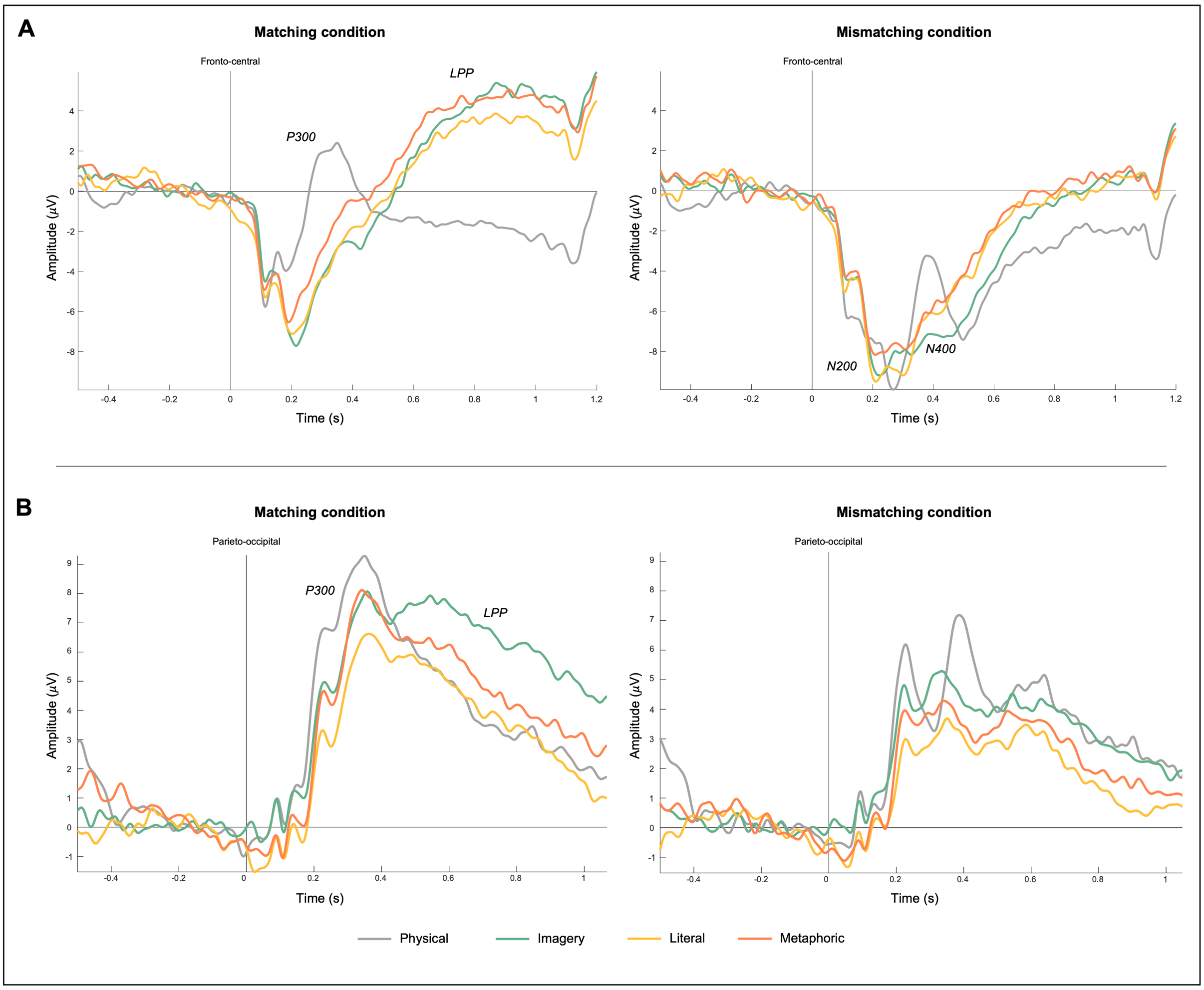


Figure 5.3. Grand means of ERP waveforms across tasks and conditions. *The panel A shows the waveforms for matching and mismatching conditions across tasks in fronto-central electrodes (Fz, FC1, FC2, F1, F2, FCz), showing P300/Late Positive Potentials (LPP) pattern for the matching condition and N200/N400 pattern for the mismatching condition. The panel B shows the waveforms for matching and mismatching conditions in parieto-occipital electrodes (CP1, Pz, CP2, P1, PO3, POz, PO4, P2, CPz), showing P300/LPP pattern in the matching condition.*

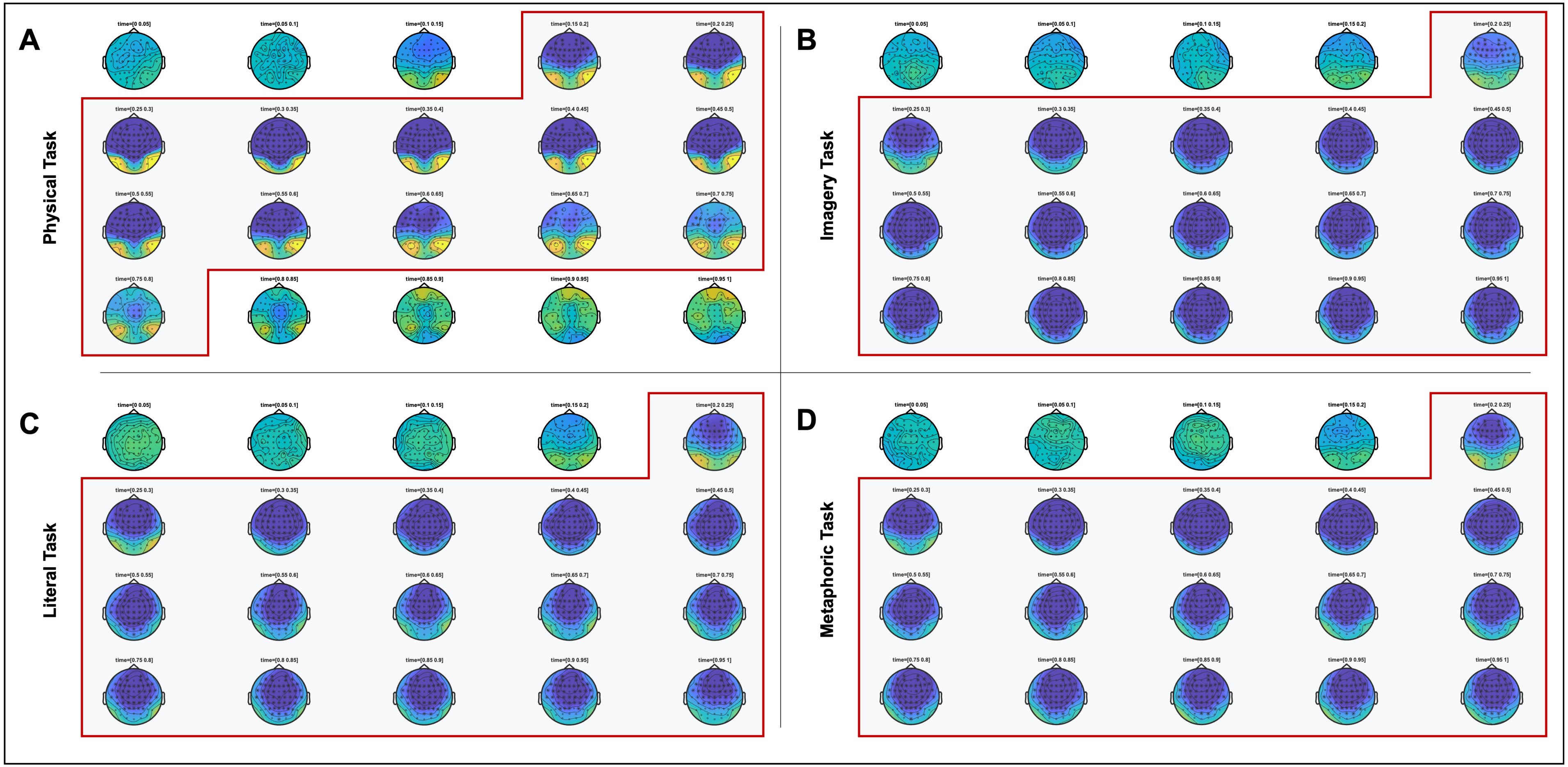


Figure 5.4. The output of cluster permutation tests comparing matching and mismatching waveforms across tasks. *The panels show the significant clusters (delimited by the red line) indicating the difference in amplitude and topographic distribution (across time windows of 50 ms each) between waveforms associated with matching and mismatching conditions for Physical (A), Imagery (B), Literal (C), and Metaphoric (D) tasks. Colormaps tending to the dark blue color indicate clusters with more negative amplitude, namely more negative mismatching waveforms compared to matching waveforms. Electrodes associated with a significant cluster for an entire 50 ms time window are represented with the symbol * (significance level:* $p < .05$*).*

5.3.2.2. Matching effects across tasks

We then tested the difference in matching waveforms among more cognitive tasks, namely the Imagery, Literal, and Metaphoric tasks (Figure 5.5). The permutation tests comparing the Metaphoric with the Literal tasks detected three significant positive clusters. Following a temporal order, Cluster 1 (t maxsum = 21829, p = .015) started from 202 ms to 440 ms, Cluster 2 (t maxsum = 7908, p = .039) was identified from 578 to 718 ms, and Cluster 3 (t maxsum = 21715, p = .015) started at 797 ms up to 1200 ms. All clusters indicated that matching waveforms in the Metaphoric task showed a more positive amplitude compared to the matching waveforms of the Literal task as well. All clusters had a prevalent central and fronto-central topographical distribution.

The permutation tests comparing the Imagery with the Metaphoric tasks detected one significant positive cluster (t maxsum = 3353.50, p = .014), from 462 to 1200 ms, indicating that the matching waveforms in the Imagery task showed a more positive amplitude compared to the matching waveforms of the Metaphoric task as well. This cluster was mostly topographically distributed in parieto-occipital sites.

Finally, the permutation tests comparing the Imagery with the Literal tasks detected one significant positive cluster (t maxsum = 101710, p < .001), starting at 0 ms and lasting up to 1200 ms, indicating that along the entire epoch the matching waveforms in the Imagery task were characterized by a significantly more positive amplitude compared to the matching waveforms of the Literal task. The topographical distribution of this cluster spreads from central sites, moving to centro-parietal sites from 200 ms, and reaching parieto-occipital sites.

5.3.2.3. Mismatching effects across tasks

Finally, we tested the difference in mismatching waveforms among Imagery, Literal, and Metaphoric tasks (Figure 5.6). The permutation tests comparing the Metaphoric with the Literal tasks detected one significant positive cluster (t maxsum = 11253, p = .002), from 183 to 336 ms, with a topographical distribution in central sites. This positive cluster indicated that in the mismatch condition, Metaphoric waveforms showed a more positive amplitude compared to the Literal task. The permutation tests comparing the Imagery with the Metaphoric tasks detected one significant negative cluster (t maxsum = -847.46, p = .047), confined to the 393-604 ms time window and mostly topographically localized in parieto-occipital sites. This cluster indicates that the mismatching waveforms of the Imagery task were characterized by a more negative amplitude compared to the mismatching waveforms of the Mataphoric task.

Finally, the permutation tests comparing the Imagery with the Literal tasks detected two significant positive clusters: Cluster 1 (t maxsum = 18327, p = .015), lasted from 11 to 460 ms, while Cluster 2 (t maxsum = 19168, p = .015) lasted from 613 to 1200 ms. The two clusters indicated that mismatching waveforms in the Imagery task were characterized by a significantly more positive amplitude compared to the mismatching waveforms of the Literal task. The topographical distribution of Cluster 1 involves mostly central and parietal sites, while Cluster 2 involved central and parieto-occipital sites.

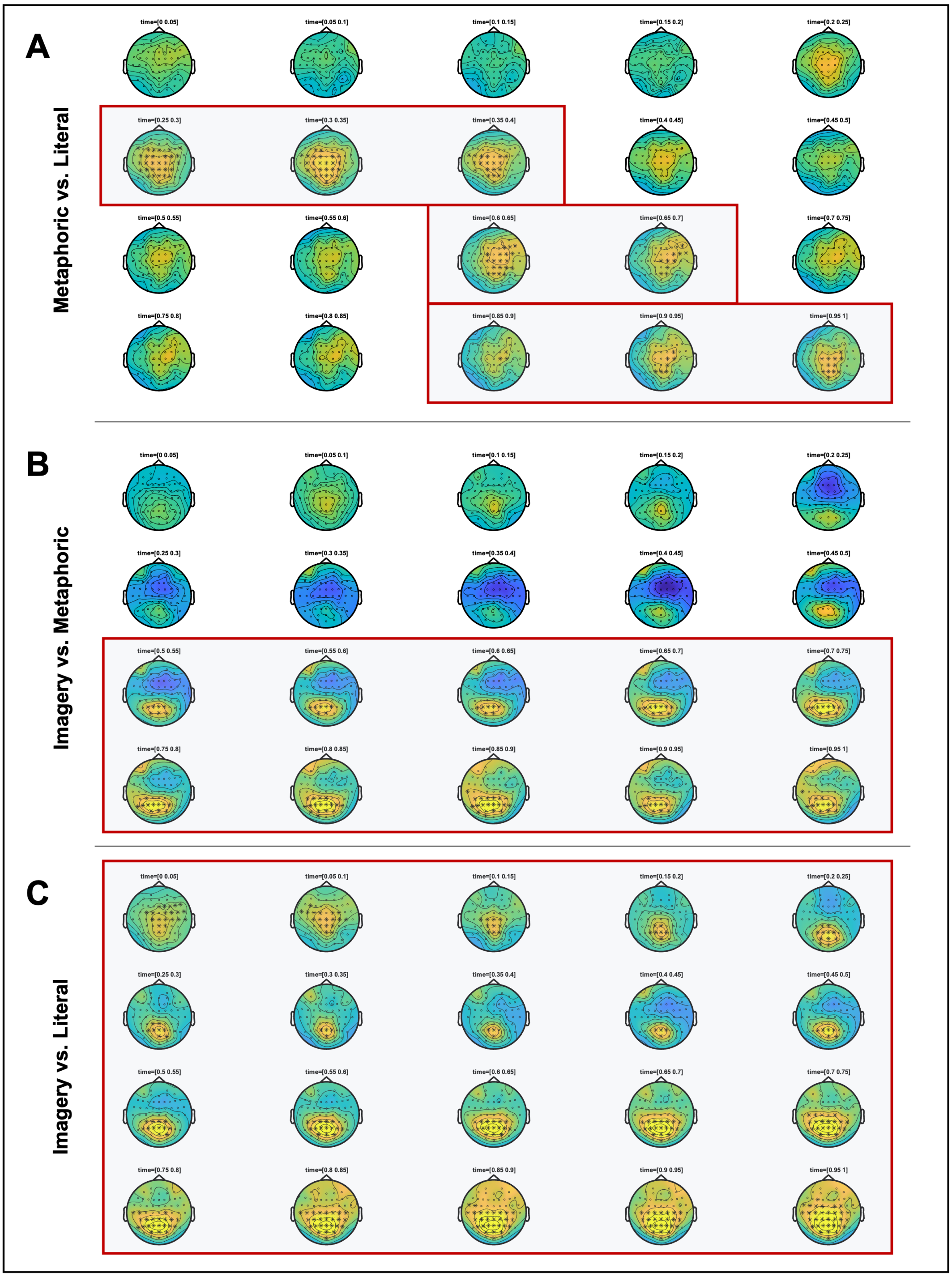


Figure 5.5. The output of cluster permutation tests comparing tasks for the matching condition. *The panels show the significant (delimited by the red line) clusters indicating the difference in amplitude and topographic distribution (across time windows of 50 ms each) among matching waveforms, contrasting in particular Metaphoric with Literal (A), Imagery with Metaphoric (B), and Imagery with Literal (C) tasks. Colormaps tending to the dark blue color indicate clusters with more negative amplitude, while colormaps tending to orange and yellow colors indicate clusters with more positive amplitude. Electrodes associated with a significant cluster for an entire 50 ms time window are represented with the symbol * (significance level:* p < *.05*).

