UNIVERSITY OF TRENTO

DOCTORAL THESIS

Left prefrontal and parietal contribution to sentence processing: a neuromodulation approach

Candidate: *Lorenzo Vercesi*

Supervisor: *Prof. Luigi Cattaneo*
Co-supervisor: *Prof. Gabriele Miceli*

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy
in the
Doctorate in Cognitive and Brain Sciences

Academic Year 2023 – 2024

Declaration

I hereby declare that the contents and organization of this dissertation constitute original work and do not compromise in any way the rights of third parties, including those relating to the security of personal data.

Lorenzo Vercesi
November, 2023

CONTENTS

STRUCTURE OF THE THESIS

The present thesis is based on two main projects. The first project was structured on the basis of the outcomes of a series of works published by our group between 2015 and 2021 that investigated the neural basis of sentence processing through a neuromodulation approach. This project has been the main focus of my PhD work, it provides the original data that will be discussed and it represents the main core of this dissertation. The manuscript of this work is in preparation. The second project (still ongoing) aimed at comparing two statistical approaches to data analysis: the standard approach of Analysis of Variance (ANOVA) and the Linear Mixed Model Effects (LMMs) method. Previously acquired data were re-analysed to explore differences and similarities of the experimental outcome in the two approaches. The resulting dissertation consists of three main chapters. The first and the second chapters are dedicated to the first project and a related pilot study; the third chapter relates to the second project. The structure of each chapter is the same: a general introduction on the topic and literature background, a description of the materials and methods, the analysis of the results, a discussion and some final conclusions and remarks.

Figure legends

All figures (figures, graphs, tables) have been renumbered from 1 within each new chapter.

Fig. = figure

Tab. = table

For graphs and tables:

* = $p < 0.05$

** = $p < 0.01$

*** = $p < 0.001$

The word has not been recognized as a virus because it has achieved a state of stable symbiosis with the host. My basis theory is that the written word was literally a virus that made the spoken word possible. Language is a virus from outer space

William S. Burroughs

The idea that science is a concentrate of truth is a philosophers' idea

Paul Feyerabend

While the commonplace can be understood as a reduction of the exceptional, the exceptional cannot be understood by diluting the commonplace

Edgar Wind

THESIS ABSTRACT

Describing a comprehensive neurofunctional model of sentence comprehension has always been a complex challenge. On one hand, disentangling the subprocesses that are necessary for computing the meaning of a sentence and their neural underpinnings is insidious. Each subprocess is closely interconnected with the others, and isolating only one as if it were separable can undermine the investigation of the overall process above. On the other hand, available data on the neural basis of sentence processing are not straightforward. This thesis explores relevant contributions and attempts to highlight open questions regarding the neural basis of two key processes in sentence comprehension, namely morphosyntactic processing and thematic role assignment. It presents and discusses original data resulting from an experiment that, to our knowledge, represents the first investigation of the neural basis of these two processes in the same sentential context. Results demonstrate that morphosyntactic and thematic processing rely on functionally distinct neural correlates in the left hemisphere. Morphosyntactic aspects are mostly processed in a left prefrontal network including the left inferior frontal gyrus (IFG) and the middle frontal gyrus (MFG), whereas thematic role assignment correlates with a left parietal node including the left intraparietal sulcus (IPS). Moreover, it is argued that results support the view that these regions play a language-related rather than domain-general role in human cognition. Finally, two statistical approaches to the analysis of the same TMS language data (ANOVA and Linear Mixed Models – LMMs) are compared. Their outcomes are discussed and an attempt is made at accounting for similarities and differences. Results suggest that the two models should not be considered on a sort of quality hierarchy according to which one has greater or lesser explanatory power than the other. Rather, they both represent legitimate and reliable approaches to account for data variability.

CHAPTER 1

Investigating the neural correlates of morphosyntactic and thematic role processing in the left hemisphere: a TMS study

ABSTRACT

Correctly assigning thematic roles (who does what to whom) and processing morphosyntactic information (agreement phenomena) are critical for sentence comprehension. Even though their neural correlates have been investigated in numerous studies, these processes have never been addressed in identical sentential context and in the same modality. In the literature, comprehending complex sentences (including subject- and object-relatives) was used to study thematic role assignment, and producing agreement phenomena in noun or verb phrases to investigate morphosyntax. Results stressed the role of posterior regions in sentence processing and questioned that of prefrontal regions in the same endeavor. The present work investigates the comprehension of thematic and morphosyntactic information in the same active and passive reversible sentences. Repetitive Transcranial Magnetic Stimulation (rTMS) was applied to brain regions that, based on previous lesion and neuroimaging reports, play a putative role in thematic and morphosyntactic processes - the left inferior frontal gyrus (l-IFG), the left middle frontal gyrus (l-MFG) and the left intraparietal sulcus (l-IPS). Results support a differential involvement of these regions. The prefrontal regions (l-IFG and l-MFG) showed selective response when morphosyntactic information was crucial for comprehension, and the l-IPS when a correct mapping of thematic roles was required. Evidence suggests a neurofunctional distinction between morphosyntactic processes and thematic role assignment and provides new insights on the distinct role of the targeted regions.

INTRODUCTION

At the heart of sentence comprehension: morphosyntactic processing and thematic role mapping

The processing of morphosyntactic aspects (agreement phenomena) and the assignment of thematic roles (who does what to whom) is fundamental to the computation of sentential meaning. Morphosyntax, as defined by **Raettig et al. (2009)**, encompasses the "syntactic operations that influence the morphology of one or more elements within a sentence." Morphosyntactic information involves extracting morphological details from the input and combining them with syntactic and grammatical rules that govern sentence structure and proper agreement.

Various types of morphosyntactic violations have been used to explore this aspect of sentence processing, mostly in behavioral (see for example **Mancini et al., 2013, 2014a,b)** and electrophysiological investigations **(Coulson et al., 1998; Tanner & van Hell, 2014; Hasting & Kotz, 2008; Molinaro et al., 2011, 2008; Mancini et al., 2011a,b)**. To identify and address these issues, different elements at the juncture of individual word morphology and the arrangement of words within a structure must be thoroughly analyzed and integrated. An inability to accurately identify these violations could arise from damage to different processing stages, ranging from single-word morphology to higher-level factors like syntactic structure and to their interaction. Accurate morphosyntactic processing leads to a correct understanding of sentence meaning. For instance, a sentence like "*The waiters serve the customer*" defines a specific event involving an action with multiple participants (at least two waiters - the doers of the action - and one customer - the target of the action), distinct from sentences like "*The waiter serves the customer*" or "*The waiter serves the customers*". Hence, proficient morphosyntactic processing plays a pivotal role in sentence comprehension.

Another crucial aspect discussed here is the assignment of thematic roles. Thematic roles have been defined as: "*semantic roles that may be played by the subcategorized complements (or arguments) of a verb*" (**Tanenhaus and Carlson, 1989**). They correspond to labels we assign to who/what is doing/receiving a specific action denoted by a verb. In a sentence like *The boy eats the apple* one has to understand that someone (*the boy*) is doing the action of eating something (*the apple*); these semantic labels are called agent and theme. Several features influence the correct assignment of thematic roles. First, thematic role mapping is modulated by word order. In S-V-O languages (such as Italian or English), word order is 'canonical' when a sentence is in the active form (agent-verb-theme). In the same languages, word order is reversed in passive sentences (theme-verb-agent). Non-canonical word order makes the processing of passive structures more demanding. Second, thematic role assignment depends on semantic features, such as semantic reversibility. This feature refers to the property of a linguistic structure or sentence whereby its meaning remains acceptable when read or processed in both its original and its reverse word order. Word order and semantic (ir)reversibility influence the ease with which a sentence can be processed by a speaker/listener. In an active irreversible sentence like *The boy eats the apple,* word order and semantic irreversibility are strong cues for thematic role assignment, since in active sentences the first noun typically is the grammatical subject and the agent. In the corresponding passive form *The apple is eaten by the boy*, word order does not provide useful information, as the first noun is the grammatical subject but is also the theme, not the agent. However, in this case semantic irreversibility constrains sentence interpretation, as semantic knowledge about the real world establishes that only a boy can eat an apple. In contrast, in semantically reversible sentences (*The girl is kissed by the boy*) both constituents can potentially be assigned the role of agent. The verb '*to kiss*' requires an animate agent and the sentence has two animate constituents competing for that role. In such sentences, the assignment of thematic roles must be driven by syntactic analysis in order to solve the competition between the constituents. It is thought that semantically reversible passive sentences require a re-analysis of the first-pass parsing due to the changed syntactic structure and semantic reversibility (**Chomsky 1965, 1981**; **Pollard and Sag 1994**; **Bresnan 2000**). This process results in the re-assignment of the previously mapped thematic roles. Due to re-analysis, reversible passive sentences are associated with higher costs of processing.

To sum up, thematic role assignment can be modulated by both syntactic and semantic aspects and this makes this process ideal to investigate sentence processing and meaning interpretation.

The processing of non-linguistic information

In addition to linguistic knowledge, sentence comprehension requires domain-general cognitive resources. When processing a sentence, many pieces of linguistic information stored in long-term memory (LTM) (syntactic/grammatical rules, general semantic knowledge, morphosyntactic information, etc.) must be retrieved and maintained active in short-term memory (STM). The contribution of STM in sentence processing has largely been debated. The role of STM and working memory (WM) in linguistic/cognitive tasks was modeled for the first time by Baddeley and his collaborators (**Baddeley et al., 1986; Baddeley & Hitch, 1994; Gathercole & Baddeley, 1993**). On this view, WM is articulated in three systems: the central executive (main component), the articulatory loop and the visual-spatial sketchpad (the so-called 'slave systems'). The central executive system controls and regulates cognitive processes. It can be considered as the link between WM and LTM: information in the WM storage is directed and targeted to the LTM storage. This dynamic WM system consists of two systems: the visuo-spatial sketchpad and the phonological loop. These 'slave systems' temporarily store verbal and visual information when the central executive is overloaded. Alternatively, **Just and Carpenter (1992)** challenge the idea that WM has a fixed storage system, proposing a more dynamic perspective in which WM capacity varies among individuals due to differences in attentional control, rather than due to a fixed storage limit. These differences stem from variations in attentional allocation strategies. Instead of including modality-specific buffers, in this view WM overlaps only with the central executive component in Baddeley and Hitch's model that is related to language comprehension.

Caplan & Waters' (2013) proposal did not include the typical verbal WM, as in the model of Baddeley & Hitch, rather a separate memory store, specific for language. This store is divided into two components: (i) an "interpretative processing" and (ii) a "post-interpretative processing". The first component would be involved in extracting meaning from a linguistic signal; the second in using that meaning to accomplish more complex tasks such as storing information in LTM, reasoning, action planning or thematic role assignment. This model is based on evidence in both healthy volunteers and patients, indicating that the store systems recruited by language processing are distinct from those measured by standard tests of STM. Another conceptualization is represented by the multi-store models (**Martin et al., 1994**; **Howard & Nickels, 2005**), which hypothesize the existence of more than one store, each linked to distinct processes. Some of these models argue that semantic and phonological stores are separated and only the semantic store would be involved in language processing **(Martin & Romani, 1994**; **Martin & He, 2004**). Other models assume the presence of a phonological buffer for the input (language comprehension) and a separated phonological buffer for the output (language production) (**Monsell, 1987**).

These models were developed on the basis of two lines of research. One line compared the effects of concurrent presentation of digits or words to examine the related cognitive load. Underlying this comparison is the idea that if STM and language processing share the same resources, an STM task that requires them should make syntactic processing more difficult (**Carpenter & Just, 1992**). The other line has focused on STM deficits on patients affected by lesions and on aphasic patients. These patients should exhibit difficulties in language comprehension due to their STM impairments (for an exhaustive presentation of these findings see **Papagno & Cecchetto, 2019**). Although some of the presented models show substantial overlaps, their differences underline the lack of consensus in interpreting existing evidence and make them converge on a comprehensive model of the link between STM/WM and language processing.

Another domain-general aspect involved in sentence comprehension is conflict monitoring. As in other cognitive domains (see **Botvinik et al., 2001** and **Coles et al., 2001** for the very first accounts of conflict monitoring in sensory attention), this mechanism intervenes when a strong expectation is violated by what is perceived in the input; to solve this conflict a reanalysis is required to re-process the input (for a review of the different models of conflict monitoring in language perception see **Van De Meerendonk et al., 2009**). Even though these models differ in some respects, a common assumption is that conflict monitoring is triggered by a discrepancy between expected and perceived representation. Consider for instance a passive, semantically reversible sentence such as: "*The boy is kissed by the girl*". To interpret its meaning, one must identify the passive structure, revise the default parsing to re-assign thematic roles (in passives the agent is to the right of the verb), and overcome the semantic ambiguity/conflict that arises from two constituents competing for the role of agent. This latter step is crucial to the process. It relies on a revising mechanism that allows "solving" the conflict between two competing constituents that results from the semantic reversibility of the action. Solving this ambiguity requires both domain-general cognitive resources and linguistic knowledge (syntactic and semantic).

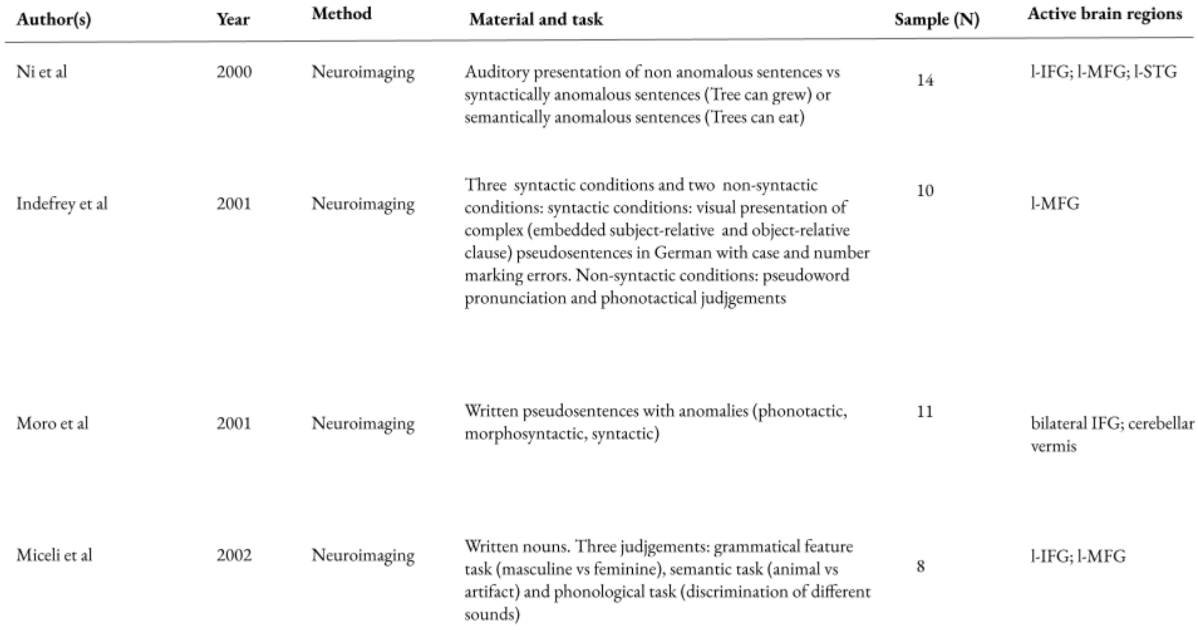
The neural basis of sentence processing

The mechanisms underlying the interpretation of sentences in real time have always attracted the attention of researchers. According to the main neurocognitive models, language is processed in a left lateralized fronto-temporo-parietal network, but the specific role of the regions that are part of this network in supporting the different processes of the form-to-meaning mapping is still debated. **Friederici**'s model (**2011, 2012**) proposed a functional dissociation of the ventral and the dorsal pathways, each divided in two subpathways. A subpath of the ventral pathway would be related to semantic processing by virtue of the connections between the IFG and middle and posterior superior (mSTG and pSTG) and the middle temporal gyri (MTG). Another ventral subpath would be linked to the initial stages of the syntactic processing through projecting to the left frontal operculum. As for the dorsal stream, a subpath connecting temporal and premotor regions through the inferior parietal cortex would be responsible for mediating sensory and motor function. Another subpath would play a top-down control role particularly recruited to the processing of syntactically complex sentences. This path links the inferior frontal regions (Broca's area) to temporal areas. Another relevant neurobiological model of language processing has been proposed by **Hagoort** (2005). This account stands on the ground that the main functional components of language are represented by Memory, Unification and Control. The Memory component refers to the linguistic information stored in LTM and to the retrieval operations. The Unification component pertains to the integration of the lexically retrieved information in higher-level structures. The Control component connects language to action. These three components would be supported by distinct but interconnected regions in the left hemisphere. The Memory component would be sustained by the left temporal cortex; in particular, the phonetic and phonological aspects are linked to the posterior superior temporal gyrus (STG) and to the superior temporal sulcus (STS), while semantic aspects seem to be related to the left middle and inferior temporal gyri. The Control component is suggested to be sustained by a network of regions including the anterior cingulate cortex (ACC) and the dorsolateral prefrontal cortex (DLPFC). Finally, Broca's area would play a crucial role in the Unification process. **Bornkessel-Schlesewsky & Schlesewsky** account (**2013**) also supported the idea of a dissociation between a dorsal and a ventral pathway that reflects a functional distinction between processing of syntactic and lexical and semantic aspects. In this model the anterior portion of the left temporal gyrus would be involved in the assignment of thematic roles. On the other hand, syntactic processing would be mediated by temporal and parietal areas in the dorsal pathway. **Matchin and Hickok** (2020) proposed a hierarchical model according to which language comprehension would be supported by an interaction between left inferior frontal nodes (namely the posterior inferior frontal gyrus (pIFG) and the posterior middle frontal gyrus (pMFG) and conceptual and semantic systems in the temporal and inferior parietal lobes. The pMFG is supposed to have a hierarchical lexical-syntactic structure that interconnects with the semantic area, whereas the pIFG would transform these representations into morphosyntactic sequences.

While these models differ in their specific neural emphasis, they collectively contribute to the understanding of the distributed and interconnected nature of neural networks involved in language processing. The accounts presented were developed on the basis of a large body of evidence reported in the literature. The next sections will review the most relevant findings, with a specific focus on morphosyntactic processing and thematic role mapping.

Neural correlates of morphosyntactic processing

According to the literature, morphosyntactic processing relies on a left-lateralized frontotemporal network. In general, two types of morphosyntactic processing have been investigated: nominal and verbal. This distinction is crucial since processing nouns or verbs has been associated with different brain regions (**Perani et al., 1999; Davis et al., 2004; Tyler et al., 2004**). Here, we summarize relevant contributions on the neural bases of this process; neuropsychological, neuroimaging, neuromodulation and electrophysiological studies will be considered (see **Tab.1**).



Author(s)	**Year**	**Method**	**Material and task**	**Sample (N)**	**Active brain regions**
Ni et al	2000	Neuroimaging	Auditory presentation of non anomalous sentences vs syntactically anomalous sentences (Tree can grew) or semantically anomalous sentences (Trees can eat)	14	l-IFG; l-MFG; l-STG
Indefrey et al	2001	Neuroimaging	Three syntactic conditions and two non-syntactic conditions: syntactic conditions: visual presentation of complex (embedded subject-relative and object-relative clause) pseudosentences in German with case and number marking errors. Non-syntactic conditions: pseudoword pronunciation and phonotactical judjgements	10	l-MFG
Moro et al	2001	Neuroimaging	Written pseudosentences with anomalies (phonotactic, morphosyntactic, syntactic)	11	bilateral IFG; cerebellar vermis
Miceli et al	2002	Neuroimaging	Written nouns. Three judjgements: grammatical feature task (masculine vs feminine), semantic task (animal vs artifact) and phonological task (discrimination of different sounds)	8	l-IFG; l-MFG

Newman et al	2003	Neuroimaging	Ungrammatical sentences (conjoined active and object relative): (i) noun-verb disagreement in number (ii) extraneous verb that cannot be incorporated in the thematic structure	13	l-IFG
Hernandez et al	2004	Neuroimaging	Gender decision with opaque vs transparent words in Spanish	9	l-IFG
Raettig et al	2011	Neuroimaging	Auditorily presented sentences with verb past-tense anomalies (eg *Mike has EAT the apple*)	15	STS
Quinones et al	2014	Neuroimaging	Grammaticality judjgement in visually presentend sentences. Ungrammatical sentences: defaul disagreement; unagreement (a phenomenon of disagreement present in Spanish) and person mismatch	21	l-IFG
Carreiras et al	2015	Neuroimaging	Number disagreement between pronoun and verb and between article and noun	32	l-IFG
Mancini et al	2017	Neuroimaging	Real sentences with person and number disagreement	25	l-IFG
Shapiro et al	2001	NIBS: TMS	Verbs and nouns production in short sentences: production of the singular or plural form of presented regular and nouns and production of the first singular or third plural person of regular verbs	14	l-MFG
Cappa et al	2002	NIBS: TMS	Verbs and nouns processing during object- and action-naming task	9	l-DLPFC
Cappelletti et al	2008	NIBS: TMS	Production of singular and plural forms of regular and irregular visually presented nouns; production of past-tense form of regular and irregular visually presented verbs	20	l-IFG; l-MFG
Lo Gerfo et al	2008	NIBS: TMS	Production of singular or plural form of a visually presented noun; production of first singular or third plural form of a visually presented verb	28	l-MFG

Acheson & Hagoort	2001	NIBS: TMS	Ambiguity detection task with sentences containing a temporarily word-class ambiguous word that could either be resolved as a noun or a verb depending on the sentence context	20	l-IFG; l-MFG
Carreiras et al	2002	NIBS: TMS	Grammaticality judjgement task with visually presented determiner-noun pairs	12	l-IFG
den Ouden et al	2008	VBLSM	Three tasks from the Northwestern Assessment of Verbs and Sentences (NAVS): ASPT, SPPT, SCT	71	left inferior FC
Kutas & Hillyard	1984	EEG	Sentence comprehension task: reading sentences and answering questions about their contents	nd	left inferior FC
Kutas & Federmeier	2000	EEG	Semantic memory use in sentence comprehension	review	left inferior FC
Friederici et al	1998	EEG	Sentence comprehension task: detecting correct or incorrect words at the end of a sentence	10	l-IFG
Hagoort et al	2003	EEG	Sentence comprehension task: detecting syntactic, semantic or syntactic and semantic violations in sentence-internal or sentence-final noun phrases	24	l-IFG
Wassenaar & Hagoort	2005	EEG	Sentence comprehension task: detecting syntactic violations	26	l-IFG

Tab.1 *Neuropsychological, neuroimaging, neuromodulation and electrophysiological evidence for a left frontal region involvement in morphosyntactic processing*

Processing of nouns and verbs

Several studies have investigated the neural underpinnings of morphosyntactic processing of nouns and verbs. **Moro and colleagues (2001)** focused on article-noun disagreement in pseudo-sentences, comparing grammaticality judgments of correct and incorrect sentences to lists of phonotactically legal and illegal words. They found greater activation for pseudo-sentences than for words lists in the bilateral inferior frontal gyrus (IFG) and the cerebellar vermis. In another study **(Ni et al., 2000)** morphosyntactic violations such as: "*Flowers can grew*" and semantic violations ("*Trees can eat*") were investigated. Greater activation for morphosyntactic violations was found in the left inferior frontal gyrus, whereas semantic violations increased hemodynamic response in left middle and superior frontal gyri (BA 44, 45, 46, 47, 6, 8), the bilateral postcentral gyrus, the right supramarginal gyrus and the head of the right caudate nucleus. Along the same lines, **Indefrey et al. (2001)** contrasted violations of verbal finiteness in pseudo-sentences with a phonological processing task; they reported stronger activity during morphosyntactic processing in the left MFG. **Newman et al. (2003)** investigated subject-verb agreement in sentences with a singular subject-plural verb inflection mismatch (eg "*The lady praises the sister and meet the artist in the night*") and in sentences with syntactic violations (eg "*The coach watched the poet and told the visitor took in the evening*"). They found different activation in the two tasks, revealing a greater activity in the left pars opercularis of the IFG for the sentences with extra verbs. Greater activation of the pars triangularis was associated with the noun-verb agreement. **Raettig et al. (2011)** investigated morphosyntactic processing in sentences that violated the verb-past tense agreement (e.g. "*Mike has EAT the apple*"). They suggested that the left posterior superior temporal sulcus (STS) is involved when processing these agreement violations. **Quiñones et al. (2014)** compared correct and incorrect sentences with subject–verb agreements or disagreements. They used a peculiar Spanish feature, the 'unagreement' phenomenon: a mismatch in subject–verb agreement that produces a grammatically legal sentence ("*Los pintores trajimos...*" [The painters (we) brought]). This construction requires the perceiver to shift the subject from 3rd-person to 1st-person, thus reinterpreting "*The painters*" as "*We painters*". They found greater activity in the pars triangularis and opercularsi of the l-IFG for both "normal" person mismatch and unagreement. Finally, a Voxel-Based-Lesion-Symptom-Mapping (VBLSM) study by **den Ouden and colleagues (2019)** investigated the neural basis of morphosyntactic processing (both at word and sentence level) in three tasks of the Northwestern Assessment of Verbs and Sentences (NAVS), providing evidence for the involvement of the left inferior frontal cortex, together with the left posterior superior temporal and angular gyrus. More specifically, the inferior frontal cortex correlated with morphosyntactic aspects, and the posterior temporo-parietal cortex with the assignment of thematic roles.

Processing gender and number

Noun gender has also been investigated; in a study by **Miceli and colleagues (2002)** a stronger response in the left middle and inferior frontal gyrus was found during a gender judgment task in which participants read written nouns varying for grammatical gender. Moreover, **Hernandez et al. (2004)** investigated grammatical gender assignment in opaque and transparent gender words. They reported decisions for opaque gender words to increase morphosyntactic processing demands in the l-IFG, as compared to transparent gender words. A study by **Carreiras et al. (2015)** contrasted number disagreement between pronoun and verb and between article and noun and found that the l-IFG is significantly activated by the processing of number agreement violations.

Neuromodulation studies also converge in outlining a role of the l-IFG and the l-MFG in aspects of morphosyntactic processing **(Shapiro et al., 2001; Lo Gerfo et al., 2008; Cappa et al., 2002; Acheson & Hagoort, 2010**). A contribution of the l-IFG to grammatical gender processing and number agreement was found in another TMS study by **Carreiras et al. (2010)**. Along these lines a study using repetitive Transcranial Magnetic Stimulation (rTMS) (**Cappelletti et al., 2008**) contrasted the processing of regular and irregular verbs with that of nouns. Distinct subparts of the left prefrontal cortex (l-MFG vs l-IFG and posterior portion of the l-MFG) were stimulated in two experiments. When stimulating the anterior l-MFG, response times increased for verbs but not for nouns. On the contrary, verb-specific effects were not found when stimulating the l-IFG and the posterior MTG.

Morphosyntactic processes have been investigated also in research using event-related potentials (ERPs). Thanks to their remarkable temporal resolution, these studies allow very accurate investigations on the timing of different morphosyntactic processes and their neural underpinnings. Therefore, the results of these electrophysiological studies must be considered notwithstanding the low spatial resolution, which does not allow precise identification of the neural source of the investigated phenomena.

Morphosyntactic agreement has been related to two main effects: the left anterior negativity (LAN) and the P600. The LAN component represents an early ERP effect, reported normally in the same time window as the N400 (i.e., 300-450 ms) (**Kutas and Hillyard, 1984; Kutas and Federmeier, 2009**), whereas the P600 has a later onset (the P600 component is divided in two subparts, an early onset-P600 (500-750 ms) and a late onset-P600 (750-1000 ms)). More than the temporal dynamics of these components, we are interested here in their topographical distribution (see **Fig. 1**, showing the cortical topography of the main ERP components that result from morphosyntactic violations. The greatest overlap between electrophysiological ERP evidence and neuroimaging studies on morphosyntactic processing is represented by the LAN, which is associated with left anterior frontal regions (**Molinaro et al., 2011; 2015; Mancini et al., 2011a,b**). This evidence is further supported by electrophysiological studies on patients with aphasia. For example, **Friederici et al (1998)** showed that patients affected by Broca's aphasia (typically resulting from damage to the left frontal lobe) did not exhibit the LAN when syntactic elements were manipulated, despite showing a preserved N400 for semantic manipulation. On the other hand, patients with Wernicke's aphasia (frequently resulting from temporo-parietal damage) showed the reverse profile. Thus, the LAN could be related to the anterior portion of the left frontal lobe, and the N400 component to the superior temporal lobe. A left anterior frontal topographical distribution of the LAN has been shown in several studies investigating different types of agreement in different languages (for a comprehensive review see Molinaro et al., 2011). Some studies also suggest a left frontal involvement in the P600 component. **Hagoort (2003)** and **Wassenaar et al (2004)** show that Broca's aphasics did not show a P600 in response to agreement violations, even though the effect correlated with the severity of syntactic comprehension impairment: patients with good syntactic comprehension showed a P600 effect, while patients with poor syntactic comprehension did not.

To summarize, ERP findings on the neural underpinnings of morphosyntactic processing converge in suggesting a left anterior frontal contribution, reflected by a LAN component.

