UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO
DI SCIENZE POLITICHE E DELLA COMUNICAZIONE

Distributional Factors in Language Processing: Evidence from Parametric and Naturalistic Functional MRI

Tutor:
Ch.mo Prof. Laudanna Alessandro (M-PSI/01)
Co-tutor:
Ch.mo Ch.mo Prof. Esposito Fabrizio (ING-INF/06)

Candidato:
Russo Andrea Gerardo
Matr. 8801400033

Anno Accademico 2019/2020

UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI SCIENZE POLITICHE E DELLA
COMUNICAZIONE

Corso di Dottorato di Ricerca in:

Scienze del linguaggio, della società, della politica e dell'educazione

XXXIII Ciclo

Coordinatore: Ch.mo Prof. Fimiani Filippo

Distributional Factors in Language Processing:

Evidence from Parametric and Naturalistic Functional MRI

Tutor: Candidato:

Ch.mo Prof. Laudanna Alessandro (M-PSI/01)   Russo Andrea Gerardo

Co-Tutor: Matr. 8801400033

Ch.mo Prof. Esposito Fabrizio (ING-INF/06)

Anno Accademico 2019/2020

Alla mia famiglia,
per avere creduto in me più di
quanto io abbia mai fatto.

Summary

Abstract

Language is a distinctive human ability that supports our daily life interactions. A deep understanding of the brain mechanisms behind language processing is fundamental to create physiological models and find possible alterations derived from specific pathologies.

In the last decades, the general technological advancement and the development of more precise and less invasive investigation techniques have increased dramatically our knowledge of the neural correlates of language processing. However, the great variety of human languages, the possibility to communicate across multiple and different channels and the uncertainty about the actual role of some linguistic features leave several open questions with a concrete possibility of neuroscientific innovation. For instance, the features of a highly inflected language like Italian can provide interesting research questions and additional insights on how our brain processes linguistic information.

The main aim of this work is to explore in greater details the influence of the linguistic distributional factors on the neural correlates of both language production and comprehension by exploiting the richness of the Italian language. Therefore, three functional magnetic resonance imaging (fMRI) experiments were performed by using both classical parametric and novel naturalistic frameworks. Two fast event-related fMRI experiments investigated the influence of the language distributional factors on the neural correlates of the inflectional process, whereas a third experiment was dedicated to providing additional insights on the linguistic prediction mechanism during natural language comprehension by modelling the neural response with two statistical language models.

The first experiment addresses the influence of the inflectional classes (i.e. the conjugations) and their distributional features (e.g. the size, the productivity, and the ortho-phonological consistency) on the generation of the past participle of the Italian verbs by analyzing both neural and behavioral data. The study reports significant effects of the conjugations on the cognitive operations and, for the first time,

differential cortical activations in the left middle frontal gyrus, left the supplementary motor area and left anterior cingulate cortex for verbs from different conjugations supporting the hypothesis that the neural correlates of the verb inflection are influenced by specific properties of the inflectional classes.

The second experiment explores the influence of the noun inflectional classes and their properties (e.g. the consistency, number and cumulative frequency of members) on the nominal inflection by analyzing the neural data of a group of participants involved in overt inflection task from the singular to plural and vice versa. The study reports an extensive bilateral cortical network involving the cingulate cortex, frontal and temporal areas, and the cerebellum, revealing that the neural activations are modulated by specific distributional features of the noun inflectional paradigm.

The third experiment investigates the neural correlates of the linguistic prediction underlying the natural language processing during narrative listening. The interest, therefore, shifts from word (verb or noun) production to structured text understanding. This is done by fitting the fMRI data with models that encode the probabilistic features of the language via the estimation of the so-called surprisal, that is a measure that quantifies the unexpectedness of a word given the previous ones. Two stochastic language models were estimated on a large written Italian corpus to obtain two versions of surprisal: a lexical-only version, based only on the lexical information of the chosen stimulus and a novel semantics-weighted model that integrates both lexical and semantic features.

Our study reports better prediction accuracy and better fitting of the fMRI data for the semantics-weighted model. The two models produced both overlapping and distinct activations: while lexical-only surprisal activated secondary auditory areas in the superior temporal gyri and the cerebellum, semantics-weighted surprisal additionally activated the left inferior frontal gyrus. The results support the usefulness of the surprisal models to describe the linguistic prediction and suggest that the proper inclusion of semantics information in the surprisal model may increase its the sensitivity in higher-order language-related areas, with possible implications for future

naturalistic fMRI studies of language under normal and (clinically or pharmacologically) modified conditions.

Besides the investigation of the influence of the language distributional factors on the neural correlates of both language production and comprehension, an additional aim of this work is the proposal of an innovative procedure in the broader field of the fMRI-neurofeedback (fMRI-NF).

In general, the fMRI-NF has been successfully applied in several cognitive domains and it is a procedure based on the possibility to self-modulate the neural signal of a brain region to explore and induce behavioral changes. The proposed method integrates the representational similarity analysis (RSA) and the fMRI-NF framework to train the subjects in modulating their mental state rather than simply regulating the neural signal of a region. The method has been tested in a pilot experiment at 7 Tesla where the subject was asked to imagine concrete objects. The results show that the presented approach is feasible suggesting further investigations and future applications in several domains, including representational and distributional aspects of language processing.

Riassunto

Il linguaggio è una caratteristica tipica del genere umano ed è alla base di tutte le nostre interazioni quotidiane. Per questo motivo lo studio delle sue basi neurali è fondamentale per la creazione di modelli fisiologici e l'individuazione di possibili alterazioni tipiche di alcune patologie.

Negli ultimi decenni, l'avanzamento tecnologico e lo sviluppo di modelli e strumenti, allo stesso tempo più precisi e meno invasivi, ci hanno permesso di aumentare le nostre conoscenze sulle basi neurali del linguaggio. Tuttavia, il grande numero di lingue presenti in tutto il mondo, la loro diversità e il fatto che il linguaggio umano può essere espresso sotto varie forme (e.g. scritta, orale) lascia aperte numerose domande, ampi spazi di ricerca scientifica e concrete possibilità di innovazione.

Questo lavoro si concentra su alcune di queste domande, ancora senza risposta, analizzando dati di risonanza magnetica funzionale acquisiti all'interno sia di classici paradigmi sperimentali parametrici sia di innovativi paradigmi di tipo naturalistico. Inoltre, l'utilizzo della lingua italiana come argomento di studio ci permette di espandere le nostre conoscenze sulle basi neurali del linguaggio e dei processi grammaticali per via delle sue caratteristiche e differenze rispetto ad altre lingue più comunemente e intensamente studiate come l'Inglese. Per questo motivo, due esperimenti di risonanza magnetica funzionale con paradigma di tipo fast event-related sono stati condotti su due gruppi di volontari sani per indagare le basi neurali del processo di flessione linguistica, mentre un esperimento con paradigma naturalistico è stato condotto per studiare la predizione linguistica durante l'ascolto di un brano.

Il primo esperimento è stato dedicato allo studio dei processi cerebrali che sono alla base della creazione del participio passato dei verbi Italiani provenienti da tutte e tre le coniugazioni. L'innovazione di questo esperimento risiede nel fatto che, rispetto agli studi sulla lingua Inglese che si sono potuti concentrare solo sulle differenze tra verbi regolari e irregolari, l'Italiano presenta tre classi flessive per i verbi caratterizzate ognuna da specifiche peculiarità nella formazione del participio passato. Questo studio riporta, per la prima volta, delle

differenze di attivazione corticale nel giro frontale medio sinistro, nell'area supplementare premotoria sinistra e nel cingolo sinistro anteriore per verbi appartenenti a differenti coniugazioni supportando l'ipotesi che le attivazioni cerebrali siano modulate da specifiche proprietà delle classi flessive.

Il secondo esperimento di risonanza magnetica funzionale è dedicato allo studio dei processi neurali che guidano la flessione dei sostantivi. Per questo motivo, i soggetti sono stati coinvolti in un compito di flessione ad alta voce del sostantivo, dalla sua forma singolare a quelle plurale e viceversa. L'ipotesi è che la flessione dei sostantivi sia influenzata dal suffisso della forma lessicale, dalla selezione del paradigma flessivo e dall'identificazione del genere grammaticale. I risultati dello studio riportano delle attivazioni nella corteccia cingolata, nelle aree frontali e temporali e nel cervelletto suggerendo che queste siano modulate dalle caratteristiche tipiche del paradigma flessivo del sostantivo.

Il terzo esperimento differisce dai primi due nel tipo di paradigma sperimentale usato e nell'argomento studiato. Infatti, lo scopo di questo esperimento è lo studio della predizione linguistica durante l'ascolto di un brano (scenario naturalistico) modellando il dato neurale con due modelli probabilistici stimati su un grande corpus dell'Italiano scritto: un modello solo lessicale, già usato in studi precedenti, e uno totalmente nuovo che combina informazione lessicale e semantica. I risultati dello studio mostrano che la combinazione di informazioni lessicali e semantiche aumenta la capacità predittiva del modello e quella di spiegare il segnale neurale. In aggiunta, oltre alle attivazioni comuni tra i due modelli nelle aree uditive secondarie, si nota come l'integrazione dell'informazione semantica con quella lessicale migliori la sensibilità in aree considerate importanti nella rete neurale del linguaggio come il giro frontale inferiore. In generale, i risultati supportano l'uso di modelli statistici del linguaggio per spiegare i processi di predizione linguistica a livello neurale e la combinazione di più informazioni linguistiche, aprendo possibili scenari per l'uso di questi paradigmi nello studio della comprensione linguistica anche a livello clinico.