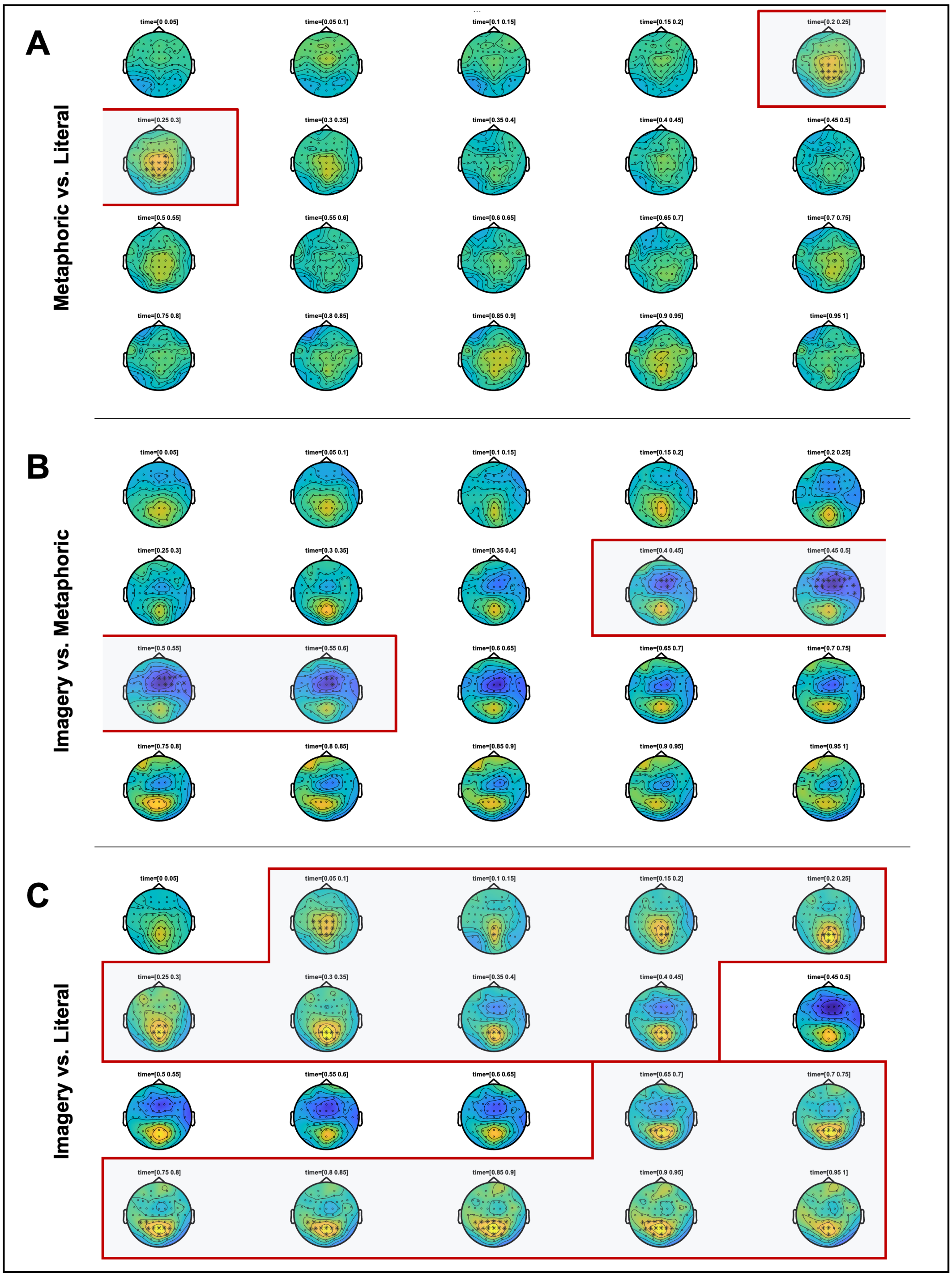


Figure 5.6. The output of cluster permutation tests comparing tasks for the mismatching condition. *The panels show the significant clusters (delimited by the red line) indicating the difference in amplitude and topographic distribution (along time windows lasting 50 ms each) among mismatching waveforms, contrasting in particular Metaphoric with Literal (A), Imagery with Metaphoric (B), and Imagery with Literal (C) tasks. Colormaps tending to the dark blue color indicate clusters with more negative amplitude, while colormaps tending to orange and yellow colors indicate clusters with more positive amplitude. Electrodes associated with a significant cluster for an entire 50 ms time window are represented with the symbol * (significance level:* $p < .05$*).*

5.4. Discussion

In this study, we aimed to track the neurophysiological markers of mental imagery involvement during metaphor processing, focusing in particular on the generation of complex mental images reflecting a more elaborated representation of the figurative meaning conveyed by metaphors.

To do so, we exploited previous evidence that images activated by sentence comprehension influence the ERP components of subsequent perceptual processing (Dudschig et al., 2016) to implement an EEG experiment including four tasks, where we measured how the neural response associated with visual image processing was affected by different imagistic processes, namely non-linguistic deliberate mental imagery (Imagery Task) and imagery generated during the reading of metaphoric sentences (Metaphoric Task) or literal descriptions (Literal Task). We also included a Physical Task, where participants had to judge the compatibility between two subsequently presented pictures, to inspect ERP components related to basic visual perceptual processes.

We used the Physical Task as the benchmark for our expected results: in particular, we predicted a P300/N200 ERP pattern during the visual elaboration of matching and mismatching picture stimuli (respectively), compatible with existing literature on picture-picture verification tasks reporting similar ERP patterns associated with the elaboration of visual stimuli either compatible or deviant from a pre-activated mental representation (Folstein & Van Petten, 2008; Friedman et al., 1988; Polich, 2007). Accordingly, we expected that linguistic-related and non-linguistic imagery processes would elicit P300/N200-like ERP responses during the elaboration of matching/mismatching visual stimuli, indicating the degree to which the pre-activated mental representation differed from the subsequent visual stimulus (Wu et al., 2012). Along this vein, we predicted two scenarios: i) if metaphor understanding – as expected by the imagistic account (Davidson, 1978; Lepore & Stone, 2010) – benefits from richer mental imagery processes compared to literal sentences, we would expect greater P300/N200-like ERP responses in the Metaphor Task compared to the Literal Task; moreover, if mental imagery processes involved in metaphor understanding are more relevantly activated than in literal sentences, we would expect metaphor-evoked P300/N200-like ERP responses not to be significantly different from the neural responses elicited in the Imagery Task; ii) if metaphors, however, do not benefit from imagistic processes to a greater extent compared to literal sentence processing, we would expect the P300/N200-like ERP responses to be similar in the Metaphoric and Literal tasks, with both tasks being different from the Imagery Task.

The visual inspection of ERP grand means across Tasks and conditions confirmed that in the Physical Task the presentation of two subsequent matching or mismatching pictures resulted in clear P300/N200 ERP responses, with centro-parieto-occipital topographic distribution for the P300 and predominant fronto-central distribution of the N200 response. Moreover, the tasks measuring the involvement of mental imagery processes (both pure and linguistic-related) showed a P300/N200-like response followed by later positive (in the matching condition) and negative (in the mismatching condition) activities. To summarize, the results showed that: i) the comparison between Metaphoric and Literal tasks indicated that metaphor-related mental representations, compared to literal-related ones, elicited a greater central positivity during subsequent visual processing in the P300 time window (between 200-400 ms) and in later time windows (600-700 ms and 800-1200 ms), whereas – in the mismatching condition – the Metaphoric Task was associated with less pronounced centrally distributed negativity in the N200 time window compared to the Literal Task; ii) the comparison between the Imagery and Metaphoric tasks revealed a significantly more pronounced later positivity in parieto-occipital sites for the matching conditions in the Imagery Task (starting from 500 ms), while for the mismatching condition the Metaphoric task exhibited less pronounced negativity in fronto-central sites in the 400-600 ms time window; iii) finally, the contrast between Imagery and Literal tasks in the matching condition was characterized by a greater and sustained positivity (especially in parieto-occipital sites) for pure imagery compared to imagery evoked by literal sentence processing, while in the mismatching condition the Imagery Task elicited a less pronounced negative response in early (up to 450 ms) and later (from 600 to 1200 ms) time windows, with predominant distribution in parieto-occipital sites.

Overall, the results in the P300/N200 time window confirmed our prediction i) for what concerns the involvement of mental imagery in the Metaphoric Task relative to the Literal one, indicating that metaphor processing engages non-propositional representations to a greater extent compared to literal sentence processing. However, the results in later time windows partially align with the prediction ii), by showing that imagistic processes involved in metaphor processing are also different from pure mental imagery generation.

To go more in depth in the interpretation of the neural responses, we will start from the most relevant contrast, namely the comparison between the Metaphor and Literal tasks for both matching and mismatching conditions. The results showed a higher P300-like response for metaphors relative to literal sentences during the elaboration of a subsequent, matching visual stimulus. This outcome seems to indicate that by reading a sentence like *Some hairstyles are bushes*, participants activated a visual representation that was perceptually closer to a subsequent matching picture (e.g., the picture of a person with messy hair) compared to the visual representation possibly activated by reading a simple non-metaphoric description of the same scenario (i.e., *Some hairstyles are uncombed*). From the perspective of the research on mental imagery involvement during pragmatic processing, these results indicate that the processes involved during metaphor understanding cannot be entirely abstract and propositional (Black, 1955; Grice, 1975; Sperber & Wilson, 2008), but rather seem to benefit from the activation of experience-based representation to a greater extent compared to literal language processing. This outcome confirms behavioral evidence documenting the activation of mental images during metaphor understanding, which were nevertheless limited to literal aspects of the metaphoric stimulus (Al-Azary & Katz, 2021), and expands such literature by showing that speakers, when reading a metaphor, might also experience the activation of a mental representation that encompasses a more elaborated interpretation of the global figurative meaning. To combine these results within a coherent frame, we can speculate that the content of mental imagery activated during metaphor reading might indeed change and get enriched as the processing goes on: while it might simply start from the content of the metaphoric vehicle, it might also blend with the metaphoric topic and result in highly composite images that represent a figurative interpretation of the sentence (Carston, 2010b). What the modulation of the P300 response seems to capture is this latter stage, where the content of mental images has expanded from the concept denoted by the vehicle to possibly incorporate the topic.

In the matching condition, the Metaphoric and the Literal tasks also showed a difference in the later time window, where ERP response for metaphoric sentences exhibited a greater positive amplitude compared to literal ones. We can capitalize on the literature on the P300 and subsequent Late Positive Potentials (LPP) to interpret this differential later positivity. As argued by Hajcak and Foti (2020), the LPP is an ERP response closely related to the P300 and seems to reflect the extent to which a stimulus significantly activates a motivational system, thus requiring the allocation of mnestic and attentional resources to orient a response (Fields, 2023). We observed in both tasks a sustained positive activity, which might indicate that the presentation of a sentence followed by a picture requires a greater effort to prepare an appropriate response to the matching/mismatching question, while in a more perceptual-based task, such as the Physical one, the attentional and motivational request is less demanding. The LPP was even greater for the Metaphoric Task, which suggests that whenever participants read a metaphor, they not only activated a richer perceptual representation – resulting in the higher positive ERP response in the P300 window – but also required an extra cognitive effort to integrate it with its linguistic, propositional representation, especially when they were asked to use it to decide whether the picture was compatible with the metaphoric meaning. This was not the case for literal sentences, where the integration between perceptual and propositional representations might be less costly as an effect of the general less effortful processing required for literal meaning compared to metaphoric ones (Bambini, Bertini, et al., 2016; Bambini et al., 2013; Noveck et al., 2001).

The Literal and Metaphoric tasks showed a different pattern when it comes to the mismatching condition: while both tasks exhibited a N200-like ERP response, the amplitude was significantly less negative for metaphors compared to literal sentences, suggesting that images aroused by metaphors are perceived as less deviant from the target (i.e., the mismatching picture) compared to images evoked by literal sentences. These results seem to suggest a property of metaphor-induced mental images that is not shared with literal counterparts: on the one hand, the mental images activated during metaphor processing are more detailed and richer in content than the ones generated by literal sentences; on the other hand, mental images of metaphors are more malleable and can be more easily accommodated to be adapted to context (see Paivio & Walsh, 1993), while the ones activated by literal sentences are more rigid. This effect in the neural response associated with metaphor-induced images mirrors the fact that figurative interpretations maintain a degree of open-endedness compared to literal sentences, for which the interpretation is more straightforward and constraining (Ketelaar et al., 2012; Shen et al., 2015). Grounded on our results, this behavior suggests that more creative uses of language come with "costs and benefits" (Bambini et al., 2023): in particular, metaphor understanding seems to be accompanied by more vivid and detailed images, which better represent the state of affair conveyed by the sentence meaning, yet it also leaves the listener with more uncertainty concerning the boundaries of such images.

Interestingly, the negativity associated with the mismatching condition was more sustained in both tasks (up to 700 ms) compared to the Physical Task, especially in frontro-central sites. Previous studies using sentence-picture verification paradigms reported increased negativity during the elaboration of the picture stimulus when the content was mismatching with respect to the sentence meaning: in some cases, such increased negative response was associated with hindered integration between the sentence-evoked mental simulation and the picture content (de Nooijer et al., 2016), whereas in others similar greater negativity was associated with the N400 ERP response indexing the integration of higher conceptual and lower perceptual processes in a more advanced stage of processing (Hirschfeld et al., 2012). We argue that this latter case might better describe the response observed for the Literal and Metaphoric tasks in our study: while in an earlier time window, compatible with the N200 ERP component, the two tasks differed, indicating a different mismatching effect between the sentence-evoked mental image and the visual stimulus, the two tasks showed a similar pattern of negativity in a later time-window, during which the neural response reflected the integration between lexical and conceptual information with an already active perceptual representation (Hirschfeld et al., 2011; see also Dudschig et al., 2016 for similar effects on sound processing).

Moving further, we were also interested in testing the difference between deliberate non-linguistic imagery and metaphor-evoked mental images. The comparison between the two tasks in the matching and mismatching conditions showed that the P300/N200-like response for pure imagery was not significantly different from metaphor-evoked imagery. These results seem to suggest that the perceptual representation activated during a deliberate imagery task and by reading metaphors did not dramatically diverge in what concerns their effect on the processing of matching and mismatching visual stimuli. In other words, the imagistic processes involved in these two tasks largely overlapped with respect to their effect on subsequent perceptual processes.