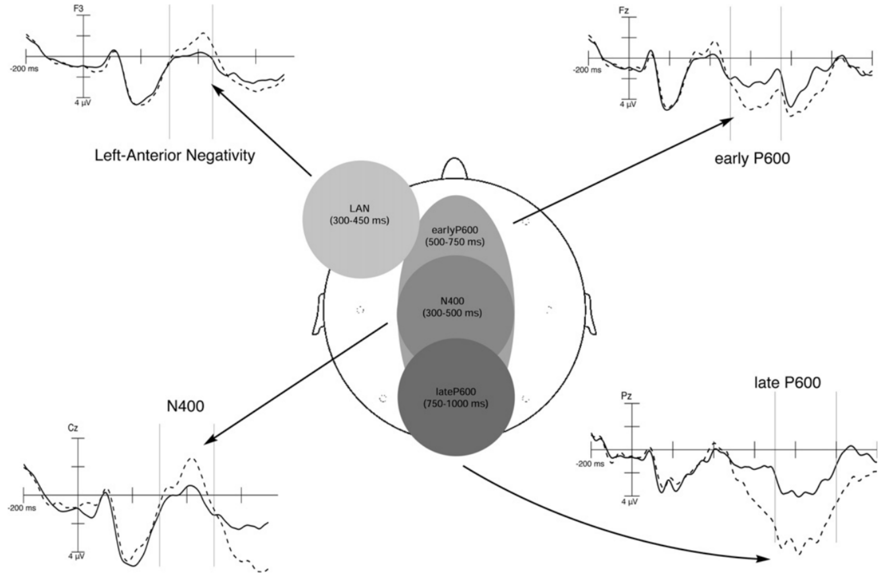
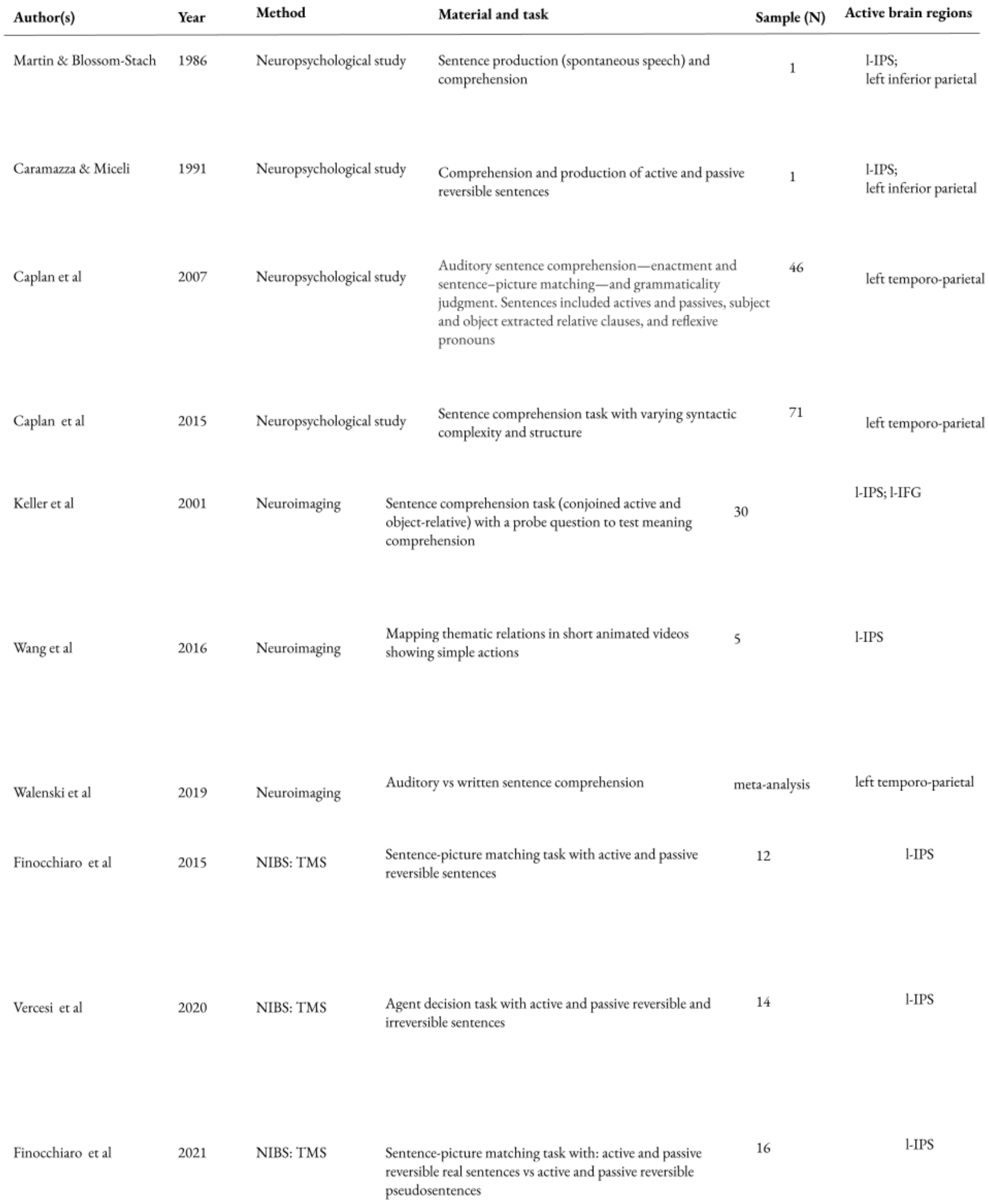


Fig. 1 *Main ERP components affected by agreement manipulations and their topographical distribution on the cortex (taken from Molinaro et al., 2011)*

Neural correlates of thematic role assignment

The literature suggests that the assignment of thematic roles relies on a left temporal and parietal network (see **Tab. 2**). Early anecdotal evidence comes from lesion studies in the nineties. **Martin & Blossom-Stach (1986)** and **Caramazza & Miceli (1991)** reported on two aphasic patients with a left parietal lesion, who were unable to comprehend and produce active and (even more) passive reversible sentences due to the inability to map syntactic roles onto thematic roles and vice versa. **Thothathiri et al (2012)** investigated the neural basis of sentence processing through voxel-based-lesion-symptom-mapping (VLSM) (**Bates et al., 2003**). They found a correlation between a left temporo-parietal damage and poor comprehension of reversible sentences, mostly when including thematic role inversions. These results are in line with other VLSM studies (see **Dronkers et al., 2004** or **Bates et al., 2003**).Other studies on aphasic patients found a correlation between temporo-parietal damage and the comprehension of sentences of varying syntactic complexity (**Caplan et al., 2007; Caplan et al., 2015**). Evidence from these studies suggests that temporo-parietal areas are involved in the comprehension of reversible sentences. Additional results were provided by a VLSM study (**Rogalsky et al., 2018)** of the comprehension of canonical and non-canonical reversible sentences in patients with chronic focal cerebral damage. Results showed that the left posterior superior temporal and inferior parietal regions overlapped with the worst performance on passive reversible sentences. . Another VLSM study (**Magnusdottír et al., 2013**) provided similar evidence of a left temporo-parieto-occipital contribution to thematic role mapping in a sentence-picture matching task with sentences of varying syntactic complexity. An involvement of the left posterior parietal cortex in the assignment of thematic roles was also shown by **den Ouden et al. (2019)**.

A link between thematic role mapping and left temporo-parietal areas has been shown also by neuroimaging studies. An fMRI project by **Keller et al. (2001)** investigated sentence comprehension through a task that required participants to read a sentence and answer a probe question on its content. Results showed greater activation in left temporal and left inferior parietal regions. **Richardson et al. (2010)** explored the relationship between semantic reversibility and sentence processing over a range of syntactic structures. When contrasting reversible and irreversible sentences, they found greater activation in a lateral portion of the left posterior-superior temporal gyrus and in an inferior parietal region. **Wang et al. (2016)** used machine-learning classifiers trained on fMRI data acquired while participants watched short videos showing agent-verb-theme propositions. Classifiers were able to identify thematic roles from the activation patterns in several areas, including the left inferior parietal sulcus. Moreover, in a neuromodulation study (**Finocchiaro et al., 2015)** the posterior portion of the left intraparietal sulcus (l-IPS) responded to rTMS during a sentence-picture matching task on reversible passive sentences. These findings were replicated in two additional neuromodulation studies (**Finocchiaro et al., 2021**; **Vercesi et al., 2020**).



Author(s)	Year	Method	Material and task	Sample (N)	Active brain regions
Martin & Blossom-Stach	1986	Neuropsychological study	Sentence production (spontaneous speech) and comprehension	1	l-IPS; left inferior parietal
Caramazza & Miceli	1991	Neuropsychological study	Comprehension and production of active and passive reversible sentences	1	l-IPS; left inferior parietal
Caplan et al	2007	Neuropsychological study	Auditory sentence comprehension—enactment and sentence–picture matching—and grammaticality judgment. Sentences included actives and passives, subject and object extracted relative clauses, and reflexive pronouns	46	left temporo-parietal
Caplan et al	2015	Neuropsychological study	Sentence comprehension task with varying syntactic complexity and structure	71	left temporo-parietal
Keller et al	2001	Neuroimaging	Sentence comprehension task (conjoined active and object-relative) with a probe question to test meaning comprehension	30	l-IPS; l-IFG
Wang et al	2016	Neuroimaging	Mapping thematic relations in short animated videos showing simple actions	5	l-IPS
Walenski et al	2019	Neuroimaging	Auditory vs written sentence comprehension	meta-analysis	left temporo-parietal
Finocchiaro et al	2015	NIBS: TMS	Sentence-picture matching task with active and passive reversible sentences	12	l-IPS
Vercesi et al	2020	NIBS: TMS	Agent decision task with active and passive reversible and irreversible sentences	14	l-IPS
Finocchiaro et al	2021	NIBS: TMS	Sentence-picture matching task with: active and passive reversible real sentences vs active and passive reversible pseudosentences	16	l-IPS

Tab.2 *Summary of studies providing evidence for the correlation between left temporo-parietal regions and the assignment of thematic roles*

Neural correlates of nonlinguistic aspects of sentence processing

In addition to specifically linguistic processes and representations, sentence comprehension is supported by domain-general processes. Many lesion studies suggest that the visuospatial sketchpad of WM is represented in a ventral path that connects the occipital and the temporal cortex, responsible for object recognition, and a dorsal path linking occipital and parietal cortices, responsible for the computation of spatial relations (**Muller & Knight, 2006**). As for the phonological loop, a distinction between a phonological short-term store (associated with the inferior parietal cortex) and an articulatory rehearsal component (relying on Broca's area and the supplementary motor association area (SMA)) has been proposed (**Vallar & Papagno, 2002**). Relevant insights came from the investigations of the case of patient PV reported by Papagno and Vallar's works (**Vallar & Papagno, 1986**; **Baddeley et al., 1998**; **Papagno & Vallar, 1992**; **Papagno & Cecchetto, 2019** and many others). This patient suffered from bilateral damage to the parietal lobes, leading to severe deficit in verbal STM. Despite having intact LTM and general cognitive abilities, patient PV showed significant impairments in retaining and recalling verbal information over short intervals. The performance notably declined with increasing verbal load, suggesting a specific deficit in STM capacity. This case suggested a critical role of the parietal lobes in supporting verbal STM processes, reinforcing the view that these regions play a crucial role in the temporary storage and manipulation of verbal information. Additionally, while some studies suggest that the prefrontal cortex (PFC) has a role in supporting cognitive control, integrating across-domain information, selecting stimuli and information monitoring **(Miller, 2000; Baldo & Shimamura, 2002**), others suggest that WM is supported by the ventral and dorsal lateral prefrontal cortex (see **Muller & Knight, 2006**).

The neurofunctional correlates of verbal WM processes were also investigated in neuroimaging studies. It has been hypothesized that Broca's area constitutes the neural substrate of WM processes involved in highly demanding sentence processing (**Caplan & Waters, 1999; Just and Carpenter, 1992**). In this regard, **Rogalsky, Matchin & Hickok (2008)** propose that the contribution of Broca's area to sentence processing stems from its link to verbal WM. In their study on sentence comprehension, the issue was investigated by varying the degree of syntactic complexity in three conditions: (i) during a baseline (no secondary task), (ii) during concurrent speech articulation and (iii) during a concurrent finger-tapping task. Results were in line with the hypothesis that Broca's area is involved in the articulatory rehearsal mechanism of verbal WM, as a complexity effect was present in all the conditions.

Finally, neurostimulation evidence provided useful insights. **Romero Lauro et al. (2010)** administered 1 Hz rTMS over the pars opercularis of the l-IFG (BA44) -which is supposed to be the neural correlate of the process of rehearsal- and over the BA40 in the parietal lobe (the neural correlate of the short-term store) while participants were engaged in a sentence comprehension task with different STM demands. A greater number of errors was found in syntactically complex sentences when stimulating the l-IFG. Reduced accuracy was found on longer but syntactically easier sentences when the BA40 was stimulated. Based on these findings, **Giustolisi et al. (2018)** applied anodal transcranial direct current stimulation (atDCS) over the l-IFG while executing an auditory comprehension task and found significant improvements in performance accuracy, regardless of load on STM. The same task was used in another complementary study (**Vergallito et al., 2020**) in which the target of atDCS was the left inferior parietal cortex. As a result of stimulation, a decrease in performance accuracy was reported, showing a critical involvement of both the l-IFG and the left inferior parietal cortex in language comprehension, despite the opposite direction of the effect found in the two connected studies.

Summary and open issues

Available evidence suggests that key aspects of sentence comprehension (including morphosyntactic and thematic role processing) are sustained by a large left fronto-temporo-parietal network. The assignment of thematic roles seems to be related to a temporo-parietal region (including the l-IPS), whereas the processing of morphosyntactic information would correlate to a prefrontal region that includes the l-IFG and the l-MFG. However, this picture is not as straightforward as it might seem. In the first place, findings come from studies that investigated the two processes separately, meaning that there is no systematic investigation of both processes via the same task. Moreover, morphosyntactic processing has been investigated mostly through tasks requiring *anomaly detection* in simple syntactic contexts (see for example **Coulson et al., 1998**; **Hasting & Kotz, 2008; Molinaro**

et al., 2011; Molinaro et al., 2008), whereas thematic role assignment has been studied mostly in the context of *sentence comprehension* tasks including structurally more complex, reversible sentences, both on patients (**Caramazza and Zurif 1976; Caplan and Futter 1986; Grodzinsky 2000; Love et al. 2008; Thompson and Choy 2009)** and on unimpaired participants (for reviews see **Meyer and Friederici 2015; Rodd et al. 2015; Martin et al. 2015; Walenski et al. 2019**). Given that in sentence comprehension the assignment of thematic roles presupposes the processing of morphosyntactic information, the relationships between the two sets of processes and their neural substrates remain unclear. Experimental tasks that test thematic role assignment entail the computation of morphosyntactic aspects to be performed correctly. Disentangling between the two processes when exploring their neural correlates requires an experimental manipulation such that the relationship between each single process and its neural underpinnings can be explored separately.

In addition, the complex anatomical organization of the targeted prefrontal regions makes it challenging to posit a specific contribution to language processing for each area. The development of probabilistic cytoarchitectonic maps has raised some important issues on the structure-function correspondence (**Amunts and Zilles, 2001**, **Zilles et al., 2002**). This approach has significantly contributed to the understanding of the structural and functional organization of Broca's area. **Amunts and Zilles (1999; 2001)** challenged the view that this region has a homogeneous structure. Rather, Broca's area would be heterogeneously organized and it would comprise different subfields with distinct cellular compositions and connectivity patterns. The authors provided evidence that the size of Broca's area (BA44 and 45) and its macroanatomical landmarks and surrounding areas significantly differ among the cytoarchitectonic maps. These interindividual differences in the cytoarchitecture were greater than cytoarchitectonic variability between these regions in individual brains. These findings make it difficult to investigate the functional role of this region in sentence processing.

Furthermore, several issues about the anatomy and structural organization of the MFG (for a detailed description of the anatomy of this region see **Petrides & Pandya, 2012**) have garnered attention in neuroscientific research. Anatomically, the MFG is part of the prefrontal cortex. However, specific delineation of functional subdivisions within the MFG remains a topic of debate, with some studies suggesting functional heterogeneity across its anterior-posterior axis (**Badre & D'Esposito, 2009**). Moreover, individual differences in MFG morphology and its association with cognitive abilities have been explored (**Burgaleta et al., 2014**), raising questions about the region's structural variability and its impact on cognitive processing. These issues underscore the complexity in understanding the precise anatomical and organizational features of the left middle frontal gyrus and its functional implications in cognitive processes. These issues make it particularly difficult to identify a precise target in this region. So far, this has not been addressed, even in studies that could benefit from a rigorous anatomical localization (such as TMS studies, in which spatial accuracy is fundamental).

In addition, neuroimaging findings are often interpreted as compatible with the view that the left fronto-temporo-parietal network is a cluster of multiple-demand regions differentially activated as a function of either language-specific or domain-general aspects (see **Santi & Grodzinski, 2007**). For these reasons, a systematic investigation of the language-related contribution of this network, with a focus on the neural basis of thematic role mapping and morphosyntactic processing would help update neurofunctional hypotheses on the role of the left hemisphere in sentence comprehension.

The project

The present work investigates the role of prefrontal (IFG and MFG) and posterior parietal (l-IPS) regions, that have emerged in the literature as possibly critical for the elaboration of morphosyntax and thematic roles. TMS was selectively administered on these regions in association with a sentence-picture matching task. This allowed both processes to be included in the same experiment to highlight and disentangle the role of the target regions. The objectives of the study can be thus summarized:

1. Demonstrating that the stimulated sites play a causal role in thematic role assignment and in morphosyntactic processing
2. Establishing if the stimulated regions play the same or a distinct role in the two processes

MATERIAL AND METHODS

Participants

23 healthy volunteers (13 F, 10 M) were recruited for the experiment. They were all students or researchers at the University of Trento. The inclusion and exclusion criteria were the following:

Inclusion criteria:

- right-handedness assessed with the Edinburgh inventory (**Oldfield, 1971**)
- Italian as the native language of the participant

- unimpaired or corrected-to-normal vision

Exclusion criteria:

- incapacity to express an informed consent to the experimental procedures
- counterindications to rTMS including:
  - prior history of neurological conditions, seizures, or psychiatric symptoms
  - ongoing treatment with psychoactive or other seizure threshold-lowering drugs (as in **Rossi et al., 2021**)
- all other data collected by the TMS safety questionnaire

Three participants could not be included for safety concerns (their resting motor thresholds were too high to allow safe administration of TMS on the frontal region). Two participants started the experiment but eventually were excluded because they responded at chance level. The final sample included 18 participants (12 F, 6 M) (mean age=24.8 SD=2.57).

Experimental task

Participants sat in front of a computer screen, at an approximately 1 meter distance. They completed a sentence-picture matching task in which they had to decide by button press whether a written sentence and a picture matched. The timeline of the task is shown in **Fig.2**. Immediately after a fixation cross, the trial appeared on the screen and remained for 900 ms. This time window was established after a pilot study designed to test different parameters (for a description of the pilot study and its results see Chapter 2). As soon as the trial disappeared from the screen participants were allowed to respond by pressing with the index finger one of two different buttons. The index finger of the left hand was used for the key associated with matching trials and that of the right hand for the key corresponding to mismatching trials. This configuration was kept fixed for all participants in all sessions. Response times (RT) and performance accuracy (ACC) were recorded starting from the appearance of the response screen. The experimental task consisted of four blocks of 32 stimuli each. Within each block stimuli were randomized so that no more than two consecutive trials of the same type could occur. Block order was randomized within each session and between sessions. Before starting the experiment, a practice session was administered to allow participants to familiarize with the task. It consisted of a short version of the task with trials like those used in the experimental task. Overall, the practice session included 40 trials divided into two blocks of 20 trials each.

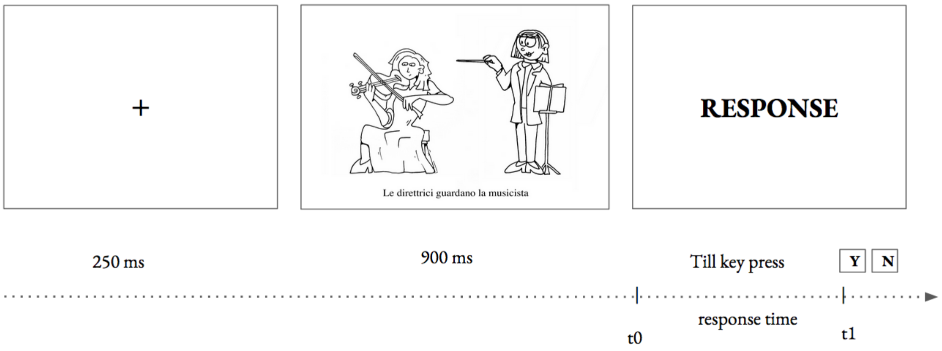


Fig.2 *Timeline of the experimental task*

Experimental stimuli: the sentences

The dataset included 128 reversible sentences (64 actives and 64 passives). Within both active and passive sentences, 32 were assigned to the morphosyntactic condition (MS) and 32 to the thematic role condition (TR). Within the MS condition, half of the sentences contained morphosyntactic alternatives on the agent and half on the theme. See **Tab.3** for examples of each type of sentence.

Given the presence of the auxiliary verb and of the *by*-phrase, passive sentences were systematically 2 or 3 syllables longer than active ones (mean length (in syllables) of active sentences = 12.4; mean length of passive sentences = 14.7). Both feminine and masculine nouns were used. In all the sentences, the two constituents could be both male or both female.

CONDITION	**EXAMPLE**
ACTIVE_TR_1 (Sg)	*Il pugile ascolta lo sciatore* (The boxer listens to the skier)
ACTIVE_TR_2 (Pl)	*Le bariste baciano le chitarriste* (The [BARTENDERS] kiss the [GUITARISTS])
ACTIVE_MS_1 (Pl)	*Le direttrici guardano la musicista* (The [DIRECTORS] watch the [MUSICIAN])

ACTIVE_MS_2 (Pl)	*I ballerini applaudono il tennista* (The dancers applaud the tennis player)
PASSIVE_TR_1 (Sg)	*Il pittore è fotografato dal barista* (The painter is photographed by the bartender)
PASSIVE_TR_2 (Pl)	*Le cantanti salutano le cameriere* (The [SINGERS] greet the [WAITRESSES])
PASSIVE_MS_1 (Pl)	*Il vigile è indicato dai falegnami* (The traffic policeman is pointed to by the the carpenters)
PASSIVE_MS_2 (Pl)	*I pirati sono toccati dal cavaliere* (The pirates are touched by the knight)

Tab.3 *Examples of the sentences for each condition. Legend: Sg=singular; Pl=plural. Words in [] correspond to the feminine*

Experimental stimuli: the pictures

The dataset included 128 black-and-white pictures, drawn by a professional illustrator.

Experimental stimuli: the trials

A trial consisted of a picture and a sentence that appeared on the screen at the same time. In half of the trials the meaning of the sentence matched the content of the picture (matching trials), whereas in the other half there was a sentence-picture mismatch (mismatching trials). Mismatching trials contained either a morphosyntactic mismatch or a reversal of thematic roles. Morphosyntactic mismatches could occur on the agent or on the theme. In the stimuli that tapped thematic role reversal, agent and theme were both singular or both plural. Half of the mismatching trials in the TR condition contained singular agent and theme and the other half contained plural agent and theme.

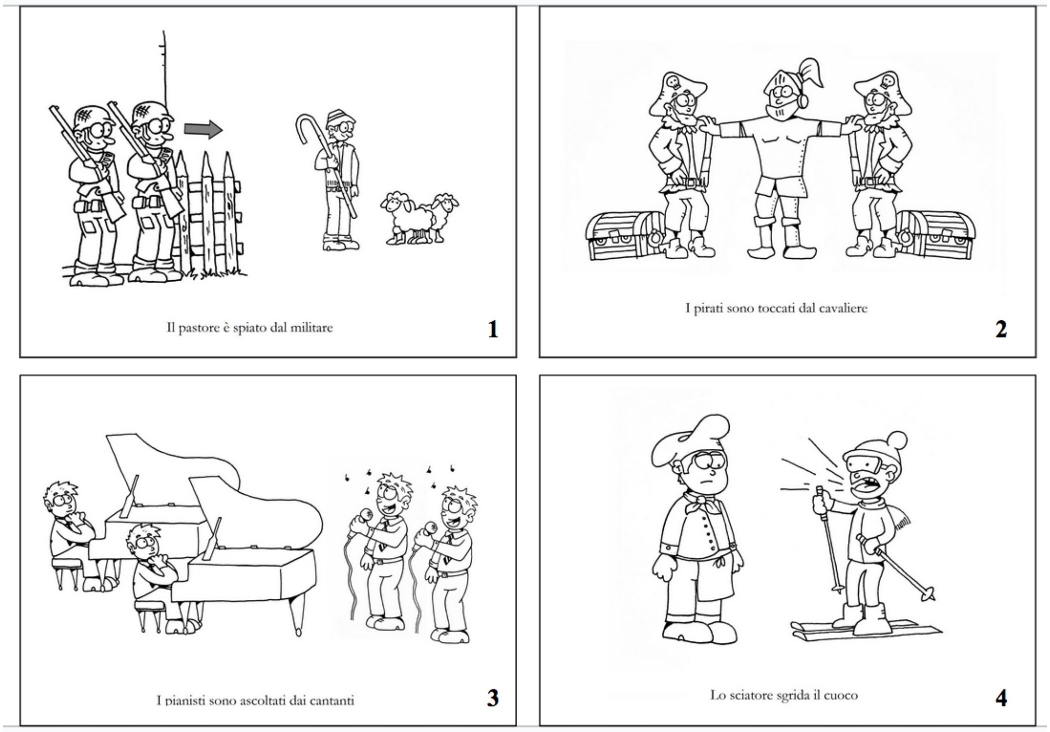


Fig.3 *Examples of matching and mismatching trials for each condition (1=mismatching (MS); 2=matching (MS); 3=mismatching (TR); 4=matching (TR))*

Structure of the experiment

The experiment was divided in two sessions, spaced by one week. Every session took approximately two hours. In each session one or two target regions were stimulated. A no-stimulation control condition (sham) was also included in each session. Several elements were counterbalanced across participants: (i) the number of regions stimulated in each session (either one region in the first session and two regions in the second session or the reverse), (ii) the order in which target regions were stimulated and (iii) the time window during which the sham stimulation was administered in each session (before stimulating the first region or after stimulating the first or the second region). Six options were created to determine how many regions and in which order to stimulate. Within each session one or two regions could be stimulated. For example, session 1 could include the stimulation of the IPS and in session 2 the IFG and the MFG could be stimulated. The order of stimulation could vary as shown in **Tab.4**

	SESS 1	**SESS 2**
Option 1	IFG, IPS	MFG
Option 2	MFG, IFG	IPS
Option 3	IPS, MFG	IFG
Option 4	IPS	IFG, MFG
Option 5	MFG	IPS, IFG
Option 6	IFG	MFG, IPS

Tab.4 *Counterbalanced conditions for the target regions and the order of stimulation for the sessions*

A second counterbalancing procedure determined when the sham stimulation was administered in each session. **Tab.5** shows the counterbalancing options:

	Option 1	*Option 2*	*Option 3*
Sess 1	real, sham	real, sham	real, sham
Sess 2	sham, real, real	real, sham, real	real, real, sham

	Option 1	*Option 2*	*Option 3*
Sess 1	sham, real	sham, real	sham, real
Sess 2	sham, real, real	real, real, sham	real, sham, real

Tab.5 *Counterbalancing of the order of administration of real-TMS and sham-TMS between and within sessions*

Participants were randomly assigned to each counterbalanced condition, so that each option was applied to three participants.

Experimental design and statistical analysis

Experimental design

Our experimental design is a 2x2x2x4 within-participants design. It includes four within-participants factors:

- DIATHESIS (*active* vs *passive*)
- TRIAL (*matching* vs *mismatching*)
- CONDITION (*MS* vs *TR*)
- REGION (*IFG* vs *IPS* vs *MFG* vs *sham*)

The main analyses only considered the factors REGION and CONDITION. The focus of the analyses was put on TMS- and condition-related effects. The relevant expected effects were those depending on the stimulation targets and on the linguistic processes. The effects and interactions between the other factors were not relevant for the experimental hypotheses.

Other analyses included the factor DIATHESIS and the factor MS MISMATCH (*1* vs *2*). This factor pertains only to the mismatching trials in the MS condition, that could contain morphosyntactic mismatches on the first (1) or on the second (2) constituent.

In the main analyses the effect of TMS as compared to the sham condition was explored for each stimulation site in the two levels of CONDITION. First, we compared the effect of TMS vs sham on all the stimulation sites in all (matching and mismatching) trials for the MS and the TR condition. Second, we explored the effect of TMS on all the stimulated regions only on mismatching trials for the MS and the TR condition. Finally, we analyzed effects of stimulation on all the target regions for the MS and the TR condition separately. This analysis was motivated by the a priori hypothesis that stimulation over the prefrontal targets (IFG and MFG) would have affected performance on MS, while stimulating the IPS would have influenced performance on TR. A control analysis has been done to see whether the order of the stimulated targets influenced the TMS-related effects. This was done to exclude any possible carry over effects of TMS over the first stimulated regions on the subsequent stimulation of another region. This analysis considered the factor TMS ORDER (*opt 1* vs *opt 2* vs *opt 3* vs *opt 4* vs *opt 5* vs *opt 6*), which had 6 levels, one for each sequence option randomly assigned to participants.

Additional analyses investigated: (i) the effect of real-TMS vs sham for each stimulation site in the two levels of CONDITION and of DIATHESIS and (ii) the effect of stimulation on mismatching trials in the MS condition, depending on whether the mismatch occurred on the first or on the second constituent.

Statistical analysis

Statistical analyses were performed using Linear Mixed Effects Model (LMMs) (for a complete guide to these models see **Raudenbush & Bryk, 2002** or **Singer & Willett, 2003**). Analyses were performed using the *General analyses for linear models (GAMLj)* jamovi (https://www.jamovi.org/) module (Gallucci M., 2019; retrieved from https://gamlj.github.io/). Response times were analyzed through the "*mixed effects*" module of the GAMLj, whereas performance accuracy was analyzed using a mixed logistic model of the same module ("*generalized mixed model*"). This model is basically a logistic regression that estimates the probability of the distribution of a dichotomous variable in relation to a set of predictors (see **Wright, 1995** for an exhaustive overview). Both models had participants as the clustering variable and the intercept across subjects as random-effect.

TMS settings

Stimulation protocol

An offline stimulation protocol was adopted. TMS was delivered before the experimental task. Low-frequency rTMS (LF-TMS) was administered at 1Hz. The LF-TMS consisted of a train of 600 stimuli at the rate of 1 stimulus per second (1Hz) (10 minutes). Stimulation intensity corresponded to 90% of the individual visible resting motor threshold (RMT). The RMT was calculated as the lowest stimulation intensity applied over the primary motor cortex which produces more than fifty percent of motor evoked potentials (MEPs) of amplitude > 50 microvolts out of ten stimuli in the right 1DI muscle. The mean intensity corresponded to 35% (SD=3.58). The most recent international guidelines (**Rossi et al., 2021**) report LF-TMS to be non-invasive and safe.

Each participant received real stimulation and sham stimulation for each target region. Real stimulation was provided via a MC-B70 figure-of-eight coil and a MagPro X100 stimulator (MagVenture). Sham stimulation was administered by placing the coil perpendicular to the scalp.