Oltre agli studi dedicati all'esplorazione dei processi neurolinguistici, questo lavoro contiene la proposta di un nuovo paradigma sperimentale nel campo più ampio ed emergente dell'interazione cervello-computer con risonanza magnetica funzionale (fMRI-neurofeedback). In generale, le tecniche di fMRI-neurofeedback sono state applicate in molti campi delle neuroscienze dimostrando la possibilità di influire su comportamenti esterni con la auto-modulazione del segnale cerebrale stimato da una regione di interesse. Il modello proposto integra il paradigma del fMRI-neurofeedback con il modello della representation similarity analysis (RSA) in modo tale da allenare il soggetto a modulare il suo stato cerebrale e non semplicemente a modificare il singolo segnale di attivazione estratto da una regione. Il metodo è stato implementato e testato sui dati di un esperimento pilota, contenente un task immaginativo, condotto con uno scanner di risonanza magnetica ad alto campo (7 Tesla). I risultati, sebbene preliminari, mostrano che la tecnica è applicabile facilmente a un esperimento di fMRI-neurofeedback e suggeriscono l'implementazione di nuovi esperimenti in altri ambiti applicativi, compreso lo studio dei processi linguistici.

Chapter 1:
Exploring the neural bases of language processing with functional MRI

1.1 The neuroanatomical organization of language processing

Language is an efficient and sophisticated communication system that allows humans to convey, in various forms, a theoretically infinite number of ideas and to describe things and events both spatially and temporally distant. During the centuries, it has become an essential part of our life, supporting the creation of thoughts, the collection of new information and knowledge, and the transmission of cultural practices (Hagoort, 2019). Besides, it provides a powerful representational medium in which concepts can be stored and manipulated efficiently (Ünal and Papafragou, 2018) and it has a mutual relationship with perception, consciousness, and memory (Perlovsky and Sakai, 2014).

It has been estimated that there are nearly 7000 spoken idioms and more than 100 sign languages in the world (Vigliocco et al., 2005) and each of these languages requires different features from its speakers by even influencing fundamental dimensions of human experience like the notions of time and space (Boroditsky, 2011). Indeed, there is evidence supporting that linguistic differences affect also our preferences (Danziger and Ward, 2010) and how we reconstruct what we have eyewitnessed (Fausey and Boroditsky, 2011). Furthermore, there are studies showing that speakers of different languages focus their attention on different features of the world while they are converting thoughts into speech (Bunger et al., 2016; Gleitman et al., 2007; Papafragou et al., 2008; Ünal and Papafragou, 2018, 2016).

Therefore, understanding the neural bases of language processing has been an interesting field of study for both linguists and neuroscientists since the 19th century.

The formulations of the earliest language models were highly influenced by the studies on aphasic syndromes induced by brain damages and were focused on which regions are most important for language processing (Phillips and Sakay, 2005). For instance, in 1874, Carl Wernicke proposed a model for language processing that classifies regions according to the tasks they accomplish. In particular, two main areas were identified: Brodmann's area 44/45 and Brodmann's area 22. According to the model, the first is specialized

in transforming the language representations in speech and it is also known as Broca's area because Paul Broca was the first, in 1861, to identify and claim a link between damages in this region and impairments in language production (Broca, 1861). The second, also known as Wernicke's area, is responsible for decoding and storing the language units, as a relation between the lesion in this area and impairments in language comprehension and semantic coherence had been observed (Wernicke, 1874). The Wernicke's model and its updated version proposed by Geschwind (Geschwind, 1965) have suggested a left-lateralized organization for the language processing with the comprehension mechanism mainly supported by the temporal cortex (in particular by the left superior temporal gyrus (L-STG)) and the production process hosted in the left inferior frontal lobe (Geschwind, 1965; Wernicke, 1874).

However, in the 1970s and 1980s, several findings questioned the bases of Wernicke's model (Hickok and Poeppel, 2007). For instance, it was observed that damages to the L-STG were linked to deficits in language production but not comprehension (Damasio and Damasio, 1980) and that disruptions in the left frontal lobes caused impairments in language tasks requiring syllable discrimination (Blumstein et al., 1977a, 1977b). Besides, the advent of modern neuroimaging techniques such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG), contributed to revisit the notions of the Wernicke's model (Tippett et al., 2014). Particularly, since Ogawa et al. (1990) introduced a way to measure the brain activity by estimating the blood oxygenation level-dependent (BOLD) signal (Ogawa et al., 1990), the number of fMRI studies focused on language processing have increased steadily (Ardila and Bernal, 2016; Willems and van Gerven, 2018), expanding our knowledge and leading, for example, to the proposal of a more comprehensive functional and anatomical model of speech organization (Hickok and Poeppel, 2007).

1.2 The influence of distributional factors in language processing

Human language is a complex high-order function and therefore, to understand its neural correlates, it has to be decomposed into basic building blocks and core operations (Hagoort, 2019): The building blocks consist of the knowledge about sound patterns, semantic and syntactic features of the lexical items (e.g. noun, verbs and grammatical gender), and orthographic forms. Core operations include all those operations that require information to be exchanged among these building blocks or retrieved from memory (e.g. word recognition), enabling to integrate small into larger arrangements (e.g. verb/noun inflection) (Hagoort, 2019). Despite this complexity, children are able to handle these core operations already in the earliest years of life without formal instructions, i.e. mainly by inferring distributional patterns and statistical schemes in the linguistic input (Romberg and Saffran, 2010) suggesting an important role of these factors in language processing.

The neural correlates of the building blocks, the basic core operations and their interactions have been extensively studied in cognitive neuroscience with both behavioral and neural (e.g. fMRI) experiments. For instance, the contribution of the grammatical information to both lexical processing and representation has been investigated by comparing the characteristics of two grammatical classes, i.e. nouns and verbs, in lexical production and comprehension processes (Laudanna et al., 2004). Previous studies showed that the two grammatical classes are handled by two different neural networks: left temporal regions for nouns and left prefrontal regions for verbs (Crepaldi et al., 2011; Vigliocco et al., 2011).

Other studies focused their attention on the role of the grammatical category information within single word classes (e.g. inflectional class, mood, tense and person for verbs and gender and number for nouns) in determining the organization of the lexical system (Laudanna et al., 2004). For instance, a large body of research explored the generation of the inflected forms of both verbs and nouns. In the first case, most of the studies were focused on a relatively poor inflectional phenomenon, such as the formation of the past participle

of the English verbs, and aimed at disclosing a clear dichotomy between the processing of regular and irregular verbs (Joanisse and Seidenberg, 2005; Kielar and Joanisse, 2009; Marslen-Wilson et al., 1995; McClelland and Patterson, 2002; Pinker and Ullman, 2002; Rumelhart and McClelland, 1986; Smolka et al., 2013; Ullman, 2001, 1993; van der Lely and Ullman, 2001). In the second case, the nominal inflection has been mainly investigated by studying the consequences on the lexical access of the variable association between the grammatical gender and the noun suffixes (Caffarra et al., 2014; Caffarra and Barber, 2015; Heim, 2008; Heim et al., 2005; Hernandez et al., 2004; Miceli et al., 2002; Padovani et al., 2005; Quiñones et al., 2018; Wang and Schiller, 2019).

However, in highly inflected languages, like Italian, the inflection of both verbs and nouns are influenced by additional factors, such as the distributional features, that were not exhaustively considered in previous studies. First, the inflection of verbs intersects other aspects of morphology, such as the presence of the conjugations, that creates paradigmatic relations among wordforms beyond the classical regular/irregular distinction (Aronoff, 1994; Carstairs-McCarthy, 1994). Second, the nominal inflection seems to be affected by distributional factors such as the suffix type frequency, the suffix predictability and the size of the inflectional neighborhood (Mirković et al., 2011; Nevat et al., 2017; Zwitserlood et al., 2000).

The distributional features of the language have a role also in the comprehension process. In fact, it has been argued that language comprehension is supported by a probabilistic prediction mechanism that exploits the recurrence in the language corpus of groups of lexical, syntactic and semantic cues to predict the upcoming word (Dikker et al., 2014; Dikker and Pylkkänen, 2013; Lau et al., 2016; Lopopolo et al., 2017; Wicha et al., 2004; Willems et al., 2016). A possible way to model and study this phenomenon is the use of a measure known as surprisal that quantifies the Shannon information received with the incoming input (Shannon, 1948) and that can be estimated with any stochastic language model (Armeni et al., 2017). Previous findings demonstrated that the surprisal is linearly correlated with the

language-related cognitive effort (Demberg and Keller, 2008; Hale, 2016, 2001; Levy, 2008; Smith and Levy, 2013).