The Imagery and the Metaphoric tasks, however, differed in later time windows, highlighting that mental images activated in a pure imagery task might exhibit differences in the subsequent integration processes compared to metaphor-evoked ones. For the matching condition, the Imagery Task was associated with a greater sustained LPP compared to the Metaphor Task, extending from central to parieto-occipital sites. We might argue that this higher and sustained positivity not only reflected the motivational response triggered by a significant stimulus, but also a greater cognitive cost required to elaborate an appropriate response for the matching/mismatching task using the pre-activated imagistic representation (Fields, 2023; Hajcak & Foti, 2020). Additionally, the topographical distribution of the LPP, which for the Imagery Task also extended to more occipital sites, suggests that imagery generation was more effortful *per se* in the Imagery Task compared to the linguistic tasks, as more occipitally distributed LPPs have been also related to an increased cognitive effort before the presentation of a visual stimulus (Matsuda & Nittono, 2015). This particular occipital LPP was not documented in other studies employing pure visual imagery generation tasks followed by the presentation of a pictorial stimulus: in these studies, participants had to evoke visual mental images of definite human referents, such as famous actors, using their name as a prompt (e.g., *Brad Pitt*), followed by either matching or mismatching photograph (Schendan & Ganis, 2012; Wu et al., 2012). Conversely, in our pure imagery task, participants were asked to evoke the mental image of an indefinite human referent based on the property denoted by an adjective used as a prompt, which might represent a more challenging task especially for further integration processes involving the generated mental representation. Notably, in the mismatching condition, the Imagery task was associated with a greater N400 response compared to the Metaphor Task. Similar N400 components were documented in sound processing preceded by pure sound imagery elicited using words (e.g., *bark*), albeit without a difference in amplitude with respect to linguistic-related imagery evoked by literal sentences (Dudschig et al., 2016). This supports the idea that pure mental images – particularly those deliberately generated from properties – might require extra effort to be integrated with more abstract conceptual representations compared to images constrained by linguistic cues.

To conclude the discussion of our results, it is also relevant to mention some issues related to our experimental paradigm, the most relevant one being the repetition of the visual stimuli across all tasks. During the experiment, all participants were presented with the same set of matching and mismatching pictures, which were then repeated in all tasks. While the order of presentation of matching and mismatching stimuli, as well as the order of tasks consistently varied among participants, this is a limitation, as it might have influenced not only the ease of mental imagery generation but also the content of mental representation in some tasks. In particular, the repetition effects might have favored mental imagery generation in the Imagery Task and might also have provided more constraints to the content of images, especially when the task was presented as the last one in the experiment. Although the variation of task order was meant to counterbalance familiarity effects due to visual stimuli repetition (Dudschig et al., 2016), additional sensitivity analyses are needed to check the effect of block order on the neural response of visual processing after pure imagery generation (see Kropotov et al., 2019).

Besides methodological limitations, our study has also significant implications. In particular, our findings have a relevant impact on the theoretical debate concerning the nature of metaphoric meaning, opposing on the one hand purely propositional accounts, arguing that metaphoric understanding involves linguistic operations applied to abstract and symbolic content only (Black, 1955; Grice, 1975; Sperber & Wilson, 2008), to non-propositional ones on the other hand, arguing that metaphoric meaning is not conveyed by verbal but rather by imagistic content (Davidson, 1978; Lepore & Stone, 2010). By bringing novel neurophysiological data, our study supports more in-between positions (Carston, 2010b, 2018; Gibbs & Matlock, 2008; Paivio & Walsh, 1993), arguing in favor of a more multimodal representation of metaphoric meaning, where both propositional and non-propositional levels are activated during processing. Within such positions, the results we brought go in the direction of the involvement of more complex and malleable mental images, which are not limited to the literal level of metaphoric vehicle (Al-Azary & Katz, 2021; Gibbs & Matlock, 2008), but rather encompass more blended and dynamic representations. It is also relevant to mention that our findings suggest that imagery evoked by metaphoric linguistic stimuli, albeit in part similar, might be qualitatively different from more deliberate and controlled imagistic processes, activated without linguistic prompts. This outcome has an impact on the characterization of metaphoric-evoked imagistic processes with respect to other instances of mental imagery, as it supports the idea that propositional representations operate in parallel with imagistic ones and might be crucial in both guiding and constraining the content of mental images during metaphor processing (Paivio & Walsh, 1993).

To conclude, our study provides unprecedented empirical evidence concerning the neural dynamics of complex mental imagery involvement during figurative language processing. Some questions, however, are not directly addressed by this work and might inspire future studies. First, the role of individual differences in mental imagery and linguistic skills, as well as stimuli properties (e.g., sentence imageability), was not included in the analysis of ERP responses, yet it might influence the neural response associated with metaphor-evoked images (Bambini et al., 2023). Second, our results cannot directly address another relevant issue, namely whether the images activated during metaphor processing are functionally part of the machinery supporting pragmatic inferential skills (for a discussion, see Casasanto & Gijssels, 2015). Both questions indicate relevant areas to focus on in future research.

CONCLUSION

The main purpose of this dissertation was to provide novel empirical evidence to support an account of language meaning as organized in different dimensions, with a specific focus on higher-level pragmatic uses of language. The motivation behind this work inevitably originated from a widely debated topic in Philosophy, Psychology, and Neuroscience – namely, whether sensorial and motor experience constitute semantic representations or not – and tried to broaden it to the case of metaphoric meaning: to what extent modality-specific dimensions contribute to our ability to understand metaphors? However, as explicitly stated in its premises, the overarching aim of this dissertation was not to endorse either *amodal* or *modal* views as mutually exclusive theoretical proposals, but rather to move forward from these traditional positions towards an integrated, *multimodal* account of language meaning (Dove, 2023). In this perspective, this work capitalized on the long-standing theoretical debate on the organization of semantic memory to open a new horizon in the field of Neuropragmatics: by conjugating behavioral and neurophysiological evidence, this work highlighted that a proper account for multimodal experience in pragmatics is indeed relevant for its empirical and theoretical implications.

Overall, across the five studies presented, this dissertation left a "trail of (empirical) breadcrumbs" on how we process literal and non-literal meanings, describing novel pieces of evidence aligning to current up-to-date multidimensional semantic accounts (Pexman, 2020). I will try to redirect the attention to the main focal points arising from the whole work, as well as their theoretical significance and their novelty.

1. Multimodality is horizontally and vertically flexible

The first obvious aspect emerging from the studies presented above is that access to experience-based information during language processing is inherently flexible, which means that is highly influenced by contextual factors and by changes in the processing demands. **Study 1**, in particular, addressed this flexibility for the semantic representation of Italian verbs, showing that information derived via bodily experience (i.e., sensory-motor, but also interoceptive experience) is more relevantly activated during verb processing when episodic memory operations are also involved (e.g., during memory recognitions), yet it seems to be negligible when more abstract operations become central (e.g., morphological properties during grammatical decisions). This "horizontal" flexibility – as it is observed at the level of single-word processing only – is not new and confirms the hypothesis that there is a trade-off between amodal and multimodal features of semantic representations that is modulated by task characteristics (Connell, 2019; Connell & Lynott, 2013). What is at least partially new is the evidence that such flexibility is also "vertical", as we can observe it also by comparatively testing the role of modality-related information when words are embedded in declarative sentences and when we move from literal to higher-level pragmatic uses. This is one of the main implications of **Study 2**, where the effect of motor grounding disruption on action language processing was tested in individuals with Motor Neuron Diseases, considering not only literal but also figurative uses of action verbs. In particular, this study has the merit of providing novel neuropsychological data supporting the role of motor-related simulations in action verb processing, as well as the somatotopic distribution of such effects. However, it also showed that the link between processing and simulations might become looser when action language is used in a more abstract sense, as in metaphors. In other words, we could assume that when pragmatic inferential skills are at stake, the involvement of motor simulations might become secondary and may not provide any behavioral advantage for the processing of action-related metaphors compared to non-action-related ones.

What Study 2 suggested is that the trade-off between "amodality" and "experience-based modalities" can be generalized to the continuum between literal and figurative meanings. Similar conclusions can be also drawn by considering the behavior of typical adults in **Study 4**: in particular, the results of the metaphor priming experiment showed that perceptual and motor information linked to the literal meaning of metaphors might not become available, while more abstract aspects (e.g., the relationship between words in the same semantic network) do. It is worth mentioning that sensory-motor activations during metaphor processing were not found in a heterogeneous population of typical adults with middle-to-low literacy levels, which might explain differences with other studies involving younger and highly educated adults (Al-Azary & Katz, 2021).

2. Inter-population and inter-individual heterogeneity

The second relevant aspect emphasized in this dissertation is that the effects of experience-based information during language processing show remarkable variability at the population and individual levels.

Starting from cross-population variations, most of the previous neuropsychological literature supporting modal accounts addressed the disruption of sensory-motor grounding of meaning by looking at conditions characterized by motor system dysfunctions (e.g., neurodegenerative motor disorders, see Birba et al., 2017), as a testing ground for the functional role of motor simulations in action language processing. **Study 2** followed that path, although it brought novel insights on the flexibility of motor disruption effects (as already discussed) and contributed to characterizing the relationship between action-language deficits and motor phenotype in the case of a rare genetic condition, such as Hereditary Spastic Paraplegia due to *SPG4* gene mutation.

One of the innovative aspects of this dissertation was the attempt to provide novel evidence of how experience-based modalities might be differently involved in metaphor processing by looking at a population mostly overlooked by the theoretical discussion, namely schizophrenia. In particular, the main implications of **Study 3** and **Study 4** were the evidence that specific semantic dimensions (e.g., information derived via perceptual experience) might intercept psychopathological features in individuals with schizophrenia, such as concretism, and might become central during pragmatic processing, while other dimensions (e.g., motor-related information) might be less relevant. In particular, Study 3 showed how dimensions related to perceptual experience (e.g., word concreteness) might reflect difficulties in pragmatic processing. Elaborating on that, Study 4 also added the evidence that difficulties in integrating perceptual-based information might be coupled with altered visual imagery processes during metaphor understanding. Taken together, the evidence from clinical populations allows the flexibility of multimodal involvement during metaphor processing to be extended to a cross-diagnostic level, showing how some dimensions might not be relevant for some groups of individuals while being intrusively activated for others.

It is also relevant to mention that in several parts of this dissertation, attention was driven to the role of inter-individual differences, which is currently a relevant topic in the literature on multimodal accounts of semantic representation (Ibáñez et al., 2023). In particular, **Study 1** highlighted how differences in vocabulary skills interfered with the involvement of bodily experience during single-word processing in typical adults. Similarly, **Study 2** showed patterns of individual variability within the cohort of participants with Hereditary Spastic Paraplegia, indicating that while some individuals exhibited clear somatotopically distributed difficulties with action language, others were undoubtedly unimpaired. This last point has also a relevant methodological implication, as it suggests that neuropsychological data testing the presence of changes in the experience-based grounding of semantic knowledge should account for brain plasticity and compensatory processes (Aiello et al., 2022).

3. The "mind's eye" on metaphors

A crucial aspect of this dissertation was the centrality given to the relationship between metaphor processing and mental imagery, an almost neglected dimension in Experimental Pragmatics. While multimodal accounts of semantic processing are now starting to consider the role of imagery (Muraki, Speed, et al., 2023), it is precisely from the opposition between symbolic (i.e., propositional) and imagistic (i.e., non-propositional) representations that the philosophical debate on metaphoric meaning started (Black, 1955; Davidson, 1978). **Study 4** and **Study 5**, in particular, capitalized on theoretical intuitions and sparse empirical evidence documenting imagery involvement during metaphor understanding (Al-Azary & Katz, 2021; Gibbs & Matlock, 2008), with the aim of addressing three major questions: i) is mental imagery involved in metaphor processing? ii) what is the content of such images? iii) could it be that mental images might interfere with pragmatic processing in some individuals?

Study 4, in particular, focused on the hypothesis that metaphor processing involves the activation of perceptual-based images related to the literal denotation of lexically encoded concepts (Paivio & Walsh, 1993) and that such images would interfere with pragmatic processing in individuals with schizophrenia. The results confirmed the involvement of visual images, triggered by the literal interpretation of the metaphoric utterances, and showed that imagistic processes can obstacle the understanding of figurative meaning in populations characterized by specific psychopathological features, as in the case of concretism in individuals with schizophrenia. Study 5 better elaborated on these findings and integrated behavioral data using neurophysiological methods (i.e., the recording of EEG activity) to test the involvement of more complex images during metaphor processing, namely the activation of visual representations of the intended figurative meaning conveyed by the metaphor. The neural response, acquired from typical adults with university-level education, confirmed the hypothesis that such representations are indeed generated during the processing of metaphors to a greater extent compared to literal sentences.

Taken together, both studies have strong implications for theoretical models of metaphor processing. First, the evidence of mental imagery involvement during pragmatic processing implies that metaphoric meaning cannot be purely amodal, namely it cannot just involve computations among symbolic representations as predicted by so-called "propositional" accounts (Black, 1955; Grice, 1975; Sperber & Wilson, 2008). Second, the finding showing that mental images are not only involved but might also become an obstacle in populations with fragile pragmatic skills, as in the case of schizophrenia, seems to suggest that such images might not be simply epiphenomenal but might rather be functionally involved in the processing itself (Carston, 2018). Third, the outcome of the EEG study prompts the endorsement of a dynamic and malleable idea of mental imagery during metaphor understanding: along this vein, mental images evoked by reading a metaphoric utterance are not crystallized to the literal meaning of lexically encoded concepts but are more likely to blend with the figurative interpretation.

Overall, the multimodal account of figurative language meaning that is supported by the studies in this dissertation might be summarized in two major properties: i) it is flexible, in the sense that it involves both propositional and non-propositional operations, namely it simultaneously exploits context-driven inferential skills and experience-based representations; ii) it is highly variable at the inter-individual level, which means that the extent to which propositional and non-propositional operations become available varies as a function of individual characteristics. Notably, this account is clearly compatible with markedly "hybrid" theories of metaphor processing, such as the "Dual Coding Theory" (Paivio & Walsh, 1993), but it is not in contradiction with recent proposals put forward in "Relevance Theory", which acknowledge the role of imagistic processes to some extent in figurative language understanding (Carston, 2010b).

4. Future directions

In several points of this dissertation I have referred to Allan Paivio and Mary Walsh's work on metaphor and mental imagery, which starts with these lines: «For the student of language and thought, metaphor is a solar eclipse: it hides the object of study and at the same time reveals some of its most salient and interesting characteristics when viewed through the right telescope» (Paivio & Walsh, 1993: 307). I could not find a less evocative metaphor to describe my three-year-long experimental research on the topic of figurative language processing and, more in general, pragmatic ability. This dissertation represents an attempt to explore the relationship between pragmatic inferencing and multimodal experience "through the right telescope". Hence, a number of questions are still open to receive empirical answers.

Starting from neuropsychological data acquired from neurological conditions, future studies are deemed to develop robust behavioral and neurofunctional models to account for the effects of the disruption of perceptual and motor grounding on pragmatic inferencing, considering recent advances in mapping behavioral changes and brain lesions in neural sites supposedly associated with action meaning processing (see Weiss et al., 2016).

Moving to the relationship between mental imagery and pragmatic processing, further studies are required to empirically investigate the functional role of imagistic processes in the inferential machinery: are the two merely parallel operations or are somehow intertwined in specific processing phases? And what is the specific role of different modalities (e.g., visual or motor) contributing to mental imagery? There are numerous ways to test the causative role of mental images at behavioral and neurophysiological levels, including mental imagery training (Bonnet et al., 2022) as well as the investigation of populations characterized by sensorial deprivation, as in congenitally blind individuals (Cattaneo et al., 2008). Such approaches might indeed help to shed new light on the exploration of the functional contribution of mental imagery to pragmatic processing.