Target regions

Three target regions were identified based on the relevant literature: a site on the l-IFG corresponding to the pars triangularis, a site on the l-MFG and the posterior portion of the l-IPS (see **Fig. 4**).

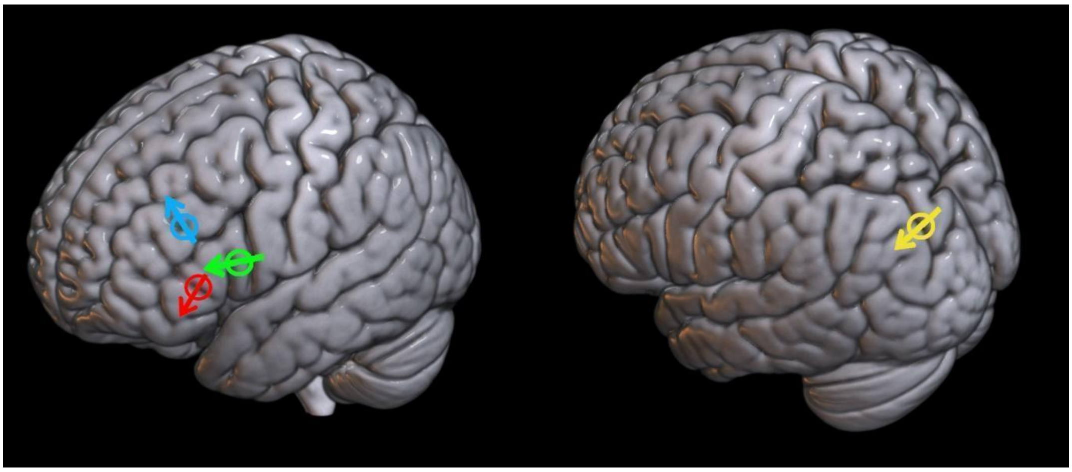


Fig. 4 *Localization and direction of the current of the stimulation targets on MNI space (template: spm152). Legenda: blue=MFG; red=pars opercularis [****NOTE: pars opercuoaris was only used as an anatomical reference to identify the MFG and <u>not</u> used as a stimulation target****]; green=pars triangularis ; yellow=IPS).*

MRI reconstruction and neuronavigation

Structural MRI was available for each participant. All the structural images (T1-weighted sequences) were acquired through a Bruker 4T scanner. The MRI images of each individual were used as a basis for a 3D reconstruction of that individual's brain. Each stimulation target was identified on these 3D reconstructions based on macroanatomical landmarks. The IFG spot was placed in the middle of the pars triangularis of the IFG. The MFG spot was identified as a spot equidistant from pars opercularis and pars triangularis, 1 cm dorsal to the IFS. The IPS spot was localized as the point between the middle and the posterior third of the length of the IPS (as in **Finocchiaro et al., 2015**; **Vercesi et al., 2020**; **Finocchiaro et al., 2021**).

Before stimulating each target, the TMS coil, the participants' head and the 3D reconstruction were co-registered in space via the Softaxic Neuronavigation System using a Polaris Spectra camera. This allows spatially accurate administration of TMS and online checking of the coil position during stimulation.

RESULTS

Data treatment

Trials corresponding to response times below 300 ms and above 4 s were eliminated. We considered these trials as unreliable responses for being either too fast or too slow. The excluded trials amounted to 8% of the entire dataset.

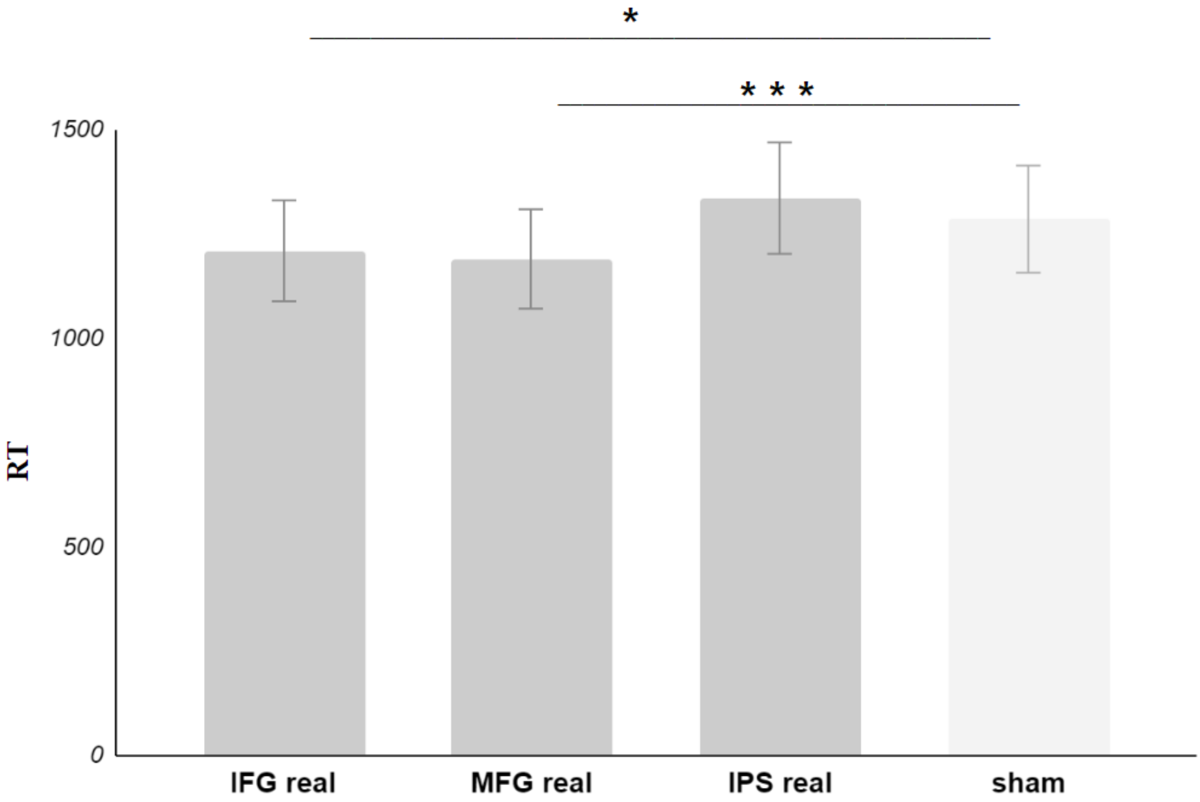
Analyses

As mentioned above, the experimental task was designed to investigate the neural underpinnings of two linguistic processes in the same sentential context, on the assumption that each process can be associated with its neural correlates based on the response to stimulation. For example, if IFG is implicated in morphosyntactic processes, TMS on the IFG should affect responses in the MS condition and not (or to a lesser extent) in the TR condition. Conversely, the reverse picture should emerge if IPS is implicated in TR. In mismatching trials of the MS condition, correct responses require the subject to recognize the number contrast between one of the characters represented in the pictured stimulus and the corresponding noun in the written sentence. In mismatching trials of the TR condition, on the other hand, the participant must understand that the agent and theme role represented in the picture are the reverse to those represented in the written sentence. For this reason, we decided to analyze the mismatching trials separately when investigating the link between the condition and the stimulated regions.

The model includes the factors REGION and CONDITION.

Statistics

Results showed a significant main effect of REGION ($F(1,3)=13.50$, $p < 0.001$) and CONDITION ($F(1,1)=13.04$, $p < 0.001$). No interaction effect was found ($F(1,1)=1.39$, $p = 0.243$). Bonferroni's Post Hoc Comparisons were run to further explore the effects. Significant differences between TMS and sham were observed on the IFG ($t(1, 5101)= -3.084$, $p = 0.012$) and on the MFG ($t(1, 5101)= -3.809$, $p < 0.001$), whereas the difference fell just short of significance on the IPS ($t(1, 5102)=2.612$, $p = 0.054$). As seen in **Graph 1**, stimulation on the IFG and the MFG significantly reduced RT compared to the sham condition. Descriptives are shown in **Tab.6**.

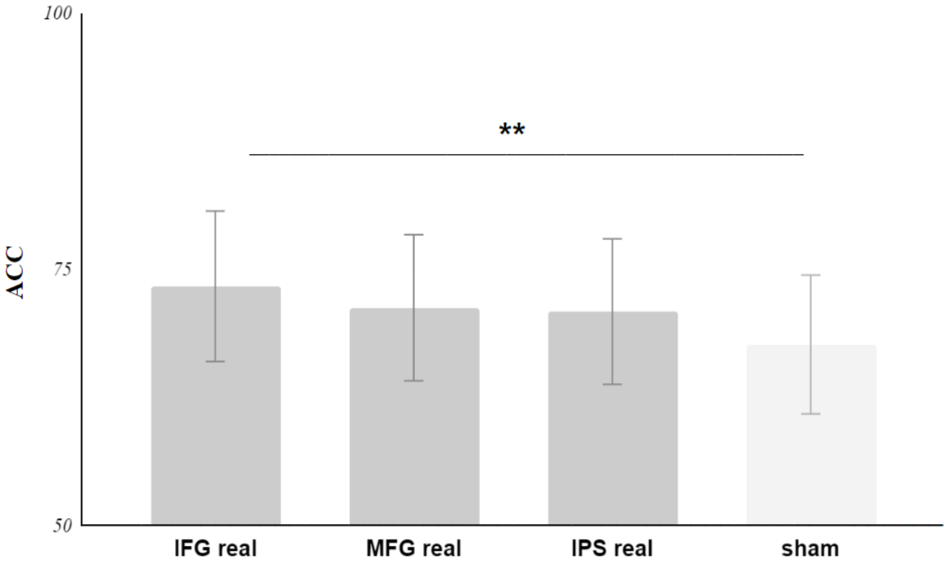


Graph 1 *Bar plot of the mean RT performance in the frontal targets (IFG real and MFG real) and the parietal target (IPS real) and in the sham condition for mismatching trials. Asterisks denote significant differences*

		IFG real	MFG real	IPS real	sham
RT	*mean*	1210.72	1191.16	1336.96	1286.67
	std	637	663	794	704
ACC	*mean*	73.39	71.29	70.92	67.71
	std	4.2	5.3	5.4	6.8

Tab.6 *Descriptive statistics (mean RT and ACC and standard deviation) for each stimulated target and the sham*

To analyze the effect of TMS on performance accuracy a logistic model was adopted with the same factors as above. Results show a main effect of REGION (X^2(1)=7.82, p = 0.005) and of CONDITION (X^2(3)=12.18, p = 0.007) and no interaction effect (X^2(3)=3.27, p = 0.351). Bonferroni's Post Hoc Comparisons showed significant differences only on the IFG (z=3.270, p = 0.006), but not on the MFG (z=2.018, p = 0.262) and on the IPS (z=1.829, p = 0.404). TMS significantly enhanced performance accuracy only on the IFG, while leaving performance on the other regions unaffected (**Graph 2**)



Graph 2 *Bar plot of the mean ACC performance in the frontal targets (IFG real and MFG real) and the parietal target (IPS real) and in the sham condition for mismatching trials*

To analyze any potential after-effect of administering TMS on more than one target per session, a control analysis was performed including the factors REGION, CONDITION and TMS ORDER as fixed-effects and the intercepts across participants and TMS ORDER as random-effects. Both on ACC and RTs, the model did not show any relevant effect related to the order of stimulation (see **Tab. 7**):

	CONDITION	**REGION**	**TIMS ORDER**	**COND*REG**	**COND*ORD**	**COND*REG*ORD**
RT	$F(1,3)=12.85$, $p < 0.001$	$F(1,3)=20.3$, $p < 0.001$	$F(1,5)=4.36$, $p = 0.095$	$F(1,3)=1.84$, $p = 0.137$	$F(1,5)=1.75$, $p = 0.122$	$F(1,15)=0.53$, $p = 0.921$
ACC	$X^2(1)=8.82$, $p = 0.002$	$X^2(3)=11.8$, $p = 0.001$	$X^2(5)=3.78$, $p = 0.067$	$X^2(1)=2.32$, $p = 0.245$	$X^2(5)=3.91$, $p = 0.301$	$X^2(15)=1.09$, $p = 0.765$

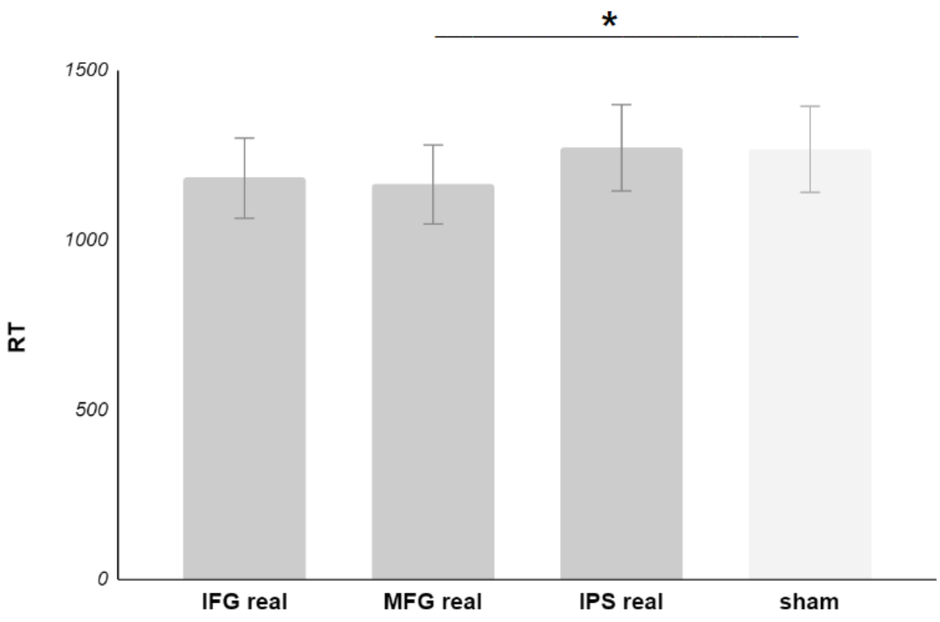
Tab. 7 *Summary of effects of the models for RT and ACC*

These results showed that there was no significant interaction between the behavioral effects of stimulation administered on two different regions in the time window of the same experimental session, independently of which specific regions were stimulated and in which order.

To explore the effect of TMS on each stimulation target we analyzed the two conditions (MS and TR) separately. The model included the factor REGION.

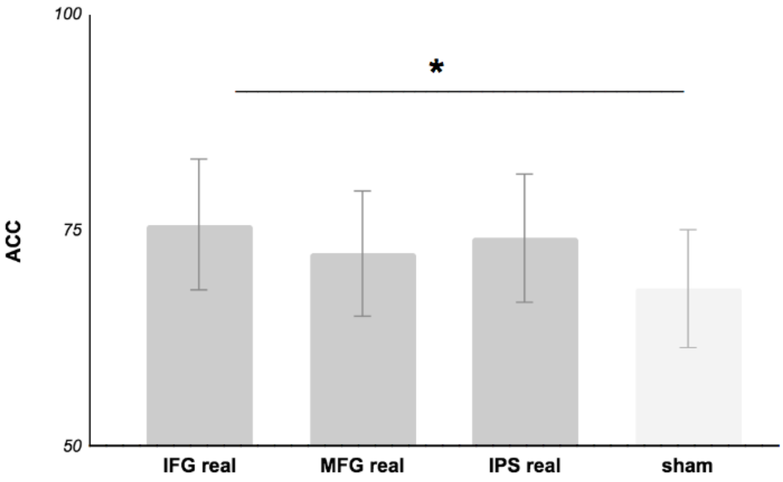
MS condition

Results showed a main effect of REGION ($F(1,3)=5.26$, $p = 0.001$) that was further investigated by means of Bonferroni's Post Hoc Comparisons. The TMS vs sham contrast reached significance only on the MFG ($t(1, 2547)= -3.087$, $p = 0.012$), on which TMS decreased RT (see **Graph 3**). No significant effect was seen on the IFG ($t(1, 2548)= - 2.420$), $p = 0.094$) and on the IPS ($t(1, 2548)=0.483$, $p = 1.000$).



Graph 3 *Barplot of the mean RT performance in all the stimulated regions and the sham only on MS*

The effect of TMS on performance accuracy was also explored. Results showed a significant effect of REGION ($X^2(3)=12.3$, $p = 0.006$). Bonferroni's Post Hoc Comparisons showed that TMS significantly improved accuracy only on the IFG ($z=3.115$, $p = 0.011$), (**Graph 4**). No significant differences in accuracy were found on the MFG ($z=0.147$, $p = 0.590$) and on the IPS ($z=2.425$, $p = 0.092$).



Graph 4 *Bar plot of the mean ACC performance in all the stimulated regions and the sham on MS condition*

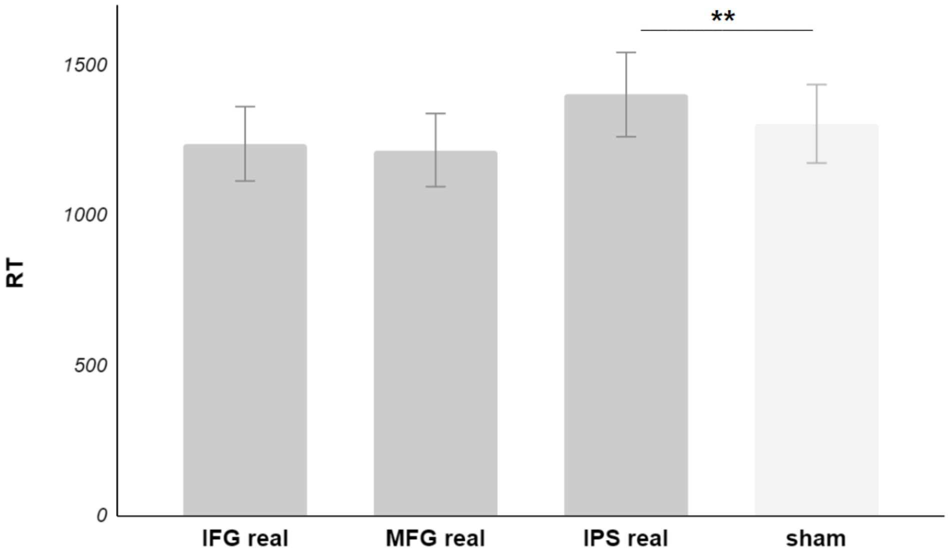
Tab.8 shows descriptives:

MS		IFG real	MFG real	IPS real	sham
RT	*mean*	1182.84	1164.46	1272.22	1268.17
	std	630	668	762	678
ACC	*mean*	75.72	72.37	74.15	68.3
	std	4.2	4.4	4.3	4.6

Tab.8 *Descriptive statistics (mean RT and ACC and standard deviation) for each stimulated target and the sham in the MS condition*

TR condition

The same analyses were performed on the TR condition. Results showed a main effect of REGION only for RT ($F(1,3)=9.56$, $p < 0.001$) and not for ACC ($X^2(3)=1.13$, $p = 0.334$). Post hoc comparisons showed that TMS yielded a significant RT increase only on the IPS ($t(1, 2539)=3.175$, $p = 0.009$) (**Graph 5**)



Graph 5 *Bar plot of the mean RT performance on all the stimulated regions and the sham on the TR condition only*

Descriptives are shown in **Tab.9**:

TR		IFG real	MFG real	IPS real	sham
RT	*mean*	1238.6	1217.79	1402.34	1305.3
	std	643	658	821	728
ACC	*mean*	71.06	70.21	67.67	67.12
	std	4.5	4.5	4.6	4.7

Tab.9 *Descriptive statistics (mean RT and ACC and standard deviation) for each stimulated target and the sham in the TR condition*

Effects of diathesis and ms mismatch (MS condition)

We also analyzed the effect of diathesis (active vs passive) and position (first vs second noun) of the morphosyntactic mismatches as a function of the stimulated region. The aim of this analysis was to see whether stimulation of the prefrontal spots had a differential effect on performance depending on diathesis, mismatch position and their possible interactions.

The model includes the factors DIATHESIS (*active* vs *passive*), MS MISMATCH (*1* vs *2*) and REGION.

Descriptives are shown in **Tab.10**:

ms mismatch 1		active	passive
RT	*mean*	1200.07	1210.22
	std	674	666
ACC	*mean*	84.1	69.1
	std	3.6	4.6

ms mismatch 2		active	passive
RT	*mean*	1180	1330
	std	648	735
ACC	*mean*	66.1	65.1
	std	4.7	4.7

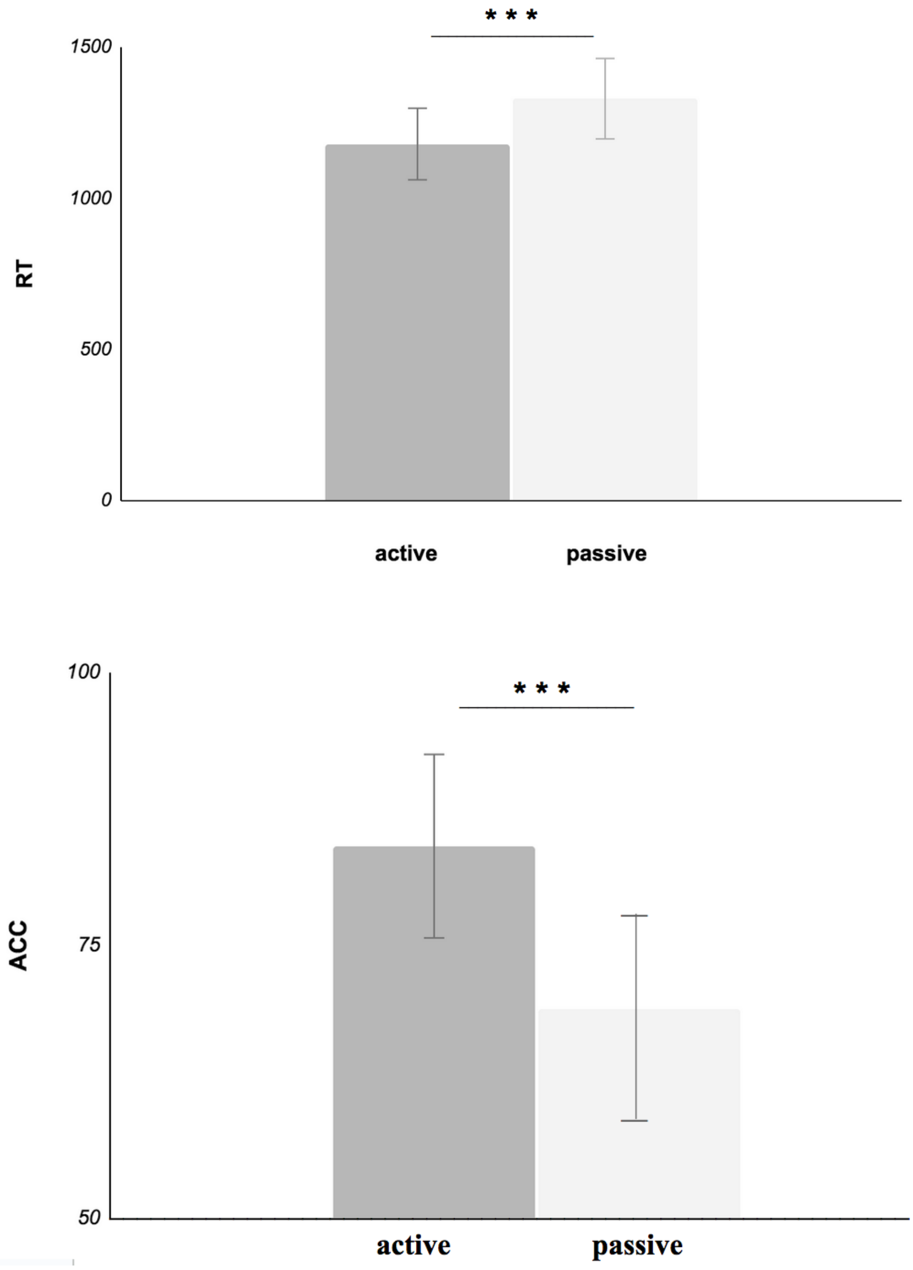
Tab.10 *Descriptive statistics (mean and standard deviation) for RT and ACC in active vs passive sentences with ms mismatch on the first (1) and on the second (2) constituent*

Results showed similar outcomes for RT and ACC. Significant main effects of all the factors and a MS MISMATCH x DIATHESIS interaction effect were found. No TMS-specific effect was found. Post hoc comparisons revealed that:

- active sentences were associated with faster RT and higher ACC than passive sentences;
- sentences with the MS mismatch on the first constituent were associated with faster RT and higher ACC than sentences with the MS mismatch on the second constituent;
- performance on sentences with the MS mismatch on the first constituent were faster and more accurate on active diathesis than on passive diathesis
  (see **Tab.11** and **Graph 6**).

		DIATHESIS	**MS MISMATCH**	**REGION**	**MS MISM x DIATH**
LMM	*RT*	$F(1,1)=11.262$, $p < 0.001$	$F(1,1)=5.594$, $p = 0.018$	$F(1,3)=5.486$, $p < 0.001$	$F(1,1)=6.410$, $p = 0.011$
	ACC	$X^2(1)=22.532$, $p < 0.001$	$X^2(1)=22.532$, $p < 0.001$	$X^2(3)=12.839$, $p = 0.005$	$X^2(1)=20.762$, $p < 0.001$
P-H	*RT*	$t= -3.36$, $p < 0.001$	$t = -2.37$, $p = 0.018$		act 2 vs pass 2 contrast $t = -4.085$, $p < 0.001$
	ACC	$z=4.75$, < 0.001	$z=6.20$, $p < 0.001$		act 1 vs pass 1 contrast $z=6.211$, $p < 0.001$

Tab.11 *Report of the results (main analyses (LMM) and Posthoc tests (P-H))*



Graph 6 *On top: significant mean RT differences between active and passive voice on sentences with ms mismatch on the first constituent (1). On bottom: significant mean ACC differences between active and passive voice on sentences with mismatch on the second constituent (2)*

Effect of diathesis (TR condition)

These analyses aimed at exploring whether diathesis (active vs passive) influenced performance on the TR condition when each TMS site was stimulated. These analyses were carried out to see whether previous findings on the role of the l-IPS in the comprehension of reversible passive sentences would replicate.

The model had the factors DIATHESIS and REGION and explored the effect of TMS vs sham on all the stimulation targets with active vs passive sentences.

Descriptives are shown in **Tab.12**:

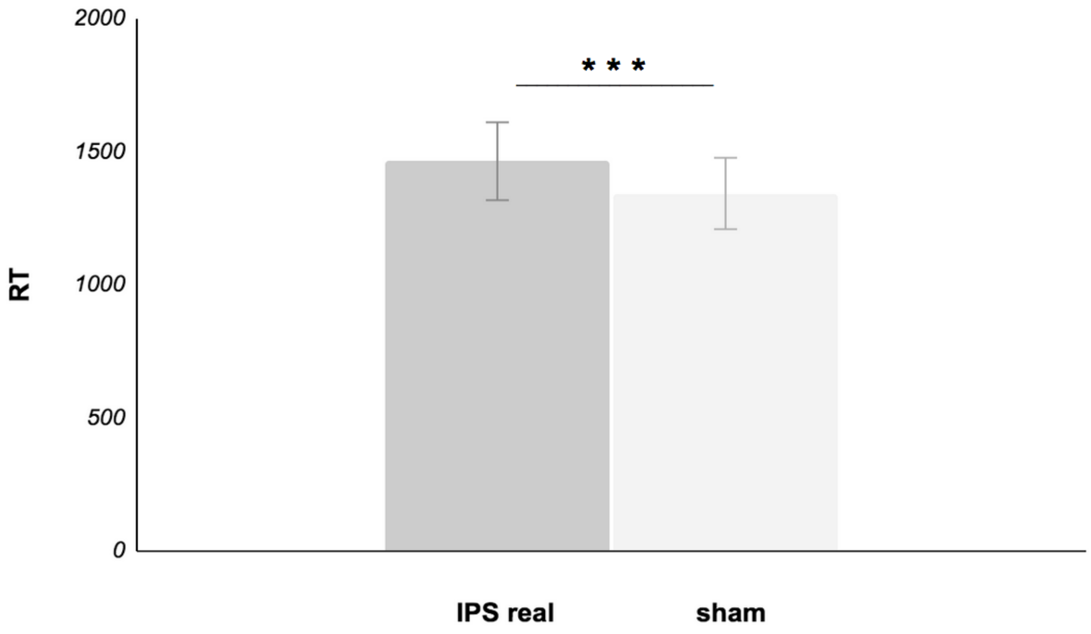
active		**IFG real**	**MFG real**	**IPS real**	**sham**
RT	*mean*	1203.44	1174.86	1349.16	1273.11
	std	640	627	794	701
ACC	*mean*	74.03	70.92	70.11	68.73
	std	4.3	4.5	4.5	4.6

passive		**IFG real**	**MFG real**	**IPS real**	**sham**
RT	*mean*	1282.16	1269.1	1465.58	1344.29
	std	646	691	850	758
ACC	*mean*	67.39	69.36	64.75	65.16
	std	4.7	4.6	4.7	4.7

Tab.12 *RT and ACC descriptives (mean and standard deviation) for active and passive sentences on each stimulation target*

Analyses of RT indicated a strong main effect of DIATHESIS ($F(1,1)=11.65$, $p < 0.001$) and of REGION ($F(1,3)=9.55$, $p < 0.001$) and no interaction effect ($F(1,3)=0.208$, $p = 0.891$). Bonferroni's Posthoc Comparisons reached significance only on the IPS ($t(1, 2535)=3.213$, $p < 0.001$) and not on the IFG ($t(1, 2534)= -1.942$, $p = 0.313$) and on the MFG ($t(1, 2535)= -2.296$, $p = 0.131$). As shown in **Graph 7**, TMS significantly increased RT on the IPS on passive sentences.

On ACC results showed a significant main effect of DIATHESIS ($X^2(1,1)=5.23$ $p = 0.022$) and not of REGION ($X^2(1,3)=3.33$ $p = 0.343$) and no interaction effect ($X^2(1,3)=0.97$ $p = 0.808$). Bonferroni's Posthoc Comparisons on DIATHESIS showed that active sentences were significantly associated with a lower amount of error than passive sentences.



Graph 7 *Bar plot of the significant real vs sham difference on the IPS for passive sentences*

Summary of the results

Prefrontal regions (IFG and MFG) were selectively involved in morphosyntactic processing and not in thematic role mapping, while the parietal region (IPS) showed the opposite pattern. On the prefrontal regions, performance on MS trials was enhanced by TMS, resulting in faster RT and increased ACC. In contrast, TMS on the parietal site affected performance on TR trials negatively, resulting in comparable ACC but longer RT.