In this work, the role of the linguistic distributional factors in both language production (verb and noun inflection) and comprehension (natural listening) processes has been explored in greater details by performing three experiments.

The first experiment investigated if the neural correlates of the Italian verb inflection are affected by the presence of inflectional classes (i.e. the conjugations) and their properties (Laudanna, 2007; Laudanna et al., 2004) by using an overt past participle generation task in both a behavioral and a rapid event-related fMRI experiment (ISI: 7.2 s; jitter: 2.4 s). The second chapter of this work is entirely focused on this experiment and its results.

The second experiment aimed at disclosing whether and how the lexical system and its neural correlates respond to distributional factors of the inflectional properties of the Italian nouns. A set of stimuli, carefully selected using a reliable database about the distribution of noun inflectional features (De Martino et al., 2019, 2018), were administered to a group of healthy subjects in a rapid event-related (ISI 7.5 s; jitter 4.5 s) fMRI experiment with an overt inflection task. The third chapter of this work is entirely focused on this experiment and its results.

The third experiment exploited the advantages of a naturalistic experimental paradigm, such as the flexibility and the ecological validity (Kandylaki and Bornkessel-Schlesewsky, 2019), and the ability of the stochastic language frameworks in modelling the richness and the complexity of the real-life linguistic experience (Armeni et al., 2017) to investigate the linguistic prediction underlying the language comprehension. In detail, two variants of a measure that quantifies the expectation of a word based on the previous ones (i.e. the context), called "surprisal", has been estimated on a large written Italian corpus: a more classical one based only on the co-occurrences of the lexical items and a novel version that integrates information from multiple linguistic levels (i.e. lexical surprisal and semantic distance). Both surprisal models were applied to the fMRI signal measured during the listening to an audiobook and their performance

in capturing the BOLD signal variance were investigated and compared. The fourth chapter of this work is focused on this experiment and its results.

1.3 A multidimensional approach to the fMRI neurofeedback

A further aspect investigated in this work, and described in the fifth chapter, is a methodological innovation in the field of the real-time fMRI neurofeedback (rt-fMRI-NF). The latter is a psychophysiological method in which the online measured BOLD signal is provided as feedback to the subject to allow manipulation and self-regulation of the neural correlates of a targeted behavior (Sitaram et al., 2017). Rt-fMRI-NF has been associated with improvements or changes in specific neural functions and/or behaviors and it has been successfully applied in a great variety of domains (Watanabe et al., 2017). In a typical rt-fMRI-NF experiment the NF signal is based on the neural signal changes of a selected ROI or on the variation of the functional connectivity of a network, whereas the feedback provided to the subject, that can be more or less complex, is visual, acoustic, haptic or electrical (Paret et al., 2019). However, even the most complex visual feedback, such as a virtual environment, is based on the variation of a single dimension of the fMRI signal although several scenarios could benefit from a different approach that evaluates multiple dimensions of the BOLD signal. Therefore, a novel experimental paradigm that provides a NF signal based not only on the current participant's mental representation of a stimulus but also on its relationships (similarity) with other mental representations, is proposed by integrating the representation similarity analysis (Kriegeskorte et al., 2008a) in an fMRI-NF paradigm. To test the feasibility and validate the method a pilot fMRI data set has been acquired at 7 Tesla. The fifth chapter of this work is focused on this pilot experiment and its results.

Chapter 2:

Producing regularly and irregularly inflected verb forms: behavioral data from the three Italian conjugations

M. De Martino, A. Mancuso, **A. G. Russo**, A. Elia, F. Di Salle, R. Saponiero, S. Vietri, F. Esposito & A. Laudanna (2020), Language, Cognition and Neuroscience, 35:4, 420-439, DOI: 10.1080/23273798.2019.1668953

2.1 Introduction

The hypothetical segregation of the grammatical information from the other forms of linguistic features, such as semantics and phonology, has been a central issue in the research on the linguistic processes (Bates and Goodman, 1997). For instance, the investigation of the differences between regular and irregular verbs has been a critical test for the cognitive status of the grammatical information.

Previous findings have been explained in the dual- (Allen and Badecker, 2002; Marslen-Wilson et al., 1995; Penke and Krause, 2002; Ullman, 1993; van der Lely and Ullman, 2001) or in the single-mechanism framework (Eddington, 2002; Justus et al., 2011; Kielar and Joanisse, 2009; Rumelhart and McClelland, 1986; Sach et al., 2004; Schreuder et al., 1999; Smolka et al., 2013). The first postulates that regular forms are generated using a rule-based procedure (regardless of the phonology and of the semantics of the stem) that is subserved by the frontal cortex and the basal ganglia, whereas irregular forms are stored in the mental lexicon, hosted by the temporal lobe, and retrieved with an associative mechanism (Pinker, 1991; Pinker and Ullman, 2002; Ullman, 2001). The second postulates that regular and irregular verbs are processed by a single neural network exploiting the phonological and semantic features of the item (McClelland and Patterson, 2002).

However, there is an increasing number of studies whose findings cannot be clearly explained by the dual or the single-mechanism models. For instance, in highly inflected languages the difference between regular and irregular word forms intersects many morphological aspects, such as the presence of inflectional classes (Aronoff, 1994; Colombo et al., 2004; Meunier and Marslen-Wilson, 2004) that have been shown to play a role in the inflectional process. Behavioral studies demonstrated that the higher is the consistency and regularity of the inflectional class the smoother and the more accurate is the inflectional process (Colombo et al., 2006, 2004; De Martino et al., 2017, 2011; Laudanna, 2007; Laudanna et al., 2004; Veríssimo and Clahsen, 2009). Therefore, linguists and psycholinguists (Burzio, 1998; Laudanna, 2007, 1999) hypothesized that the inflectional process is influenced by the features of inflectional class such as the

size, the productivity, and the ortho-phonological consistency. However, the understanding of the role of the inflectional classes during morphological encoding is limited and inconclusive as previous research focused on a relative morphologically poor phenomenon like the inflection in the English language (Bordag and Pechmann, 2009; Colombo et al., 2004; Veríssimo and Clahsen, 2009).

This study aims to explore the behavioral and neural correlates of the formation of the past participle by exploiting the characteristics of the Italian inflectional system and test whether the differences in the inflectional process between verbs belonging to different inflectional classes, can be explained with the dual or single-mechanism or by the features of the classes. Therefore, two groups of Italian native speakers were involved respectively in a behavioral and in a rapid event-related fMRI experiment where they produced overtly the past participle of a set of both regular and irregular Italian verbs belonging to the three conjugations.

The effect of the conjugations and its characteristics has been analyzed by contrasting all the regular verbs belonging to the three conjugations. In light of previous findings (Colombo et al., 2006, 2004; Laudanna et al., 2004), if the consistency of the inflectional class affects the inflectional process, slower response latencies and more errors in the behavioral task and, higher cortical activations in the fMRI signal are expected for the 2nd conjugation compared to the 1st and, to a lesser extent, to the 3rd. On the other hand, if the 1st conjugation acts as the default inflectional class, thus recreating the regular-irregular dichotomy, differences with both the 2nd and the 3rd conjugations should be observed (Say and Clahsen, 2002; Veríssimo and Clahsen, 2009).

The comparison of the stimuli of the 2nd conjugations to the stimuli of the 3rd conjugations allowed us to explore the possible interactions between inflectional class and the regularity. In light of previous observations (Colombo et al., 2004; Slioussar et al., 2014) if the lower consistency of the conjugation affects the inflection, a disadvantage (longer response latencies, higher error rates, and higher cortical activations) for the 2nd conjugation compared to the 3rd is expected.

Concerning the possible brain areas involved in the fMRI task, if the inflectional process can be explained with the dichotomy regular-irregular, activations in the temporal lobe for the irregulars (non-default classes: 2nd and 3rd conjugations) and, in the frontal lobe and basal ganglia for the regulars (default class: 1st conjugation) are expected (Ullman, 2001). On the other hand, if the inflectional process is influenced by the characteristics of the inflectional classes, we expect higher activations for the 2nd conjugation compared to the 1st and the 3rd (to a lesser extent). Moreover, we expect an interaction effect between conjugation and regularity in a more widespread network including areas previously observed to be involved in inflectional operations such the left inferior frontal gyrus (L-IFG) (Bozic and Marslen-Wilson, 2010; Carota et al., 2016), the anterior cingulate cortex (ACC)(Carter and van Veen, 2007), the supplementary motor area (SMA)(Sahin et al., 2006) and the postcentral gyrus (PostCG) (Indefrey and Levelt, 2004; Miceli et al., 1983).

2.2 Materials and methods

2.2.1 Behavioral experiment

2.2.1.1 Participants

Forty-two healthy native Italian speakers (8 males, minimum age: 19 years old, maximum age: 31 years old), all students from the University of Salerno were included in the behavioral experiment. All participants had a normal or corrected-to-normal vision and served for one session lasting about 15 min.

2.2.1.2 Stimuli

Italian verbs can be classified, with few exceptions (Thornton et al., 1997), in three main inflectional classes called conjugations (Serianni, 1988) that differ in terms of the thematic vowel (TV) of the infinitive

form, numerosity, productivity and presence of regular and irregular inflection (Table 2.1).