Finally, it is also relevant to address the importance of the role of individual differences in mental imagery to characterize behavioral and neural investigations of its role in pragmatic inferencing. Mental imagery is a highly variable phenomenon at the inter-individual level, with individuals experiencing no imagery at all in daily living (i.e., in the case of *aphantasia*) and others reporting extremely vivid and rich mental images (i.e., *hyperphantasia*). It is then up to future studies encompassing such extreme variability with respect to the role of imagery in pragmatic processing.

I started my dissertation with a question from Carl Wernicke's *Recent works on aphasia* (1885), asking which processes would explain our ability to understand the word "bell". Wernicke replied to his own question by saying: «The acoustic message must stimulate the memory images of a bell which are deposited in the cortex and located according to the sensory organs. These would then include the acoustic imagery aroused by the sound of the bell, visual imagery established by means of form and color, tactile imagery acquired by cutaneous sensation, and finally, motor imagery gained by exploratory movements of the fingers and eyes» (Wernicke, 1885: 103).

This answer elicits many other questions concerning the exact role of these images, whether we use them to better grasp non-literal meanings or whether we just experience them with the "mind's eye" when our brain has already silently completed its computations. Addressing these questions will be, hopefully, part of my future research.

BIBLIOGRAPHY

Aiello, E. N., Grosso, M., Di Liberto, A., Andriulo, A., Buscone, S., Caracciolo, C., Ottobrini, M., & Luzzatti, C. (2022). Disembodying language: Actionality does not account for verb processing deficits in Parkinson's disease. *Journal of Neurolinguistics*, *61*, 101040. https://doi.org/10.1016/j.jneuroling.2021.101040

Aiello, E. N., Pain, D., Gallucci, M., Feroldi, S., Guidotti, L., Mora, G., & Luzzatti, C. (2023). Rethinking motor region role in verb processing: Insights from a neurolinguistic study of noun-verb dissociation in amyotrophic lateral sclerosis. *Journal of Neurolinguistics*, *66*, 101124. https://doi.org/10.1016/j.jneuroling.2023.101124

Al-Azary, H., & Katz, A. N. (2021). Do metaphorical sharks bite? Simulation and abstraction in metaphor processing. *Memory & Cognition*, *49*(3), 557–570. https://doi.org/10.3758/s13421-020-01109-2

Almeida, V. N., & Radanovic, M. (2021). Semantic priming and neurobiology in schizophrenia: A theoretical review. *Neuropsychologia*, *163*, 108058. https://doi.org/10.1016/j.neuropsychologia.2021.108058

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th Ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Anselmetti, S., Poletti, S., Ermoli, E., Bechi, M., Cappa, S., Venneri, A., Smeraldi, E., & Cavallaro, R. (2008a). The Brief Assessment of Cognition in Schizophrenia. Normative data for the Italian population. *Neurological Sciences*, *29*(2), 85–92. https://doi.org/10.1007/s10072-008-0866-9

Anselmetti, S., Poletti, S., Ermoli, E., Bechi, M., Cappa, S., Venneri, A., Smeraldi, E., & Cavallaro, R. (2008b). The Brief Assessment of Cognition in Schizophrenia. Normative data for the Italian population. *Neurological Sciences*, *29*(2), 85–92. https://doi.org/10.1007/s10072-008-0866-9

Antonietti, A., & Crespi, M. (1995). *Analysis of three questionnaires for assessing the vividness of mental image*.

Arcara, G., & Bambini, V. (2016). A Test for the Assessment of Pragmatic Abilities and Cognitive Substrates (APACS): Normative Data and Psychometric Properties. *Frontiers in Psychology*, *7*, 70. https://doi.org/10.3389/fpsyg.2016.00070

Argiris, G., Budai, R., Maieron, M., Ius, T., Skrap, M., & Tomasino, B. (2020). Neurosurgical lesions to sensorimotor cortex do not impair action verb processing. *Scientific Reports*, *10*(1), 523. https://doi.org/10.1038/s41598-019-57361-3

Assink, E. M. H., Bergen, F. Van, Teeseling, H. Van, & Knuijt, P. P. N. A. (2004). Semantic Priming Effects in Normal Versus Poor Readers. *The Journal of Genetic Psychology*, *165*(1), 67–80. https://doi.org/10.3200/GNTP.165.1.67-80

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bak, T. H., & Chandran, S. (2012). What wires together dies together: Verbs, actions and neurodegeneration in motor neuron disease. *Cortex*, *48*(7), 936–944. https://doi.org/10.1016/j.cortex.2011.07.008

Bak, T. H., & Hodges, J. R. (2004). The effects of motor neurone disease on language: Further evidence. *Brain and Language*, *89*(2), 354–361. https://doi.org/10.1016/S0093-934X(03)00357-2

Bak, T. H., Yancopoulou, D., Nestor, P. J., Xuereb, J. H., Spillantini, M. G., Pulvermüller, F., & Hodges, J. R. (2006). Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing. *Brain*, *129*(2), 321–332. https://doi.org/10.1093/brain/awh701

Ballot, C., Mathey, S., & Robert, C. (2021). Word imageability and orthographic neighbourhood effects on memory: a study in free recall and recognition. *Memory*, *29*(6), 829–834. https://doi.org/10.1080/09658211.2021.1921216

Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Lawrence Erlbaum Associates, Inc.

Bambini, V., Agostoni, G., Buonocore, M., Tonini, E., Bechi, M., Ferri, I., Sapienza, J., Martini, F., Cuoco, F., Cocchi, F., Bischetti, L., Cavallaro, R., & Bosia, M. (2022). It is time to address language disorders in schizophrenia: A RCT on the efficacy of a novel training targeting the pragmatics of communication (PragmaCom). *Journal of Communication Disorders*, *97*, 106196. https://doi.org/10.1016/j.jcomdis.2022.106196

Bambini, V., Arcara, G., Bechi, M., Buonocore, M., Cavallaro, R., & Bosia, M. (2016). The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry*, *71*, 106–120. https://doi.org/10.1016/j.comppsych.2016.08.012

Bambini, V., Arcara, G., Bosinelli, F., Buonocore, M., Bechi, M., Cavallaro, R., & Bosia, M. (2020). A leopard cannot change its spots: A novel pragmatic account of concretism in schizophrenia. *Neuropsychologia*, *139*, 107332. https://doi.org/10.1016/j.neuropsychologia.2020.107332

Bambini, V., Arcara, G., Martinelli, I., Bernini, S., Alvisi, E., Moro, A., Cappa, S. F., & Ceroni, M. (2016). Communication and pragmatic breakdowns in amyotrophic lateral sclerosis patients. *Brain and Language*, *153–154*, 1–12. https://doi.org/10.1016/j.bandl.2015.12.002

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, *7*(559). https://doi.org/10.3389/fpsyg.2016.00559

Bambini, V., Bischetti, L., Bonomi, C. G., Arcara, G., Lecce, S., & Ceroni, M. (2020). Beyond the motor account of amyotrophic lateral sclerosis: Verbal humour and its relationship with the cognitive and pragmatic profile. *International Journal of Language & Communication Disorders*, *55*(5), 751–764. https://doi.org/10.1111/1460-6984.12561

Bambini, V., & Canal, P. (2021). Neurolinguistic Research on the Romance Languages. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.443

Bambini, V., & Ceroni, M. (2021). Amyotrophic Lateral Sclerosis. In L. Cummings (Ed.), *Handbook of Pragmatic Language Disorders* (pp. 435–459). Springer International Publishing. https://doi.org/10.1007/978-3-030-74985-9_17

Bambini, V., & Frau, F. (2022). Neuropragmatica. In F. Domaneschi & V. Bambini (Eds.), *Pragmatica sperimentale* (pp. 259–269). il Mulino.

Bambini, V., Frau, F., Bischetti, L., Cuoco, F., Bechi, M., Buonocore, M., Agostoni, G., Ferri, I., Sapienza, J., Martini, F., Spangaro, M., Bigai, G., Cocchi, F., Cavallaro, R., & Bosia, M. (2022). Deconstructing heterogeneity in schizophrenia through language: a semi-automated linguistic analysis and data-driven clustering approach. *Schizophrenia*, *8*, 102. https://doi.org/https://doi.org/10.1038/s41537-022-00306-z

Bambini, V., Ghio, M., Moro, A., & Schumacher, P. B. (2013). Differentiating among pragmatic uses of words through timed sensicality judgments. *Frontiers in Psychology*, *4*, 938. https://doi.org/10.3389/fpsyg.2013.00938

Bambini, V., Ranieri, G., Bischetti, L., Scalingi, B., Bertini, C., Ricci, I., Schaeken, W., & Canal, P. (2023). The costs of multimodal metaphors: comparing ERPs to figurative expressions in verbal and verbo-pictorial formats. *Discourse Processes*, 1–25. https://doi.org/10.1080/0163853X.2023.2282895

Bara, B. G. (2010). *Cognitive Pragmatics. The Mental Processes of Communication*. MIT Press.

Barattieri di San Pietro, C., Luzzatti, C., Ferrari, E., de Girolamo, G., & Marelli, M. (2023). Automated clustering and switching algorithms applied to semantic verbal fluency data in schizophrenia spectrum disorders. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2023.2178662

Barch, D. M., Cohen, J. D., Servan-Schreiber, D., Steingard, S., Steinhauer, S. S., & van Kammen, D. P. (1996). Semantic priming in schizophrenia: An examination of spreading activation using word pronunciation and multiple SOAs. *Journal of Abnormal Psychology*, *105*(4), 592–601. https://doi.org/10.1037/0021-843X.105.4.592

Bardzokas, V. (2023). Creative metaphors and non-propositional effects. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*. https://doi.org/10.1075/prag.22034.bar

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., & Mazzoleni, M. (2004). Introducing the la Repubblica Corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*, *January 2004*, 1771–1774.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226. https://doi.org/10.1007/s10579-009-9081-4

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/S1364-6613(02)00029-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumard, J., Lesourd, M., Remigereau, C., Jarry, C., Etcharry-Bouyx, F., Chauviré, V., Osiurak, F., & Le Gall, D. (2018). Tool use in neurodegenerative diseases: Planning or technical reasoning? *Journal of Neuropsychology*, *12*(3), 409–426. https://doi.org/10.1111/jnp.12121

Bäumer, D., Talbot, K., & Turner, M. R. (2014). Advances in motor neurone disease. *Journal of the Royal Society of Medicine*, *107*(1), 14–21. https://doi.org/10.1177/0141076813511451

Bechi, M., Bosia, M., Buonocore, M., Agostoni, G., Bosinelli, F., Silvestri, M. P., Bianchi, L., Cocchi, F., Guglielmino, C., Spangaro, M., & Cavallaro, R. (2020). Stability and generalization of combined theory of mind and cognitive remediation interventions in schizophrenia: Follow-up results. *Psychiatric Rehabilitation Journal*, *43*(2), 140–148. https://doi.org/10.1037/prj0000379

Bechi, M., Bosia, M., Spangaro, M., Buonocore, M., Cocchi, F., Pigoni, A., Piantanida, M., Guglielmino, C., Bianchi, L., Smeraldi, E., & Cavallaro, R. (2015). Combined social cognitive and neurocognitive rehabilitation strategies in schizophrenia: neuropsychological and psychopathological influences on Theory of Mind improvement. *Psychological Medicine*, *45*(15), 3147–3157. https://doi.org/10.1017/S0033291715001129

Bell, V., & Halligan, P. W. (2010). Additional data on whether vividness of visual mental imagery is linked to schizotypal traits in a non-clinical population. *Psychiatry Research*, *178*(3), 568–569. https://doi.org/10.1016/j.psychres.2009.05.003

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berg, A. O., Melle, I., Zuber, V., Simonsen, C., Nerhus, M., Ueland, T., Andreassen, O. A., Sundet, K., & Vaskinn, A. (2017). Modelling difficulties in abstract thinking in psychosis: the importance of socio-developmental background. *Cognitive Neuropsychiatry*, *22*(1), 39–52. https://doi.org/10.1080/13546805.2016.1259998

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and Linguistic Aspects of Visual Imagery in Sentence Comprehension. *Cognitive Science*, *31*(5), 733–764. https://doi.org/10.1080/03640210701530748

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. M. (2005a). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*. http://linguistica.sns.it/CoLFIS/Home.htm

Bertinetto, P. M., Finocchiaro, C., & Rastelli, C. (2020). Looking for the Kennformen of the Italian verb paradigm. An experimental study. *Acta Linguistica Academica*, *67*(2), 219–246. https://doi.org/10.1556/2062.2020.00015

Binz, B., & Brüne, M. (2010). Pragmatic language abilities, mentalising skills and executive functioning in schizophrenia spectrum disorders. *Clinical Neuropsychiatry*, *7*(3), 91–99.

Birba, A., García-Cordero, I., Kozono, G., Legaz, A., Ibáñez, A., Sedeño, L., & García, A. M. (2017). Losing ground: Frontostriatal atrophy disrupts language embodiment in Parkinson's and Huntington's disease. *Neuroscience & Biobehavioral Reviews*, *80*, 673–687. https://doi.org/10.1016/j.neubiorev.2017.07.011

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Bischetti, L., Frau, F., & Bambini, V. (2024). Neuropragmatics. In M. J. Ball, N. Müller, & E. Spencer (Eds.), *The Handbook of Clinical Linguistics, Second Edition* (pp. 41–54). Wiley. https://doi.org/10.1002/9781119875949.ch4

Bischetti, L., Pompei, C., Scalingi, B., Frau, F., Bosia, M., Arcara, G., & Bambini, V. (2023). Assessment of pragmatic abilities and cognitive substrates (APACS) brief remote: a novel tool for the rapid and tele-evaluation of pragmatic skills in Italian. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-023-09667-y

Black, M. (1955). Metaphor. *Proceedings of the Aristotelian Society*, *55*(1), 273–294. https://doi.org/10.1093/aristotelian/55.1.273

Bleuler, E. (1911). *Dementia Praecox oder Gruppe der Schizophrenien*. Deuticke.

Bocanegra, Y., García, A. M., Lopera, F., Pineda, D., Baena, A., Ospina, P., Alzate, D., Buriticá, O., Moreno, L., Ibáñez, A., & Cuetos, F. (2017). Unspeakable motion: Selective action-verb impairments in Parkinson's disease patients without mild cognitive impairment. *Brain and Language*, *168*, 37–46. https://doi.org/10.1016/j.bandl.2017.01.005

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer (Version 6.1.36)* (Version 6.1.36). Institute of Phonetic Sciences of the University of Amsterdam.