Passive diathesis and the position of the MS mismatches, even though they did not interact with stimulation, influenced both RT and ACC. Sentences in the passive voice with MS mismatch on the second constituent were systematically associated with slower and less accurate performance. Passive sentences were also significantly affected by stimulation on the IPS, in line with previous findings on the link between this region and thematic reanalysis.

DISCUSSION

This study investigated the neural correlates of morphosyntactic and thematic role processes in the same sentential context. Since in the critical experimental trials a picture and a sentence contrasted in specific morphosyntactic or thematic features, the study allowed exploring the role played by two prefrontal regions (IFG and MFG) and a parietal region (IPS) in these two fundamental aspects of sentence comprehension.

Results: (i) confirm previous findings on the role of these left hemisphere regions in sentence processing; (ii) suggest a selective role of the IFG and the MFG in morphosyntactic processing; (iii) confirm a selective role of the IPS in thematic role assignment and (iii) demonstrate a functional distinction between left prefrontal and parietal networks in sentence processing.

TMS on the frontal regions (IFG and MFG) influenced the processing of morphosyntactic information and did not affect thematic role assignment. Conversely, TMS on the parietal region (IPS) affected thematic role mapping and not morphosyntactic processing. This suggests that morphosyntactic and thematic processing rely on at least partially distinct neural correlates. The prefrontal region (including IFG and MFG) correlates with morphosyntax and the parietal region (IPS) with argument labeling. Stimulation of the IPS affected particularly performance reversible passive sentences, thus confirming previous TMS studies (**Finocchiaro, 2015; 2021 and Vercesi, 2020**).

The left fronto-temporo-parietal network in sentence processing

The role of the IFG and the MFG

In our study TMS on both the IFG and the MFG selectively influenced performance on MS trials. Even though our data do not provide sufficient information to support a selective role for one or the other region, they allow some insights. In all the analyses that showed an effect of TMS on these two regions, performance was enhanced. The stimulation on the MFG

systematically speeded execution, while that on the IFG reduced error rate. The interpretation of this outcome is not straightforward; however, these results suggest that both the regions are involved in processing morphosyntactic features (number processing in particular).
Our results are in line with the main neurocognitive models of language comprehension that connect language processing to a left lateralized fronto-temporo-parietal network (**Friederici, 2011, 2012**; **Matchin & Hickok, 2020**; **Bornkessel-Schlesewsky & Schlesewsky, 2013**). Whether the involvement of left frontal regions in sentence processing is specifically linguistic is still a matter of debate. In literature two main views on the functional role of Broca's area have been put forward: a language-selective hypothesis and a domain-general hypothesis. On the first account, Broca's region would be involved in syntactic processing either in language (**Caramazza & Zurif, 1976**; **Friederici, 2018**) or across multiple domains (**Thompson-Schill et al., 1997; Tettamanti & Weniger, 2006**). On the second account, this region is involved in executive functions such as working memory (WM) or cognitive control (**Thompson-Schill et al., 1997; Novick et al., 2005**). Other neurofunctional studies have suggested that these regions function as a 'multiple-demand‘ system, presenting a shared activation pattern in response to different cognitive demands (MD) (**Duncan & Owen 2000**). Findings in the neuroimaging literature are often interpreted as being compatible with the view that the left fronto-temporo-parietal network is a cluster of multiple-demand regions differentially activated as a function of either language-specific or domain-general aspects (see **Santi & Grodzinski, 2007**). This view is also supported by a recent exploratory review by **Fedorenko and Blank (2020)**. They argued that the distinction between neural substrates devoted to language-specific and to domain-general mechanisms should not be approached as antinomic, and integrate the two alternatives in a comprehensive model postulating that Broca's area, by virtue of its dense anatomical and functional connections, is involved both in language-specific and in domain-general MD processes. Fedorenko & Blank claim that, since Broca's area is structurally and functionally heterogeneous, it is divided in sub-regions with distinct neural patterns that correspond to a language-selective frontotemporal network and a domain-general multiple-demand frontoparietal network, respectively (for a detailed description of this account see Fedorenko & Blank, 2020).

Although referring to our results as supporting one or the other hypothesis is beyond the scope of this work, results are compatible with the view that the involvement of the targeted regions in sentence processing is triggered by linguistic aspects, such as morphosyntax and thematic role mapping. On one hand, detecting a morphosyntactic mismatch requires recognizing morphological elements and their link with the syntactic structure (e.g., processing the auxiliary verb and inflected arguments in passive sentences). On the other hand, identifying a thematic role reversal relies on a combination of syntactic and semantic features, particularly in passive reversible sentences. Moreover, our results suggest that the two processes rely on at least partially distinct neural correlates. This finding could support the language-specific account. If the involvement of the targeted regions in sentence comprehension were not related to the specificity of the material to be processed, but by the general multi-demand resources required by the task, it would not explain why each site was associated with one process and not the other. On a domain-general view, both regions should have responded to TMS regardless of the thematic or morphosyntactic condition. Selective involvement in MS and TR processes supports the language-specific role of these regions in sentence processing.

The role of the l-IPS

The IPS responded to stimulation particularly on reversible passive sentences. This finding confirms previous TMS studies investigating the role of the l-IPS in thematic role assignment, in passive reversible sentences (**Finocchiaro et al., 2015; 2021** and **Vercesi et al., 2020**). These sentences are supposed to require a thematic reanalysis of the first-pass encoding of thematic roles due to co-occurrence of non-canonical syntactic structure and semantic reversibility, that makes them harder to process. In our investigation, the IPS did not respond to stimulation during MS trials. Even though these trials require the detection of the MS mismatch, they still require a correct mapping of thematic roles. The selective involvement of the IPS on TR trials reinforces the hypothesis of its specific role in thematic role assignment.

These data are consistent with the view that the l-IPS responds to language-specific dimensions. It has been associated with language comprehension, and in particular with thematic reanalysis (**Finocchiaro et al., 2015; 2021**; **Vercesi et al., 2020**; and the present study), needed when the first-pass parsing of thematic roles must be revised based on the changed syntactic structure and semantic reversibility. However, the specific nature of the l-IPS involvement in sentence processing is yet unclear. On one hand, it could support language comprehension by activating specific linguistic knowledge stored in LTM to make it available in WM. On the other hand, it could be associated with a domain-general function, providing the neural substrate for highly demanding tasks such as revising the initial thematic role assignment. In other words, it could function as a language-selective or domain-general region. Even though some studies suggest a role of the l-IPS in language comprehension, other investigations postulate a domain-general function of this region in human cognition. A review by **Lambon-Ralph et al. (2017)** suggests that the l-IPS is part of a network that underlies semantic control. This network: "*is thought to support working memory and executive representations that encode information about the temporal, situational and task context relevant to the current behaviour*". In support of this view, the authors present data from patients with temporo-parietal lesions, who showed problems in actively manipulating knowledge; these patients were first described with the term 'semantic aphasia' (SA) by **Luria (1964)**. SA has been shown to differ from semantic dementia (SD) in both verbal and non-verbal domains (**Corbett et al., 2009; Noonan et al., 2010**). In particular, deficits associated with SD seem to be connected to a damage to the semantic representation network, whereas SA affects performance in tasks related to the semantic control network. Patients suffering from SA typically show poorest performance in highly demanding tasks and tasks that require ambiguity and conflict monitoring (for a study on SA patients that shows a link between semantic control deficits and temporo-parietal damage see **Noonan et al., 2010**). In further support, neuromodulation studies show that rTMS on the temporo-parietal cortex (including the IPS) significantly influences performance on semantic functioning, especially when cognitive control is needed (**Hoffman et al., 2011; Whitney et al., 2011b; Whitney et al., 2012**).

Overall, our data support the view that the IFG, the MFG and the IPS are involved in language comprehension and play selective roles in processing different linguistic aspects. The functional distinction between the frontal and the parietal regions clearly demonstrates role selectivity and suggests that the contribution of these sites is language-specific rather than domain-general. The domain-general account assumes that these regions are part of a multiple-demand network, activated by tasks of different nature that share cognitive demands. If this were the case, we should have observed a generic TMS response of all three stimulated regions for both investigated processes, which was not the case. In contrast, the different response to stimulation according to the process primarily required by the mismatching trials shows that the selective response depends at least in part on the linguistic features of the trial. Although not yet sufficient to demonstrate the selectivity for language of these regions, our data add further evidence of their involvement in critical aspects of language processing.

CONCLUSIONS

This study provides insights on the causal role of the IFG, MFG and IPS in morphosyntactic and thematic role processing. It sought to establish if each region provided a selective contribution to processing different sentential features in the context of the same task. The task was designed to explore the specific role of each target by requiring the processing of either MS or TR aspects in order to recognize MS or TR mismatches between sentences and pictures. Data show that the two processes rely on functionally distinct neural correlates: the prefrontal regions (l-IFG and l-MFG) are related to the processing of morphosyntactic features and are not involved in thematic role assignment, while the parietal region (l-IPS) shows the opposite behavior. Our results replicate and reinforce previous findings correlating the l-IPS to the assignment of thematic roles in reversible sentences with non-canonical word order, that require a re-analysis of the first-pass encoding of thematic roles. Even though they do not allow detailed conclusions on the nature of the role played by the brain regions targeted in our experiment, results strongly suggest that the IFG, the MFG and the IPS are involved in morphosyntactic and thematic processing.

LIMITATIONS

Even though this experiment yielded clear results, it has limitations. In the first place, it does provide evidence that IFG and MFG (but not IPS) are involved in morphosyntactic processing, but the experimental paradigm only probed one morphosyntactic feature, namely number. Future research should explore the contribution of the key nodes of the left prefrontal network to additional aspects of MS processing.

The experiment also has methodological limitations. Even though the order of target stimulation within and between sessions was counterbalanced and randomized, more than one target region was stimulated, and both sham and TMS conditions were delivered during the same session. Real/Sham TMS of only one region per session would improve the investigation of each region’s role by canceling out the potential aftereffects of the stimulation of a region on the stimulation of another. Time constraints did not allow this approach. Another limitation is represented by the fact that the differential effects of TMS on

RTs found on MS and TR depending on the stimulation site did not emerge from an interaction between condition and site of stimulation. Rather, these effects were found when looking at the influence of TMS on the different targets in the two linguistic conditions separately. This was done to test the a priori hypothesis that the prefrontal regions are involved in MS processing and the inferior parietal region plays a role in mapping TR.

CHAPTER 2

A pilot study for behavioral validation of the linguistic task used in the TMS experiment

To finalize the experimental paradigm and the materials to be used for the just-reported rTMS experiment, a behavioral pilot study was conducted. This study aimed to determine the most adequate time window for stimuli presentation in the TMS experiment. The selected time window had to be sufficiently short as to prevent ceiling effects, but at the same time sufficiently long as to prevent random responses due to insufficient time to respond. In addition, even though performed on an underpowered sample, several preliminary analyses were done to provide a qualitative observation of the effects of the factors to be included in the TMS experiment.

Participants

10 participants (5 female, 5 male; mean age=28.5 (+-2.87 SD)) with unimpaired or corrected-to-normal vision, no prior history of neurological conditions, seizures, or psychiatric symptoms took part in the pilot. All were recruited at the University of Trento and participated on a voluntary basis.

Task

The task was the same as that described in **Chapter 1**. Response time and accuracy were collected. Before starting the experimental procedure, participants completed a practice run to get acquainted with the task. They were presented with 16 trials, divided in two blocks of 8 trials each. After the training phase, requests for clarification were invited. Subsequently, the experiment was started. The experiment included 128 trials, divided in 4 blocks of 32 trials each. After each block participants were allowed a short break.

Experimental paradigm

To test the experimental material, participants' performance was analyzed and the effect of all the factors included in the protocol was evaluated, even those of dimensions expected to be outcome-neutral or less explanatory. The factors correspond to: DIATHESIS, TRIAL and CONDITION. Four different trial presentation times were used: 875, 900, 925 and 950 ms. The reason for this manipulation resides in the wish to identify a trial presentation window that allowed task execution with an accuracy between 70 and 80% - that is, a time window that would make the task neither too easy nor too hard, thus preventing floor and ceiling effects.

Times were counterbalanced and randomly assigned to the trials so that each block had the same number of trials of the same condition assigned to each presentation time.

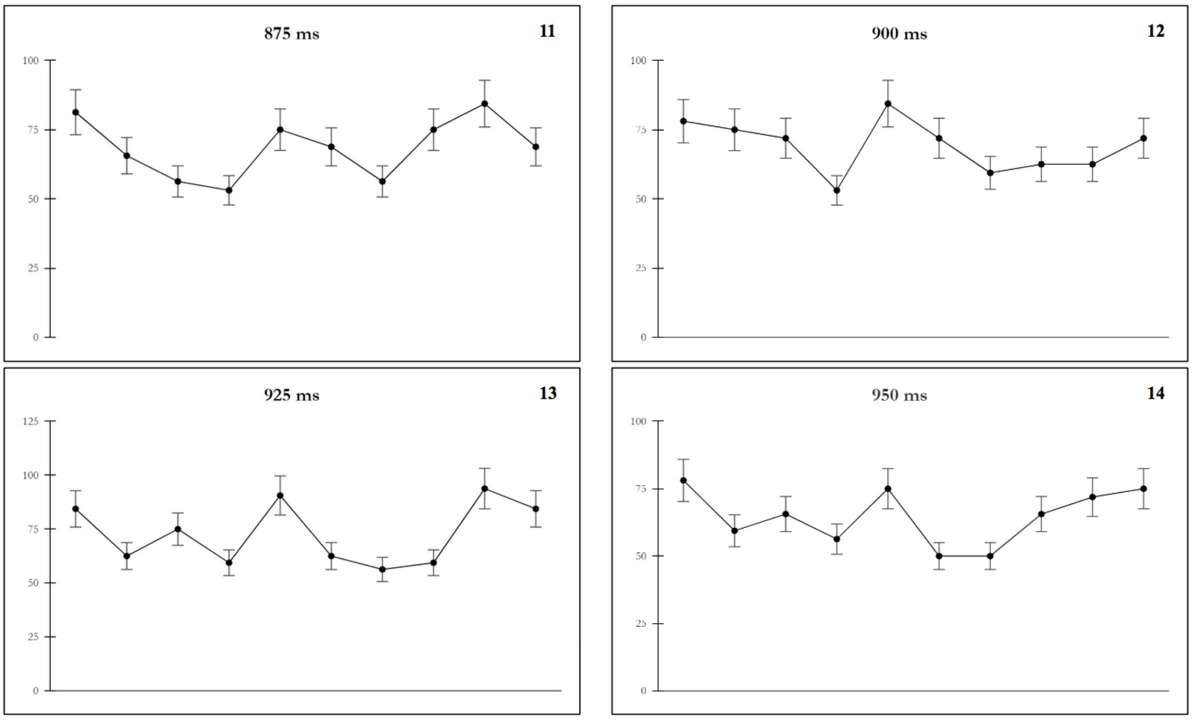
The pilot had a 2x2x2x4 design, with four factors: DIATHESIS (*active* vs *passive*), TRIAL (*matching* vs *mismatching*), CONDITION (*TR* vs *MS*) and TIMING (*875* vs *900* vs *925* vs *950*).

Statistical analyses

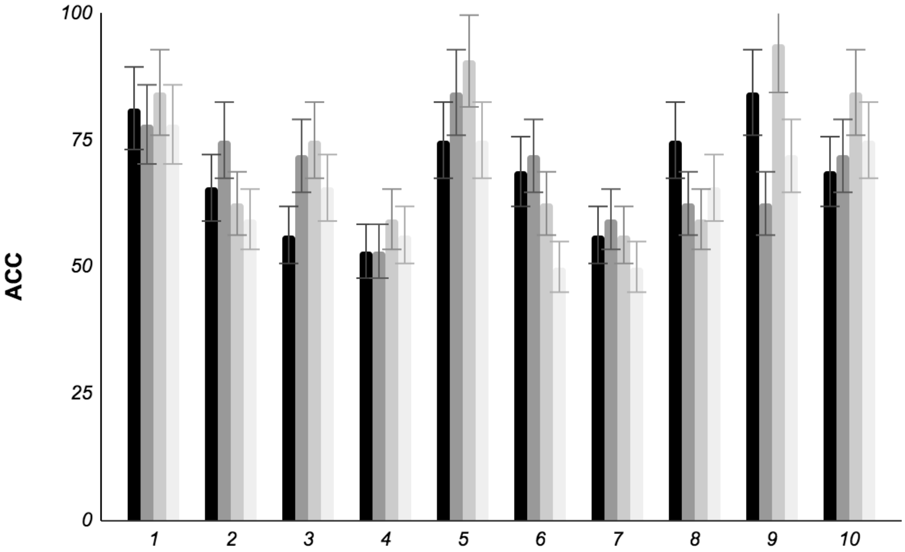
As mentioned above, this was a pilot study designed to test some elements to include in the TMS experiment. Given the small sample size, all the analyses that will be presented could only provide exploratory results to be interpreted as preliminary considerations.

TIMING

We first analyzed performance accuracy based on the four levels of TIMING. We wished to establish the trial presentation time that resulted in performance accuracy between 70 and 80%. Overall, mean performance accuracy for the four levels of TIMING was 68.4% (SD=10.8) for 875 ms, 69.1 % (SD= 9.48) for 900 ms, 72.8% (SD=14.4) for 925 ms and 64.6% (SD=10.4) for 950 ms. **Graph 1** and **2** shows mean performance accuracy for each stimulus presentation time:



Graph 1 *X axis: Mean accuracy scores of all the participants in each stimulus presentation time (875, 900, 925, 950 ms); Y axis: performance accuracy (%)*



Graph 2 *Bar plot of mean accuracy scores of all the participants in each stimulus presentation time (875, 900, 925, 950 ms). Color legenda:*

█ 875 ▓ 900 ▒ 925 ░ 950

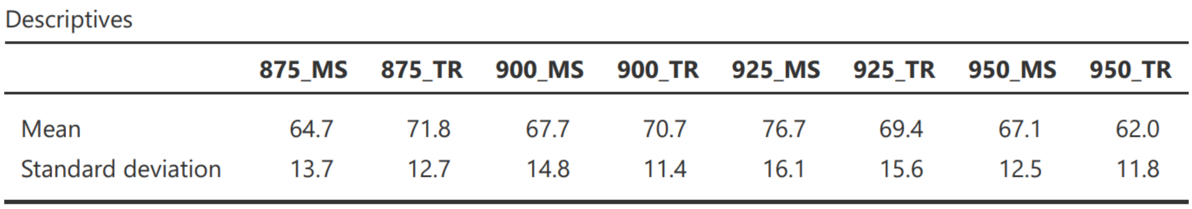
:

Data were analyzed through a one-way repeated-measures ANOVA with the within-participants factor TIMING on four levels (875 vs 900 vs 925 vs 950). No significant effect of TIMING on performance accuracy was shown (*F*(3, 27)=2.16, *p*=0.116), meaning that there were no significant differences between scores on the four trial presentation times. Statistically, thus, there was no specific reason to choose a particular timing. However, looking at the performance of all subjects (Graphs 1 and 2), greater uniformity and consistency of distribution is observed in the 900 ms window.

In addition, we looked at the interaction between the factor TIMING and the other factors DIATHESIS, TRIAL and CONDITION.

TIMING x CONDITION

First, the interaction between the factors TIMING and CONDITION was explored. Mean performance accuracy of participants in each time window for both TR and MS conditions is reported in **Tab.1**:

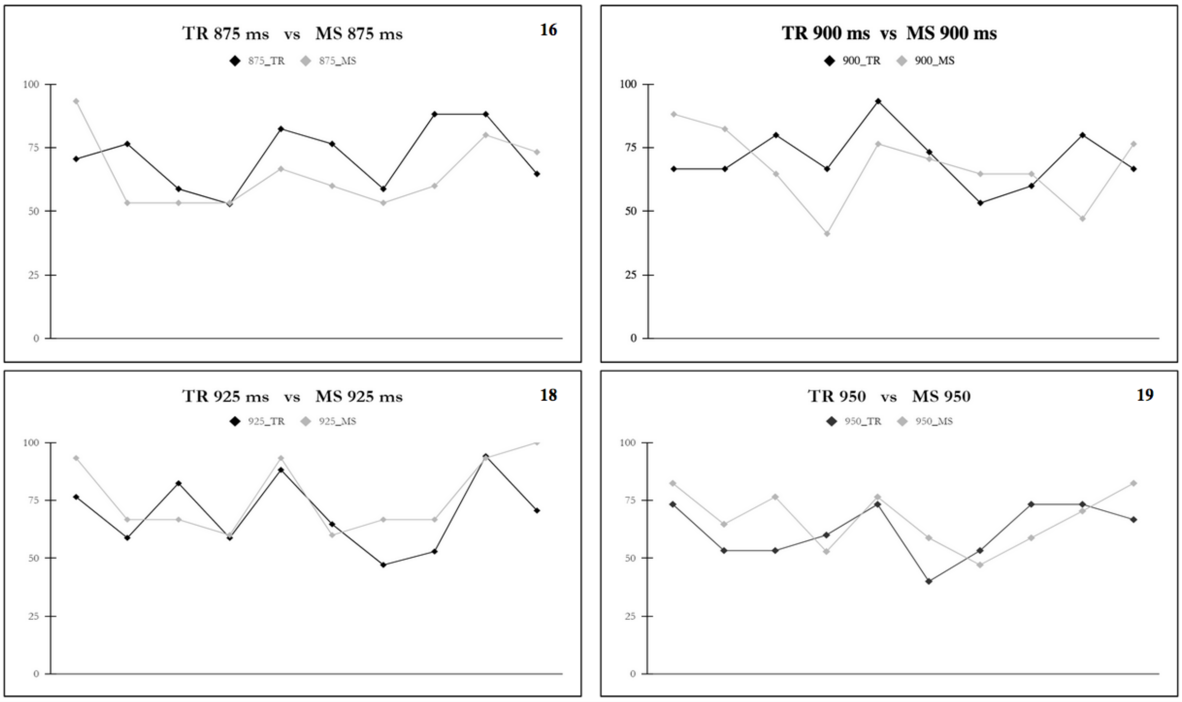


Descriptives

	875_MS	875_TR	900_MS	900_TR	925_MS	925_TR	950_MS	950_TR
Mean	64.7	71.8	67.7	70.7	76.7	69.4	67.1	62.0
Standard deviation	13.7	12.7	14.8	11.4	16.1	15.6	12.5	11.8

Tab.1 *Descriptives of mean accuracy performance for each condition (MS and TR) in each time window (875, 900, 925, 950 ms)*

Graph 3 shows performance at each time window for the TR and MS condition:

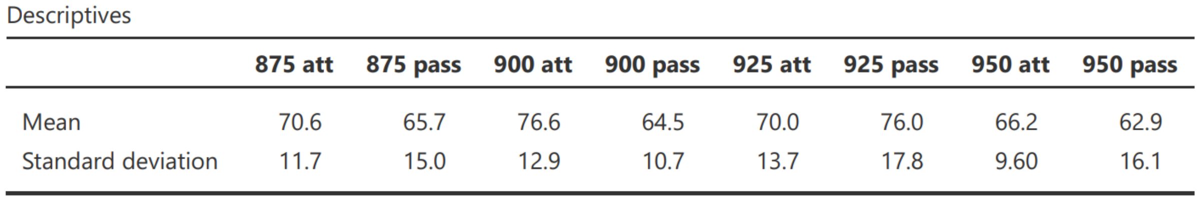


Graph 3 *X axis: Mean accuracy scores of all the participants in each stimulus presentation time (875, 900, 925, 950 ms) for MS and TR conditions; Y axis: performance accuracy (%)*

Statistical analyses were conducted via a two-way repeated measures ANOVA. Results showed no significant main effect of either TIMING ($F(3, 27)=2.42$, $p=0.087$) or CONDITION ($F(1, 9)=0.03$, $p=0.857$) and no interaction effect TIMING x CONDITION ($F(3, 27)=2.63$, $p=0.070$). Performance accuracy was not influenced by the two experimental conditions (TR and MS) in any time window. Even though some slight differences can be observed, TR and MS do not differ substantially at a behavioral level.

TIMING x DIATHESIS

The interaction between TIMING and DIATHESIS was also explored. Mean accuracy of participants for both active and passive sentences in all the time windows are shown in **Tab.2**:

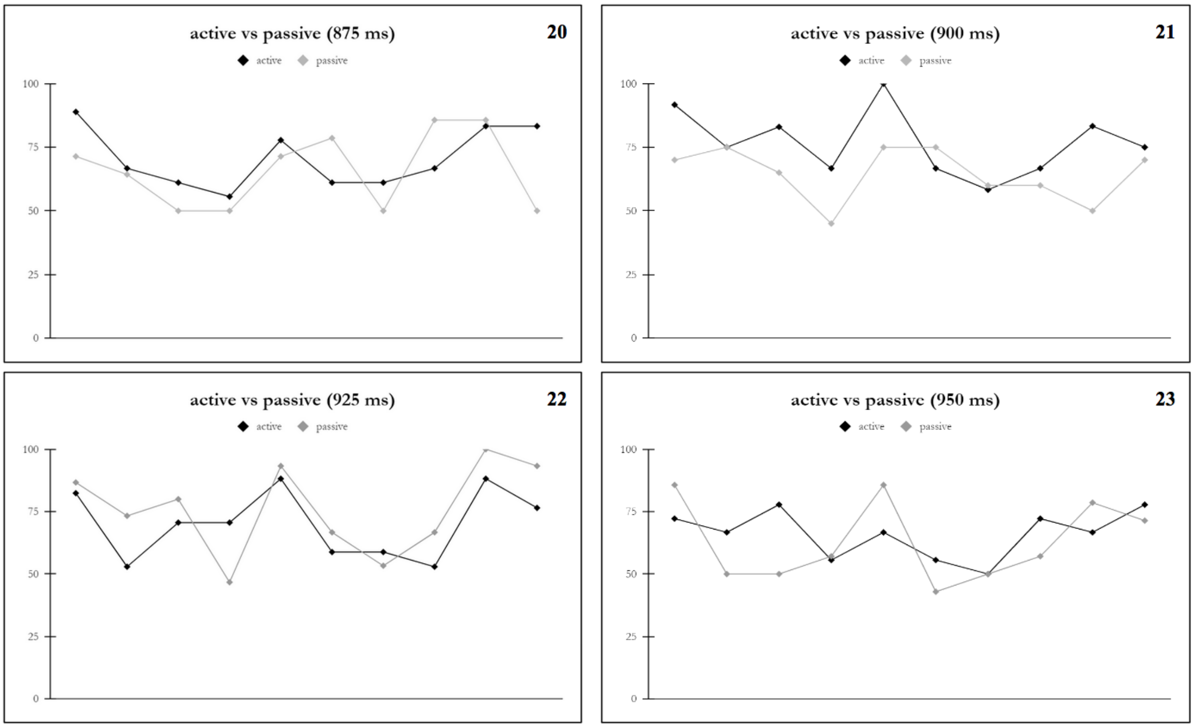


Descriptives

	875 att	875 pass	900 att	900 pass	925 att	925 pass	950 att	950 pass
Mean	70.6	65.7	76.6	64.5	70.0	76.0	66.2	62.9
Standard deviation	11.7	15.0	12.9	10.7	13.7	17.8	9.60	16.1

Tab.2 *Descriptives of mean accuracy performance for active and passive sentences in each time window (875, 900, 925, 950 ms)*

See **Graph 4** for a description of performance in this comparison:



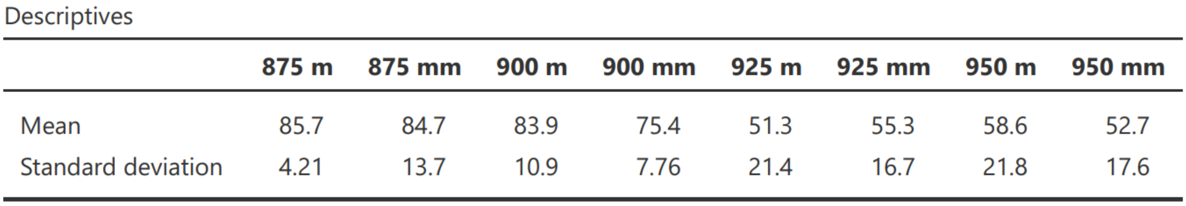
Graph 4 *X axis: Mean accuracy scores of all the participants in each stimulus presentation time (875, 900, 925, 950 ms) for active and passive sentences; Y axis: performance accuracy (%)*

Statistical analyses were conducted through a two-way repeated measures ANOVA. No significant main effect of either TIMING (*F*(3, 27)=2.60, *p*=0.073) or DIATHESIS (*F*(1, 9)=3.84, *p*=0.082) emerged, and no interaction effect TIMING x DIATHESIS (*F*(3, 27)=2.44, *p*=0.086). No significant differences were found between active and passive trials in all time windows. At an observational qualitative level, some difference between active and passive sentences can be seen, with passive sentences being generally associated with lower accuracy. However, these differences are in line with the expectations and with the literature and they reasonably should not reflect unexpected effects in the TMS experiment.

Another factor to be considered is TRIAL. As for TIMING, the interactions between TRIAL and the other factors were explored.

TRIAL x TIMING

Mean performance accuracy of the participants on both matching (m) and mismatching (mm) trials for all the time windows is shown in **Tab.3**:



Descriptives

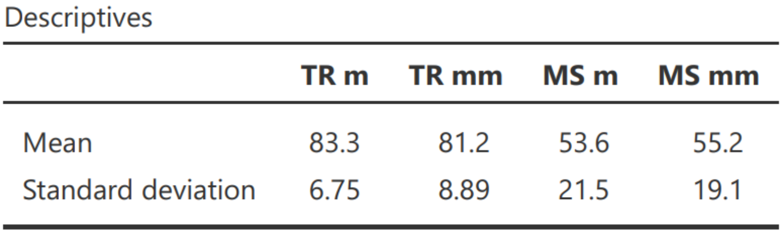
	875 m	875 mm	900 m	900 mm	925 m	925 mm	950 m	950 mm
Mean	85.7	84.7	83.9	75.4	51.3	55.3	58.6	52.7
Standard deviation	4.21	13.7	10.9	7.76	21.4	16.7	21.8	17.6

Tab.3 *Mean and standard deviation for mean ACC performance for matching (m) and mismatching (mm) trials*

Analyses were carried out as in previous contrasts. There was a significant main effect of TIMING ($F(3, 27)=18.95$, $p < 0.001$), but no significant main effect of TRIAL ($F(1, 9)=1.55$, $p=0.244$) and no TRIAL x TIMING interaction ($F(3, 27)=2.27$, $p=0.104$).