	Total	Regular inflection	Irregular inflection
1st conjugation	71%	97%	3%
2nd conjugation	18%	22%	78%
3rd conjugation	11%	84%	16%
verbs with regular inflection	84%		
verbs with irregular inflection	16%		

Table 2.1 Distribution of Italian verbs along the regularity/irregularity dimension and within different inflectional classes; count based on the Italian basic dictionary, a corpus of 7076 Italian words (Thornton et al., 1997)

Considering the regularity of the inflection, the 1st conjugation (TV: "-a") is highly regular, the 2nd conjugation (TV: "-e") is mainly irregular although it contains 13 regular sub-paradigms (count based on the BDVDB database (Thornton et al., 1997)) and the 3rd conjugation (TV: "-i") is mainly regular with a small presence of irregular verbs and 5 related sub-paradigms (BDVDB database (Thornton et al., 1997)).

In this experiment, a set of one 168 Italian verbs were selected from all the three conjugations. For each conjugation, the 56 verbs selected were divided into two lists: the first with 28 regular verbs and the second with 28 irregular verbs. The two lists of the 1st conjugation contained both regular verbs, as in this conjugation there are only four irregular verbs. This arrangement replicated as much as possible the distributions of verbs in real-life scenarios. The 6 lists were matched for the psycholinguistics variables reported in Table 2.2.

	1st conjugation	2nd conjugation	3rd conjugation

	reg	reg II	reg	irr	reg	irr
letters (infinitive form)	8.8 (1)	8.7 (1.4)	8.6 (1.5)	9.2 (1.4)	8.3 (1.4)	8.4 (1.5)
written type frequency (Bertinetto et al., 2005)	250 (350) (63 pm)	283 (358) (71 pm)	246 (283) (62 pm)	276 (838) (59 pm)	239 (258) (60 pm)	306 (454) (77 pm)
written token frequency (Bertinetto et al., 2005)	51 (55) (13 pm)	50 (70) (13 pm)	44 (74) (12 pm)	44 (107) (12 pm)	43 (51) (11 pm)	57 (74) (15 pm)
spoken token frequency (past participle form) (De Mauro et al., 1993)	3 (6) (6 pm)	6 (11) (12 pm)	3 (5) (6 pm)	4 (8) (8 pm)	3 (5) (6 pm)	7 (19) (14 pm)
n-count (infinitive form)	2 (2)	2 (2)	3 (2)	2 (1)	2 (1)	3 (1)
n-count (past participle form)	6 (2)	5 (1)	3 (1)	4 (3)	3 (2)	4 (3)
number of consonant clusters (infinitive form)	1 (1)	1 (1)	1 (1)	2 (1)	1 (1)	1 (1)
number of consonant clusters (past participle form)	1 (1)	1 (1)	1 (1)	2 (1)	1 (1)	2 (1)

Table 2.2 Mean values (standard deviations in parentheses, pm=per million) for the relevant lexical parameters controlled in the behavioral experiment.

Moreover, to avoid any strategic bias and to match the distribution of conjugation/regularity through the Italian verbal system, 42 filler verbs (36 regular verbs of 1st conjugation and 6 regular verbs 3rd conjugation) were added to the experimental list. Thus, the list of 210 stimuli was arranged in 7 blocks composed of 30 stimuli and visually delivered to each participant. Both stimuli and blocks were fully randomized.

2.2.1.3 Experimental procedure

Participants were provided with the infinitive form of the verb (e.g. giocare, to play) and were requested to produce overtly the corresponding past participle (e.g. giocato, played) as soon as possible. The stimuli appeared in lower-case letters (12-point size) in the center of a computer screen, preceded by a period of 300 ms with a fixation point and a blank period of 300 ms. Participant's onset verbal response was detected by a microphone using a voice-key detector, which was connected to the stimulation computer. The presentation of the stimuli was controlled with E-Prime 2.2 experimental control shell (Psychology Software Tools, Sharpsburg, PA, USA, www.pstnet.com).

As soon as a voice response was detected, the stimulus disappeared from the computer screen. Each stimulus remained on the screen for a maximum of 800 ms and if participants did not produce any answer within that time, the feedback-mask "Fuori tempo" ("Out of time"), appeared. Reaction times from word presentation to the beginning of each vocalization (onset) were measured. Oral responses were recorded and carefully checked for accuracy by an experimenter. Before the experimental session, participants performed a brief training simulation.

The experimental session consisted of 7 blocks of 30 stimuli interleaved by a 30 s break.

2.2.2 Functional MRI experiment

2.2.2.1 Participants

Twenty-three healthy volunteers (11 males; minimum age: 19 years old, maximum age: 34 years old) were enrolled in the fMRI experiment. All participants had normal vision and no history of neurological, psychiatric, developmental, or linguistic disorders. Participants provided written informed consent.

The experimental procedure conformed to the principles of the Declaration of Helsinki and was approved by the local ethics committee.

2.2.2.2 Stimuli

The stimuli used in the fMRI experiment were identical to the stimuli used in the behavioral experiment. However, to reduce the acquisition time the filler stimuli were not included. The stimuli presentation was synchronized with the fMRI acquisition and randomized in each scan using the software Presentation (Version 1.8, Neurobehavioral System, www.neurobs.com).

2.2.2.3 Experimental procedure

Functional MRI acquisition was performed using a 3 Tesla scanner (Magnetom Skyra, Siemens Healthcare, Erlangen, Germany) equipped with a 20-channel parallel head coil. The fMRI scan consisted of 600 volumes of a repeated gradient-echo echo-planar imaging sequence (repetition time (TR) = 2200 ms, echo time (TE) = 30 ms, number of axial slices = 28, matrix = 128×128, field of view (FOV) = 240 mm, thickness = 4 mm, interslice gap = 0 mm). Three-dimensional T1-weighted sagittal images (MPRAGE sequence, voxel size = $1.0 \times 1.0 \times 1.0$ mm3) were acquired in the same session to have a high-resolution anatomical reference for registration and normalization of the functional images.

During the fMRI scan, the list of 168 verbs was serially and randomly provided to the subject in their infinitive form using a video display unit (connected to a back-projection screen in the MRI room) controlled by a personal computer (synchronized with the fMRI acquisition). A row of white fixation crosses (++++++) was maintained on the center of the visual field against a black

background. Using a rapid event-related paradigm (ISI: 7.2 s; jitter: 2.4 s), stimuli (white lower-case letters, Arial font, size 48) replaced the fixation crosses for a period of 800 ms. Participants were asked to covertly inflect the verb on the screen in the past participle form. Stimuli were replaced by a row of fixation crosses (++++++) for 1400 ms. When the crosses were replaced by hashtags (########) for 1000 ms, participants were then required to produce aloud the past participle. Stimuli presentation and the word production timings were uncoupled to minimize possible artifacts associated with small head movements due to the word production. Vocal responses were reproduced in the MRI console via loudspeakers connected to the stimulation computer, recorded via an MRI compatible voice microphone laying on their mouth (Serene Sound, Resonance Technology, USA, www.mrivideo.com) and then checked for accuracy by an experimenter. Besides, a row of fixation crosses/hashtags rather than a simple fixation cross/hashtag was used to avoid spurious activations in visual brain areas.

Before starting the experiment, participants were instructed and performed a brief offline simulation of the task using a small set of stimuli not included in the experimental list.

2.2.3 Data analysis

2.2.3.1 Behavioral data analyses

Statistical analyses were carried on both response latencies and response accuracy.

Concerning the response latencies, the effect of the inflectional class was tested using an ANOVA by participants (F1) and by items (F2). In both cases, Conjugation was considered the independent variable with three levels (1st, 2nd, and 3rd), and the response time of the regular verbs from all the conjugations was considered as the dependent variable.

Response time, elicited only the 2nd and the 3rd conjugations, were used as the dependent variable in an ANOVA by participants (F1) and by items (F2) where the Conjugation (two levels: 2nd and 3rd) and the Regularity (two levels: regular and irregular) were considered as

independent variables to test the main effect of Conjugation, Regularity and their interaction.

Incorrect responses (yes/no) were analyzed using a Generalized linear mixed model (GLMM), where Conjugation, Regularity, and their interactions were considered as fixed effects (Jaeger, 2008; Quené and van den Bergh, 2008). Significant effects and parameters were evaluated using a Wald chi-square test (Bates et al., 2015).

A qualitative analysis was performed on the incorrect responses to explore whether regular and irregular paradigms could influence morphological errors. The following distinction has been made:

- Regularizations:
  1. a regular inflection is used to inflect an irregular verb-stem.
  2. the regular inflection of a given conjugation is used to inflect a verb-stem from a different conjugation.
- Irregularizations:
  1. an irregular inflection is used to inflect a regular stem.
  2. irregular verbs are inflected by using an irregular inflectional affix not compatible with the input stem.

Within the categories "Regularizations" and "Irregularizations" additional subclasses have been identified:

- conjugation change: errors resulting from the combination of verb-stems with suffixes belonging to other conjugations.
- no change: errors resulting from the combination of verb-stems with a suffix which is present within the conjugation of the target.