Bonnet, C., Bayram, M., El Bouzaïdi Tiali, S., Lebon, F., Harquel, S., Palluel-Germain, R., & Perrone-Bertolotti, M. (2022). Kinesthetic motor-imagery training improves performance on lexical-semantic access. *PLOS ONE*, *17*(6), e0270352. https://doi.org/10.1371/journal.pone.0270352

Bora, E., Tang, E., Kuperberg, G. R., Sommer, I., Palaniyappan, L., Alonso, M. F., Mota, N. B., & Hinzen, W. (2021). *Diverse International Scientific Consortium for Research in Thought, Language and Communication (DISCOURSE) in Psychosis*. https://discourseinpsychosis.org/

Borghi, A. M., & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. https://books.google.com/books?id=BTq3BAAAQBAJ&pgis=1

Borghi, A. M., & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, *48*(3), 763–773. https://doi.org/10.1016/j.neuropsychologia.2009.10.029

Bosco, F. M., Gabbatore, I., Gastaldo, L., & Sacco, K. (2016). Communicative-Pragmatic Treatment in schizophrenia: A pilot study. *Frontiers in Psychology*, *7*, 166. https://doi.org/10.3389/fpsyg.2016.00166

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Bradley, W. G., Andrew, A. S., Traynor, B. J., Chiò, A., Butt, T. H., & Stommel, E. W. (2018). Gene-Environment-Time Interactions in Neurodegenerative Diseases: Hypotheses and Research Approaches. *Annals of Neurosciences*, *25*(4), 261–267. https://doi.org/10.1159/000495321

Brooks, B. R., Miller, R. G., Swash, M., & Munsat, T. L. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, *1*(5), 293–299. https://doi.org/10.1080/146608200300079536

Brüne, M. (2003). Theory of mind and the role of IQ in chronic disorganized schizophrenia. *Schizophrenia Research*, *60*(1), 57–64. https://doi.org/10.1016/S0920-9964(02)00162-7

Brüne, M., & Bodenstein, L. (2005). Proverb comprehension reconsidered - "Theory of mind" and the pragmatic use of language in schizophrenia. *Schizophrenia Research*, *75*(2–3), 233–239. https://doi.org/10.1016/j.schres.2004.11.006

Bucci, W. (1984). Linking words and things: Basic processes and individual variation. *Cognition*, *17*(2), 137–153. https://doi.org/10.1016/0010-0277(84)90016-7

Buck, B., Minor, K. S., & Lysaker, P. H. (2015). Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. *Comprehensive Psychiatry*, *58*, 138–145. https://doi.org/10.1016/j.comppsych.2014.12.015

Buonocore, M., Spangaro, M., Bechi, M., Baraldi, M. A., Cocchi, F., Guglielmino, C., Bianchi, L., Mastromatteo, A., Bosia, M., & Cavallaro, R. (2018). Integrated cognitive remediation and standard rehabilitation therapy in patients of schizophrenia: persistence after 5 years. *Schizophrenia Research*, *192*, 335–339. https://doi.org/10.1016/j.schres.2017.05.022

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. https://doi.org/10.1037/0033-295X.114.2.340

Cacciari, C., Bolognini, N., Senna, I., Pellicciari, M. C., Miniussi, C., & Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: A

TMS study. *Brain and Language*, *119*(3), 149–157. https://doi.org/10.1016/j.bandl.2011.05.004

Canal, P., Bischetti, L., Bertini, C., Ricci, I., Lecce, S., & Bambini, V. (2022). N400 differences between physical and mental metaphors: The role of Theories of Mind. *Brain and Cognition*, *161*, 105879. https://doi.org/10.1016/j.bandc.2022.105879

Carston, R. (2010a). Lexical pragmatics, ad hoc concepts and metaphor: A Relevance Theory perspective. *Italian Journal of Linguistics*, *22*(1), 153–180.

Carston, R. (2010b). Metaphor: Ad hoc concepts, literal meaning and mental images. *Proceedings of the Aristotelean Society*, *110*(3), 295–321. https://doi.org/10.1111/j.1467-9264.2010.00288.x

Carston, R. (2018). Figurative Language, Mental Imagery, and Pragmatics. *Metaphor and Symbol*, *33*(3), 198–217. https://doi.org/10.1080/10926488.2018.1481257

Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistics Vanguard*, *1*(1), 327–337. https://doi.org/10.1515/lingvan-2014-1015

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1346–1360. https://doi.org/10.1016/j.neubiorev.2008.05.002

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., & Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, *169*(1–2), 13–21. https://doi.org/10.1016/S0022-510X(99)00210-5

Chamard, L., Ferreira, S., Pijoff, A., Silvestre, M., Berger, E., & Magnin, E. (2016). Cognitive Impairment Involving Social Cognition in SPG4 Hereditary Spastic Paraplegia. *Behavioural Neurology*, *2016*, 1–6. https://doi.org/10.1155/2016/6423461

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Colle, L., Angeleri, R., Vallana, M., Sacco, K., Bara, B. G., & Bosco, F. M. (2013). Understanding the communicative impairments in schizophrenia: A preliminary study. *Journal of Communication Disorders*, *46*(3), 294–308. https://doi.org/10.1016/j.jcomdis.2013.01.003

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–550. https://doi.org/10.3758/s13423-012-0368-x

Corcoran, C. M., Mittal, V. A., Bearden, C. E., E. Gur, R., Hitczenko, K., Bilgrami, Z., Savic, A., Cecchi, G. A., & Wolff, P. (2020). Language as a biomarker for psychosis: A natural language processing approach. *Schizophrenia Research*, *226*, 158–166. https://doi.org/10.1016/j.schres.2020.04.032

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cousins, K. A. Q., Ash, S., & Grossman, M. (2018). Production of verbs related to body movement in amyotrophic lateral sclerosis (ALS) and Parkinson's Disease (PD). *Cortex*, *100*, 127–139. https://doi.org/10.1016/j.cortex.2017.08.030

Covington, M. A., He, C., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S., Semple, J., & Brown, J. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research*, *77*(1), 85–98. https://doi.org/10.1016/j.schres.2005.01.016

Crawford, J. R., Garthwaite, P. H., & Ryan, K. (2011). Comparing a single case to a control sample: Testing for neuropsychological deficits and dissociations in the presence of covariates. *Cortex*, *47*(10), 1166–1178. https://doi.org/10.1016/j.cortex.2011.02.017

Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist*, *12*(4), 482–486. https://doi.org/10.1076/clin.12.4.482.7241

Cuccio, V. (2022). The figurative brain. In A. M. García & A. Ibáñez (Eds.), *The Routledge Handbook of Semiosis and the Brain*. Routledge.

Damasio, A. R. (1989). Concepts in the Brain. *Mind & Language*, *4*(1–2), 24–28. https://doi.org/10.1111/j.1468-0017.1989.tb00236.x

Davidson, D. (1978). What Metaphors Mean. *Critical Inquiry*, *5*(1), 31–47. https://doi.org/10.1086/447971

de Boer, J. N., van Hoogdalem, M., Mandl, R. C. W., Brummelman, J., Voppel, A. E., Begemann, M. J. H., van Dellen, E., Wijnen, F. N. K., & Sommer, I. E. C. (2020). Language in schizophrenia: relation with diagnosis, symptomatology and white matter tracts. *Npj Schizophrenia*, *6*(1), 1–10. https://doi.org/10.1038/s41537-020-0099-3

de Boer, J. N., Voppel, A. E., Brederoo, S. G., Wijnen, F. N. K., & Sommer, I. E. C. (2020). Language disturbances in schizophrenia: the relation with antipsychotic medication. *Npj Schizophrenia*, *6*(1), 24. https://doi.org/10.1038/s41537-020-00114-3

De Brigard, F., Umanath, S., & Irish, M. (2022). Rethinking the distinction between episodic and semantic memory: Insights from the past, present, and future. *Memory & Cognition*, *50*(3), 459–463. https://doi.org/10.3758/s13421-022-01299-x de Carvalho, M., Dengler, R., Eisen, A., England, J. D., Kaji, R., Kimura, J., Mills, K., Mitsumoto, H., Nodera, H., Shefner, J., & Swash, M. (2008). Electrodiagnostic criteria for diagnosis of ALS. *Clinical Neurophysiology*, *119*(3), 497–503. https://doi.org/10.1016/j.clinph.2007.09.143

de Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio-visual integration in schizophrenia. *Schizophrenia Research*, *59*(2–3), 211–218. https://doi.org/10.1016/S0920-9964(01)00344-9

de Jong, J. J., Hodiamont, P. P. G., & de Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: Reduced regulatory effects. *Schizophrenia Research*, *122*(1–3), 136–143. https://doi.org/10.1016/j.schres.2010.04.010

de Nooijer, J. A., Gootjes, L., van Gog, T., Paas, F., & Zwaan, R. A. (2016). Picturing meaning: an ERP study on the integration of left or right-handed first-person perspective pictures into a sentence context. *Language, Cognition and Neuroscience*, *31*(2), 167–178. https://doi.org/10.1080/23273798.2015.1062901

Deamer, F., Palmer, E., Vuong, Q. C., Ferrier, N., Finkelmeyer, A., Hinzen, W., & Watson, S. (2019). Non-literal understanding and psychosis: Metaphor comprehension in individuals with a diagnosis of schizophrenia. *Schizophrenia Research: Cognition*, *18*, 100159. https://doi.org/10.1016/j.scog.2019.100159

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The Neural Career of Sensory-motor Metaphors. *Journal of Cognitive Neuroscience*, *23*(9), 2376–2386. https://doi.org/10.1162/jocn.2010.21596

Di Nuovo, S., Castellano, S., & Guarnera, M. A. (2014). *Mental Imagery Test - MIT*. Hogrefe.

Dillon, B., & Keshev, M. (n.d.). Syntactic dependency formation in sentence processing: A comparative perspective. In S. Barbiers, N. Corver, & M. Polinsky (Eds.), *The Cambridge Handbook of Comparative Syntax*. Cambridge University Press.

Diveica, V., Muraki, E. J., Binney, R. J., & Pexman, P. M. (2023). Socialness Effects in Lexical-Semantic Processing. *PsyArXiv*.

Diveica, V., Pexman, P. M., & Binney, R. J. (2022). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods*, *55*(2), 461–473. https://doi.org/10.3758/s13428-022-01810-x

Domaneschi, F., & Bambini, V. (2020). Pragmatic competence. In E. Fridland & C. Pavese (Eds.), *Routledge Handbook of Skill and Expertise* (pp. 1–12). Routledge.

Dove, G. (2014). Thinking in words: Language as an embodied medium of thought. *Topics in Cognitive Science*, *6*(3), 371–389. https://doi.org/10.1111/tops.12102

Dove, G. (2023). Concepts require flexible grounding. *Brain and Language*, *245*, 105322. https://doi.org/10.1016/j.bandl.2023.105322

Dudschig, C., Mackenzie, I. G., Strozyk, J., Kaup, B., & Leuthold, H. (2016). The Sounds of Sentences: Differentiating the Influence of Physical Sound, Sound Imagery, and Linguistically Implied Sounds on Physical Sound Processing. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(5), 940–961. https://doi.org/10.3758/s13415-016-0444-1

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology*, *33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Elvevåg, B., Foltz, P. W., Weinberger, D. R., & Goldberg, T. E. (2007). Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research*, *93*(1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001

Faber, I., Branco, L. M. T., & França Júnior, M. C. (2016). Cognitive dysfunction in hereditary spastic paraplegias and other motor neuron disorders. *Dementia & Neuropsychologia*, *10*(4), 276–279. https://doi.org/10.1590/s1980-5764-2016dn1004004

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Felsenheimer, A. K., & Rapp, A. M. (2023). Proverb comprehension in schizophrenia: A comprehensive review and meta-analysis. *Comprehensive Psychiatry*, 152444. https://doi.org/10.1016/j.comppsych.2023.152444

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia*, *51*(8), 1510–1517. https://doi.org/10.1016/j.neuropsychologia.2013.04.008

Fields, E. C. (2023). The P300, the LPP, context updating, and memory: What is the functional significance of the emotion-related late positive potential? *International Journal of Psychophysiology*, *192*, 43–52. https://doi.org/10.1016/j.ijpsycho.2023.08.005

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S. J., & Dolan, R. J. (1995). The Mind's Eye—Precuneus Activation in Memory-Related Imagery. *NeuroImage*, *2*(3), 195–200. https://doi.org/10.1006/nimg.1995.1025

Fodor, J. A. (1975). *The Language of Thought: Vol. Vol. 5*. Harvard University Press.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Frau, F., Bischetti, L., Campidelli, L., Tonini, E., Muraki, E. J., Pexman, P. M., & Bambini, V. (*under review*). Understanding with the body? Testing the role of verb relative embodiment across tasks at the interface of language and memory. Submitted to *Journal of Memory and Language*.

Frau, F., Cerami, C., Dodich, A., Bosia, M., & Bambini, V. (2024). Weighing the role of social cognition and executive functioning in pragmatics in the schizophrenia spectrum: A systematic review and meta-analysis. *Brain and Language*, *252*, 105403.

Friedman, D., Sutton, S., Putnam, L., Brown, C., & Erlenmeyer-Kimling, L. (1988). ERP Components in Picture Matching in Children and Adults. *Psychophysiology*, *25*(5), 570–590. https://doi.org/10.1111/j.1469-8986.1988.tb01893.x

Frith, C. D. (1979). Consciousness, Information Processing and Schizophrenia. *British Journal of Psychiatry*, *134*(3), 225–235. https://doi.org/10.1192/bjp.134.3.225

Gallese, V., & Cuccio, V. (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: Insights from the study of action verbs processing and motor disorders in Parkinson's disease. *Cortex*, *100*, 215–225. https://doi.org/10.1016/j.cortex.2018.01.010

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479. https://doi.org/10.1080/02643290442000310

Gamer, M., Lemon, J., & Singh, P. (2022). *irr: Various Coefficients of Interrater Reliability and Agreement* (R package version 0.84.1).

García, A. M., Sedeño, L., Trujillo, N., Bocanegra, Y., Gomez, D., Pineda, D., Villegas, A., Muñoz, E., Arias, W., & Ibáñez, A. (2017). Language Deficits as a Preclinical Window into Parkinson's Disease: Evidence from Asymptomatic Parkin and Dardarin Mutation Carriers. *Journal of the International Neuropsychological Society*, *23*(2), 150–158. https://doi.org/10.1017/S1355617716000710

Gardner, M. K., Rothkopf, E. Z., Lapan, R., & Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition*, *15*(1), 24–28. https://doi.org/10.3758/BF03197709

Gianelli, C., Marzocchi, M., & Borghi, A. M. (2017). Grasping the agent's perspective: A kinematics investigation of linguistic perspective in Italian and German. *Frontiers in Psychology*, *8*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2017.00042

Gibbs, R. W. (2005). Embodiment in Metaphorical Imagination. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition. The Role of Perception and Action in Memory, Language, and Thinking* (pp. 65–92). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.004

Gibbs, R. W., & Matlock, T. (2008). Metaphor, imagination, and simulation. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 161–176). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.011

Glenberg, A. M. (2015). Few Believe the World Is Flat: How Embodiment Is Changing the Scientific Understanding of Cognition. *Canadian Journal of Experimental Psychology*, *69*(2), 165–171. https://doi.org/10.1037/cep0000056

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/BF03196313

Glucksberg, S. (1989). Metaphors in Conversation: How Are They Understood? Why Are They Used? *Metaphor and Symbolic Activity*, *4*(3), 125–143. https://doi.org/10.1207/s15327868ms0403_2

Goldstein, K. (1944). Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia* (pp. 19–39). University of California Press.

Goldstein, K. (1959). Concerning the concreteness in schizophrenia. *The Journal of Abnormal and Social Psychology*, *59*(1), 146–148. https://doi.org/10.1037/h0045400

Goslin, J., Galluzzi, C., & Romani, C. (2013). PhonItalia: a phonological lexicon for Italian. *Behavior Research Methods*, *46*(3), 872–886. https://doi.org/10.3758/s13428-013-0400-8

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological Society*, *16*(5), 748–753. https://doi.org/10.1017/S1355617710000676

Gregory, S., Long, J. D., Klöppel, S., Razi, A., Scheller, E., Minkova, L., Papoutsi, M., Mills, J. A., Durr, A., Leavitt, B. R., Roos, R. A. C., Stout, J. C., Scahill, R. I., Langbehn, D. R., Tabrizi, S. J., & Rees, G. (2017). Operationalizing compensation over time in neurodegenerative disease. *Brain*, *140*(4), 1158–1165. https://doi.org/10.1093/brain/awx022

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3, Speech Acts* (pp. 41–58). Academic Press.