TRIAL x CONDITION

TRIAL and CONDITION interaction was explored as in the other contrasts. **Tab.4** shows mean accuracy of the participants in MS vs TR conditions for matching and mismatching trials.

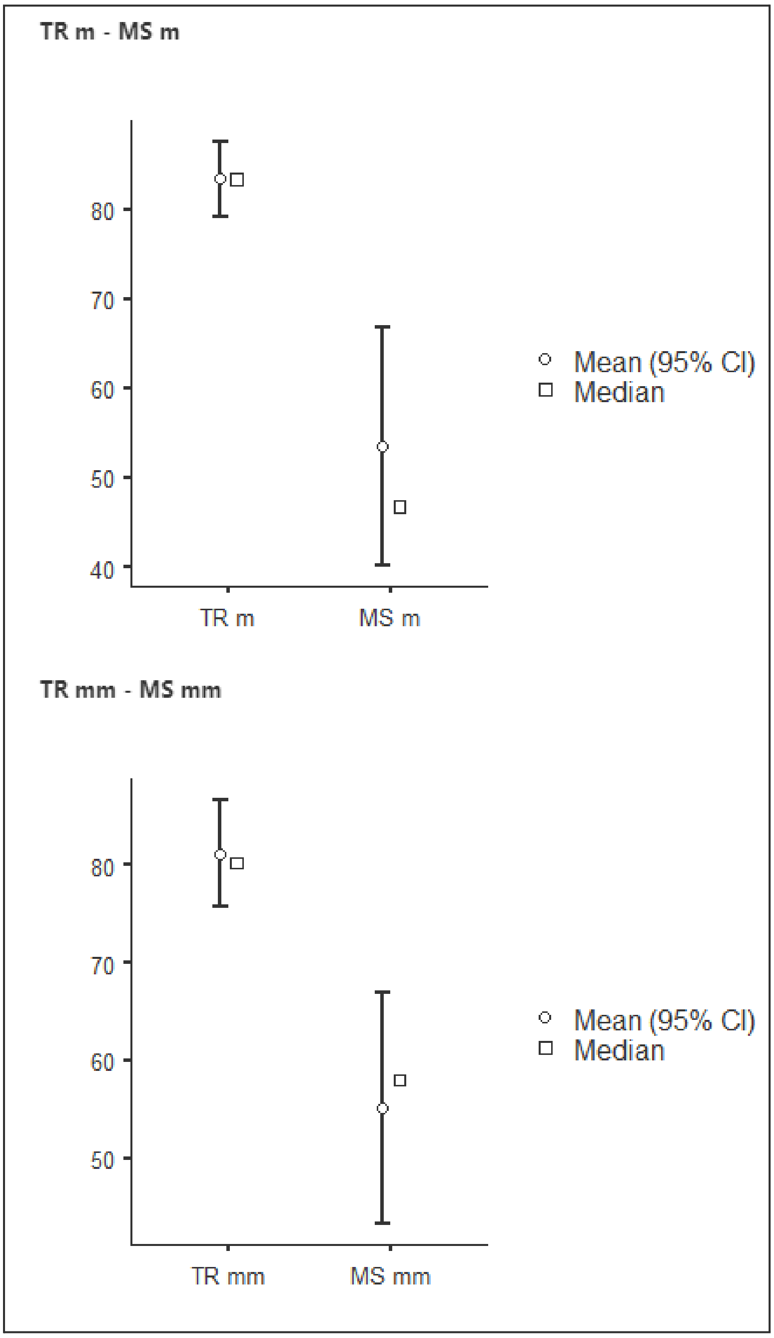


Descriptives

	TR m	TR mm	MS m	MS mm
Mean	83.3	81.2	53.6	55.2
Standard deviation	6.75	8.89	21.5	19.1

Tab.4 *Mean and standard deviation for mean ACC performance for matching (m) and mismatching (mm) trials in MS and TR conditions*

ANOVA results showed a significant main effect of CONDITION (($F(1, 9)=26.81$, $p < 0.001$) and no significant main effect of TRIAL ($F(1, 9)=0.006$, $p = 0.935$) and no TRIAL x CONDITION ($F(1, 9)=0.19$, $p = 0.669$). The main effect of CONDITION was further explored through two paired t-tests that contrasted the two conditions (MS and TR) within the same type of trial (matching vs mismatching). Results showed that performance in the TR vs MS contrast significantly varied in both matching ($t(9)=4.43$, p=0.002) and mismatching trials (*t*(9)=3.74, p=0.003). Performance accuracy was significantly lower on MS trials in both matching and mismatching trials (**Graph 5**).

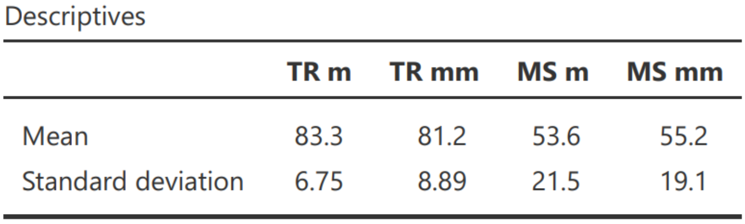


Graph 5 *Statistical comparison between mean ACC performance matching (m) and mismatching (mm) trials for TR and MS conditions*

We interpreted these differences as task-dependent random effects, whose effect on performance should not be alarming.

TRIAL x DIATHESIS

Finally, TRIAL x DIATHESIS interaction was explored. **Tab.5** below shows performance on active vs passive trials for both matching and mismatching trials:



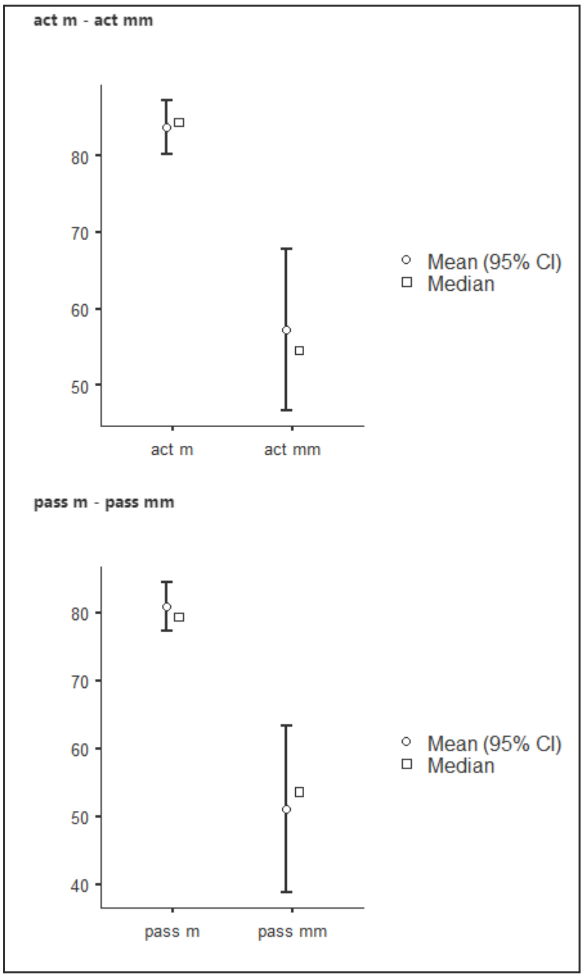
Descriptives

	TR m	TR mm	MS m	MS mm
Mean	83.3	81.2	53.6	55.2
Standard deviation	6.75	8.89	21.5	19.1

Tab.5 *Mean and standard deviation of ACC performance for TR and MS matching (m) and mismatching (mm) trials*

ANOVA showed a marginally significant main effect of DIATHESIS (($F(1, 9)=5.63$, $p=0.046$) and a significant main effect of TRIAL (($F(1, 9)=27.03$, $p < 0.001$). The DIATHESIS x TRIAL interaction was not significant ($F(1, 9)=0.941$, $p=0.357$). The main effect of TRIAL was explored as in the previous analysis. The paired t-tests that reached significance were those contrasting matching vs mismatching trials within the same diathesis. Accuracy in both active and passive trials was significantly lower for mismatching than matching trials ($t=5.14$,$p<0.001$) and in the passive trials ($t=4.83$, $p < 0.001$) as shown in

Graph 6:

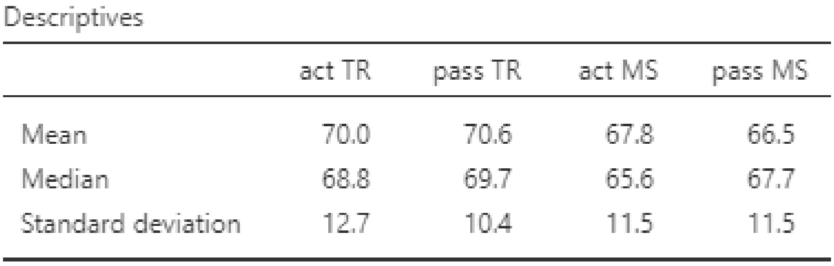


Graph 6 *Statistical comparison between mean ACC performance for active and passive sentences in matching and mismatching trials*

Since DIATHESIS did not represent a substantial source of variability in the dataset (i.e. active vs passive trials were not significantly different), this result was not interpreted as relevant for the TMS experiment.

DIATHESIS x CONDITION

The last comparison was between DIATHESIS and CONDITION (**Tab.6**):



Descriptives

	act TR	pass TR	act MS	pass MS
Mean	70.0	70.6	67.8	66.5
Median	68.8	69.7	65.6	67.7
Standard deviation	12.7	10.4	11.5	11.5

Tab.6 *Mean and standard deviation for active and passive sentences in MS and TR conditions*

ANOVA showed no significant main effects of DIATHESIS (*F*(1, 9)=0.01, *p*=0.903) and CONDITION (*F*(1, 9)=2.82, *p*=0.12) and no DIATHESIS x CONDITION interaction (*F*(1, 9)=0.03, *p*=0.595).

Summary and conclusions

The goal of the pilot study was to test and finalize the task and to preliminary explore all experimental variables in order to obtain useful information in view of the forthcoming TMS experiment.

The task proved to be effective in comparing thematic role assignment and morphosyntactic processing, with the two conditions showing no significant differences in terms of accuracy. The other variables were also explored, alone and in interactions with each other. No statistically significant differences were found in any comparisons, except for task-dependent effects of a random and unimportant nature.

Another goal of the pilot was to determine the stimulus presentation window that allowed response accuracy between 70-80%. Since the most suitable time window appears to be 900 ms, it was chosen for the TMS experiment. In this time window mean performance accuracy was globally in the expected range and the distribution of the scores was more consistent and uniform.

LIMITATIONS

The reported study was a pilot experiment designed to test some elements to finalize the task to be used in the TMS experiment. Given the small sample size (N=10), it was not possible to draw any certain conclusion from statistical analyses. As a consequence, all the reported results should be treated as preliminary. This strongly limited the impact that the information it provided on the effects of the experimental factors could have on formulating expectations for the TMS study.

—

CHAPTER 3

Are Linear Mixed Models (LMM) systematically better than Analysis of Variance (ANOVA) in investigating data variability in TMS studies of language?

ABSTRACT

Analysis of variance (ANOVA) and Linear Mixed Models (LMMs) are used to explore variability in the data. Even though both statistical approaches investigate variance, there are important differences to be considered. While ANOVA is a fixed-effects model, able to explore the variability of the predictor variables, LMMs include both fixed- and random-effects, thus managing to account for multiple sources of random variability that ANOVA cannot include. For this reason, LMMs are increasingly used in the pipelines of data analysis. However, the two models should not be treated as being alternative and neither is intrinsically more explanatory, hence preferable in general. Rather, they should be considered as providing different levels of information content, depending on the type and the nature of the data. This work addresses this issue by comparing the use of both ANOVA and LMMs on the same dataset. Two language and neurostimulation experiments previously published have been re-analyzed through a pipeline that included both models and their outcomes have been

compared. Results confirm that ANOVA and LMMs do not differ in their outcomes, but in the information about variability they allow to obtain.

INTRODUCTION

The link between the pipeline and the outcomes

The relationship between the pipelines used in data analysis and the conclusions drawn on their basis is a well-known issue. It has been suggested that analytical flexibility plays a crucial role in causing inflated rates of false positive results (**Ioannidis, 2005**; see also **Carp, 2012** for an inquiry on the link between analytical flexibility and experimental outcomes). As analysis procedures have become more complex, analytical flexibility has increased. Several studies highlighted the problem of the impact of choosing a specific pipeline for data analysis on outcomes. The debate has been ongoing especially in the neuroimaging domain. For example, **Weissenbacher et al., 2009** and **Carp, 2011** investigated the link between outcomes and the order of analytical steps. Other investigations addressed the issues raised by selecting a specific software for data analysis (**Smith et al., 2005; Poline et al., 2006**). Yet other studies explored the conjoined impact of these dimensions (**Churchill et al., 2012**).

This trend may have nontrivial consequences. A recent study (**Botvinik-Nezer, 2020**) underlined the importance of questioning the relationship between analysis pipeline and results. The same neuroimaging dataset was given to 70 teams of researchers, who were asked to analyze the dataset and test nine *a priori* hypotheses. Each team was instructed to use the procedures customarily used in their laboratory. As it happened, no team used the same analysis pipeline, and the results of hypothesis testing showed detectable variation, even in teams whose statistical maps showed a high correlation at intermediate stages of the analyses. The authors underline the importance of contrasting such pipeline-dependent effects with proper approaches, and encourage groups to share data, codes and materials in order to increase reproducibility. Even though conducted on fMRI data, the study addresses an issue that can be extended to any kind of dataset. In the present study the problem is addressed in the context of behavioral and neurostimulation data.

Two neurostimulation experiments conducted and published by our group (**Finocchiaro et al., 2015** and **2021**) will be taken as study cases to explore any potential pipeline-dependent effects in the analyses. Experiments used Non-Invasive Brain Stimulation (NIBS) and

Transcranial Magnetic Stimulation (TMS) to investigate the neural basis of sentence processing. The resulting datasets were analyzed via both ANOVA and LMMs. The results of each experiment were compared and discussed.

Despite investigating and exploring the same parameter (*i.e.* the variance in the data), the two models differ importantly in their theoretical assumptions and in the level of information they can provide. In the literature it is frequently claimed that LMMs should replace the more traditional ANOVA approach, as they can include and explain sources of variability in the data that ANOVA cannot take into account (see **Krueger et al., 2004**; **Boisgontier & Cheval, 2016**; **de Melo et al., 2022**). The present methodological study is informed by this crucial question: should ANOVA be abandoned? If the answer were to be affirmative, we would be authorized/forced not only to eliminate ANOVA as a pipeline, but also to reconsider or challenge conclusions drawn on ANOVA-based experiments. The present study will compare the two models to address this issue.

As in other domains, in the field of neurostimulation the study of the link between basic sensorimotor skills and their neural correlates is simpler than for higher-level processes, such as language. For example, the study of object grasping is facilitated by the behavioral outcomes that directly reflect the process (the sequences of organized hand movements performed to grasp the object). In contrast, studying sentence comprehension is not as straightforward and behavioral outcomes must be measured indirectly (reaction times and errors in pressing a response key following the presentation of a stimulus in a forced-choice context). As the complexity of the process increases, so does the number of variables to be factored in the experimental paradigm to examine and measure the phenomena under investigation. A larger number of parameters inevitably affects the interpretability of the observed statistical results. Furthermore, the link between statistical significance and reliability of inference is not always linear. In the case of low-level processes, high significance generally corresponds to a greater reliability of results. Studying complex processes characterized by many variables may originate results that are statistically less powerful in terms of significance. Increased complexity may make data interpretation less straightforward and result in weaker inferences. However, the implications of these differences should be treated cautiously. The link between levels of significance and quality and/or reliability of the inference is not linear; models that allow multiple levels of informativeness, that improve the quality of the inference, are often associated with less significant outcomes, and *vice versa*. The present work aims at investigating the relationship between the adopted pipelines and their outcomes, with a specific focus on their interpretation.

The models: ANOVA

ANOVA is used to explore mean differences between experimental groups, and in particular the relationship between a continuous dependent variable (DV) and one or more categorical independent variables (IVs) across multiple experimental groups (for an introductory overview see **Sawyer, 2013**). Since its very beginning, ANOVA has represented a cornerstone in the exploration of variance in different kinds of data. It is the approach adopted most frequently to account for differences in the dataset. However, it is constrained by specific assumptions that are often taken for granted or even ignored. These assumptions clarify the limitations and pitfalls of ANOVA and should be considered as important caveats for data interpretation.

Firstly, ANOVA assumes that the population from which the dataset is obtained has a normal distribution (assumption of *normality*). The shape of data distribution can be checked through statistical normality tests, of which the Kolmogorov-Smirnof or the Shapiro-Wilk are the most frequently used. For non-normally distributed data, a non-parametric version of ANOVA is available (Kruskal-Wallis for one-way ANOVA and Feldman for repeated-measures ANOVA). However, there is some evidence based on simulations of data from a wide variety of non-normal distributions that show that the incidence of false positives is not as high when the assumption of normality is violated (**Glass et al. 1972**, **Harwell et al. 1992**, **Lix et al. 1996**). In addition, non parametric tests may not lead to better results than the parametric equivalent if the shape of the distribution of the different groups varies greatly (e.g. one is skewed to the right and one to the left). Consequently, ANOVA can fit many datasets with non-normal distributions. Hence, the assumption of normality does not represent a full constraint to the use of ANOVA.

Secondly, ANOVA presupposes that the measures observed in the samples are independent of each other (assumption of *independence*). This assumption is particularly important in repeated measures designs, in which multiple measures are taken for each participant at different points in time, as it implies that the repeated observations on the same data unit must not be correlated. A correlation between repeated measures signals that the measures are dependent, violating the independence assumption.

Finally, the most critical assumption of ANOVA is represented by the *homogeneity of variance*, according to which the variance across groups should be approximately equal. This assumption implies that the size of the effect of the experimental treatment is similar across participants. This is tantamount to saying that the amount of variance resulting from the differences in 'responding' to the treatment cannot be estimated by ANOVA, which *a priori* takes it as being equal. For example, consider an experiment designed to test the relationship between a given brain region and a given cognitive function through TMS. The function under study is evaluated through a cognitive task that is administered both during TMS and in a control condition in the absence of stimulation (placebo or sham). TMS is administered over a given brain region that is supposedly related to the cognitive function under investigation. TMS represents the experimental treatment (the IV), while the DV is represented by the performance in the task. To demonstrate that the cognitive function is causally associated with the targeted brain region, performance in the cognitive task during stimulation should differ from performance in the absence of stimulation. Suppose the two conditions (real and placebo stimulation) show significant differences. The assumption of homogeneity of variance implies that the observed mean differences are equal across participants. The treatment effect is then calculated as the average of the treatment response of all participants. This means that each participant is supposed to have on average an equal response to treatment. Since it is highly likely that a treatment influences each participant differently -albeit slightly-, including the mean response to the treatment could result in losing information on individual variability. The difference in each participant's response represents a source of random variability that could be of interest when exploring the overall variance in the data. Similarly, other sources of random variability can be present in the data and missed by ANOVA-based analyses.

Note that, while implications and limitations of ANOVA must be kept in mind, the statistical reliability of legitimate inferences is not affected by ANOVA's failure to estimate all sources of variability. On the other hand, the ANOVA does not allow to estimate the multiple sources of random variability, except by 'dissolving' them into the average variability of all participants. When these sources of variability must be estimated, LMMs can be used.

The models: LMMs

LMMs is a statistical tool that can incorporate both fixed and random effects (for a broad and complete theoretical view on LMM see **Jiang & Nguyen, 2007;** see also **Raudenbush & Bryk, 2001** or **Singer & Willett, 2003**). They represent a particularly useful approach for repeated measures designs, in which multiple measures are made on the same data unit. Moreover, LMMs are specifically suited for the analysis of non-independent data and allow one to deal with missing values. For these reasons, they are often preferred to the standard ANOVA approach.

As other fixed-effects models, ANOVA can only estimate one source of random variability, that comes from the random sampling to measure the variables. The across-subject variability is called "residual" and corresponds to the variance that is not explained by the predictors of the model (*i.e.*, the fixed effects). Conversely, mixed models' effects can account for more than one source of random variability in the data by including random effects in the model. A mixed model thus includes both fixed and random effects. While on one hand fixed effects are expected to affect the DV, random effects can be considered as grouping factors to be controlled (see **Clark & Linzer, 2015** for a scrutiny of uses of fixed- and random-effects models and differences between them). For example, one could consider each participant as a cluster and use the variable 'participants' as a grouping factor. This would allow one to estimate the random variance coming from each participant to better explore the effect of the fixed effects on the response variable. This can be done with other categorical variables that putatively contribute to an account of random variability in the data.

The purposes of the study

Exploring multiple sources of random variability in the data can be crucial in complex designs, in which several uncontrolled variables are expected to influence the amount of variance in the sample. Nevertheless, cautiousness is critical in assigning LMMs an absolute dominance over standard models of analysis of variance. What LMMs can offer, in fact, should be intended as a refinement and specification of ANOVA, rather than a replacement.

When approaching data analysis, it remains crucial to question the nature and type of the data and whether a more in-depth modeling of individual behavior is more informative than what ANOVA alone can provide. Starting from these considerations, the present work aims at comparing the two approaches from premises to outcomes and exploring differences and similarities between them (for a similar approach see **Koerner & Zhang, 2017**; **Yu et al., 2021**). Since both ANOVA and LMMs are fully legitimate models, this comparison does not intend to attribute greater or lesser explanatory power to one or the other, nor to suggest that one is inherently better than the other. Rather, the two methods are compared based on the information they can provide on data variability.

MATERIAL AND METHODS: THE DATASET

Experiment 1 (Finocchiaro et al., 2015)

Overview

The main goal of this experiment was to provide causal evidence for the relationship between the l-IPS and thematic role mapping. To this aim, repetitive TMS (rTMS) was applied over three sites along the l-IPS: an anterior, a middle and a posterior site (P1, P2 and P3 as in **Fig. 1**) while participants were engaged in a sentence-picture matching task (**Fig. 2**) that included truncated active and passive reversible sentences (**Fig. 3**). Participants had to decide by key press if the event represented in a picture matched the meaning of a written sentence. Stimulation was administered at the frequency of 5 Hz starting from stimulus onset. The stimulation session (TMS) was compared to a baseline control condition in which stimulation was not administered. Both response times (RTs) and performance accuracy (ACC) were collected as behavioral measures.

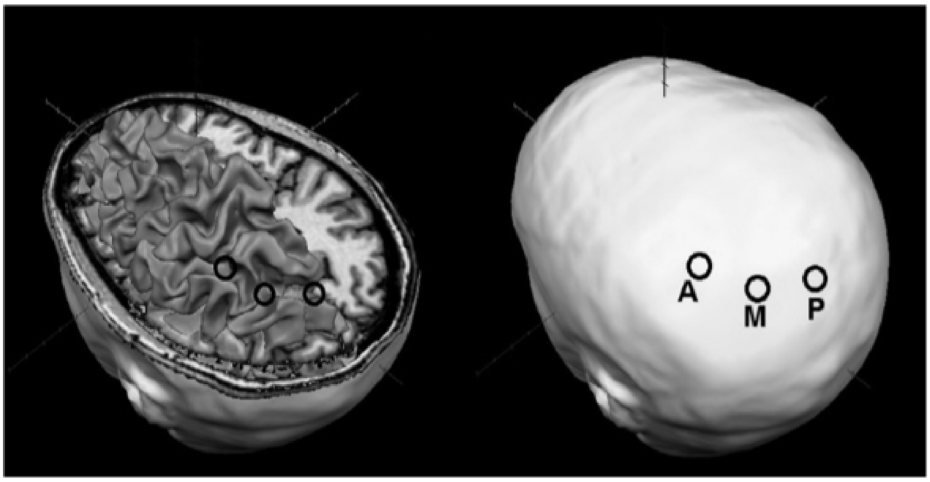


Fig. 1 *On the left: localization of the three stimulation sites (anterior, middle, posterior) on the native space of and individual estimated MRI (taken from Finocchiaro et al., 2015)*

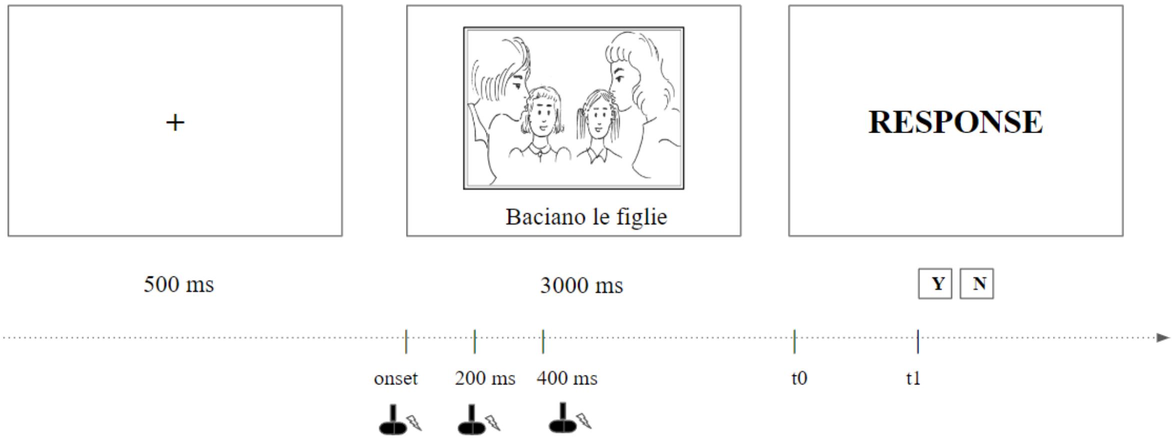


Fig.2 *Timeline of the experimental task*

Sentences

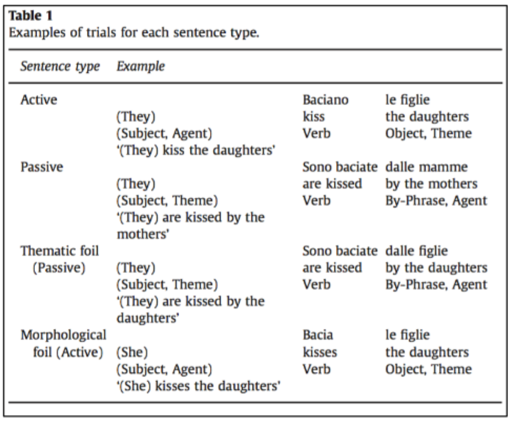


Table 1
Examples of trials for each sentence type.

Sentence type	*Example*		
Active		Baciano	le figlie
	(They)	kiss	the daughters
	(Subject, Agent)	Verb	Object, Theme
	'(They) kiss the daughters'		
Passive		Sono baciate	dalle mamme
	(They)	are kissed	by the mothers
	(Subject, Theme)	Verb	By-Phrase, Agent
	'(They) are kissed by the mothers'		
Thematic foil (Passive)		Sono baciate	dalle figlie
	(They)	are kissed	by the daughters
	(Subject, Theme)	Verb	By-Phrase, Agent
	'(They) are kissed by the daughters'		
Morphological foil (Active)		Bacia	le figlie
	(She)	kisses	the daughters
	(Subject, Agent)	Verb	Object, Theme
	'(She) kisses the daughters'		

Pictures

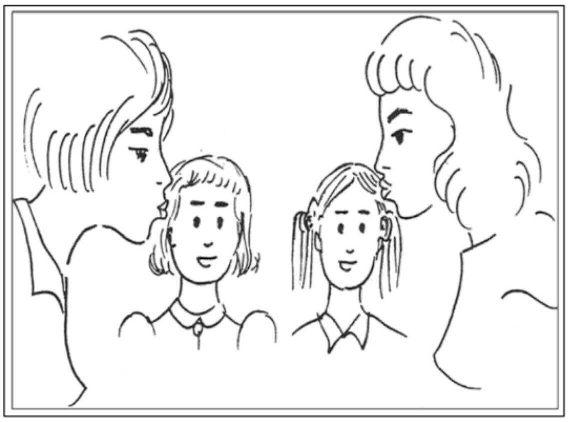


Fig.3 *Examples of sentences and pictures included in the task (adapted from Finocchiaro et al., 2015)*

Material

Each stimulus consisted of a black-and-white drawing and an active or passive reversible sentence. The sentence could correspond to the picture (matching stimuli), or to a thematic or morphological foil (mismatching stimuli). Foils were structurally identical to the experimental stimuli but were not included in the analyses. Overall, 128 stimuli (picture + sentence) were considered for the analyses, divided in four experimental blocks of 32 stimuli each.

Design

The experiment followed a 2x2x3 design with three factors: DIATHESIS (*active* vs *passive*), CORRESPONDENCE (*matching* vs *mismatching*) and STIMULATION SITE (*P1* vs *P2* vs *P3*).

TMS protocol

Three consecutive pulses were administered at a frequency of 5 Hz starting from stimuli onset. Stimulation targets were localized on individual MRI-based 3D reconstructions to drive neuronavigated administration of TMS.

Summary of results

TMS influenced performance only when applied over the posterior parietal site (P3). Accuracy on passive sentences was significantly increased by stimulation; conversely, accuracy on active sentences was significantly decreased by stimulation. RTs were not affected by TMS.

Experiment 2 (Finocchiaro et al., 2021)

Overview

This experiment aimed at investigating the role of verb semantics in the assignment of thematic roles. rTMS was applied over the posterior third of the l-IPS (P3 in Experiment 1) while participants were undergoing a sentence comprehension task that included both real sentences and pseudo-sentences. The real and pseudo-sentences were followed by two alternatives, one corresponding to the agent and one to the theme. The task required to identify either the agent or the theme by pressing two different keys, one for the alternative on top and one for the alternative on bottom. The position of the alternative that matched with the agent or the theme was counterbalanced and randomized. Stimulation was administered in two temporal windows: T1 (between 200 and 400 ms) and T2 (between 600 and 800 ms) (**Fig.4**). Real-TMS was compared to sham-TMS. Both reaction times and performance accuracy were collected as behavioral measures.