These classifications allowed us to explore to what extent participants followed the conjugation rules in making errors. In fact, according to the dual-mechanism the 1st conjugation, acting as the default class, tends to attract the errors (Say and Clahsen, 2002).

2.2.3.2 Image data analysis

Slice scan time correction, motion correction, spatial smoothing (Gaussian kernel with 6 mm full width half maximum), and high-pass filtering (cut-off to 0.008 Hz) were applied to the fMRI time-series using BrainVoyager QX (Version 2.8, Brain Innovation, The Netherlands, www.brainvoyager.com). Then, the fMRI images were registered to the MPRAGE images, and the resulting realigned data

were then transformed into Talairach space before the computation of a random-effects multi-subject GLM.

This hierarchical analysis entailed a first-level analysis in which all experimental conditions for each subject were modelled as separate predictors. The resulting GLM fits thus contained 6 beta weights per subject (Reg1, Irr1, Reg2, Irr2, Reg3, Irr3) corresponding to the 6 predictors of interest in the first-level GLM. Each predictor was convolved with a canonical double-gamma hemodynamic response function (HRF) peaking 5 s after the offset of word presentation. Motion correction parameters (3 translation and 3 rotation parameters) were included as covariates of no interest in the first level (individual) GLM analysis to account for residual motion-related effects.

In the second-level analyses, the Conjugation effect in the regular verbs and the effect of Regularity, Conjugation, and their interaction for the 2nd and 3rd conjugations verbs were analyzed. First, the beta values of all the regular verbs entered in a random-effects one-way (1x3) ANOVA, where the Conjugation factor had three levels (1st, 2nd, and 3rd). Second, to jointly model the main effects of Regularity and Conjugation on verb inflection, as well as their interaction, the beta values were analyzed in a two-way 2x2 (Regularity x Conjugation) random-effects analysis of variance (ANOVA) model. For this model, only 2nd and 3rd conjugation verbs were considered.

The statistical parametric maps (F statistics) resulting from both ANOVA models were overlaid in pseudo color on the average Talairach-normalized anatomical scan after correction for multiple comparisons using cluster-level thresholding (Forman et al., 1995). More specifically, maps were initially thresholded at a voxel-wise p-value of 0.005 (uncorrected) and then subjected to a whole-brain (no mask) correction procedure based on the estimate of the intrinsic spatial smoothness of the map and 10000 iterations of a Monte Carlo simulation, to determine the minimum cluster size threshold ensuring a corrected p-value of 0.05 (cluster-level corrected). Small volume corrections were also considered for subcortical nuclei and hippocampus based on a priori regional hypotheses (Worsley et al., 1996). These analyses were carried out to test specific predictions of the declarative-procedural model about the role of these areas in

processing regular and irregular verb-forms (Ullman, 2001). For compact clusters exhibiting significant effects, all activated voxels were selected, and the corresponding individual beta values were averaged. For each cluster, the averaged beta values (for all the experimental conditions) were extracted and plotted (mean betas with standard errors).

2.3 Results

2.3.1 Behavioral experiment

2.3.1.1 Response latencies: Conjugation effect in regular verbs

Two participants were excluded from the ANOVA by participants (F1) and by items (F2) because of the high rate of non-valid responses (trials during which the voice-key was accidentally activated). Data from three items from the 2nd conjugation ("intessere", to interweave, "mietere", to harvest, and "precedere", to precede) were excluded from the analyses for their elicited error rates higher than 2.5 standard deviations over the mean. The effect of the variable Conjugation was significant: $F1(2, 78) = 17.33$, $MSE = 7049$, $p < 0.00001$; $F2(2, 78) = 3.89$, $MSE = 3986$, $p = 0.02$. Tuckey HSD post hoc tests revealed that 2nd conjugation verbs were responded to significantly slower when compared to 1st (+27 ms; $p = 0.0001$) and 3rd conjugation verbs (+17 ms; $p = 0.001$). No significant differences were detected between the verbs of the 1st and the 3rd conjugation.

2.3.1.2 Response latencies: main effects of Conjugation and Regularity and Conjugation by Regularity Interaction in 2nd and 3rd conjugation regular and irregular verbs

Two participants were excluded from these analyses because of the high rate of non-valid responses. Data from three items ("consistere", to consist, and "redigere", to draw up, 2nd conjugation, irregulars, "intessere", to interweave, 2nd conjugation, regular) were excluded

from the analyses due to error rates higher than 2.5 standard deviations over the mean. The ANOVAs by participants performed on response latencies evidenced (i) a main effect of Conjugation ($F (1, 39) = 17.71$, MSE = 6210, p = 0.0001) with slower response latencies observed for the 2nd conjugation verbs (+13 ms) compared to 3rd conjugation verbs; (ii) a main effect of Regularity ($F (1, 39) = 4.55$, MSE = 1651, p = 0.04) with slower responses for irregular verbs (+7 ms) than regular verbs; (iii) a significant Conjugation by Regularity interaction ($F (1, 39) = 4.41$, MSE = 1534, p = 0.04) with faster responses for the 3rd regular verbs (−13 ms, p = 0.004).

No significant effects were found in the ANOVA by items. A summary of results is provided in Table 2.3

Main effect of Conjugation					
	reaction times			error rate	
1st conjugation	548 ms (42)			6%	
2nd conjugation	575 ms (48)			16%	
3rd conjugation	558 ms (46)			13%	
Regularity by Conjugation Interaction					
	regular verbs		irregular verbs		total
	reaction times	error rate	reaction times	error rate	reaction times	error rate
2nd conjugation	577 ms (47)	20%	577 ms (46)	30%	577 ms (46)	25%
3rd conjugation	558 ms (46)	13%	571 ms (49)	27%	564 ms (48)	20%
total	567 ms (48)	16%	574 ms (47)	28%		

Table 2.3 Behavioral experiment: Mean reaction times (standard deviations in parentheses) and error rates.

2.3.1.3 Error analyses

The incorrect responses (whose latencies were excluded from the reaction time analyses) constituted 11% of the whole dataset and were distributed in 10 categories (Table 2.4).

Regular verbs of the 1st conjugation elicited a significantly lower error rate (5%) compared to both the regular verbs of the 2nd (13%, p = 0.001) and the 3rd (12%, p < 0.01) conjugations revealing a significant

effect of Conjugation (Wald chi-square [df = 2] = 14.62, p < 0.001). No significant differences between 2nd and 3rd conjugation verbs were observed (p>0.05).

The analyses performed on the errors of the verbs from the 2nd and the 3rd conjugations revealed a significant main effect of Regularity (Wald chi-square [df = 1] = 19.29, p < 0.001) with the irregular verbs showing a significantly higher percentage of errors (25% vs. 15%). Concerning the distributions of the incorrect responses in the category of the morphological errors, the analysis revealed no errors for the 1st conjugation. Errors due to irregularization were significantly higher for 2nd conjugation irregular verbs; regularizations errors were mostly elicited by 3rd conjugation irregular verbs; the 4% of morphological errors resulted in a conjugation change and they mostly interested the 3rd conjugation regular verbs that were attracted by the pattern of the 2nd conjugation. (Table 2.5)

Error description	Error type rate
Missing Responses (no response given)	19%
Fragments (the intended target was not entirely articulated)	30%
Hesitations (pause made during response articulation)	16%
Uninflected Stimuli (input stimulus read aloud instead of being inflected)	0.8%
Visual Errors (response that can be ascribed to a misreading of the input stimulus)	17%
Phonological Errors (target misspelt and resulting in either a word or a non-word	3%
Morphological Errors (responses morphologically related to the intended target)	11%
Mixed Errors (responses related to the intended target for more than one feature, misreading of the input stimulus and erroneous inflection)	2%
Other (responses not included in the described categories)	0.3%

Table 2.4 Description and distribution of error in different categories.

Morphological errors						
	1st conj reg	1st conj reg II	2nd conj reg	2nd conj irr	3rd conj reg	3rd conj irr

	1st conj reg	1st conj reg II	2nd conj reg	2nd conj irr	3rd conj reg	3rd conj irr
Irregularizations	0%	0%	2.9%	18.8%	1.4%	7.2%
Regularizations	0%	0%	0%	8.7%	4.3%	56.5%
Conjugation changes						
Irregularizations	1st conj reg	1st conj reg II	2nd conj reg	2nd conj irr	3rd conj reg	3rd conj irr
towards 2nd conjugation	-	-	0%	0%	0%	0%
towards 3rd conjugation	-	-	0%	0%	0%	0%
no change	-	-	3%	19%	1%	7%
Regularizations						
Regularizations	1st conj reg	1st conj reg II	2nd conj reg	2nd conj irr	3rd conj reg	3rd conj irr
towards 1st conjugation	-	-	0%	0%	0%	0%
towards 2nd conjugation	-	-	0%	0%	4%	0%
towards 3rd conjugation	-	-	0%	0%	0%	0%
no change	-	-	0%	9%	0%	57%

Table 2.5 Distribution of morphological errors across conjugations and regularity categories, with a focus on conjugation changes.