Gröhn, C., Norgren, E., & Eriksson, L. (2022). A systematic review of the neural correlates of multisensory integration in schizophrenia. *Schizophrenia Research: Cognition*, *27*, 100219. https://doi.org/10.1016/j.scog.2021.100219

Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, *71*(18), 1396–1401. https://doi.org/10.1212/01.wnl.0000319701.50168.8c

Guazzotti, P., & Oddera, M. F. (2006). *Il Grande Dizionario dei Proverbi Italiani*. Zanichelli.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Hajcak, G., & Foti, D. (2020). Significance?... Significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: An integrative review. *Psychophysiology*, *57*(7). https://doi.org/10.1111/psyp.13570

Hardiman, O. (2011). Amyotrophic Lateral Sclerosis. In O. Hardiman & C. P. Doherty (Eds.), *Neurodegenerative Disorders* (pp. 143–166). Springer London. https://doi.org/10.1007/978-1-84996-011-3_7

Harrow, M. (1972). Concrete and Idiosyncratic Thinking in Acute Schizophrenic Patients. *Archives of General Psychiatry*, *26*(5), 433. https://doi.org/10.1001/archpsyc.1972.01750230043008

Harrow, M. (1974). Abstract and Concrete Thinking in Schizophrenia During the Prechronic Phases. *Archives of General Psychiatry*, *31*(1), 27–33. https://doi.org/10.1001/archpsyc.1974.01760130013002

Hartig, F. (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (R package version 0.4.6).

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *The Journal of Neuroscience*, *27*(52), 14365–14374. https://doi.org/10.1523/JNEUROSCI.4549-07.2007

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological Distinction of Action Words in the Fronto-Central Cortex. *Human Brain Mapping*, *21*(3), 191–201. https://doi.org/10.1002/hbm.10157

Hinzen, W., & Palaniyappan, L. (2024). The 'L-factor': Language as a transdiagnostic dimension in psychopathology. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *131*, 110952. https://doi.org/10.1016/j.pnpbp.2024.110952

Hirschfeld, G., Feldker, K., & Zwitserlood, P. (2012). Listening to "flying ducks": Individual differences in sentence–picture verification investigated with ERPs. *Psychophysiology*, *49*(3), 312–321. https://doi.org/10.1111/j.1469-8986.2011.01315.x

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing – MEG dissociates early perceptual and late N400 effects. *Brain and Language*, *116*(2), 91–96. https://doi.org/10.1016/j.bandl.2010.07.002

Hitczenko, K., Mittal, V. A., & Goldrick, M. (2021). Understanding Language Abnormalities and Associated Clinical Markers in Psychosis: The Promise of Computational Methods. *Schizophrenia Bulletin*, *47*(2), 344–362. https://doi.org/10.1093/schbul/sbaa141

Hoffman, R. E. (1986). Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences*, *9*(3), 503–517. https://doi.org/10.1017/S0140525X00046781

Hovhannisyan, M., Clarke, A., Geib, B. R., Cicchinelli, R., Monge, Z., Worth, T., Szymanski, A., Cabeza, R., & Davis, S. W. (2021). The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. *Memory & Cognition*, *49*(4), 712–731. https://doi.org/10.3758/s13421-020-01130-5

Hugdal, K. (2009). "Hearing voices": Auditory hallucinations as failure of top-down control of bottom-up perceptual processes. *Scandinavian Journal of Psychology*, *50*(6), 553–560. https://doi.org/10.1111/j.1467-9450.2009.00775.x

Humphries, S., Klooster, N., Cardillo, E., Weintraub, D., Rick, J., & Chatterjee, A. (2019). From action to abstraction: The sensorimotor grounding of metaphor in Parkinson's disease. *Cortex*, *121*, 362–384. https://doi.org/10.1016/j.cortex.2019.09.005

Ibáñez, A., Kühne, K., Miklashevsky, A., Monaco, E., Muraki, E., Ranzini, M., Speed, L. J., & Tuena, C. (2023). Ecological Meanings: A Consensus Paper on Individual Differences and Contextual Influences in Embodied Language. *Journal of Cognition*, *6*(1). https://doi.org/10.5334/joc.228

Jamrozik, A., McQuire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review*, *23*(4), 1080–1089. https://doi.org/10.3758/s13423-015-0861-0

Just, M. A., Newman, S. D., Keller, T. A., McEleney, A., & Carpenter, P. A. (2004). Imagery in sentence comprehension: an fMRI study. *NeuroImage*, *21*(1), 112–124. https://doi.org/10.1016/j.neuroimage.2003.08.042

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 Literary and 260 Nonliterary Metaphors on 10 Psychological Dimensions. *Metaphor and Symbolic Activity*, *3*(4), 191–214. https://doi.org/10.1207/s15327868ms0304_1

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261–276. https://doi.org/10.1093/schbul/13.2.261

Keefe, R. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research*, *68*(2–3), 283–297. https://doi.org/10.1016/j.schres.2003.09.011

Kemmerer, D. (2023). Grounded Cognition Entails Linguistic Relativity: A Neglected Implication of a Major Semantic Theory. *Topics in Cognitive Science*, *15*(4), 615–647. https://doi.org/10.1111/tops.12628

Ketelaar, P. E., Gisbergen, M. S. Van, & Beentjes, J. (2012). Interpretation of Highly Visual 'Open' Advertisements in Dutch Magazines. *Journal of Visual Literacy*, *31*(1), 23–52. https://doi.org/10.1080/23796529.2012.11674693

Khanna, M. M., & Cortese, M. J. (2021). How well imageability, concreteness, perceptual strength, and action strength predict recognition memory, lexical decision, and reading aloud performance. *Memory*, *29*(5), 622–636. https://doi.org/10.1080/09658211.2021.1924789

Kim, M. K., Müller, H. M., & Weiss, S. (2021). What you "mean" is not what I "mean": Categorization of verbs by Germans and Koreans using the semantic differential. *Lingua*, *252*. https://doi.org/10.1016/j.lingua.2020.103012

Kircher, T. T. J., Leube, D. T., Erb, M., Grodd, W., & Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *NeuroImage*, *34*(1), 281–289. https://doi.org/10.1016/j.neuroimage.2006.08.044

Kluwe-Schiavon, B., Sanvicente-Vieira, B., Kristensen, C. H., & Grassi-Oliveira, R. (2013). Executive functions rehabilitation for schizophrenia: A critical systematic review. *Journal of Psychiatric Research*, *47*(1), 91–104. https://doi.org/10.1016/j.jpsychires.2012.10.001

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. https://doi.org/10.1038/35090055

Kropotov, Yu. D., Ponomarev, V. A., Pronina, M. V., & Polyakova, N. V. (2019). Effects of Repetition and Stimulus Mismatch in Sensory Visual Components of Event-related Potentials. *Human Physiology*, *45*(4), 349–355. https://doi.org/10.1134/S0362119719040066

Kuperberg, G. R. (2010). Language in Schizophrenia Part 1: An Introduction. *Language and Linguistics Compass*, *4*(8), 576–589. https://doi.org/10.1111/j.1749-818X.2010.00216.x

Kuperman, V., Schroeder, S., & Gnetov, D. (2024). Word length and frequency effects on text reading are highly similar in 12 alphabetic languages. *Journal of Memory and Language*, *135*, 104497. https://doi.org/10.1016/j.jml.2023.104497

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

La Corte, V., Ferrieux, S., Abram, M., Bertrand, A., Dubois, B., Teichmann, M., & Piolino, P. (2021). The role of semantic memory in prospective memory and episodic future thinking: new insights from a case of semantic dementia. *Memory*, *29*(8), 943–962. https://doi.org/10.1080/09658211.2021.1936069

Laeng, B., Bloem, I. M., D'Ascenzo, S., & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, *131*(2), 263–283. https://doi.org/10.1016/j.cognition.2014.01.003

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

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

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenges to Western Thought*. University of Chicago Press.

Lam, K. J. Y., Dijkstra, T., & Rueschemeyer, S.-A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Frontiers in Psychology*, *6*, 659. https://doi.org/10.3389/fpsyg.2015.00659

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

Langdon, R., Coltheart, M., Ward, P. B., & Catts, S. V. (2002). Disturbed communication in schizophrenia: The role of poor pragmatics and poor mind-reading. *Psychological Medicine*, *32*(7), 1273–1284. https://doi.org/10.1017/S0033291702006396

Langdon, R., Davies, M., & Coltheart, M. (2002). Understanding Minds and Understanding Communicated Meanings in Schizophrenia. *Mind & Language*, *17*(1-2), 68–104. https://doi.org/10.1111/1468-0017.00189

Lecardeur, L., Giffard, B., Laisney, M., Brazo, P., Delamillieure, P., Eustache, F., & Dollfus, S. (2007). Semantic hyperpriming in schizophrenic patients: Increased facilitation or impaired inhibition in semantic association processing? *Schizophrenia Research*, *89*(1–3), 243–250. https://doi.org/10.1016/j.schres.2006.08.025

Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language*, *46*(2), 393–407. https://doi.org/10.1017/S030500091800048X

Lepore, E., & Stone, M. (2010). Against Metaphorical Meaning. *Topoi*, *29*(2), 165–180. https://doi.org/10.1007/s11245-009-9076-1

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.

Li, X., Hu, D., Deng, W., Tao, Q., Hu, Y., Yang, X., Wang, Z., Tao, R., Yang, L., & Zhang, X. (2017). Pragmatic Ability Deficit in Schizophrenia and Associated Theory of Mind and Executive Function. *Frontiers in Psychology*, *8*, 2164. https://doi.org/10.3389/fpsyg.2017.02164

Lindenmayer, J.-P., McGurk, S. R., Khan, A., Kaushik, S., Thanju, A., Hoffman, L., Valdez, G., Wance, D., & Herrmann, E. (2013). Improving Social Cognition in Schizophrenia: A Pilot Intervention Combining Computerized Social Cognition Training With Cognitive Remediation. *Schizophrenia Bulletin*, *39*(3), 507–517. https://doi.org/10.1093/schbul/sbs120

Ljubešić, N., Fišer, D., & Peti-Stantić, A. (2018). Predicting Concreteness and Imageability of Words Within and Across Languages via Word Embeddings. *Proceedings of The Third Workshop on Representation Learning for NLP*, 217–222. https://doi.org/10.18653/v1/w18-3028

Lucarini, V., Cangemi, F., Daniel, B. D., Lucchese, J., Paraboschi, F., Cattani, C., Marchesi, C., Grice, M., Vogeley, K., & Tonna, M. (2022). Conversational metrics, psychopathological dimensions and self-disturbances in patients with schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *272*(6), 997–1005. https://doi.org/10.1007/s00406-021-01329-w

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Maggio, M. G., Piazzitta, D., Andaloro, A., Latella, D., Sciarrone, F., Casella, C., Naro, A., Manuli, A., & Calabrò, R. S. (2022). Embodied cognition in neurodegenerative disorders: What do we know so far? A narrative review focusing on the mirror neuron system and clinical applications. *Journal of Clinical Neuroscience*, *98*, 66–72. https://doi.org/10.1016/j.jocn.2022.01.028

Maher, B. A., Manschreck, T. C., & Molino, M. A. C. (1983). Redundancy, Pause Distributions and Thought Disorder in Schizophrenia. *Language and Speech*, *26*(2), 191–199. https://doi.org/10.1177/002383098302600207

Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *69*(2), 172–178. https://doi.org/10.1037/cep0000060

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

Mahon, B. Z., & Hickok, G. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bulletin & Review*, *23*(4), 941–958. https://doi.org/10.3758/s13423-016-1045-2

Makowski, D. (2018). The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Makri, A., & Jarrold, C. (2021). Investigating the underlying mechanisms of the enactment effect: The role of action–object bindings in aiding immediate memory performance. *Quarterly Journal of Experimental Psychology*, *74*(12), 2084–2096. https://doi.org/10.1177/17470218211019026

Manschreck, T. C., Maher, B. A., & Ader, D. N. (1981). Formal Thought Disorder, the Type-Token Ratio, and Disturbed Voluntary Motor Movement in Schizophrenia. *British Journal of Psychiatry*, *139*(1), 7–15. https://doi.org/10.1192/bjp.139.1.7

Manschreck, T. C., Maher, B. A., Hoover, T. M., & Ames, D. (1984). The type—token ratio in schizophrenic disorders: clinical and research value. *Psychological Medicine*, *14*(1), 151–157. https://doi.org/10.1017/S0033291700003159

Manschreck, T. C., Maher, B. A., Milavetz, J. J., Ames, D., Weisstein, C. C., & Schneyer, M. L. (1988). Semantic priming in thought disordered schizophrenic patients. *Schizophrenia Research*, *1*(1), 61–66. https://doi.org/10.1016/0920-9964(88)90041-2

Marini, A., Spoletini, I., Rubino, I. A., Ciuffa, M., Bria, P., Martinotti, G., Banfi, G., Boccascino, R., Strom, P., Siracusano, A., Caltagirone, C., & Spalletta, G. (2008). The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates. *Schizophrenia Research*, *105*(1–3), 144–155. https://doi.org/10.1016/j.schres.2008.07.011

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Marmolejo-Ramos, F., Correa, J. C., Sakarkar, G., Ngo, G., Ruiz-Fernández, S., Butcher, N., & Yamada, Y. (2017). Placing joy, surprise and sadness in space: a cross-linguistic study. *Psychological Research*, *81*(4), 750–763. https://doi.org/10.1007/s00426-016-0787-9

Marschark, M., & Hunt, R. R. (1985). On memory for metaphor. *Memory & Cognition*, *13*(5), 413–424. https://doi.org/10.3758/BF03198454

Mashal, N., Vishne, T., & Laor, N. (2014). The role of the precuneus in metaphor comprehension: evidence from an fMRI study in people with schizophrenia and healthy participants. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00818

Mashal, N., Vishne, T., Laor, N., & Titone, D. (2013). Enhanced left frontal involvement during novel metaphor comprehension in schizophrenia: Evidence from functional neuroimaging. *Brain and Language*, *124*(1), 66–74. https://doi.org/10.1016/j.bandl.2012.11.012

Matsuda, I., & Nittono, H. (2015). Motivational significance and cognitive effort elicit different late positive potentials. *Clinical Neurophysiology*, *126*(2), 304–313. https://doi.org/10.1016/j.clinph.2014.05.030

Mayer, S. (2021). *imagefluency: Image Statistics Based on Processing Fluency* (R package version 0.2.4).