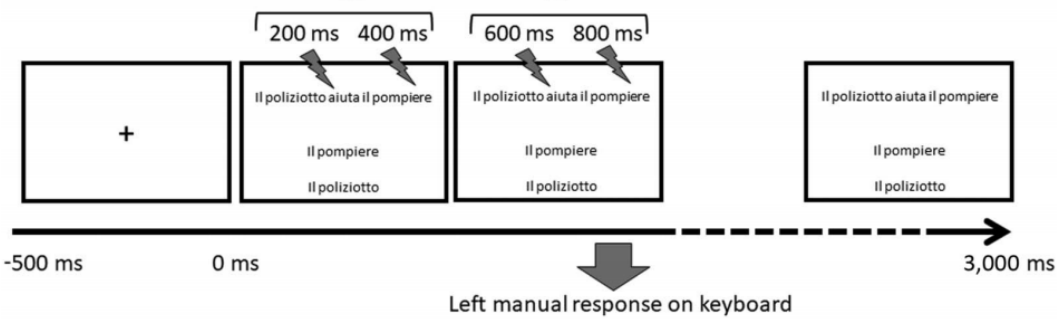


Fig.4 *Timeline of the experimental task (taken from Finocchiaro et al., 2021)*

Material

The dataset included 120 active and passive reversible real and pseudo-sentences divided in 3 blocks of 40 stimuli. All the stimuli were included in the analyses.

Design

The experiment was based on a 2x2x2x2 repeated-measures design with 4 factors: DIATHESIS (*active* vs *passive*), SEMANTICS (*real sentences* vs *pseudo-sentences*), TMS (*real* vs *sham*) and TIME (*T1* vs *T2*).

TMS protocol

Repetitive double pulse TMS was administered on the l-IPS during task execution at two temporal windows (T1 and T2). The stimulation target was determined using the 10-20 EEG system coordinates on the scalp and based on coordinates resulting from probabilistic atlases (**Fig.5**) and TMS administration was driven by neuronavigation.

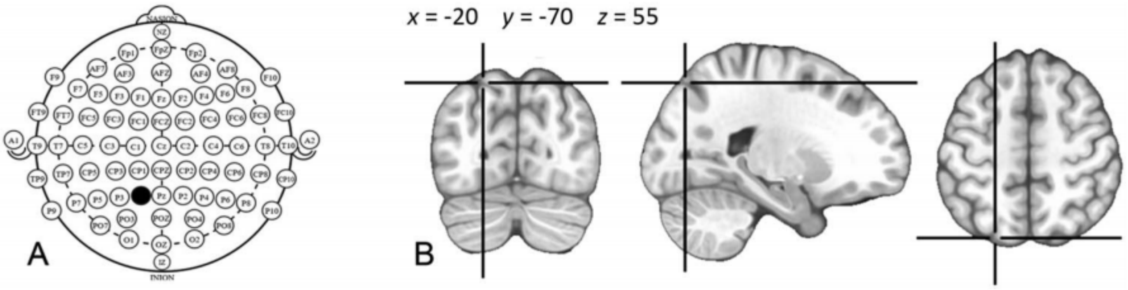
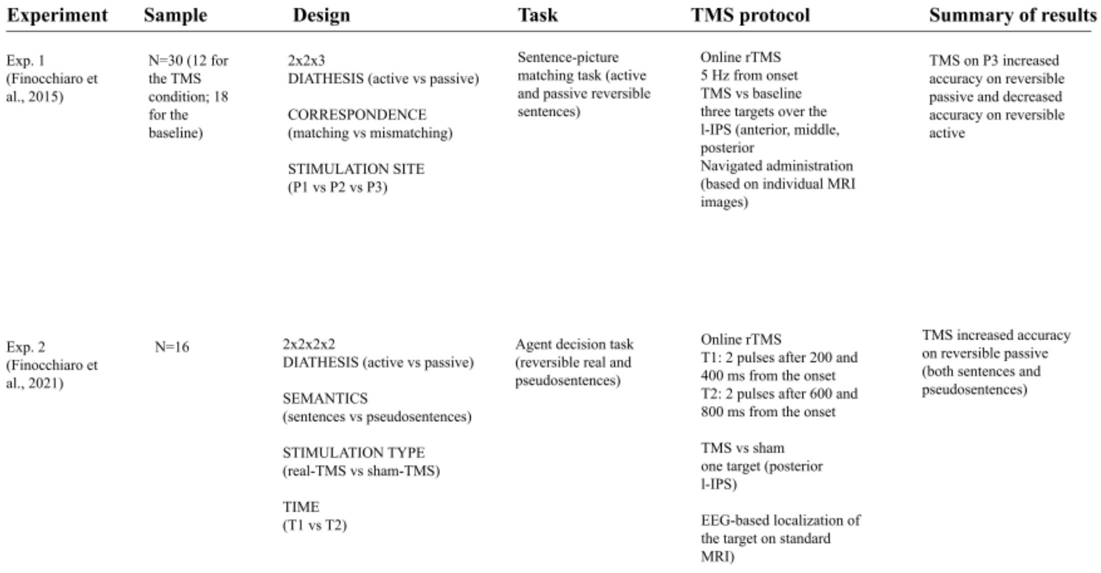


Fig.5 *Localization of the stimulation target on the 10-20 system (A) and based on statistical maps retrieved from probabilistic atlases (B) (taken from Finocchiaro et al., 2021)*

Summary of results

TMS improved performance accuracy only on passive diathesis within both real and pseudo-sentences, independently of the timing of stimulation. Analyses on RTs did not show any significant effects.

Fig. 6 summarizes the main parameters of the two experiments:



Experiment	Sample	Design	Task	TMS protocol	Summary of results
Exp. 1 (Finocchiaro et al., 2015)	N=30 (12 for the TMS condition; 18 for the baseline)	2x2x3 DIATHESIS (active vs passive) CORRESPONDENCE (matching vs mismatching) STIMULATION SITE (P1 vs P2 vs P3)	Sentence-picture matching task (active and passive reversible sentences)	Online rTMS 5 Hz from onset TMS vs baseline three targets over the l-IPS (anterior, middle, posterior Navigated administration (based on individual MRI images)	TMS on P3 increased accuracy on reversible passive and decreased accuracy on reversible active
Exp. 2 (Finocchiaro et al., 2021)	N=16	2x2x2x2 DIATHESIS (active vs passive) SEMANTICS (sentences vs pseudosentences) STIMULATION TYPE (real-TMS vs sham-TMS) TIME (T1 vs T2)	Agent decision task (reversible real and pseudosentences)	Online rTMS T1: 2 pulses after 200 and 400 ms from the onset T2: 2 pulses after 600 and 800 ms from the onset TMS vs sham one target (posterior l-IPS) EEG-based localization of the target on standard MRI)	TMS increased accuracy on reversible passive (both sentences and pseudosentences)

Fig. 6 *Summary of the two experiments parameters*

STATISTICAL ANALYSES: OVERVIEW

Assuming that in each of the two experiments the data were originally analyzed either only with ANOVA (Experiment 1) or only with LMMs (Experiment 2), our choice was to analyze the data with the model that had not been used in the original analyses and to re-analyze the data with the same model that was originally used. In order to do this, two conditions had to be met:

1. The dataset to be analyzed had to be completely identical to the original dataset, without any changes in data treatment whatsoever
2. The software used had to be the same for both models, both those new to the original work and those already used in the original work

The choice to re-analyze the data with the same model as originally used followed the principle of reproducibility: if starting from the same data and using the same analytical and statistical approach yields the same results, it means that those results are reproducible and replicable. In addition, redoing the analyses with the same software used to perform LMMs allowed us to reduce the risk of confounds related to the type of analysis software.

The choice of software (*jamovi*) was due to several factors: (i) it is an open source program freely downloadable and usable, (ii) it is a GUI-based software that is very intuitive and easy to use, and (iii) it offers the possibility of installing a package specifically developed for LMMs, an option not available in other GUI-based software such as JASP or SPSS.

For LMMs, several models of decreasing complexity were run, but a single model was chosen and reported. This was the best fitting model that best explained the variability in the data. In all the analyses, the final model had all the fixed factors as fixed-effects and the fixed factors and the intercept across participants as random-effects. In one analysis (mixed logistic model on ACC, p. 73), the factor SITE was added to the intercept across participants as random-effects.

RESULTS

Experiment 1

ANOVA

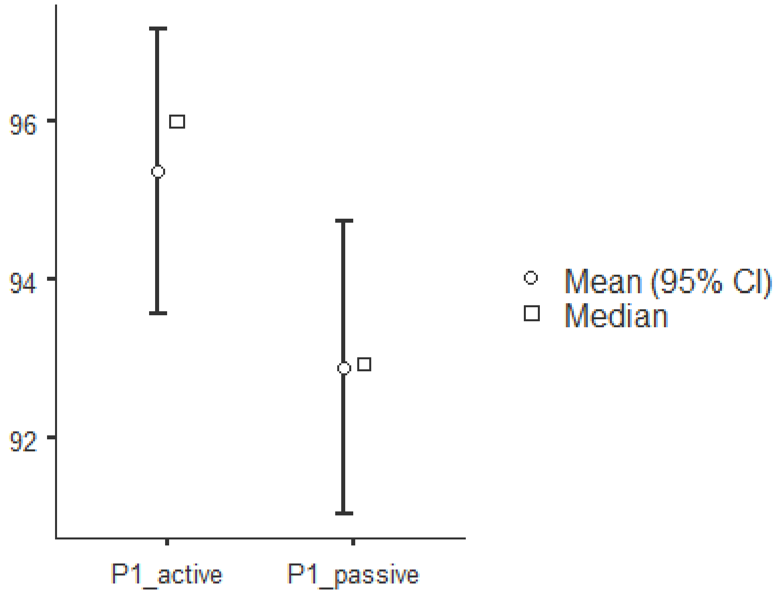
The model includes three within-participants factors:
DIATHESIS (*active* vs *passive*)
CORRESPONDENCE (*matching* vs *mismatching*)
SITE (*P1* vs *P2* vs *P3*)

The focus of the analyses was the relationship between TMS and the other factors to see whether stimulation had an effect depending on the stimulated site. To this end, a 3-way repeated-measures ANOVA was performed both on RT and ACC. On RT, no main effect of SITE or any interaction was found. On ACC, TMS interacted with DIATHESIS ($F(2,22)=8.44$, $p=0.002$). This interaction was further explored via a 2-way repeated-measures ANOVA with DIATHESIS and SITE as factors. Descriptives are shown in **Tab.1**. Results showed an interaction between the two factors ($F(2,22)=10.29$, $p < 0.001$).

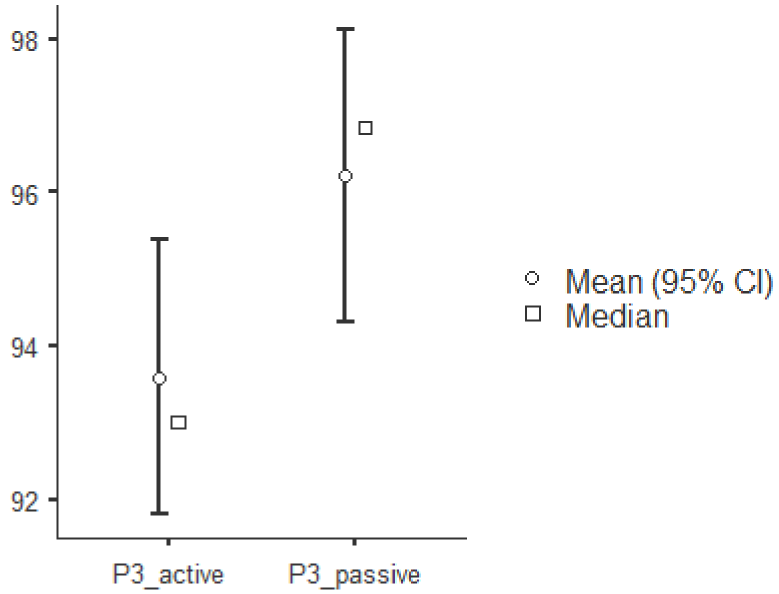
	P1 active	**P1 passive**	**P2 active**	**P2 passive**	**P3 active**	**P3 passive**
mean	95.4	92.9	94.9	93.4	93.6	96.2
std	3.1	3.2	3.4	5.5	3.1	3.3

Tab.1 *Descriptives of ACC (mean and standard deviation) for active and passive sentences in each stimulation site*

The interaction was explored through paired t-tests that showed performance on active vs passive sentences to differ significantly in P1 (t=2.76, p=0.018) and in P3 (t= -3.14, p=0.009). As shown in **Graphs 1** and **2**, ACC on active sentences was decreased by stimulation on P1, and ACC on passive sentences was improved by stimulation on P3.



Graph 1 *Plot of the mean ACC of active and passive sentences on P1*



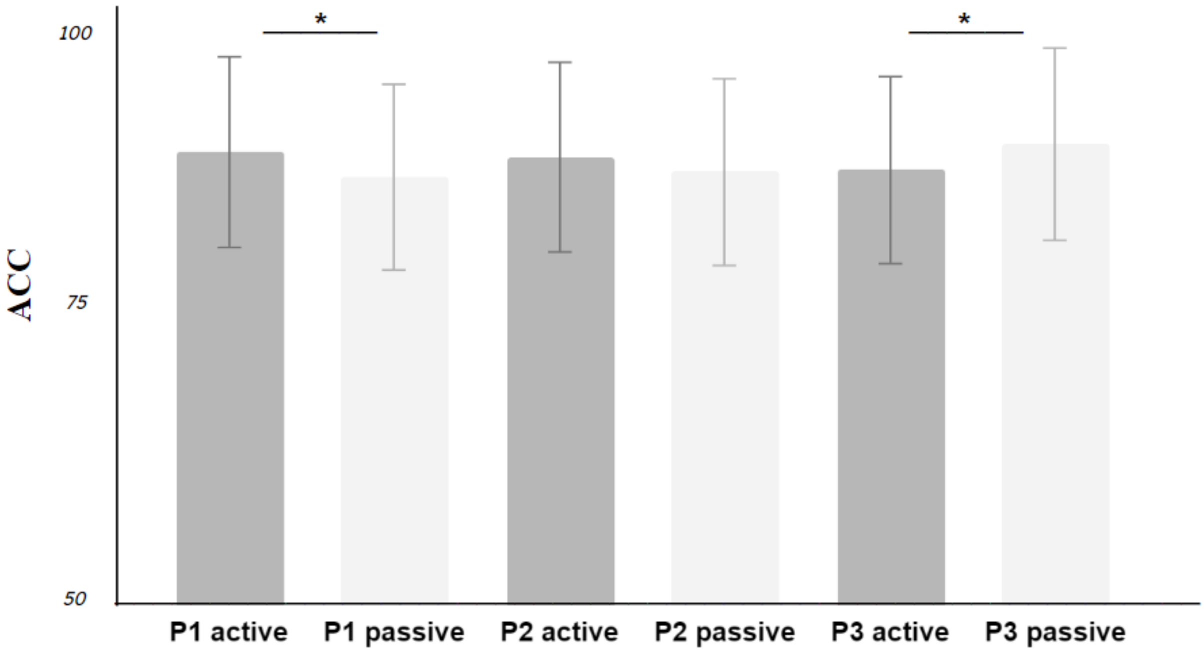
Graph 2 *Plot of the mean ACC of active and passive sentences on P3*

LMMs

Response times were analyzed through the *mixed effects* module of the GAMLj, whereas performance accuracy was analyzed using a logistic model of the same module (*generalized mixed model*).

The model had the factors DIATHESIS, SITE and CORRESPONDENCE as fixed-effects and the intercepts across subjects as random-effects. The other factors and the interactions were not included as random-effects since they did not contribute to explain the overall variability or gave rise to correlations that were too high to estimate the model.

Results were in line with the outcomes of ANOVA. RTs were not influenced by stimulation on any site. On ACC the factor SITE interacted with DIATHESIS alone ($X^2(2)=12.739$, $p=0.002$) and with DIATHESIS and CORRESPONDENCE ($X^2(2)=7.818$, $p=0.020$). This 3-way interaction was not detected by ANOVA, but the 2-way interaction was identical in both models. As done above, a similar model including SITE and DIATHESIS was used to further explore the interaction. Results showed a significant interaction ($X^2(2)=10.938$, $p=0.004$). Post-hoc comparisons showed significant differences between active and passive sentences on P1 ($z=2.073$, $p=0.038$) and on P3 ($z= -2.358$, $p=0.018$), in the same direction as indicated in ANOVA. When stimulating P1 ACC was decreased on active sentences, whereas ACC improved on passive sentences when TMS was applied over P3 (see **Graph 3**).

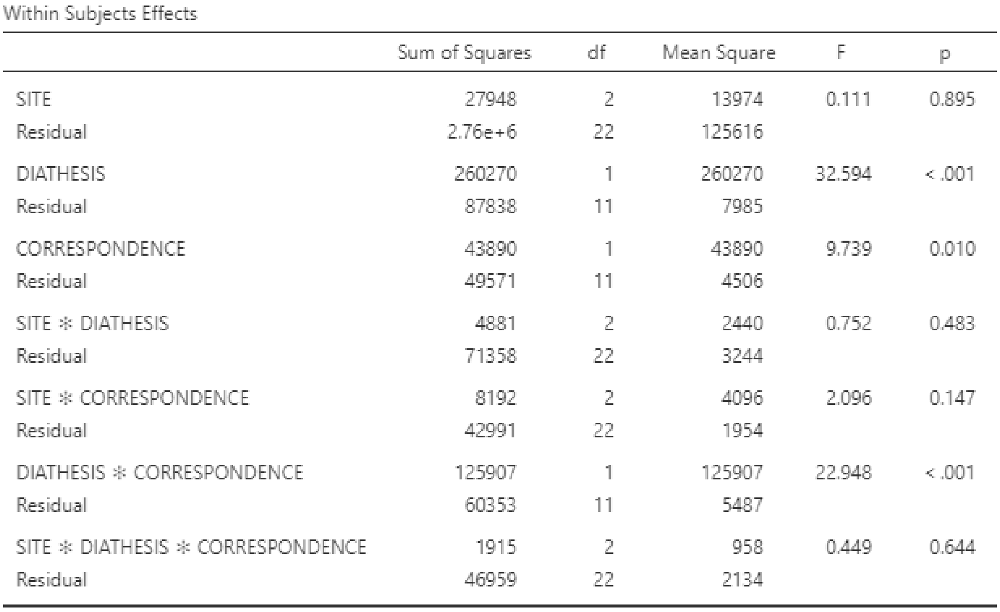


Graph 3 *Barplot of the contrast between mean ACC in active and passive sentences on each stimulation target (P1, P2, P3)*

Comparison between models

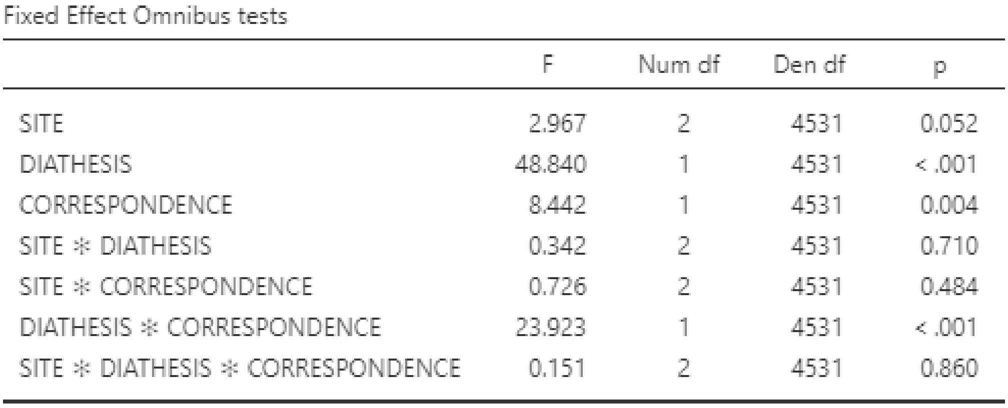
ANOVA and LMMs provided similar results and allow the same inferences. However, there are also differences between the two models that are worth focusing on.

First, we compare the outputs of the analyses with all the three factors (DIATHESIS, CORRESPONDENCE, SITE). RTs analyses are shown in **Tab.2**.



Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
SITE	27948	2	13974	0.111	0.895
Residual	2.76e+6	22	125616		
DIATHESIS	260270	1	260270	32.594	< .001
Residual	87838	11	7985		
CORRESPONDENCE	43890	1	43890	9.739	0.010
Residual	49571	11	4506		
SITE ✻ DIATHESIS	4881	2	2440	0.752	0.483
Residual	71358	22	3244		
SITE ✻ CORRESPONDENCE	8192	2	4096	2.096	0.147
Residual	42991	22	1954		
DIATHESIS ✻ CORRESPONDENCE	125907	1	125907	22.948	< .001
Residual	60353	11	5487		
SITE ✻ DIATHESIS ✻ CORRESPONDENCE	1915	2	958	0.449	0.644
Residual	46959	22	2134		

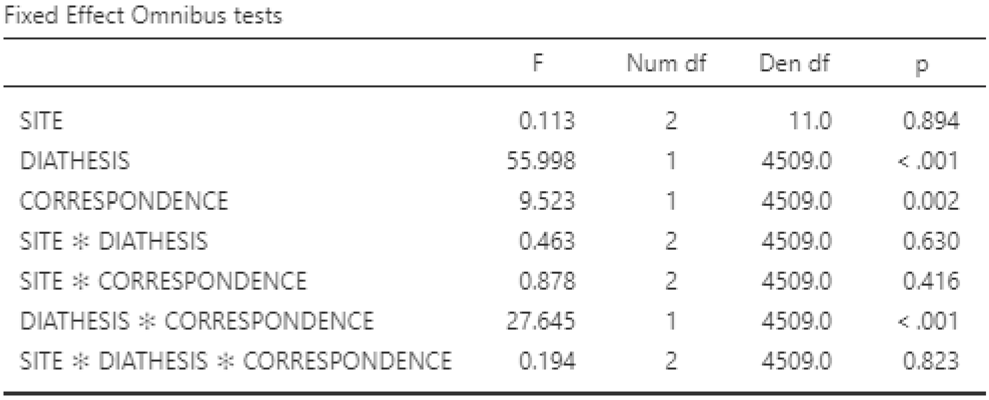


Fixed Effect Omnibus tests

	F	Num df	Den df	p
SITE	2.967	2	4531	0.052
DIATHESIS	48.840	1	4531	< .001
CORRESPONDENCE	8.442	1	4531	0.004
SITE ✻ DIATHESIS	0.342	2	4531	0.710
SITE ✻ CORRESPONDENCE	0.726	2	4531	0.484
DIATHESIS ✻ CORRESPONDENCE	23.923	1	4531	< .001
SITE ✻ DIATHESIS ✻ CORRESPONDENCE	0.151	2	4531	0.860

Tab.2 *Output of the 3-way repeated-measures ANOVA (on top) and LMMs (on bottom) performed on RT with SITE, DIATHESIS and CORRESPONDENCE as factors*

The outcomes of the two models substantially overlap. Neither ANOVA nor LMMs show significant TMS-related effects. In contrast to ANOVA, in LMMs the main effect of SITE falls just short of significance (*p* = 0.052). This effect is not particularly relevant for the issues investigated in the experiment, but represents the main difference between the two models. A possible explanation for the discrepancy could be that in our mixed model only the intercept across subjects was included as a random effect. Other sources of across-subject variability (such as stimulation site, sentence diathesis and sentence/picture correspondence) were not included and the model could not take them into account. The effect of SITE could be related to the fact that, if the differences between stimulation on P1, P2 and P3 were not homogeneous across subjects, the amount of variability carried out by these differences is dispersed in the global variability between the subjects. If this variability is added to the model as a random-effect, the main effect moves significantly away from significance, as shown in **Tab.3**:

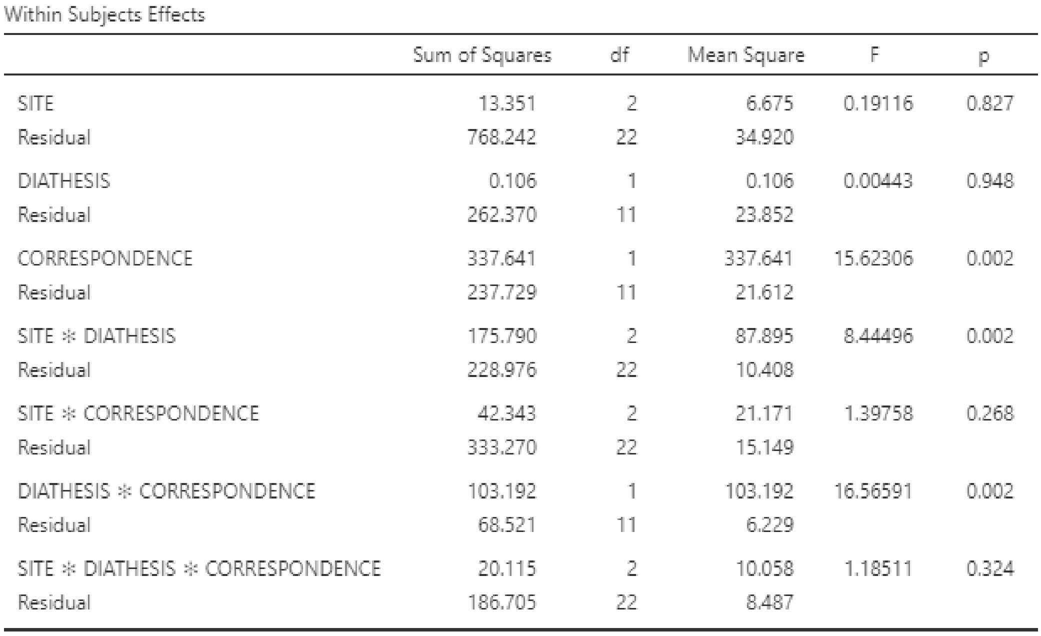


Fixed Effect Omnibus tests

	F	Num df	Den df	p
SITE	0.113	2	11.0	0.894
DIATHESIS	55.998	1	4509.0	< .001
CORRESPONDENCE	9.523	1	4509.0	0.002
SITE ✻ DIATHESIS	0.463	2	4509.0	0.630
SITE ✻ CORRESPONDENCE	0.878	2	4509.0	0.416
DIATHESIS ✻ CORRESPONDENCE	27.645	1	4509.0	< .001
SITE ✻ DIATHESIS ✻ CORRESPONDENCE	0.194	2	4509.0	0.823

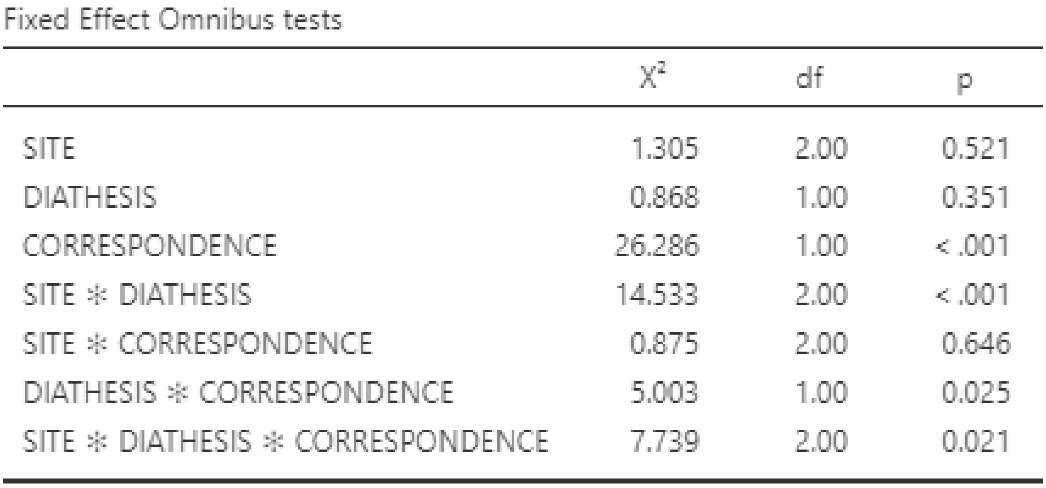
Tab.3 *Output of the LMM on RT with the intercept and SITE across subjects as random-effects*

This outcome suggests that the model that includes the amount of variability of SITE does not account for the overall variability in the data better than the simpler model that includes only the intercept. In contrast, on ACC both the models show relevant effects of TMS (**Tab.4**):



Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
SITE	13.351	2	6.675	0.19116	0.827
Residual	768.242	22	34.920		
DIATHESIS	0.106	1	0.106	0.00443	0.948
Residual	262.370	11	23.852		
CORRESPONDENCE	337.641	1	337.641	15.62306	0.002
Residual	237.729	11	21.612		
SITE ✻ DIATHESIS	175.790	2	87.895	8.44496	0.002
Residual	228.976	22	10.408		
SITE ✻ CORRESPONDENCE	42.343	2	21.171	1.39758	0.268
Residual	333.270	22	15.149		
DIATHESIS ✻ CORRESPONDENCE	103.192	1	103.192	16.56591	0.002
Residual	68.521	11	6.229		
SITE ✻ DIATHESIS ✻ CORRESPONDENCE	20.115	2	10.058	1.18511	0.324
Residual	186.705	22	8.487		



Fixed Effect Omnibus tests

	X²	df	p
SITE	1.305	2.00	0.521
DIATHESIS	0.868	1.00	0.351
CORRESPONDENCE	26.286	1.00	< .001
SITE ✻ DIATHESIS	14.533	2.00	< .001
SITE ✻ CORRESPONDENCE	0.875	2.00	0.646
DIATHESIS ✻ CORRESPONDENCE	5.003	1.00	0.025
SITE ✻ DIATHESIS ✻ CORRESPONDENCE	7.739	2.00	0.021

Tab.4 *Output of the 3-way repeated-measures ANOVA (on top) and LMM (on bottom) performed on ACC with SITE, DIATHESIS and CORRESPONDENCE as factors*

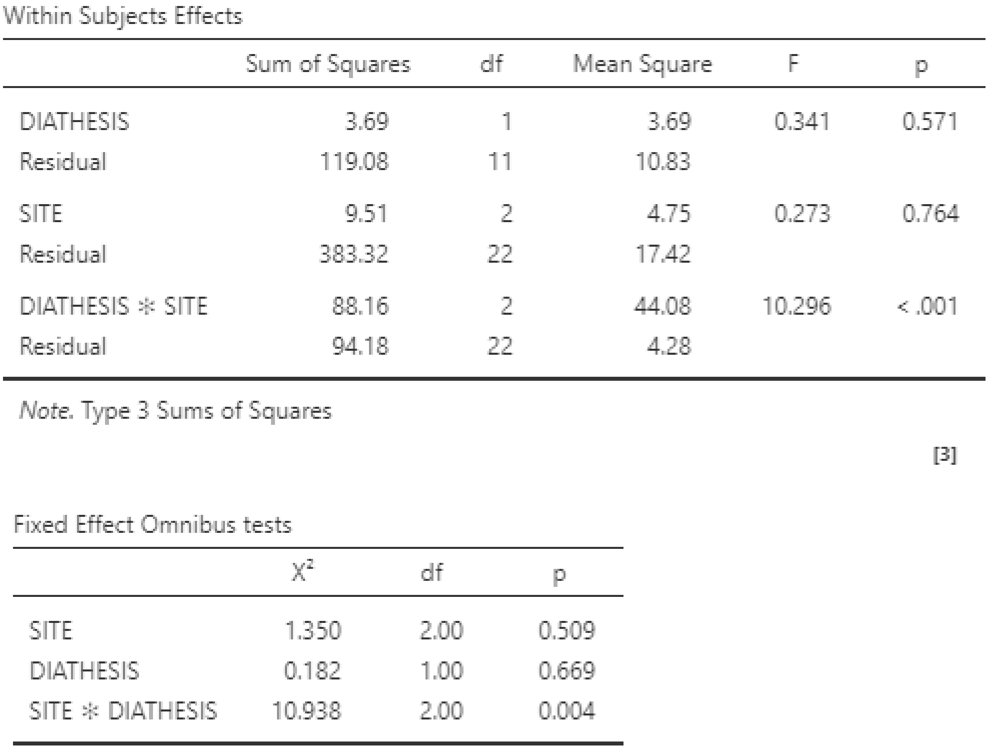
Both models show a 2-way interaction between SITE and DIATHESIS. A difference between the two models is represented by the outcome of the 3-way interaction SITE x DIATHESIS x CORRESPONDENCE. This interaction is not significant in the ANOVA, whereas it reaches significance in LMM. Perhaps this discrepancy stems from the fact that the interaction detected by ANOVA is not strong enough to survive a more in-depth investigation of data variability, provided by the logistic model. See **Tab.5** for a summary of the effects and interactions of each factor in the two models.