2.3.2 Functional MRI experiment

In the analysis of the activations elicited only by regular verbs (1-way ANOVA), statistically significant main effects of the Conjugation were observed in the left middle frontal gyrus (L-MFG) and left pre-supplementary motor area (L-PreSMA) ($p \leq 0.05$, cluster-level corrected) (Figure 2.1). Using a small volume correction, bilateral activation of the caudate nuclei was also detected ($p \leq 0.05$, small volume corrected) (Figure 2.1). All these regions were found significantly less activated for verbs belonging to the 1st relative to the 2nd or the 3rd conjugations (post hoc t-test: $p < 0.05$), albeit equally activated for verbs belonging to the 2nd and the 3rd conjugations (post hoc t-test: $p > 0.05$) (Figure 2.1).
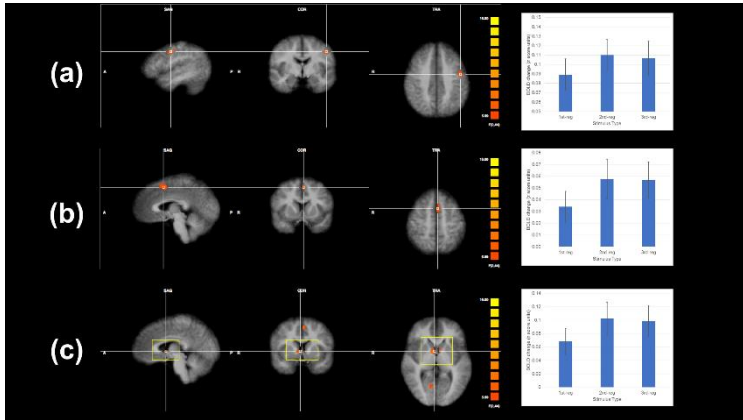
Figure 2.1 Activation pattern elicited by regular verbs: a main effect of the variable Conjugation. Clusters with statistical significant effects (overlaid on a tri-planar view of average normalized anatomical scan) and corresponding bar plots of BOLD signal changes (mean and standard error in z-score units) are shown for: (a) left middle frontal gyrus (L-MFG) ($p<=0.05$, cluster-level corrected), (b) left pre-supplementary motor area (L-PreSMA) ($p<=0.05$, cluster-level corrected) and (c) left and right caudate nuclei ($p<=0.05$, small volume corrected). The yellow box indicates the correction volume.

In the analysis of the activations elicited only by the verbs of the 2nd and 3rd conjugations (both regular and irregular) significantly higher activations for the irregular verbs compared to the regular ones in the left prefrontal cortex (L-PFC) and in the anterior part of the L-IFG was observed (Figure 2.2). Using a small volume correction, additional main effects of Regularity were also detected in L-PreSMA, in the left caudate nucleus and the left hippocampus (Figure 2.3). However, while both L-PreSMA and left caudate were significantly more activated for irregular, compared to regular verbs, the opposite pattern was observed for the hippocampus (post hoc t-test: $p < 0.05$). Independently of regularity, a cluster was detected in the perigenual part of the left ACC (L-ACC) exhibiting a significant main effect of Conjugation in the 2-way ANOVA model. This cluster was significantly more activated for verbs belonging to the 2nd relative to the 3rd conjugation (post hoc t-test: $p = 0.001$) (Figure 2.4). Finally, a significant Regularity by Conjugation interaction effect was found in the left and right postcentral gyri (R-, L-PostCG) (Figure 2.5).

Irregular verbs of the 2nd conjugation elicited significantly weaker activation than regular ones; 3rd conjugation verbs elicited the opposite pattern.
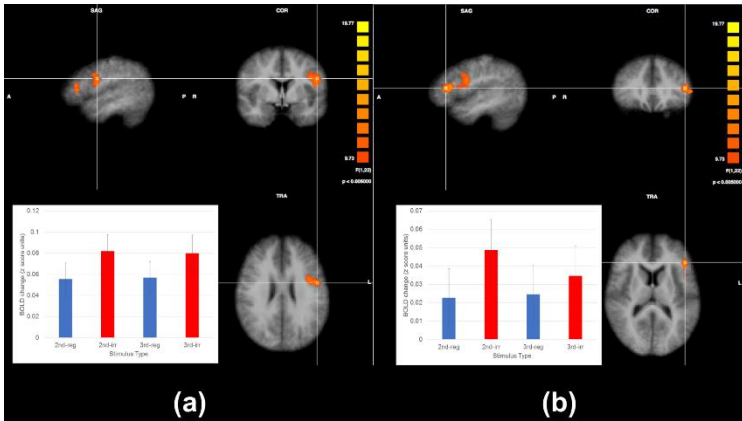

Figure 2.2 Activation pattern elicited by regular or irregular verbs from the 2nd and the 3rd conjugation: main effect of Regularity. Clusters with statistical significant effects (overlaid on a tri-planar view of average normalized anatomical scan) and corresponding bar plots of BOLD signal changes (mean and standard error in z-score units) are shown for: (a) left prefrontal cortex (L-PFC) ($p<=0.05$, cluster-level corrected), (b) left inferior frontal gyrus (L-IFG) ($p<=0.05$, cluster-level corrected).
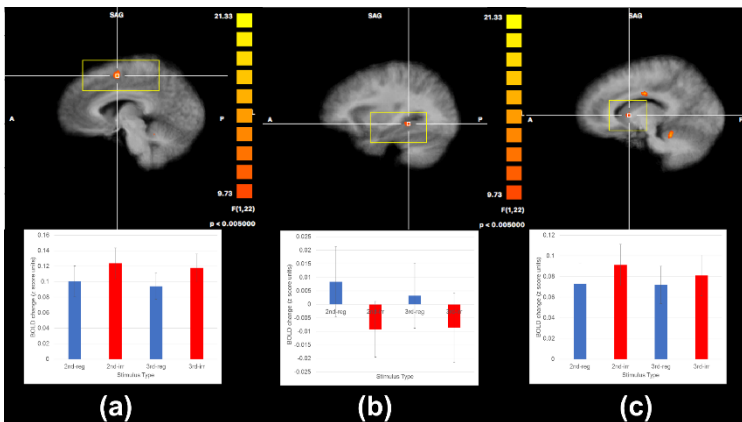
Figure 2.3 Activation pattern elicited by regular or irregular verbs from the 2nd and the 3rd conjugation: main effects of Regularity. Clusters with statistically significant effects (overlaid on a sagittal view of average normalized anatomical scan) and corresponding bar plots of BOLD signal changes (mean and standard error in z-score units) are shown for: (a) left PreSMA (p<=0.05, small volume corrected), (b) left caudate nucleus (p<=0.05, small volume corrected), (c) left hippocampus (p<=0.05, small volume corrected). The yellow boxes indicate the correction volumes.
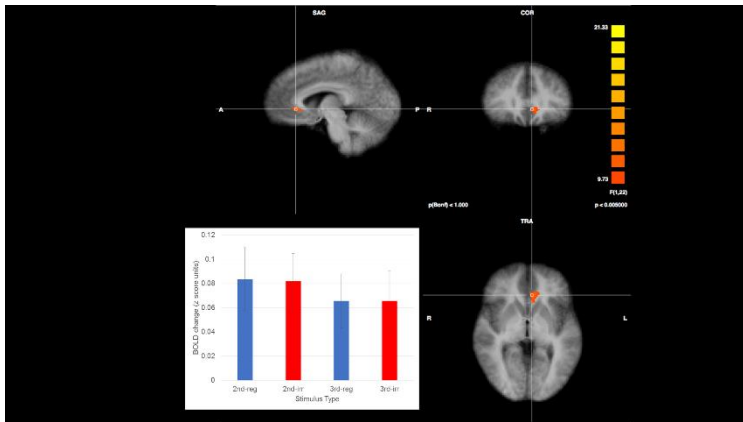

Figure 2.4 Activation pattern elicited by regular or irregular verbs from the 2nd and the 3rd conjugation: main effects of Conjugation. A cluster with statistical significant effects (overlaid on a tri-planar view of average normalized anatomical scan) and the corresponding bar plot of BOLD signal changes (mean and standard error in z-score units) is shown in the perigenual part of the left anterior cingulate cortex (L-ACC) (p<=0.05, cluster-level corrected).
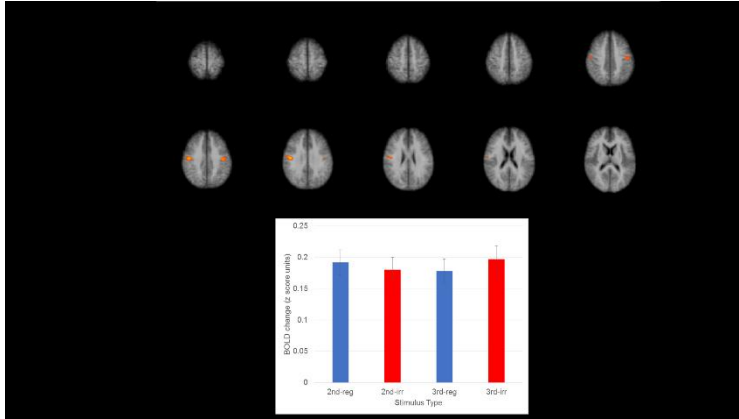
Figure 2.5 Activation pattern elicited by irregular or verbs from the 2nd and the 3rd conjugation: Regularity x Conjugation interaction. A cluster with statistically significant effects (overlaid on multiple axial slices from the average normalized anatomical scan) and the corresponding bar plot of BOLD signal changes (mean and standard error in z-score units) is shown in the bilateral post-central gyri (R-, L-PostCG).