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McDermott, C. J., & Shaw, P. J. (2008). Diagnosis and management of motor neurone disease. *BMJ*, *336*(7645), 658–662. https://doi.org/10.1136/bmj.39493.511759.BE

Messina, A., Basilico, S., Bottini, G., & Salvato, G. (2022). Exploring the role of interoception in autobiographical memory recollection. *Consciousness and Cognition*, *102*, 103358. https://doi.org/10.1016/j.concog.2022.103358

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meyyazhagan, A., & Orlacchio, A. (2022). Hereditary Spastic Paraplegia: An Update. *International Journal of Molecular Sciences*, *23*(3), 1697. https://doi.org/10.3390/ijms23031697

Michel, C. (2021). Overcoming the modal/amodal dichotomy of concepts. *Phenomenology and the Cognitive Sciences*, *20*(4), 655–677. https://doi.org/10.1007/s11097-020-09678-y

Minor, K. S., Bonfils, K. A., Luther, L., Firmin, R. L., Kukla, M., MacLain, V. R., Buck, B., Lysaker, P. H., & Salyers, M. P. (2015). Lexical analysis in schizophrenia: How emotion and social word use informs our understanding of clinical presentation. *Journal of Psychiatric Research*, *64*, 74–78. https://doi.org/10.1016/J.JPSYCHIRES.2015.02.024

Mo, C., Hannan, A. J., & Renoir, T. (2015). Environmental factors as modulators of neurodegeneration: Insights from gene–environment interactions in Huntington's disease. *Neuroscience & Biobehavioral Reviews*, *52*, 178–192. https://doi.org/10.1016/j.neubiorev.2015.03.003

Møller, M. L. H., Høj, S. H., Østergaard, K., Wallentin, M., & Højlund, A. (2023). No selective action verb impairment in patients with Parkinson's disease: Evidence from Danish patients reading naturalistic texts, a Commentary on García et al., 2018. *Cortex*, *158*, 176–180. https://doi.org/10.1016/j.cortex.2022.07.006

Moro, A. (2014). Response to Pulvermüller: the syntax of actions and other metaphors. *Trends in Cognitive Sciences*, *18*(5), 221. https://doi.org/10.1016/j.tics.2014.01.012

Moro, A., Bambini, V., Bosia, M., Anselmetti, S., Riccaboni, R., Cappa, S. F., Smeraldi, E., & Cavallaro, R. (2015). Detecting syntactic and semantic anomalies in schizophrenia. *Neuropsychologia*, *79*, 147–157. https://doi.org/10.1016/j.neuropsychologia.2015.10.030

Mossaheb, N., Aschauer, H. N., Stoettner, S., Schmoeger, M., Pils, N., Raab, M., & Willinger, U. (2014). Comprehension of metaphors in patients with schizophrenia-spectrum disorders. *Comprehensive Psychiatry*, *55*(4), 928–937. https://doi.org/10.1016/j.comppsych.2013.12.021

Muraki, E. J., Dahm, S. F., & Pexman, P. M. (2023). Meaning in hand: Investigating shared mechanisms of motor imagery and sensorimotor simulation in language processing. *Cognition*, *240*, 105589. https://doi.org/10.1016/j.cognition.2023.105589

Muraki, E. J., & Pexman, P. M. (2021). Simulating semantics: Are individual differences in motor imagery related to sensorimotor effects in language processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(12), 1939–1957. https://doi.org/10.1037/xlm0001039

Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2022). Heterogenous abstract concepts: is "ponder" different from "dissolve"? *Psychological Research*, *86*(8), 2478–2494. https://doi.org/10.1007/s00426-020-01398-x

Muraki, E. J., Speed, L. J., & Pexman, P. M. (2023). Insights into embodied cognition and mental imagery from aphantasia. *Nature Reviews Psychology*, *2*(10), 591–605. https://doi.org/10.1038/s44159-023-00221-9

Murphy, G. (2002). *The Big Book of Concepts*. The MIT Press. https://doi.org/10.7551/mitpress/1602.001.0001

Murphy, S., Gorman, G., Beetz, C., Byrne, P., Dytko, M., McMonagle, P., Kinsella, K., Farrell, M., & Hutchinson, M. (2009). Dementia in SPG4 hereditary spastic paraplegia: Clinical, genetic, and neuropathologic evidence. *Neurology*, *73*(5), 378–384. https://doi.org/10.1212/WNL.0b013e3181b04c6c

Nanay, B. (2021). Unconscious mental imagery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817), 20190689. https://doi.org/10.1098/rstb.2019.0689

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., & Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–1554. https://doi.org/10.1212/WNL.51.6.1546

Nelis, S., Holmes, E. A., Griffith, J. W., & Raes, F. (2014). Mental imagery during daily life: Psychometric evaluation of the Spontaneous Use of Imagery Scale (SUIS). *Psychologica Belgica*, *54*(1), 19–32. https://doi.org/10.5334/pb.ag

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of Emotional and Sensorimotor Knowledge in Semantic Processing of Concrete and Abstract Nouns. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00275

Nickels, L., Fischer-Baum, S., & Best, W. (2022). Single case studies are a powerful tool for developing, testing and extending theories. *Nature Reviews Psychology*, *1*(12), 733–747. https://doi.org/10.1038/s44159-022-00127-y

Noveck, I. (2018). Experimental Pragmatics: The Making of a Cognitive Science. In *Behavioral and Brain Sciences* (Vol. 31, Issue 1). https://www.cambridge.org/core/books/experimental-pragmatics/865CCF285D5330628E73340AA86405D3

Noveck, I., Bianco, M., & Castry, A. (2001). The Costs and Benefits of Metaphor. *Metaphor and Symbol*, *16*(1–2), 109–121. https://doi.org/10.1080/10926488.2001.9678889

Oertel, V., Rotarska-Jagiela, A., van de Ven, V., Haenschel, C., Grube, M., Stangier, U., Maurer, K., & Linden, D. E. J. (2009). Mental imagery vividness as a trait marker across the schizophrenia spectrum. *Psychiatry Research*, *167*(1–2), 1–11. https://doi.org/10.1016/j.psychres.2007.12.008

Oh, J., Chun, J.-W., Joon Jo, H., Kim, E., Park, H.-J., Lee, B., & Kim, J.-J. (2015). The neural basis of a deficit in abstract thinking in patients with schizophrenia. *Psychiatry Research: Neuroimaging*, *234*(1), 66–73. https://doi.org/10.1016/j.pscychresns.2015.08.007

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oomen, P. P., de Boer, J. N., Brederoo, S. G., Voppel, A. E., Brand, B. A., Wijnen, F. N. K., & Sommer, I. E. C. (2022). Characterizing speech heterogeneity in schizophrenia-spectrum disorders. *Journal of Psychopathology and Clinical Science*, *131*(2), 172–181. https://doi.org/10.1037/abn0000736

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Orsini, A., & Laicardi, C. (1997). *WAIS–R: Contributo alla taratura italiana*. Organizzazioni Speciali.

Paivio, A. (1979). *Imagery and Verbal Processes*. Psychology Press. https://doi.org/10.4324/9781315798868

Paivio, A., & Clark, J. (1986). The role of topic and vehicle imagery in metaphor comprehension. *Communication and Cognition*, *19*, 367–387.

Paivio, A., & Walsh, M. (1993). Psychological processes in metaphor comprehension and memory. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 307–328). Cambridge University Press. https://doi.org/10.1017/CBO9781139173865.016

Palaniyappan, L., Benrimoh, D., Voppel, A., & Rocca, R. (2023). Studying Psychosis Using Natural Language Generation: A Review of Emerging Opportunities. *Biological Psychiatry:*

Cognitive Neuroscience and Neuroimaging, *8*(10), 994–1004. https://doi.org/10.1016/j.bpsc.2023.04.009

Palmiero, M., Olivetti Belardinelli, M., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A., & Romani, G. L. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing*, *10*(S2), 268–271. https://doi.org/10.1007/s10339-009-0324-5

Panza, E., Meyyazhagan, A., & Orlacchio, A. (2022). Hereditary spastic paraplegia: Genetic heterogeneity and common pathways. *Experimental Neurology*, *357*, 114203. https://doi.org/10.1016/j.expneurol.2022.114203

Papeo, L., Cecchetto, C., Mazzon, G., Granello, G., Cattaruzza, T., Verriello, L., Eleopra, R., & Rumiati, R. I. (2015). The processing of actions and action-words in amyotrophic lateral sclerosis patients. *Cortex*, *64*, 136–147. https://doi.org/10.1016/j.cortex.2014.10.007

Papoutsi, M., Labuschagne, I., Tabrizi, S. J., & Stout, J. C. (2014). The cognitive burden in Huntington's disease: Pathology, phenotype, and mechanisms of compensation. *Movement Disorders*, *29*(5), 673–683. https://doi.org/10.1002/mds.25864

Parodi, L., Fenu, S., Stevanin, G., & Durr, A. (2017). Hereditary spastic paraplegia: More than an upper motor neuron disease. *Revue Neurologique*, *173*(5), 352–360. https://doi.org/10.1016/j.neurol.2017.03.034

Parola, A., Berardinelli, L., & Bosco, F. M. (2018). Cognitive abilities and theory of mind in explaining communicative-pragmatic disorders in patients with schizophrenia. *Psychiatry Research*, *260*, 144–151. https://doi.org/10.1016/j.psychres.2017.11.051

Parola, A., Lin, J. M., Simonsen, A., Bliksted, V., Zhou, Y., Wang, H., Inoue, L., Koelkebeck, K., & Fusaroli, R. (2022). Speech disturbances in schizophrenia: Assessing cross-linguistic generalizability of NLP automated measures of coherence. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2022.07.002

Parola, A., Salvini, R., Gabbatore, I., Colle, L., Berardinelli, L., & Bosco, F. M. (2020). Pragmatics, Theory of Mind and executive functions in schizophrenia: Disentangling the puzzle using machine learning. *PLoS ONE*, *15*(3), 1–17. https://doi.org/10.1371/journal.pone.0229603

Pearson, D. G., Deeprose, C., Wallace-Hadrill, S. M. A., Heyes, S. B., & Holmes, E. A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review*, *33*(1), 1–23. https://doi.org/10.1016/j.cpr.2012.09.001

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences*, *19*(10), 590–602. https://doi.org/10.1016/j.tics.2015.08.003

Pearson, J., & Westbrook, F. (2015). Phantom perception: voluntary and involuntary nonretinal vision. *Trends in Cognitive Sciences*, *19*(5), 278–284. https://doi.org/10.1016/j.tics.2015.03.004

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pesciarelli, F., Gamberoni, T., Ferlazzo, F., Lo Russo, L., Pedrazzi, F., Melati, E., & Cacciari, C. (2014). Is the comprehension of idiomatic sentences indeed impaired in paranoid Schizophrenia? A window into semantic processing deficits. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00799

Pexman, P. M. (2020). How does meaning come to mind? Four broad principles of semantic processing. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *74*(4), 275–283. https://doi.org/10.1037/cep0000235

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words. *Behavior Research Methods*, *49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, *9*(3), 542–549. https://doi.org/10.3758/BF03196311

Pexman, P. M., & Yap, M. J. (2018). Individual differences in semantic processing: Insights from the Calgary semantic decision project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1091–1112. https://doi.org/10.1037/xlm0000499

Phillips, J. B., Grenoble, L. A., & Mason, P. (2023). The unembodied metaphor: comprehension and production of tactile metaphors without somatosensation. *Frontiers in Communication*, *8*. https://doi.org/10.3389/fcomm.2023.1144018

Pinto-Grau, M., Hardiman, O., & Pender, N. (2018). The Study of Language in the Amyotrophic Lateral Sclerosis - Frontotemporal Spectrum Disorder: a Systematic Review of Findings and New Perspectives. *Neuropsychology Review*, *28*(2), 251–268. https://doi.org/10.1007/s11065-018-9375-7

Poletti, B., Solca, F., Carelli, L., Madotto, F., Lafronza, A., Faini, A., Monti, A., Zago, S., Calini, D., Tiloca, C., Doretti, A., Verde, F., Ratti, A., Ticozzi, N., Abrahams, S., & Silani, V. (2016). The validation of the Italian Edinburgh Cognitive and Behavioural ALS Screen

(ECAS). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *17*(7–8), 489–498. https://doi.org/10.1080/21678421.2016.1183679

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pomarol-Clotet, E., Oh, T. M. S. S., Laws, K. R., & McKenna, P. J. (2008). Semantic priming in schizophrenia: Systematic review and meta-analysis. *British Journal of Psychiatry*, *192*(2), 92–97. https://doi.org/10.1192/bjp.bp.106.032102

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–862. https://doi.org/10.3758/BF03192979

Pulvermüller, F. (2012). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, *25*(5), 423–459. https://doi.org/10.1016/j.jneuroling.2011.03.004

Putnam, H. (1975). The Meaning of "Meaning." *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, *7*, 131–193.

Pylyshyn, Z. W. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. The MIT Press. https://doi.org/10.7551/mitpress/2004.003.0007

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Rasmussen, K. W., & Berntsen, D. (2014). "I can see clearly now": The effect of cue imageability on mental time travel. *Memory & Cognition*, *42*(7), 1063–1075. https://doi.org/10.3758/s13421-014-0414-1

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From Knowing to Remembering: The Semantic–Episodic Distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008

Roberts, A., Nguyen, P., Orange, J. B., Jog, M., Nisbet, K. A., & McRae, K. (2017). Differential impairments of upper and lower limb movements influence action verb processing in Parkinson disease. *Cortex*, *97*, 49–59. https://doi.org/10.1016/j.cortex.2017.09.022

Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement Imagery Ability: Development and Assessment of a Revised Version of the Vividness of Movement

Imagery Questionnaire. *Journal of Sport and Exercise Psychology*, *30*(2), 200–221. https://doi.org/10.1123/jsep.30.2.200

Rofes, A., Zakariás, L., Ceder, K., Lind, M., Johansson, M. B., de Aguiar, V., Bjekić, J., Fyndanis, V., Gavarró, A., Simonsen, H. G., Sacristán, C. H., Kambanaros, M., Kraljević, J. K., Martínez-Ferreiro, S., Mavis, İ., Orellana, C. M., Sör, I., Lukács, Á., Tunçer, M., … Howard, D. (2018). Imageability ratings across languages. *Behavior Research Methods*, *50*(3), 1187–1197. https://doi.org/10.3758/s13428-017-0936-0

Romero Lauro, L. J., Mattavelli, G., Papagno, C., & Tettamanti, M. (2013). She runs, the road runs, my mind runs, bad blood runs between us: Literal and figurative motion verbs: An fMRI study. *NeuroImage*, *83*, 361–371. https://doi.org/10.1016/j.neuroimage.2013.06.050

Rosen, C., Harrow, M., Tong, L., Jobe, T. H., & Harrow, H. (2021). A word is worth a thousand pictures: A 20-year comparative analysis of aberrant abstraction in schizophrenia, affective psychosis, and non-psychotic depression. *Schizophrenia Research*, *238*, 1–9. https://doi.org/10.1016/j.schres.2021.09.005

Rosi, A., Bruine de Bruin, W., Del Missier, F., Cavallini, E., & Russo, R. (2017). Decision-making competence in younger and older adults: which cognitive abilities contribute to the application of decision rules? *Aging, Neuropsychology, and Cognition*, *26*(2), 174–189. https://doi.org/10.1080/13825585.2017.1418283

Rossi, S., Rubegni, A., Riso, V., Barghigiani, M., Bassi, M. T., Battini, R., Bertini, E., Cereda, C., Cioffi, E., Criscuolo, C., Dal Fabbro, B., Dato, C., D'Angelo, M. G., Di Muzio, A., Diamanti, L., Dotti, M. T., Filla, A., Gioiosa, V., Liguori, R., … Casali, C. (2022). Clinical-Genetic Features Influencing Disability in Spastic Paraplegia Type 4. *Neurology Genetics*, *8*(2), e664. https://doi.org/10.1212/NXG.0000000000000664

Rubio Fernández, P. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics*, *24*(4), 345–371. https://doi.org/10.1093/jos/ffm006

Rueschemeyer, S.-A., Glenberg, A. M., Kaschak, M. P., Mueller, K., & Friederici, A. D. (2010). Top-Down and Bottom-Up Contributions to Understanding Sentences Describing Objects in Motion. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00183

Ruini, C., Ottolini, F., Rafanelli, C., Ryff, C., & Fava, G. A. (2003). La validazione italiana delle Psychological Well-being Scales (PWB). *Rivista Di Psichiatria*, *38*(3), 117–130.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

Sabatelli, M., Conte, A., & Zollino, M. (2013). Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clinical Genetics*, *83*(5), 408–416. https://doi.org/10.1111/cge.12117