RT		*Effects*						
	Factors	SITE	DIATH	CORRESP	SITE * DIATH	SITE * CORRESP	DIATH * CORRESP	SITE * DIATH*CORRESP
Model	ANOVA	no	yes	yes	no	no	yes	no
	LMMs	yes	yes	yes	no	no	yes	no

ACC		*Effects*						
	Factors	SITE	DIATH	CORRESP	SITE * DIATH	SITE * CORRESP	DIATH * CORRESP	SITE * DIATH*CORRESP
Model	ANOVA	no	no	yes	yes	no	yes	no
	LMMs	no	no	yes	yes	no	yes	yes

Tab.5 *Summary of the effects of ANOVA and LMM. Legenda: yes=significant effect; no=non significant effect)*

The significant interaction was further explored by a 2-way ANOVA with the factors SITE and DIATHESIS and a mixed logistic model with the same factors. The outputs of the two models are shown in **Tab.6**:



Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
DIATHESIS	3.69	1	3.69	0.341	0.571
Residual	119.08	11	10.83		
SITE	9.51	2	4.75	0.273	0.764
Residual	383.32	22	17.42		
DIATHESIS ✻ SITE	88.16	2	44.08	10.296	< .001
Residual	94.18	22	4.28		

Note. Type 3 Sums of Squares

[3]

Fixed Effect Omnibus tests

	X^2	df	p
SITE	1.350	2.00	0.509
DIATHESIS	0.182	1.00	0.669
SITE ✻ DIATHESIS	10.938	2.00	0.004

Tab.6 *Output of the 2-way repeated-measures ANOVA (on top) and LMM (on bottom) performed on ACC with SITE and DIATHESIS as factors*

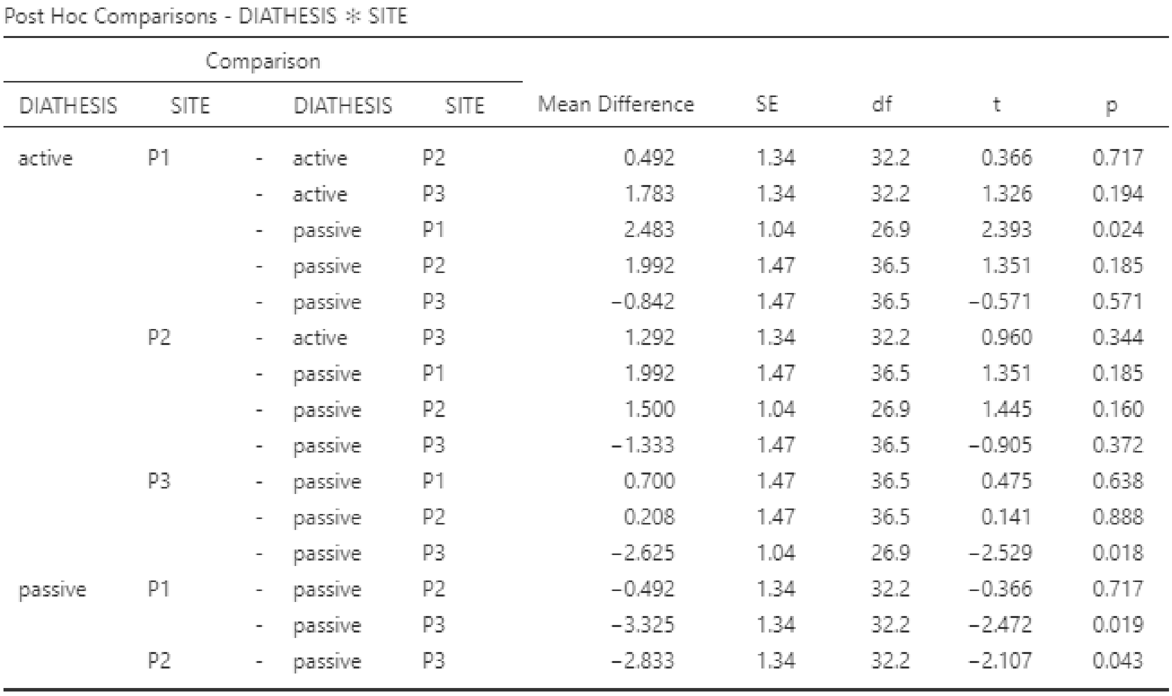
The outcomes of the two models are similar, as they both show the 2 way interaction (see **Tab.7** for a summary).

RT		*Effect*		
	Factors	SITE	DIATH	SITE * DIATH
Model	ANOVA	no	no	no
	LMMs	no	no	no

ACC		*Effect*		
	Factors	SITE	DIATH	SITE * DIATH
Model	ANOVA	no	no	yes
	LMMs	no	no	yes

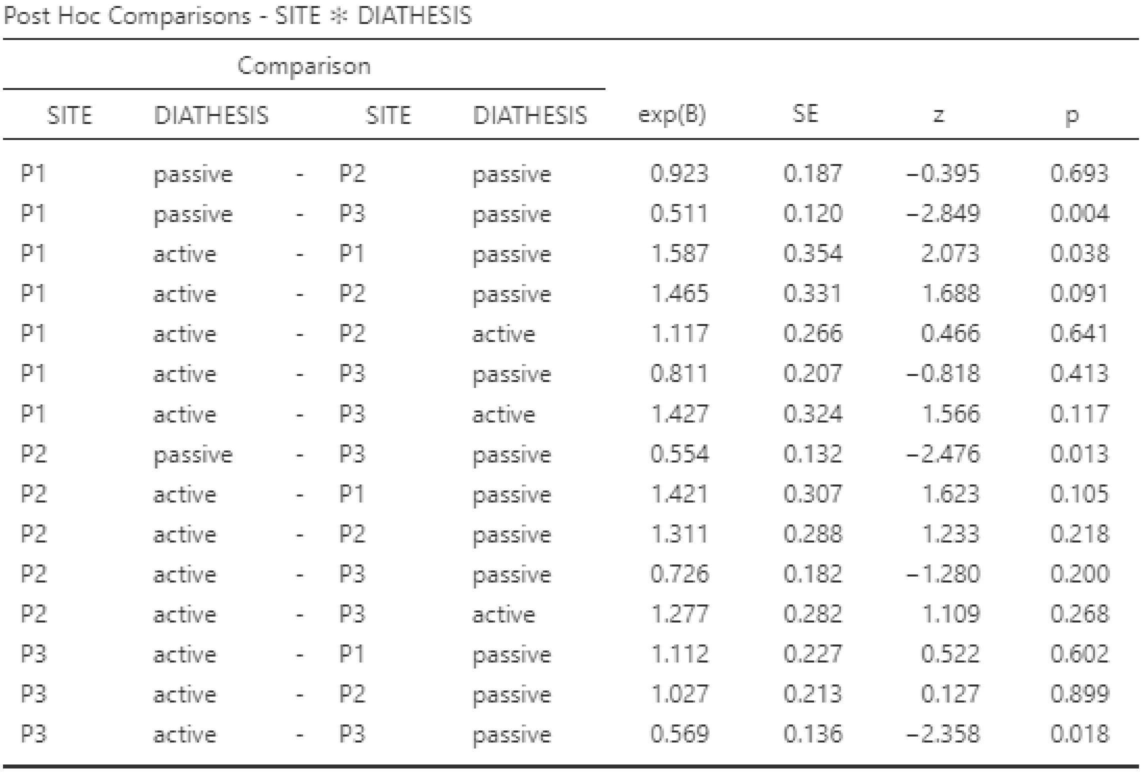
Tab.7 *Summary of the effects of ANOVA and LMM. Legenda: yes=significant effect; no=non significant effect)*

Post hoc tests also show converging results. When stimulating P1, active sentences were systematically more accurate than passive sentences and TMS on P3 significantly improved ACC on passive sentences. Outputs are shown in **Tab.8**.



Post Hoc Comparisons - DIATHESIS ✻ SITE

Comparison									
DIATHESIS	SITE		DIATHESIS	SITE	Mean Difference	SE	df	t	p
active	P1	-	active	P2	0.492	1.34	32.2	0.366	0.717
		-	active	P3	1.783	1.34	32.2	1.326	0.194
		-	passive	P1	2.483	1.04	26.9	2.393	0.024
		-	passive	P2	1.992	1.47	36.5	1.351	0.185
		-	passive	P3	−0.842	1.47	36.5	−0.571	0.571
	P2	-	active	P3	1.292	1.34	32.2	0.960	0.344
		-	passive	P1	1.992	1.47	36.5	1.351	0.185
		-	passive	P2	1.500	1.04	26.9	1.445	0.160
		-	passive	P3	−1.333	1.47	36.5	−0.905	0.372
	P3	-	passive	P1	0.700	1.47	36.5	0.475	0.638
		-	passive	P2	0.208	1.47	36.5	0.141	0.888
		-	passive	P3	−2.625	1.04	26.9	−2.529	0.018
passive	P1	-	passive	P2	−0.492	1.34	32.2	−0.366	0.717
		-	passive	P3	−3.325	1.34	32.2	−2.472	0.019
	P2	-	passive	P3	−2.833	1.34	32.2	−2.107	0.043



Post Hoc Comparisons - SITE ✻ DIATHESIS

Comparison								
SITE	DIATHESIS		SITE	DIATHESIS	exp(B)	SE	z	p
P1	passive	-	P2	passive	0.923	0.187	−0.395	0.693
P1	passive	-	P3	passive	0.511	0.120	−2.849	0.004
P1	active	-	P1	passive	1.587	0.354	2.073	0.038
P1	active	-	P2	passive	1.465	0.331	1.688	0.091
P1	active	-	P2	active	1.117	0.266	0.466	0.641
P1	active	-	P3	passive	0.811	0.207	−0.818	0.413
P1	active	-	P3	active	1.427	0.324	1.566	0.117
P2	passive	-	P3	passive	0.554	0.132	−2.476	0.013
P2	active	-	P1	passive	1.421	0.307	1.623	0.105
P2	active	-	P2	passive	1.311	0.288	1.233	0.218
P2	active	-	P3	passive	0.726	0.182	−1.280	0.200
P2	active	-	P3	active	1.277	0.282	1.109	0.268
P3	active	-	P1	passive	1.112	0.227	0.522	0.602
P3	active	-	P2	passive	1.027	0.213	0.127	0.899
P3	active	-	P3	passive	0.569	0.136	−2.358	0.018

Tab.8 *Outputs of the Posthoc comparisons (SITE x DIATHESIS) of ANOVA (on top) and LMMs (on bottom)*

The outputs of the two models overlap and show similar outcomes on the significant contrasts (P1 active vs P1 passive and P3 active vs P3 passive). The mixed logistic model is associated with a slightly lower level of significance on P1, as LMMs are generally more conservative than ANOVA in exploring data variability. In addition, the two models can also be compared based on their appropriateness to explore accuracy. In this sense, a work by **Jaeger (2008)** openly challenged the use of ANOVA for categorical outcome variables and suggested adopting mixed logistic models as a valid alternative. According to the author, while for continuous outcomes ANOVA is an reliable tool to assess whether the experimental conditions have the same means given the variances (both between and within), it would not be the case for binomially distributed variables. These variables do not generally have homogeneous variances and their exploration would then violate an ANOVA assumption. To better account for non-homogeneous variances it is suggested to use mixed logistic models. These models present great advantages compared to ANOVA. First, they can modelize random effects to take into account random sources of variability. Second, they have greater power, resulting in a better probability of detecting true effects. Finally, mixed logistic models allow us to test how much a random effect contributes to explain variability by comparing the likelihood of the model in presence or in absence of such random effect.

To conclude, the observed differences between ANOVA and LMMs can be explained by referring to the way the two models explore and account for variance in a binomially distributed outcome like accuracy.

Overall, in Experiment 1 the outcomes of ANOVA and LMMs largely overlap.

Experiment 2

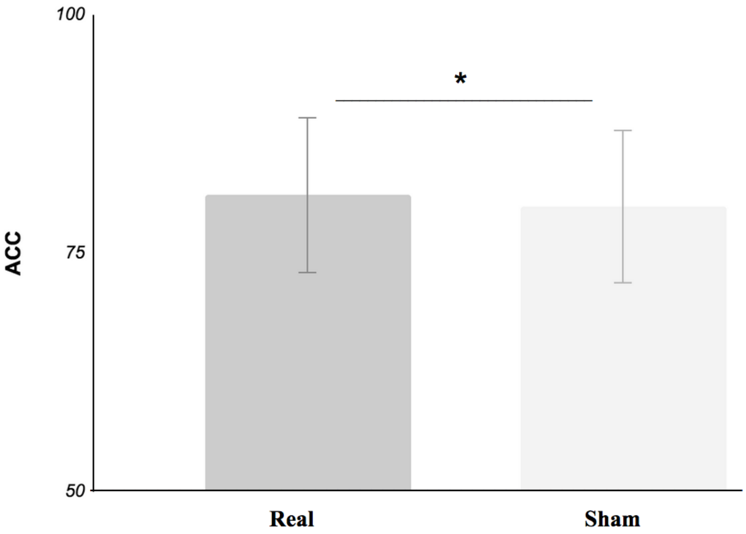
ANOVA

The analysis included all the four within-participants factors (DIATHESIS, SEMANTICS, TMS, TIME). The effects of interest were those related to the use of TMS; real-TMS was contrasted with sham-TMS to explore significant differences.

Results showed no significant TMS-related effects on RTs. On ACC the DIATHESIS*TMS ($F(1,15)=8.21$, $p=0.012$) and the DIATHESIS*SEMANTICS*TMS*TIME ($F(1,15)=7.62$, $p=0.015$) interactions were significant. This 4-way interaction was further explored by running two 3-way repeated-measures ANOVA with 3 factors (TIME, SEMANTICS, TMS), one for active and one for passive sentences. Results show a significant main effect of TMS only on passive sentences ($F(1,15)=7.08$, $p=0.018$), independent of the timing of stimulation and on the semantics of the verb (descriptives are shown in **Tab.9**). Performance accuracy was improved by TMS only on passive sentences (**Graph 4**).

ACC	**Real active**	**Real passive**	**Sham active**	**Sham passive**
mean	80	81.1	87.8	79.9
std	12.8	16.6	17.3	14.4

Tab.9 *Descriptives (mean accuracy and standard deviation) for both active and passive sentences in the real vs sham contrast*



Graph 4 *Mean accuracy in the real vs sham contrast for passive sentences*

LMMs

The model included all the within-participants factors (DIATHESIS, TIME, SEMANTICS, TMS) as fixed-effects and the intercept as random-effect. Analyses of RT did not show significant effects. On ACC only the TMS*DIATHESIS interaction reached significance ($X^2(1)=14.1$, $p < 0.001$). Post hoc analyses showed that TMS influenced performance accuracy only on passive sentences ($z= -3.29$, $p=0.006$). In these sentences ACC was systematically improved by stimulation.

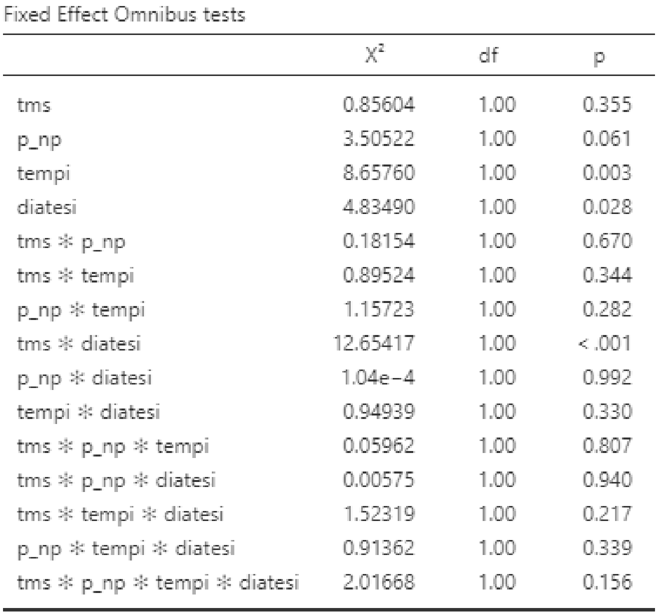
Comparison of ANOVA and LMM

The outcomes of the two models were very similar. Both showed the same TMS-related effects on performance accuracy and not on response times. The outputs of the models are shown in **Tab.10**:

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
TEMPI	0.05109	1	0.05109	7.5025	0.015
Residual	0.10215	15	0.00681		
DIATESI	0.02545	1	0.02545	0.8197	0.380
Residual	0.46570	15	0.03105		
TMS	0.01501	1	0.01501	0.5736	0.461
Residual	0.39256	15	0.02617		
SEMANTICA	0.02211	1	0.02211	2.0619	0.172
Residual	0.16084	15	0.01072		
TEMPI ✻ DIATESI	0.00205	1	0.00205	0.5906	0.454
Residual	0.05216	15	0.00348		
TEMPI ✻ TMS	0.00673	1	0.00673	1.0827	0.315
Residual	0.09318	15	0.00621		
DIATESI ✻ TMS	0.08752	1	0.08752	8.2147	0.012
Residual	0.15981	15	0.01065		
TEMPI ✻ SEMANTICA	7.65e−4	1	7.65e−4	0.2338	0.636
Residual	0.04909	15	0.00327		
DIATESI ✻ SEMANTICA	2.16e−4	1	2.16e−4	0.0637	0.804
Residual	0.05089	15	0.00339		
TMS ✻ SEMANTICA	0.00273	1	0.00273	0.5512	0.469
Residual	0.07421	15	0.00495		
TEMPI ✻ DIATESI ✻ TMS	0.00323	1	0.00323	1.7705	0.203
Residual	0.02738	15	0.00183		
TEMPI ✻ DIATESI ✻ SEMANTICA	0.00563	1	0.00563	2.0451	0.173
Residual	0.04128	15	0.00275		
TEMPI ✻ TMS ✻ SEMANTICA	0.00272	1	0.00272	0.5658	0.464
Residual	0.07208	15	0.00481		
DIATESI ✻ TMS ✻ SEMANTICA	0.00255	1	0.00255	0.3083	0.587
Residual	0.12402	15	0.00827		
TEMPI ✻ DIATESI ✻ TMS ✻ SEMANTICA	0.01128	1	0.01128	7.6264	0.015
Residual	0.02219	15	0.00148		

Note. Type 3 Sums of Squares



Fixed Effect Omnibus tests

	X²	df	p
tms	0.85604	1.00	0.355
p_np	3.50522	1.00	0.061
tempi	8.65760	1.00	0.003
diatesi	4.83490	1.00	0.028
tms ✻ p_np	0.18154	1.00	0.670
tms ✻ tempi	0.89524	1.00	0.344
p_np ✻ tempi	1.15723	1.00	0.282
tms ✻ diatesi	12.65417	1.00	< .001
p_np ✻ diatesi	1.04e−4	1.00	0.992
tempi ✻ diatesi	0.94939	1.00	0.330
tms ✻ p_np ✻ tempi	0.05962	1.00	0.807
tms ✻ p_np ✻ diatesi	0.00575	1.00	0.940
tms ✻ tempi ✻ diatesi	1.52319	1.00	0.217
p_np ✻ tempi ✻ diatesi	0.91362	1.00	0.339
tms ✻ p_np ✻ tempi ✻ diatesi	2.01668	1.00	0.156

Tab.10 *Outputs of the 4-way repeated measures ANOVA (on top) and the mixed logistic regression (on bottom) on ACC*

Both the models detected a significant main effect of TIME and an interaction effect of TMS*DIATHESIS. Compared to ANOVA, LMMs also showed a main effect of DIATHESIS, but did not show a TMS*DIATHESIS*TIME*SEMANTICS evidenced by ANOVA (see **Tab.11** for a summary of the outcomes of the two models).

ACC		*Main Effects*			
	Factors	TIME	TMS	DIATH	SEMANT
Model	ANOVA	yes	no	no	no
	LMMs	yes	no	yes	no

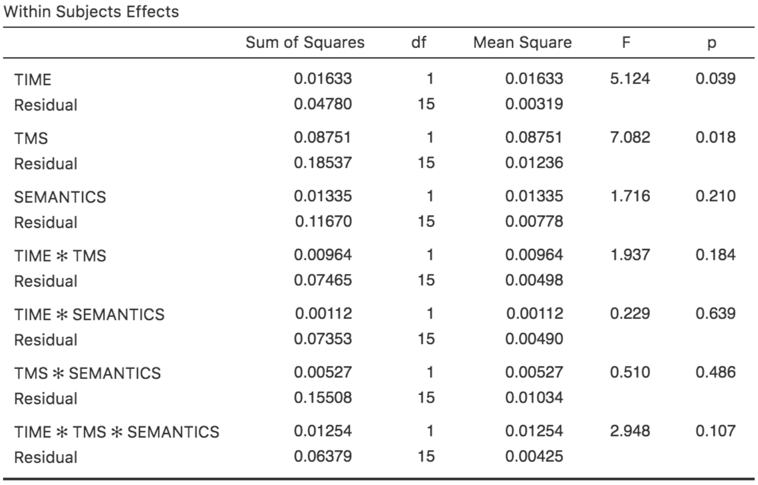
ACC		*Interactions*		
	Factors	TIME*TMS	TIME*DIATH	TIME*SEMANT
Model	ANOVA	no	no	no
	LMMs	no	no	no

ACC		*Interactions*		
	Factors	DIATH* SEMANT	TMS*DIATH	TMS* SEMANT
Model	ANOVA	no	yes	no
	LMMs	no	yes	no

ACC		*Interactions*			
	Factors	TIME*DIATH *SEMANT	TMS*DIATH*S EMANT	TMS*TIME* SEMANT	TMS*TIME*SEMANT *DIATH
Model	ANOVA	no	no	no	yes
	LMMs	no	no	no	no

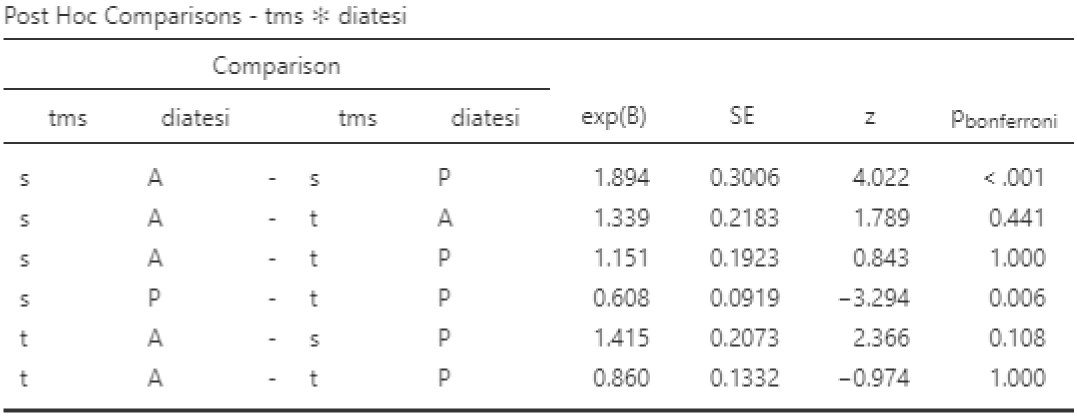
Tab.11 *Summary of the two models outcomes for each effect and interaction (yes=significant effect; no=non significant effect)*

ANOVA and LMM also converge in showing a facilitatory effect of TMS on passive diathesis, independent of verb semantics (real vs pseudo-sentences) and on the timing of stimulation (**Tab.12**):



Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
TIME	0.01633	1	0.01633	5.124	0.039
Residual	0.04780	15	0.00319		
TMS	0.08751	1	0.08751	7.082	0.018
Residual	0.18537	15	0.01236		
SEMANTICS	0.01335	1	0.01335	1.716	0.210
Residual	0.11670	15	0.00778		
TIME ✻ TMS	0.00964	1	0.00964	1.937	0.184
Residual	0.07465	15	0.00498		
TIME ✻ SEMANTICS	0.00112	1	0.00112	0.229	0.639
Residual	0.07353	15	0.00490		
TMS ✻ SEMANTICS	0.00527	1	0.00527	0.510	0.486
Residual	0.15508	15	0.01034		
TIME ✻ TMS ✻ SEMANTICS	0.01254	1	0.01254	2.948	0.107
Residual	0.06379	15	0.00425		



Post Hoc Comparisons - tms ✻ diatesi

Comparison								
tms	diatesi		tms	diatesi	exp(B)	SE	z	$p_{bonferroni}$
s	A	-	s	P	1.894	0.3006	4.022	< .001
s	A	-	t	A	1.339	0.2183	1.789	0.441
s	A	-	t	P	1.151	0.1923	0.843	1.000
s	P	-	t	P	0.608	0.0919	−3.294	0.006
t	A	-	s	P	1.415	0.2073	2.366	0.108
t	A	-	t	P	0.860	0.1332	−0.974	1.000

Tab.12 *Outputs of the two models (ANOVA on top, LMMs on bottom)*

The two outputs describe the same situation. As mentioned, two 3-way repeated-measures ANOVA were run (TIME, SEMANTICS and TMS), one on active and one on passive diathesis. The output reported on top comes from the analysis on passive diathesis only, which showed a significant effect of TMS on ACC. Since no interaction effects reached significance, it can be concluded that TMS influenced performance on passive diathesis in all sentence types (real and pseudo-sentences) and stimulation times (T1 and T2). The output at the bottom reports the post hoc analyses that were run to further explore the significant TMS*DIATHESIS interaction detected by LMMs. It shows that TMS affected ACC only on passive sentences.

Overall -as in Experiment 1- the outcomes of the two models substantially matched.

DISCUSSION

This work compared the outcomes of two commonly used methodological approaches to data analysis (ANOVA and LMMs) applied to the same dataset. Building up from the evidence reported by a number of neuroimaging studies on the link between the pipeline used and the results obtained (**Ioannidis, 2005**; **Carp, 2012**; **Weissenbacher et al., 2009**; **Carp, 2011**; **Smith et al., 2005**; **Poline et al., 2006**; **Churchill et al., 2012a,b**; **Botvinik-Nezer, 2020**), the aim of this comparison was to explore the implications of the problem for behavioral and neurostimulation data.

The comparison between the two models revealed substantial commonalities, but also discrepancies. Observed similarities highlight that while ANOVA and LMM do not share basic assumptions, they retrieve and return similar information content. Differences must be interpreted with caution. The greatest risk lies in placing the two models on a 'qualitative hierarchy', in which one is taken to be in principle preferable to the other. Rather, the focus should be placed on the information that each model can provide on data variability based on the basic assumptions and the type of data analyzed.

Similarities between models

Experiment 1 showed that TMS over the posterior parietal site of the l-IPS (P3) improved performance accuracy only on passive reversible sentences. Both ANOVA and LMMs detected TMS-related effects: the main effect of SITE and the SITE*DIATHESIS interaction. The stronger significance found in LMMs indicated that the individual variability coming from the random variation of the intercepts across participants played a role in explaining the overall variability of the data.

Posthoc analyses carried out to further characterize the interaction effect also provided similar outcomes, associated with the same level of significance (p=0.018) when comparing sham and real conditions on P3. Contrasts on P1 also showed fully compatible levels of significance (p=0.024 in ANOVA and p=0.038 in LMMs). Both values fell in the first range of significance ($p < 0.005$), with the outcomes of LMMs returning lower significance values than in ANOVA. This difference could derive from the fact that LMMs are less constrained and take into account more sources of variability, increasing the probability of error.
Overall, the outcomes of the two models were a substantial match.

Also in Experiment 2 the comparison between ANOVA and LMMs yielded largely similar results. Both models showed that stimulation over the posterior l-IPS influenced performance accuracy on both real and pseudo-sentences in the passive voice, independent of the timing of TMS administration. A significant TMS*DIATHESIS interaction effect was detected by both ANOVA and LMMs. The same was true for the contrast between real and sham on passive diathesis, which reached significance in both cases. As in Experiment 1, this outcome confirms that the inferences drawn on the effect as scrutinized by a more liberal model and by a more conservative model, which takes into account more factors involved in accounting for global variability, are equally reliable.