2.4 Discussions

In this study, the neural correlates of the production of the past participle of the Italian verbs have been investigated using a behavioral and an fMRI rapid event-related experiment. Our results support the idea that the inflection is influenced by the presence of the inflectional classes.

2.4.1 Main effect of Conjugation

A main effect of Conjugation has been observed in both the behavioral and the fMRI analyses. In the former, the 2nd conjugation verbs showed significantly higher response latencies compared to both the 1st and the 3rd conjugation verbs, that showed similar responses. In the latter, significantly higher activations for the 2nd and 3rd conjugations compared to the 1st have been observed in the L-MFG, L-preSMA, and bilaterally in the caudate nucleus. Both results support our hypothesis that inflectional class information affects the inflectional process. In particular, behavioral results do not conform with the

32

single- or the dual-mechanism and they are in line with previous studies (Bordag and Pechmann, 2009; Colombo et al., 2006, 2004; De Martino et al., 2017; Laudanna et al., 2004; Veríssimo and Clahsen, 2009). Similarly, the results of the fMRI analysis do not follow the predictions either of the dual-mechanism (Ullman, 2001) or the single-mechanism (Joanisse and Seidenberg, 2005). The lower activation for the 1st conjugation verbs in the L-MFG can be explained by a lower need for control and response monitoring (Miceli et al., 1983; Perani et al., 1999; Shapiro et al., 2006; Willms et al., 2011) due to the high consistency (Laudanna et al., 2004). The higher activations elicited by the verbs from the 2nd and the 3rd conjugations compared to the 1st in the L-preSMA is coherent with the idea that this area supports the articulation planning process (Kielar et al., 2011) that is needed to select the correct items among competing alternatives (de Diego Balaguer et al., 2006). Finally, the analyses on the morphological errors show that both regularizations and irregularizations are coherent with the conjugation paradigm rather than be biased towards the 1st conjugation, thereby ruling out the possibility to explain the observed results with the dual-mechanism model, where the largest, the most regular and productive inflectional class (the 1st conjugation in Italian) acts as default (Say and Clahsen, 2002).

2.4.2 Main effect of Conjugation in 2nd and 3rd conjugation regular and irregular verbs

A Conjugation effect was observed in both behavioral and fMRI analyses in the comparison between 2nd and 3rd conjugations, with the former that elicited higher response latencies and higher BOLD activations in the L-ACC. These results support the role of the inflectional classes during the inflectional process and do not follow the predictions of the dual- or single-mechanism. The direction of the behavioral results and the widely acknowledged role of the ACC in the error detection process (Carter and van Veen, 2007) are in line with the idea that the different regularities of the 2nd and 3rd conjugations influence speakers' performances (Colombo et al., 2006, 2004; Laudanna, 2007)

2.4.3 Main effect of Regularity in 2nd and 3rd conjugation verbs

An effect of the Regularity has been observed in both the fMRI and the behavioral analyses. A disadvantage for the irregular verbs (i.e. higher error rates and slower response latencies) and higher cortical responses in the L-PFC, in the anterior part of the L-IFG, in the L-preSMA and the left caudate have been detected. On the other hand, the regional analysis performed on the subcortical nuclei and hippocampus revealed a higher cortical response for the regular verbs compared to the irregular in the left hippocampus. Although differences between the regular and irregular verbs are postulated by the dual mechanism model, the observed brain areas are not coherent with its neuroanatomical predictions (Ullman, 2001).

The higher cortical responses for the irregular verbs and its counterpart in the behavioral data can be explained by the need for higher computational processing compared to the regular ones. For instance, the activation in the anterior part of the L-IFG (partly overlapping with the BA45 area) is in line with previous studies that suggest a role of this area in controlling task-relevant features (Badre and Wagner, 2002) and in supporting the morphological processing (Carota et al., 2016). Besides, this finding is in line with a lesion study on Parkinson's disease patient showed a correlation between the thickness of the anterior parts of L-IFG and the executive control ability to select a target among competing alternatives (Di Tella et al., 2018).

All these findings suggest that the inflection of the irregular verbs is affected by the necessity of selecting among competing alternatives.

2.4.4 Conjugation by Regularity interaction in the 2nd and 3rd conjugation verbs

An interaction effect between Conjugation and Regularity has observed bilaterally in the PostCG and the behavioral analysis. For instance, 3rd conjugation irregular verbs elicited significantly higher activation and higher response latencies than 3rd conjugation regular verbs, while 2nd conjugation verbs elicited the opposite pattern. The observed brain area has been already linked with differential

2.4.3 Main effect of Regularity in 2nd and 3rd conjugation verbs

An effect of the Regularity has been observed in both the fMRI and the behavioral analyses. A disadvantage for the irregular verbs (i.e. higher error rates and slower response latencies) and higher cortical responses in the L-PFC, in the anterior part of the L-IFG, in the L-preSMA and the left caudate have been detected. On the other hand, the regional analysis performed on the subcortical nuclei and hippocampus revealed a higher cortical response for the regular verbs compared to the irregular in the left hippocampus. Although differences between the regular and irregular verbs are postulated by the dual mechanism model, the observed brain areas are not coherent with its neuroanatomical predictions (Ullman, 2001).

The higher cortical responses for the irregular verbs and its counterpart in the behavioral data can be explained by the need for higher computational processing compared to the regular ones. For instance, the activation in the anterior part of the L-IFG (partly overlapping with the BA45 area) is in line with previous studies that suggest a role of this area in controlling task-relevant features (Badre and Wagner, 2002) and in supporting the morphological processing (Carota et al., 2016). Besides, this finding is in line with a lesion study on Parkinson's disease patient showed a correlation between the thickness of the anterior parts of L-IFG and the executive control ability to select a target among competing alternatives (Di Tella et al., 2018).

All these findings suggest that the inflection of the irregular verbs is affected by the necessity of selecting among competing alternatives.

2.4.4 Conjugation by Regularity interaction in the 2nd and 3rd conjugation verbs

An interaction effect between Conjugation and Regularity has observed bilaterally in the PostCG and the behavioral analysis. For instance, 3rd conjugation irregular verbs elicited significantly higher activation and higher response latencies than 3rd conjugation regular verbs, while 2nd conjugation verbs elicited the opposite pattern. The observed brain area has been already linked with differential

processing load in articulation planning (Indefrey and Levelt, 2004; Miceli et al., 1983) and the differences can be explained in terms of the number of members and consistency of the inflectional paradigms. In fact, the 2nd conjugation is highly inconsistent as it contains 13 competing alternative paradigms and the regular verbs constitute only 22% of the class, whereas the 3rd includes only 4 different paradigms and the regular verbs are the majority (84%). Therefore, both the higher cortical response and higher response latencies for the irregular verbs of the 3rd conjugation and the regular verbs of the 2nd, compared respectively to the 3rd conjugation regular verbs and the 2nd conjugation irregular verbs, are expected and supported by the inflectional class properties. Finally, this finding is in line with previous behavioral production studies supporting the hypothesis that word representations in the output mental lexicon are informed by inflectional class properties (Laudanna et al., 2004).

2.5 Conclusions and future perspectives

Our findings showed that the inflection of the Italian verbs reflects the properties of the conjugations rather than the dichotomy regularity/irregularity supporting the hypothesis of the lexical representation of the inflectional class and suggesting further developments of both the dual and the single-mechanism theory, at least for highly inflected languages. However, future studies are needed to understand the role of different properties of the inflectional classes on the inflectional processes, whereas clinical studies on pathological populations, the use of different investigation techniques, and different grammatical classes like the nouns, could provide interesting experimental frameworks to test the lexical representation of the inflectional classes.

Chapter 3:

The neural substrate of noun morphological inflection: A Rapid Event-related fMRI Study in Italian

Andrea G. Russo, Fabrizio Esposito, Alessandro Laudanna, Azzurra Mancuso, Francesco Di Salle, Annibale Elia, Maria De Martino

3.1 Introduction

Grammar is the ability to combine words within larger structures and manipulate linguistic units, such as stems and morphemes, to generate both novel words and word forms (Hagoort, 2019). Word stems convey the core meaning, whereas word morphemes specify the inflectional properties (e.g. tense for verbs and number for nouns) by adapting words to specific contexts without modifying the meaning. For example, inflectional morphemes are used to express number information for nouns ("one dog" vs. "two or more dogs" in English) (Marslen-Wilson and Tyler, 2007). The generation of the noun inflected forms intersects both semantic (Baggio and Hagoort, 2011; Bybee, 1985; Carreiras et al., 2010; Corbett, 1991; Dahl, 2008; Fedorenko et al., 2012; Franzon et al., 2014; Friederici et al., 2000; Luzzatti and De Bleser, 1996; Molinaro et al., 2015; Strickland, 2017) and inflectional features such as the noun grammatical gender, the noun inflectional morpheme and the inflectional class. The first is a morpho-syntactic feature that in many languages participates in the generation of the noun inflectional paradigm (Aronoff, 1994). The second is usually identified with the gender suffixes and it can be considered transparent, opaque or even irregular when it is respectively, highly, equally or unreliably associated with a specific gender on distributional grounds. The third is a group of words that generate their inflected forms following the same pattern.