Salavera, C., Puyuelo, Antoñanzas, & Teruel. (2013). Semantics, pragmatics, and formal thought disorders in people with schizophrenia. *Neuropsychiatric Disease and Treatment*, 177. https://doi.org/10.2147/NDT.S38676

Schendan, H. E., & Ganis, G. (2012). Electrophysiological Potentials Reveal Cortical Mechanisms for Mental Imagery, Mental Simulation, and Grounded (Embodied) Cognition. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00329

Schettino, A., Lauro, L. R., Crippa, F., Anselmetti, S., Cavallaro, R., & Papagno, C. (2010). The comprehension of idiomatic expressions in schizophrenic patients. *Neuropsychologia*, *48*(4), 1032–1040. https://doi.org/10.1016/j.neuropsychologia.2009.11.030

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schule, R., Holland-Letz, T., Klimpe, S., Kassubek, J., Klopstock, T., Mall, V., Otto, S., Winner, B., & Schols, L. (2006). The Spastic Paraplegia Rating Scale (SPRS): A reliable and valid measure of disease severity. *Neurology*, *67*(3), 430–434. https://doi.org/10.1212/01.wnl.0000228242.53336.90

Schwarzkopf, D. S. (2023). What is the true range of mental imagery? *Cortex*. https://doi.org/10.1016/j.cortex.2023.09.013

Sela, T., Lavidor, M., & Mitchell, R. L. C. (2015). A possible contributory mechanism for impaired idiom perception in schizophrenia. *Psychiatry Research*, *229*(1–2), 1–11. https://doi.org/10.1016/j.psychres.2015.07.021

Sheehan, P. W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of Clinical Psychology*, *23*(3), 386–389. https://doi.org/10.1002/1097-4679(196707)23:3<386::AID-JCLP2270230328>3.0.CO;2-S

Shen, Z.-Y., Tsai, Y.-T., & Lee, C.-L. (2015). Joint Influence of Metaphor Familiarity and Mental Imagery Ability on Action Metaphor Comprehension: An Event-Related Potential Study. *Language and Linguistics*, *16*(4), 615–637. https://doi.org/10.1177/1606822X15583241

Shojaie, A., Rota, S., Al Khleifat, A., Ray Chaudhuri, K., & Al-Chalabi, A. (2023). Non-motor symptoms in amyotrophic lateral sclerosis: lessons from Parkinson's disease. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *24*(7–8), 562–571. https://doi.org/10.1080/21678421.2023.2220748

Siakaluk, P. D., Newcombe, P. I., Duffels, B., Li, E., Sidhu, D. M., Yap, M. J., & Pexman, P. M. (2016). Effects of Emotional Experience in Lexical Decision. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01157

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body–object interaction effect. *Cognition*, *106*(1), 433–443. https://doi.org/10.1016/j.cognition.2006.12.011

Siciliano, M., Trojano, L., Trojsi, F., Greco, R., Santoro, M., Basile, G., Piscopo, F., D'Iorio, A., Patrone, M., Femiano, C., Monsurrò, M., Tedeschi, G., & Santangelo, G. (2017). Edinburgh Cognitive and Behavioural ALS Screen (ECAS)-Italian version: regression based norms and equivalent scores. *Neurological Sciences*, *38*(6), 1059–1068. https://doi.org/10.1007/s10072-017-2919-4

Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica*, *149*(January), 32–39. https://doi.org/10.1016/j.actpsy.2014.02.009

Sidhu, D. M., & Pexman, P. M. (2016). Is Moving More Memorable than Proving? Effects of Embodiment and Imagined Enactment on Verb Memory. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01010

Silge, J., & Robinson, D. (2017). *Text Mining with R. A Tidy Approach*. O'Reilly Media, Inc.

Silverstein, S. M., & Lai, A. (2021). The Phenomenology and Neurobiology of Visual Distortions and Hallucinations in Schizophrenia: An Update. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.684720

Solowska, J. M., & Baas, P. W. (2015). Hereditary spastic paraplegia SPG4: what is known and not known about the disease. *Brain*, *138*(9), 2471–2484. https://doi.org/10.1093/brain/awv178

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd edition). Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, *17*(1-2), 3–23. https://doi.org/10.1111/1468-0017.00186

Sperber, D., & Wilson, D. (2008). A Deflationary Account of Metaphors. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–106). Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.007

Sperber, R. D., McCauley, C., Ragain, R. D., & Weil, C. M. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition*, *7*(5), 339–345. https://doi.org/10.3758/BF03196937

Spitzer, M. (1993). The psychopathology, neuropsychology, and neurobiology of associative and working memory in schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *243*(2), 57–70. https://doi.org/10.1007/BF02191566

Spitzer, M. (1997). A Cognitive Neuroscience View of Schizophrenic Thought Disorder. *Schizophrenia Bulletin*, *23*(1), 29–50. https://doi.org/10.1093/schbul/23.1.29

Spitzer, M., Braun, U., Hermle, L., & Maier, S. (1993). Associative semantic network dysfunction in thought-disordered schizophrenic patients: Direct evidence from indirect semantic priming. *Biological Psychiatry*, *34*(12), 864–877. https://doi.org/10.1016/0006-3223(93)90054-H

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stephan-Otto, C., Siddi, S., Senior, C., Cuevas-Esteban, J., Cambra-Martí, M. R., Ochoa, S., & Brébion, G. (2017). Remembering verbally-presented items as pictures: Brain activity underlying visual mental images in schizophrenia patients with visual hallucinations. *Cortex*, *94*, 113–122. https://doi.org/10.1016/j.cortex.2017.06.009

Sterelny, K. (1986). The Imagery Debate. *Philosophy of Science*, *53*(4), 560–583. https://doi.org/10.1086/289340

Stern, C. E., & Hasselmo, M. E. (2009). Recognition Memory. In *Encyclopedia of Neuroscience* (pp. 49–54). Elsevier. https://doi.org/10.1016/B978-008045046-9.00784-1

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2021). partR2: partitioning R ^{2} in generalized linear mixed models. *PeerJ*, *9*, e11414. https://doi.org/10.7717/peerj.11414

Straka, M., & Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies.*

Strong, M. J., Grace, G. M., Orange, J. B., Leeper, H. A., Menon, R. S., & Aere, C. (1999). A prospective study of cognitive impairment in ALS. *Neurology*, *53*(8), 1665–1670. https://doi.org/10.1212/wnl.53.8.1665

Tabossi, P., Arduino, L., & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, *43*(1), 110–123. https://doi.org/10.3758/s13428-010-0018-z

Talamo, L., Celata, C., & Bertinetto, P. M. (2016). DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure*, *9*(1), 72–102. https://doi.org/10.3366/word.2016.0087

Tallaksen, C. M. E., Guichart-Gomez, E., Verpillat, P., Hahn-Barma, V., Ruberg, M., Fontaine, B., Brice, A., Dubois, B., & Dürr, A. (2003). Subtle Cognitive Impairment but No Dementia in Patients With Spastin Mutations. *Archives of Neurology*, *60*(8), 1113. https://doi.org/10.1001/archneur.60.8.1113

Tavano, A., Sponda, S., Fabbro, F., Perlini, C., Rambaldelli, G., Ferro, A., Cerruti, S., Tansella, M., & Brambilla, P. (2008). Specific linguistic and pragmatic deficits in Italian patients with schizophrenia. *Schizophrenia Research*, *102*(1–3), 53–62. https://doi.org/10.1016/j.schres.2008.02.008

Taylor, L. J., Brown, R. G., Tsermentseli, S., Al-Chalabi, A., Shaw, C. E., Ellis, C. M., Leigh, P. N., & Goldstein, L. H. (2013). Is language impairment more common than executive dysfunction in amyotrophic lateral sclerosis? *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(5), 494–498. https://doi.org/10.1136/jnnp-2012-303526

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to Action-related Sentences Activates Fronto-parietal Motor Circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281. https://doi.org/10.1162/0898929053124965

Tettamanti, M., & Moro, A. (2012). Can syntax appear in a mirror (system)? *Cortex*, *48*(7), 923–935. https://doi.org/10.1016/j.cortex.2011.05.020

The MathWorks Inc. (2022). *MATLAB* (version: 9.13.0, R2022b). The MathWorks Inc. https://www.mathworks.com.

Thompson, W. L., & Kosslyn, S. M. (2000). Neural Systems Activated during Visual Mental Imagery. In *Brain Mapping: The Systems* (pp. 535–560). Elsevier. https://doi.org/10.1016/B978-012692545-6/50020-9

Thurstone, T. G., & Thurstone, L. L. (1963). *Primary mental ability*. Science Research Associates.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*, 314. https://doi.org/10.3389/fnhum.2012.00314

Titone, D., Holzmanb, P. S., & Levya, D. L. (2002). Idiom processing in schizophrenia: Literal implausibility saves the day for idiom priming. *Journal of Abnormal Psychology*, *111*(2), 313–320. https://doi.org/10.1037/0021-843X.111.2.313

Tjuka, A. (2020). General patterns and language variation: Word frequencies across {E}nglish, {G}erman, and {C}hinese. *Proceedings of the Workshop on the Cognitive Aspects of the Lexicon*, 23–32. https://www.aclweb.org/anthology/2020.cogalex-1.3

Tousignant, C., & Pexman, P. M. (2012). Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00053

Tschacher, W., Giersch, A., & Friston, K. (2017). Embodiment and Schizophrenia: A Review of Implications and Applications. *Schizophrenia Bulletin*, *43*(4), 745–753. https://doi.org/10.1093/schbul/sbw220

Tseng, H.-H., Bossong, M. G., Modinos, G., Chen, K.-M., McGuire, P., & Allen, P. (2015). A systematic review of multisensory cognitive–affective integration in schizophrenia. *Neuroscience & Biobehavioral Reviews*, *55*, 444–452. https://doi.org/10.1016/j.neubiorev.2015.04.019

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. Academic Press.

Unger, C. (2019). Allegory in Relation to Metaphor and Irony. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, Pragmatics and Interpretation* (pp. 240–252). Cambridge University Press. https://doi.org/10.1017/9781108290593.021

van der Gaag, M., Hoffman, T., Remijsen, M., Hijman, R., de Haan, L., van Meijel, B., van Harten, P. V., Valmaggia, L., de Hert, M., Cuijpers, A., & Wiersma, D. (2006). The five-factor model of the Positive and Negative Syndrome Scale II: A ten-fold cross-validation of a revised model. *Schizophrenia Research*, *85*(1–3), 280–287. https://doi.org/10.1016/j.schres.2006.03.021

van Es, M. A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R. J., Veldink, J. H., & van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *The Lancet*, *390*(10107), 2084–2098. https://doi.org/10.1016/S0140-6736(17)31287-4

Vander Stichele, G., Durr, A., Yoon, G., Schüle, R., Blackstone, C., Esposito, G., Buffel, C., Oliveira, I., Freitag, C., van Rooijen, S., Hoffmann, S., Thielemans, L., & Cowling, B. S. (2022). An integrated modelling methodology for estimating global incidence and prevalence of hereditary spastic paraplegia subtypes SPG4, SPG7, SPG11, and SPG15. *BMC Neurology*, *22*(1), 115. https://doi.org/10.1186/s12883-022-02595-4

Vigliocco, G., & Vinson, D. P. (2012). Semantic representation. In *The Oxford Handbook of Psycholinguistics* (pp. 195–216). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0012

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*, *94*(3), B91–B100. https://doi.org/10.1016/j.cognition.2004.06.004

Voeten, C. C. (2023). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression* (R package version 2.10.1).

Wang, Y., Cui, L., Wang, H., Tian, S., & Zhang, X. (2004). The sequential processing of visual feature conjunction mismatches in the human brain. *Psychophysiology*, *41*(1), 21–29. https://doi.org/10.1111/j.1469-8986.2003.00134.x

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale - Revised*. Psychological Corp.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*(AUG), 1–17. https://doi.org/10.3389/fnhum.2014.00583

Weiss, P. H., Ubben, S. D., Kaesberg, S., Kalbe, E., Kessler, J., Liebig, T., & Fink, G. R. (2016). Where language meets meaningful action: a combined behavior and lesion analysis of aphasia and apraxia. *Brain Structure and Function*, *221*(1), 563–576. https://doi.org/10.1007/s00429-014-0925-3

Wernicke, C. (1885). Recent works on aphasia. In G. Eggert (Ed.), *Wernicke's Works on Aphasia* (pp. 100–111). Mouton.

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural Dissociations between Action Verb Understanding and Motor Imagery. *Journal of Cognitive Neuroscience*, *22*(10), 2387–2400. https://doi.org/10.1162/jocn.2009.21386

Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48*(10), 3128–3136. https://doi.org/10.1016/j.neuropsychologia.2010.06.028

Wilson, D., & Carston, R. (2007). A unitary approach to Lexical Pragmatics: Relevance, Inference and Ad hoc concepts. In *Pragmatics* (Issue January). https://doi.org/10.1057/978-1-349-73908-0_12

Wilson, D., & Sperber, D. (2004). Relevance theory. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607–632). Blackwell.

Wolfson, C., Gauvin, D. E., Ishola, F., & Oskoui, M. (2023). Global Prevalence and Incidence of Amyotrophic Lateral Sclerosis. *Neurology*, *101*(6), e613–e623. https://doi.org/10.1212/WNL.0000000000207474

Woolley, S. C., & Rush, B. K. (2017). Considerations for Clinical Neuropsychological Evaluation in Amyotrophic Lateral Sclerosis. *Archives of Clinical Neuropsychology*, *32*(7), 906–916. https://doi.org/10.1093/arclin/acx089

Wright, D. M. (1975). Impairment in abstract conceptualization in schizophrenia. *Psychological Bulletin*, *82*(1), 120–127. https://doi.org/10.1037/h0076256

Wu, J., Duan, H., Tian, X., Wang, P., & Zhang, K. (2012). The effects of visual imagery on face identification: an ERP study. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00305

Yang, F. G., Edens, J., Simpson, C., & Krawczyk, D. C. (2009). Differences in task demands influence the hemispheric lateralization and neural correlates of metaphor. *Brain and Language*, *111*(2), 114–124. https://doi.org/10.1016/j.bandl.2009.08.006

Yang, J., & Shu, H. (2016). Involvement of the Motor System in Comprehension of Non-Literal Action Language: A Meta-Analysis Study. *Brain Topography*, *29*(1), 94–107. https://doi.org/10.1007/s10548-015-0427-5

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 53–79. https://doi.org/10.1037/a0024177

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: The role of feedback. *Memory & Cognition*, *43*(8), 1148–1167. https://doi.org/10.3758/s13421-015-0536-0

Yap, M. J., & Pexman, P. M. (2016). Semantic Richness Effects in Syntactic Classification: The Role of Feedback. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01394

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*(APRIL 2012), 1–10. https://doi.org/10.3389/fnhum.2012.00072

Yentes, R., & Wilhelm, F. (2023). *careless: Procedures for computing indices of careless responding* (R package version 1.2.2).

York, C., Olm, C., Boller, A., McCluskey, L., Elman, L., Haley, J., Seltzer, E., Chahine, L., Woo, J., Rascovsky, K., McMillan, C., & Grossman, M. (2014). Action verb comprehension in amyotrophic lateral sclerosis and Parkinson's disease. *Journal of Neurology*, *261*(6), 1073–1079. https://doi.org/10.1007/s00415-014-7314-y

Zou, Z.-Y., Zhou, Z.-R., Che, C.-H., Liu, C.-Y., He, R.-L., & Huang, H.-P. (2017). Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(7), 540–549. https://doi.org/10.1136/jnnp-2016-315018