Differences between the models

In both experiments the two models also showed some discrepancies. In Experiment 1 LMM detected a 3-way interaction effect among all the fixed factors (DIATHESIS, CORRESPONDENCE and SITE), which was not shown in ANOVA. Similarly, in Experiment 2 ANOVA showed a significant interaction among all the four fixed factors (DIATHESIS, SEMANTICS, TIME, TMS) which LMM did not detect. The interpretation of such differences is challenging. They could arise from the different ways in which the two approaches model the data. LMMs were conducted using a mixed logistic regression model which explores the relationship between a dichotomous variable (0-1/yes-no type) and a set of predictors (the fixed factors in the design). The model estimates the probability of distribution of each value of the DV in relation to the predictors. A single cell in this model is represented by a 0/1 value of the dichotomous variable depending on the considered predictors. For example, in Experiment 1 a single cell could be filled by the value 1 in correspondence to a single participant's trial that consists of an active sentence, associated with a mismatching picture when stimulating P3:

SUBJECT	DIATHESIS	TRIAL	SITE	ACC score
7	active	mismatching	P3	1

The model estimates the probability of the accuracy to be equal to 1 when the trial is active, mismatching and on P3 for subject 7.

In contrast, ANOVA considers the DV as a continuous variable and represents it as a percentage value (for example 85 out of 100). In ANOVA, each cell is filled by a numerical value that corresponds to the mean accuracy (expressed as the percentage of correct responses out of overall responses) of a participant in a specific condition. For example, in Experiment 2 a cell could be a single participant's mean accuracy score (72) on a trial corresponding to a pseudo sentence, in the passive diathesis and stimulated at T1:

SUBJECT	DIATHESIS	SEMANTICS	TIMING	ACC score
12	passive	real sentence	T1	72

This representation implies that the DV is modeled as a numeric value that can change as a function of the IVs and that the differences measured in DV are numerically expressed as mean differences.
These different ways of representing the data could have played a role in explaining these discrepancies (see **Jaeger, 2008** for a detailed view on the differences between the two models when accuracy is considered a continuous or a categorical variable).

Another possible explanation refers to the characteristics of the model that returned the outcomes. ANOVA is more constrained and can only assume that the variances are homogeneous. At the same time, it is more liberal in estimating the probability of error, being more likely to show higher levels of significance. On the other hand, LMMs represent a more conservative model which can account for multiple sources of variability, resulting in a higher probability of error (and lower significance levels). The opposite scenario, in which ANOVA shows lower significance levels than LMMs, could be accounted for by the fact that probably the effect detected by the first was not strong enough to survive a more in-depth exploration of variability that the second is able to provide. This further highlights that the link between the level of significance of an effect and its statistical reliability is not linear. Rather, it depends on the intrinsic characteristics of the chosen model.

Be this as it may, we take a cautious approach. We do not wish to exploit the differences between models to assign greater explanatory value to one model over the other. Rather, we merely point out that the outcomes of the two models present some differences and try to describe the contribution of ANOVA and of LMM in terms of how they can account for data variability.

ANOVA vs LMMs

In the Introduction we mentioned the problem of the link between pipeline and results in neuroimaging data. This work started with the goal of exploring it in behavioral and neurostimulation data. Overall, our results suggest that, for the dataset considered, the differences between the two analytical approaches are not big enough to risk pipeline dependency. Rather, these differences consist of different information about variability in the data. The two models explore the common parameter (*i.e.* the variance) starting from different premises. ANOVA treats it as homogeneous by estimating the amount of variance explained by the fixed factors. This allows significant effects to be detected in a more constrained, but also more liberal way; the significance threshold is more easily reached by fixed effects, so the probability of error is lower. On the other hand, LMMs can detect, in addition to the variance arising from fixed effects, the variance originating from random effects. In this way, they can consider the variance as nonhomogeneous and capture multiple sources that contribute to explaining it, ensuring a more thorough exploration of variability. The estimation of the significance of effects is more conservative (and thus more robust), but at the same time given the greater amount of sources of variability considered the probability of error is also higher. This implies that LMMs can display lower levels of significance without decreasing their statistical reliability.

Ultimately, the results obtained suggest that the two models are interchangeable from an outcome perspective and that their use does not depend on the inherent superiority of one over the other. Rather, the choice of approach must be related to the type of data and the characteristics of the protocol and design of the experiment. If they include several potential sources of variability additional to those from fixed factors, then LMMs may be more appropriate. In cases where variability is more controlled and assumed to be homogeneous, ANOVA should be as reliable as LMMs.

CONCLUSION

This study set out to explore the problem of the relationship between analysis methods and results in behavioral and neurostimulation data. Two models were used to analyze the same dataset of two published experiments and their results were compared. In both experiments ANOVA and LMMs provided largely overlapping outcomes. Notwithstanding ANOVA's stricter assumptions, the same relevant TMS-related effects were shown by both the models. Substantially similar outcomes were shown within the assumption of the homogeneity of variance (as in ANOVA), according to which variance is homogeneous for each participant. Similarly, results did not change when setting the intercept across participants as a source of random variability (as in LMMs). This similarity can be interpreted as evidence that the effects found by the more constrained model (ANOVA) survived a less rigid model (LMMs) capable of estimating more sources of variability within the data. The results obtained from the comparison suggest that neither model is inherently better or worse than the other. Rather, each model accounts for variability in the data in different ways, depending on their assumptions. Ultimately, our results show that the link between the significance level of the effect and statistical reliability is not linear, but depends on the characteristics of the model that has explored that effect.

LIMITATIONS

Although relying on data from two published studies focused on the same topic and strongly interrelated, the dataset used is still not large enough. Including other experiments would increase the sample and the likelihood of making reliable inferences on the results. Moreover, accurately modeling the overall direction of behavior and identifying precisely the most representative patterns is not easy. Without a baseline, exploring variability is much more complex. Therefore, a possible future direction could be on one hand to extend the sample to more experiments and, on the other, to compare actual behavioral data with simulated data that establish the expected effects *a priori* and allow comparing them with observed variations.

CONCLUSIONS

In this Ph. dissertation, the results of two main projects were presented, described and discussed. The first project (**Chapter 1**) investigated the neural basis of sentence processing, with a focus on two key processes: the processing of morphosyntactic features and the assignment of thematic roles. To this end, TMS was delivered to regions identified in the literature as being putatively involved in either process. This investigation offers new insights on the study of the neural basis of language processing for several reasons. First, it reports a functional distinction among the regions that are part of the left-lateralized network associated with language comprehension. It is demonstrated that the frontal regions (the l-IFG and l-MFG) are involved in morphosyntax processing and not (or to a lesser extent) in thematic role assignment. Conversely, the parietal region (the l-IPS) has a role in thematic role assignment and not in morphosyntactic processing. Second, to our knowledge, it represents the first investigation of the two processes of MS and TR and their neural underpinnings in the context of the same experiment. Finally, though the specific nature of the involvement of these regions in the highlighted linguistic processes is still not entirely clear, this study provides evidence in support of the view that their involvement is language-specific rather than domain-general in nature.

The materials and parameters used in this experiment were finalized and established in a pilot study described in **Chapter 2**. The pilot provided critical information on the best duration of stimulus presentation, such that performance accuracy fell in an ideal range between 70 and 80%, thus avoiding possible ceiling effects. The pilot study also allowed us to preliminary explore the effects of the experimental factors that would be included in the main experiment and to gain qualitative observational information on them.

The second project (**Chapter 3**) investigated the problem of the link between analysis pipeline and results. This problem has been reported mainly from studies in the field of neuroimaging and has become a debated issue. The aim of the project was then to extend the discussion to another type of data, namely behavioral and neurostimulation data. The results of two frequently used models (ANOVA and LMMs) in analyzing the same dataset were compared, and similarities and discrepancies were explored. The dataset was provided by two experiments published by our group. This methodological study represents a relevant contribution for two reasons. In the first place, it provides ground for the investigation of data replicability and reproducibility. The exact same analyses performed in the original published papers were replicated and results completely matched. Furthermore, to our knowledge this work represents the very first inquiry on the similarities and on the differences between the standard approach of ANOVA and LMMs performed on a language and stimulation dataset. Ultimately, our results show that applying the two models yielded similar and fully compatible outcomes and suggest that neither is to be preferred as a matter of principle. Rather, ANOVA and LMM should be seen as different ways of dealing with data variability. The choice of which model to use is not constrained by the inherent superiority of one over the other, but by the type of data to be analyzed and the sources of variability to be explored.

Overall, this dissertation makes an original contribution to the study of language comprehension processes and their neural correlates. First, it provides experimental evidence on the specific role played by some key regions of the network responsible for language processing, highlighting a functional distinction between frontal areas (IFG and MFG) and parietal areas (IPS) in supporting morphosyntax processing and thematic role assignment, respectively. This evidence adds to the vast number of findings suggesting the involvement of these regions in some aspects of language processing, also indicating a more precise functional differentiation. Even though not sufficient to advocate for the domain-general or for the language-specific account, these data point out that these regions' differential activity depends on the specificity of the linguistic process, suggesting that they are involved in sentence processing to different extents.

Another contribution this paper makes is an extended exploration of the differences and similarities between statistical approaches to the analysis of experimental variance (ANOVA AND LMMs) that are often considered as alternatives. The results shown and discussed suggest that the picture is more complex and that the scope of either model is largely shared; rather, the main differences lie in their adaptability to the type of data to be analyzed and in the ways in which experimental variability is explored and explained.

Authorship Statement

(according to CRediT (Contributor Roles Taxonomy) - Elsevier)

Chapter 1

L.Vercesi: *Conceptualization, Resources, Methodology, Writing, – original draft, Visualization, Software, Formal Analysis, Data Curation, Writing – review and editing*
G. Miceli: *Supervision, Writing – review and editing*
L. Cattaneo: *Supervision, Writing – review and editing*
C. Finocchiaro: *Writing – review and editing*

Chapter 2

L. Vercesi: *Conceptualization, Methodology, Resources, Software, Formal Analysis, Data Curation, Writing –original draft, Writing – review and editing*
G. Miceli: *Supervision, Writing – review and editing*
L. Cattaneo: *Supervision, Writing – review and editing*
C. Finocchiaro: *Writing – review and editing*

Chapter 3

L. Vercesi: *Conceptualization, Methodology, Resources, Software, Formal Analysis, Data Curation, Writing –original draft, Writing – review and editing*
M. Gallucci: *Supervision, Conceptualization, Methodology, Resources*
G. Miceli: *Supervision, Writing – review and editing*
L. Cattaneo: *Supervision, Writing – review and editing*
C. Finocchiaro: *Writing – review and editing*

Outcomes and Activities

Publications in peer-reviewed journals

- Vercesi, L., Sabnis, P., Finocchiaro, C., Cattaneo, L., Tonolli, E., and Miceli, G. (2020): "**The role of the l-IPS in the comprehension of reversible and irreversible sentences: an rTMS study**". *Brain Struct. Funct.* 225, 2403–2414. doi: 10.1007/s00429-020-02130-6

Publications on conference proceedings

- Vercesi, L., Sabnis, P., Finocchiaro, C., Miceli, G., Cattaneo, L., Tonolli, E. (2019) : **"Who does what to whom: the role of the l-IPS in the comprehension of reversible and irreversible sentences"** in *Proceedings the 20th International Science of Aphasia Conference*, Ginevra. doi: 10.13140/RG.2.2.29186.22727

Conference and poster presentations

- Vercesi, L., Sabnis, P., Finocchiaro, C., Miceli, G., Cattaneo, L., Tonolli, E. (2018) :**"Who does what to whom: the role of the l-IPS in the comprehension of reversible and irreversible sentences"** *Transcranial Brain Stimulation in Cognitive Neuroscience Workshop* (CIMeC; University of Trento)[poster presentation]

- Vercesi, L., Sabnis, P., Finocchiaro, C., Miceli, G., Cattaneo, L., Tonolli, E. (2019) : **"Who does what to whom: the role of the l-IPS in the comprehension of reversible and irreversible sentences"** *20th International Science of Aphasia* [poster presentation]

Talks

- Vercesi, L. (2022): "**Who does what to whom (in the brain): the neural basis of thematic role assignment**" Neurolinguistics meeting (University of Groningen) [invited speaker]

Manuscripts in preparation (as first author)

- Vercesi, L., Cattaneo, L., Finocchiaro C., Miceli, G. "**Left prefrontal and parietal contributions to sentence processing: a TMS study**"

- Vercesi, L., Gallucci M., Miceli, G. “**Are Linear Mixed Models (LMM) systematically better than Analysis of Variance (ANOVA) in investigating data variability in TMS studies of language?**”

REFERENCES

1. Acheson, D. J., & Hagoort, P. (2013). Stimulating the brain's language network: Syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. Journal of Cognitive Neuroscience, 25(10), 1664–1677.

2. Amunts, K., & Zilles, K. (2001). Advances in cytoarchitectonic mapping of the human cerebral cortex. *Neuroimaging Clinics of North America*, *11*(2), 151-69.

3. Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. Journal of comparative neurology, 412(2), 319-341.

4. Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. Neuropsychology, 8, 485– 493.

5. Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986). Dementia and working memory. Quarterly Journal of Experimental Psychology Section A, 38, 603–618.

6. Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience*, *10*(9), 659-669.

7. Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. Nature neuroscience, 6(5), 448-450.

8. Boisgontier, M.P., & Cheval, B. (2016). The anova to mixed model transition. Neuroscience & Biobehavioral Reviews, 68, 1004–1005

9. Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain and*

language, *125*(1), 60-76.

10. Botvinick, M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108, 624–652.

11. Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J. Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Benoit, R. G., ... & Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature, 582, 84–88

12. Bresnan, J. (2000). Lexical-functional syntax. Blackwell, Oxford, UK.

13. Bryk, A., & Raudenbush, S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage Publications.

14. Burgaleta, M., MacDonald, P. A., Martínez, K., Roman, F. J., Álvarez‐Linera, J., González, A. R., ... & Colom, R. (2014). Subcortical regional morphology correlates with fluid and spatial intelligence. *Human brain mapping*, *35*(5), 1957-1968.

15. Caplan, D., & Futter, C. (1986). Assignment of thematic roles by an agrammatic aphasic patient. Brain and Language, 27, 117–135.

16. Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. Psychonomic Bulletin & Review, 20, 243–268.

17. Caplan, D., Fyndanis, V., Arcara, G., & Christidou, P. (2018). Morphosyntactic production and verbal working memory: Evidence from Greek aphasia and healthy aging. Journal of Speech, Language, and Hearing Research, 61(5), 1171-1187.

18. Cappa, S. F., Sandrini, M., Rossini, P. M., Sosta, K., & Miniussi, C. (2002). The role of the left frontal lobe in action naming: rTMS evidence. Neurology, 59(5),

720-723.

19. Cappelletti, M., Fregni, F., Shapiro, K., Pascual-Leone, A., & Caramazza, A. (2008). Processing nouns and verbs in the left frontal cortex: a transcranial magnetic stimulation study. Journal of cognitive neuroscience, 20(4), 707-720.

20. Caramazza A, Miceli G (1991) Selective impairment of thematic role assignment in sentence processing. Brain Lang 41:402–436

21. Caramazza, A., & Zurif, E. (1976). Dissociations of algorithmic and heuristic processes in language comprehension: evidence from aphasia. Brain Lang 3:572–582

22. Carreiras, M., Quiñones, I., Mancini, S., Hernández-Cabrera, J. A. & Barber, H. Verbal and nominal agreement: An fMRI study. *Neuroimage* **120**, 88–103 (2015)

23. Carp, J. (2011). Optimizing the order of operations for movement scrubbing: comment on Power et al. Neuroimage, 55(4), 1919–1920.

24. Carp, J. (2012). On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. Frontiers in Neuroscience, 6, 149.

25. Chomsky, N. (1965). Aspects of the theory of syntax. MIT Press, Cambridge, MA.

26. Chomsky, N. (1981). Lectures on government and binding. Foris, Dordrecht, The Netherlands.

27. Churchill, N., Oder, A., Abdi, H., Tam, F., Lee, W., Thomas, C., et al. (2012). Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. Human Brain Mapping, 33, 609–627.

28. Clark, T. S., & Linzer, D. A. (2015). Should I use fixed or random effects? Political Science Research and Methods

29. Coles, M.G.H., Scheffers, M.K., & Holroyd, C.B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. Biological Psychology, 56, 173–189.

30. Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. Brain, 132, 2593–2608.

31. Coulson, S., King, J.K., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. Language and Cognitive Processes, 13, 21-58.

32. Davis, M. H., Meunier, F., & Marslen-Wilson, W. D. (2004). Neural responses to morphological, syntactic, and semantic properties of single words: an fMRI study. Brain and Language, 89, 439–449.

33. de Melo, M.B., Daldegan-Bueno, D., Menezes Oliveira, M.G., & de Souza, A.L. (2022). Beyond ANOVA and MANOVA for repeated measures: Advantages of generalized estimated equations and generalized linear mixed models and its use in neuroscience research. European Journal of Neuroscience, 56, 6089–6098.

34. den Ouden DB, Malyutina S, Basilakos A, Bonilha L, Gleichgerrcht E, Yourganov G, Hillis AE, Hickok G, Rorden C, Fridriksson J. (2019). Cortical and structural-connectivity damage correlated with impaired syntactic processing in aphasia. Human brain mapping, 40(7): 2153–2173.

35. Dronkers NF, Wilkins DP, Van Valin RD Jr, Redfern BB, Jaeger J (2004) Lesion analysis of the brain areas involved in language comprehension. Cognition 92:145–177

36. Duncan, J., Owen A.M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in Neurosciences, 23(10), 475-483.

37. Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., & Amunts, K.(2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage, 36(3), 511-521.

38. Fedorenko, E., & Blank, I. A. (2020). Broca’s area is not a natural kind. Trends in cognitive sciences, 24(4), 270-284.

39. Finocchiaro C, Capasso R, Cattaneo L, Zuanazzi A, Miceli G (2015) Thematic role assignment in the posterior parietal cortex: a tms study. Neuropsychologia 77:223–232.

40. Finocchiaro, C., Capasso, R., Cattaneo, L., Zuanazzi, A., & Miceli, G. (2015). Thematic role assignment in the posterior parietal cortex: a TMS study. Neuropsychologia, 77, 223-232.

41. Finocchiaro, C., Cattaneo, L., Lega, C., & Miceli, G. (2021). Thematic reanalysis in the left posterior parietal sulcus: A TMS study. Neurobiology of Language, 2(3), 416-432.

42. Friederici, A. D. (2018). The neural basis for human syntax: Broca's area and beyond. Curr. Opin. Behav. Sci. 21, 88–92.

43. Friederici, A. D., Hahne, A., & von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke’s and a Broca’s aphasic: electrophysiological evidence for a double dissociation. Brain Lang, 62, 311-341.

44. Gallucci, M., Olson, I. R., & Fellows, L. K. (2019). General Analyses for Linear Models: A Geometric Approach. Frontiers in psychology, 10, 2743.

45. Gathercole, S.E., Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of Memory and Language, 28(2), 200-213.

46. Gathercole, S.E., Baddeley, A.D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? European Journal of Psychology of Education, 8(3), 259-272.

47. Gerfo, E. L., Oliveri, M., Torriero, S., Salerno, S., Koch, G., & Caltagirone, C. (2008). The influence of rTMS over prefrontal and motor areas in a morphological task: grammatical vs. semantic effects. Neuropsychologia, 46(2), 764-770.

48. Giustolisi, B., Vergallito, A., Cecchetto, C., Varoli, E., & Lauro, L. J. R. (2018). Anodal transcranial direct current stimulation over left inferior frontal gyrus enhances sentence comprehension. *Brain and Language*, *176*, 36-41.

49. Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. Behavioral and Brain Sciences, 23, 1–21.

50. Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. Journal of Cognitive Neuroscience, 15(6), 883–899.

51. Hagoort, P. (2005). On Broca, brain, and binding: A new framework. Trends in Cognitive Sciences, 9(9), 416–423.

52. Hasting, A.S., & Kotz, S.A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. Journal of Cognitive Neuroscience, 20, 1207–1219.

53. Hernandez, A. E., Kotz, S. A., Hoffman, J., Valentin, V. V., Dapretto, M., & Bookheimer, S. Y. (2004). The neural correlates of grammatical gender decisions in

Spanish. NeuroReport, 15, 863–866.

54. Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the “missing” word frequency effect in stroke aphasia: insights using a novel method to quantify contextual variability in meaning. Journal of cognitive neuroscience, 23(9), 2432-2446.
55. Howard, D., & Nickels, L. (2005). Separating input and output phonology: Semantic, phonological and orthographic effects in short-term memory impairment. Cognitive Neuropsychology, 22, 42-77.

56. Indefrey, P., Hagoort, P., Herzog, H., Seitz, R., & Brown, C. (2001). Syntactic processing in left prefrontal cortex is independent of lexical meaning. Neuroimage, 14, 546-555.

57. Ioannidis, J.P.A. (2005). Why most published research findings are false. PLOS Medicine, 2, e124.

58. Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

59. Jiang, J., & Nguyen, T. (2021). Linear and Generalized Linear Mixed Models and Their Applications (Second Edition). Springer.

60. Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. Cerebral cortex, 11(3), 223-237.

61. Koerner, T. K., & Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: a comparison with Pearson correlation in two auditory electrophysiology studies. Brain Sciences, 7, 26.

62. Krueger, C. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biological Research for Nursing, 6, 151–157.

63. Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological Psychology, 11, 99–116.

64. Lauro, L. J. R., Reis, J., Cohen, L. G., Cecchetto, C., & Papagno, C. (2010). A case for the involvement of phonological loop in sentence comprehension. *Neuropsychologia*, *48*(14), 4003-4011.

65. Love, T., Swinney, D., Walenski, M., & Zurif, E. (2008). How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. Brain and Language, 107, 203–219.

66. Luria, A.R. (1964). Factors and forms of aphasia. In: de Reuck AVS, O'Connor M, editors. Disorders of language. London: Churchill. 143–67.

67. Magnusdottir, S., Fillmore, P., Den Ouden, D. B., Hjaltason, H., Rorden, C., Kjartansson, O., ... & Fridriksson, J. (2013). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: A lesion‐symptom mapping study. Human brain mapping, 34(10), 2715-2723.

68. Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011a). A person is not a number: Discourse involvement in subject-verb agreement computation. Brain Research,1410,64−76.

69. Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011b). When persons disagree: An ERP study of Unagreement in Spanish.Psychophysiology,48, 1361−1371

70. Mancini, S., Molinaro, N., & Carreiras, M. (2013). Anchoring agreement in comprehension. Language and linguistics compass, 7(1), 1-21.

71. Mancini, S., Molinaro, N., Davidson, D., Avilés, A., and Carreiras, M. (2014a). Person and the syntax–discourse interface: an eye-tracking study of agreement. *J. Mem. Lang.* 76, 141–157

72. Mancini, S., Postiglione, F., Laudanna, A., and Rizzi, L. (2014b). On the person-number distinction: subject-verb agreement processing in Italian. *Lingua* 146, 28–38

73. Martin, R. C. & He, T. (2004) Semantic STM and its role in sentence processing: A replication. Brain and Language 89:76–82

74. Martin, R. C. & Romani, C. (1994) Verbal working memory and sentence processing: A multiple component view. Neuropsychology 8:506–23.

75. Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. Human Brain Mapping, 36, 1963–1981.

76. Martin, R. C., & Blossom-Stach, C. (1986). Evidence of syntactic deficits in a fluent aphasic. Brain and Language, 28(2), 196-234.

77. Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*(3), 1481-1498.

78. Meyer, L., & Friederici, A.D. (2015). Neural systems underlying the processing of complex sentences. In G. Hickok & S. Small (Eds.), Neurobiology of language (pp. 597–606). Academic Press.

79. Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., & Caramazza, A. (2002). The neural correlates of grammatical gender: An fMRI investigation. Journal of Cognitive Neuroscience, 14, 618–628.

80. Miller, E. K. (2000). The prefrontal cortex and cognitive control. Nat Rev Neurosci, 1(1), 59-65.

81. Molinaro, N., Barber, H. A., Caffarra, S., and Carreiras, M. (2015). On the left anterior negativity (LAN): the case of morphosyntactic agreement. Cortex 66, 156–159.

82. Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011). Why brother and sister are not just siblings: Repair processes in agreement computation. Journal of Memory and Language, 64(3), 211-232.

83. Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In D.A. AlIport, D.G. MacKay, W. Prinz, & E. Scheerer (Eds.), Language perception and production: Relationships among listening, speaking, reading, and writing. London: Academic Press.

84. Morris, J., Frank, S. L., & Graesser, A. C. (2003). Computational models of reading comprehension: Inferential processes and findings. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), Reading as a Perceptual Process (pp. 693-720). Amsterdam, Netherlands: Elsevier.

85. Müller, N.G., Knight, R.T. (2006). The functional neuroanatomy of working memory: contributions of human brain lesion studies. Neuroscience, 139(1), 51-58.

86. Newman, S. D., Just, M. A., Keller, T. A., Roth, J., & Carpenter, P. A. (2003). Differential effects of syntactic and semantic processing on the subregions of Broca's area. Cognitive Brain Research, 16, 297-307.

87. Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., Shaywitz, B. A., Gore, J. C. & Shankweiler, D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. Journal of Cognitive Neuroscience, 12(1), 120-133.

88. Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: evidence for the roles of prefrontal and temporo-parietal cortices. J. Cogn. Neurosci. 22, 1597–1613.

89. Novick, J.M., Trueswell, J.C., Thompson-Schill, S.L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cognitive, Affective, & Behavioral Neuroscience, 5(3), 263-281.

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

91. Papagno, C., & Cecchetto, C. (2019). Is STM involved in sentence comprehension? Cortex, 112, 80–90.
92. Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology*, *44*(1), 47-67.

93. Perani D, Cappa SF, Schnur T, Tettamanti M, Collina S, Rosa MM, Fazio F. (1999). The neural correlates of verb and noun processing, Brain, 122(12), 2337-2344.
94. Petrides, M., & Pandya, D. N. (2012). The frontal cortex. *The human nervous system*, 988-1011.

95. Poline, J.-B., Strother, S., Dehaene-Lambertz, G., Egan, G., & Lancaster, J. (2006). Motivation and synthesis of the FIAC experiment: Reproducibility of fMRI results across expert analyses. Human Brain Mapping, 27, 351–359.

96. Pollard, C., & Sag, I. A. (1994). Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press.

97. Raettig, T., Frisch, S., Friederici, A. D., & Kotz, S. A. (2009). Neural correlates of morphosyntactic and verb-argument structure processing: an EfMRI study. Cortex.

98. Richardson, F. M., Thomas, M. S., & Price, C. J. (2010). Neuronal activation for semantically reversible sentences. Journal of cognitive neuroscience, 22(6), 1283-1298.

99. Rodd, J.M., Vitello, S., Woollams, A.M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. Brain and Language, 141, 89–102.

100. Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmöller, J., ... & Hallett, M. (2021). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert Guidelines. Clinical Neurophysiology, 132(1), 269-306.

101. Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical Neurophysiology, 120(12), 2008-2039.

102. Santi, A., & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. Neuroimage, 37(1), 8-17.

103. Sawyer, F. S. (2009). Analysis of variance: the fundamental concepts. Journal of Manual & Manipulative Therapy, 17(2)

104. Shapiro, K. A., Pascual-Leone, A., Mottaghy, F. M., Gangitano, M., & Caramazza, A. (2001). Grammatical distinctions in the left frontal cortex. Journal of cognitive neuroscience, 13(6), 713-720.

105. Singer, J.D., Willett, J.B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press.

106. Smith, S., Beckmann, C., Ramnani, N., Woolrich, M., Bannister, P., Jenkinson, M., et al. (2005). Variability in fMRI: A re-examination of inter-session

differences. Human Brain Mapping, 24, 248–257.

107. Tanenhaus, M. K., Carlson, G., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. Language and Cognitive Processes, 4(3/4), SI 211–234.

108. Tanner, D., & Van Hell, J.G. (2014). ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia, 56, 289–301.

109. Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., Farah, M. J., (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. Proceedings of the National Academy of Sciences of the United States of America, 94(26), 14792-14797.

110. Thompson, C.K., & Choy, J.J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. Journal of Psycholinguistic Research, 38, 255–283.

111. Tyler, LK, Bright P, Fletcher P, Stamatakis EA. (2004). Neural processing of nouns and verbs: the role of inflectional morphology. Neuropsychologia, 42(5), 512-523.

112. Vallar, G., & Papagno, C. (1986). Phonological short-term store and the nature of the recency effect: Evidence from neuropsychology. *Brain and Cognition*, *5*(4), 428-442

113. Vergallito, A., Petilli, M. A., & Marelli, M. (2020). Perceptual modality norms for 1,121 Italian words: A comparison with concreteness and imageability scores and an analysis of their impact in word processing tasks. *Behavior Research Methods*, *52*, 1599-1616.

114. Vercesi, L., Sabnis, P., Finocchiaro, C., Cattaneo, L., Tonolli, E., & Miceli, G. (2020). The role of the l-IPS in the comprehension of reversible and irreversible

sentences: an rTMS study. Brain Structure and Function, 225(8), 2403-2414.

115. Walenski, M., Europa, E., Caplan, D., & Thompson, C.K. (2019). Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. Human Brain Mapping, 39, 1–30.

116. Wang, J., Cherkassky, V. L., Yang, Y., Chang, K. M. K., Vargas, R., Diana, N., & Just, M. A. (2016). Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. Cognitive neuropsychology, 33(3-4), 257-264.

117. Wassenaar, M., Brown, C.M., Hagoort, P. (2004). ERP effects of subject–verb agreement violations in patients with Broca's aphasia. Journal of Cognitive Neuroscience, 16(4), 553-576.

118. Wechsler, S., & Zlatic, L. (2003). The Many Faces of Agreement. Stanford, CA: Center for the Study of Language and Information.

119. Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., & Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: A quantitative comparison of preprocessing strategies. Neuroimage, 47, 1408–1416.

120. Weniger, D. & Friederici, A. D. (2010). Syntactic processing in the human brain is supported by a global hierarchical organization. Neuroscience, 165(1), 109-120.

121. Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. Frontiers in Psychology, 7

122. Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. American Psychological

Association.

123. Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T. C., & Xu, X. (2021). Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. Neuron

124. Zilles, K., Palomero-Gallagher, N., Grefkes, C., Scheperjans, F., Boy, C., Amunts, K., & Schleicher, A. (2002). Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. *European neuropsychopharmacology*, *12*(6), 587-599.

LIST OF FIGURES

Chapter 1

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

LIST OF GRAPHS

Chapter 1

Chapter 2

Chapter 3