In general, the brain mechanisms underlying the generation of inflectional morphology have been mainly studied by comparing words with a regular and irregular inflection (see (Leminen et al., 2019) for a review) and by contrasting nouns and verbs (Benetello et al., 2016; Finocchiaro et al., 2010; Miozzo et al., 2010; Shapiro et al., 2005; Tsigka et al., 2014; Tyler et al., 2004). In addition, researchers focused also on nominal inflection by mainly investigating the consequences on the lexical access of the gender-to-morpheme transparency, i.e. the variable consistent association between gender and specific morphemes (typically the noun suffixes). It has been shown that noun forms with a gender transparent suffix are processed faster compared to nouns with opaque (i.e. morphemes that can be equally associated with more than one gender) and/or irregular

suffixes (i.e. morphemes that can be associated with more than one gender although are strongly biased towards one specific gender). Moreover, the comparison between the brain responses to gender transparent and gender opaque words elicited regionally specific effects mostly in temporal and frontal cortical areas (Caffarra et al., 2014; Caffarra and Barber, 2015; Heim, 2008; Heim et al., 2005; Hernandez et al., 2004; Miceli et al., 2002; Padovani et al., 2005; Quiñones et al., 2018; Wang and Schiller, 2019). However, it has not been yet fully clarified whether such effects are genuinely due to the suffix transparency or other to gender-related inflectional properties. Indeed, behavioral, neural and simulation studies have shown that nominal inflection is affected by distributional factors such as suffix type frequency, suffix predictability and inflectional neighborhood size (Mirković et al., 2011; Nevat et al., 2017; Zwitserlood et al., 2000), at least in highly inflected languages where multiple parameters contribute to the lexical representation of words and word-forms (Baayen et al., 2011; Hendrix, 2016; Milin et al., 2009). For instance, the Italian inflectional paradigm of nouns consists of two forms (singular and plural) that are distinguished by the inflectional suffix (i.e. the final vowel) encoding both gender and number information (i.e. letto/letti, bed/beds, masculine, casa/case, house/houses, feminine). The final vowel of the singular form (the citation form) acts as a criterion for the generation of number-inflected forms and the distinction between nominal inflectional classes (i.e. the different singular-to-plural mapping possibilities) ((Salvi and Vanelli, 2004), see Table 3.1) that are classified by descriptive grammars according to their reliability of the gender-to-ending association of the citation form (Table 3.2):

- **transparent classes**, where gender is highly associated with the citation form suffix, (i.e. masculine nouns from the -o/-i class and feminine nouns from the -a/-e class);
- **opaque classes**, where the gender suffix can be equally probable for masculine and feminine nouns in the citation form (masculine and feminine nouns from the -e/-i class);

- **irregular classes**, where the gender suffix could be untrustworthy in the citation form (masculine nouns from the -a/-i class and feminine nouns ending in -o).

Inflectional class	Inflectional endings (singular/plural)	Prevailing gender	Examples
1	-o/i	masculine	*orologio/orologi* (watch/watches)
2	-a/-e	feminine	*mela/mele* (apple/apples)
3	-e/-i	masculine and feminine	*ponte/ponti* (bridge/bridges) *nave/navi (ship/ships)*
4	-a/-i	masculine	*clima/climi* (climate/climates)
5	singular=plural	masculine and feminine	*città, crisi* (city/cities, crisis/crisis)

Table 3.1: Inflectional Classes of Italian nouns (Salvi and Vanelli, 2004)

Distribution of Italian nouns across inflectional classes		
Singular/plural suffix	Type Frequency (percentages)	Type Frequency (raw values)
o/i	31.3%	7,391
a/e	23.4%	5,528
a/i	0.6%	140
e/i	20.1%	4,737
other	24.7%	5,823
	100%	23,619

Singular/plural suffix	Masculine Nouns		Feminine Nouns	
	Type Frequency (percentages)	Type Frequency (raw values)	Type Frequency (percentages)	Type Frequency (raw values)
o/i	60.8%	7,389	0.02%	2
a/e	0.1%	8	48%	5,520
a/i	1%	138	0.02%	2

e/i	20%	2,449	20%	2,288
other	18%	2,165	32%	3,658
	100%	12,149	100%	11,470
Distribution of inflectional endings in Italian nouns				
Singular				

Suffix	Type Frequency (percentages)		Type Frequency (raw values)	
-o	33%		7,763	
-a	27%		6,459	
-e	26%		6,032	
other	14%		3,365	
	100%		23,619	

Suffix	Masculine Nouns		Feminine Nouns	
	Type Frequency (percentages)	**Type Frequency (raw values)**	**Type Frequency (percentages)**	**Type Frequency (raw values)**
-o	57%	7,726	0.4%	37
-a	4%	578	58.9%	5,881
-e	25%	3,429	26.1%	2,603
other	14%	1,906	14.6%	1,459
	100%	13,639	100%	9,980

Plural				

Suffix	Type Frequency (percentages)		Type Frequency (raw values)	
-i	54%		12,831	
-e	30%		7,048	
other	16%		3,740	
	100%		23,619	

Suffix	Masculine Nouns		Feminine Nouns	
	Type Frequency (percentages)	**Type Frequency (raw values)**	**Type Frequency (percentages)**	**Type Frequency (raw values)**
-i	75,27%	10,266	25,7%	2,565
-e	8,1%	1,099	60%	5,949
other	16,7%	2,274	15%	1,466
	100%	13,639	100%	9,980

Table 3.2: Distribution of noun inflectional suffixes and inflectional classes across grammatical genders in written Italian. Counts were performed on the DeGNI database (De Martino et al., 2019, 2018) that provides gender and inflectional class annotations for noun-types (23,619 records) extracted from the CoLFIS corpus (Bertinetto et al., 2005).

However, in Italian, there are additional features of the inflectional suffixes that might play a critical role in the lexical access.

First, the same suffix can be present in both masculine and feminine nouns, but it can be differently associated with the gender and the noun inflectional pattern. For example, the suffix –a is present in the citation form of two kinds of non-opaque (i.e. transparent and irregular) nouns, however, it is gender transparent for feminine nouns (i.e. class #2 in Table 3.1) and irregular for masculine nouns (i.e. class #4 in Table 3.1). It has been shown that these associations influence the lexical processing (Cubelli et al., 2005; De Martino et al., 2017, 2011) and, likely, they affect also the inflectional mechanism.

Second, in some Italian inflectional classes, the transparency of the gender suffix is not fixed as it changes within the paradigm of the noun. In fact, it is possible that a noun suffix is transparent in one inflectional variant of the noun (i.e. the plural form of the class #3 in Table 3.1) and not transparent in the other one (i.e. the singular form of the class #3 in Table 3.1), thus possibly influencing the inflectional operations. This variability in the degree of transparency of the gender suffix across different inflectional variants of the same noun constitutes the index of consistency of Italian nominal inflectional classes. It has been hypothesized, but not fully clarified, that the consistency of inflectional classes, their type-frequency (number of members) and token frequency (cumulative frequency of members) could affect lexical processing, and specifically inflectional operations, independently of, or in interaction with, the gender suffix transparency (De Martino et al., 2019; but see also Mirković et al., 2011; Nevat et al., 2017; Zwitserlood et al., 2000).

In this work, we aimed at verifying whether and how the lexical system and its neural correlates respond to distributional parameters of inflectional properties of nouns by exploiting the Italian rich and complex morphology and the availability of reliable corpus-based quantitative data about the distribution of noun inflectional

features (De Martino et al., 2019, 2018). Therefore, 50 healthy young adults were enrolled in a fast event-related functional MRI (fMRI) experiment where they were asked to inflect nouns aloud from singular-to-plural and vice versa upon visual presentation of an input form. Noun stimuli were selected to create a factorial experimental design highlighting the specific effect of the gender-to-suffix association (transparency) and the consistency and size of the inflectional class. The assumption underlying the task is that, during inflectional processes, the suffix of the input form triggers the selection of the inflectional paradigm through the identification of the grammatical gender. If the transparency of the gender suffix of the input form is the key factor in inflectional processing, we expect that the process should be less demanding for transparent input noun forms, regardless of the size and consistency of their inflectional class. For example, masculine nouns from the -a/-i class should be processed more effortlessly when being inflected from P2S than from S2P. In contrast, if the consistency and/or of the frequency distribution of the inflectional class affect inflectional processing, the inflectional operations are expected to be globally more demanding for nouns from small and/or less consistent inflectional classes. More specifically, in light of previous studies, regionally specific differences in fMRI responses between opaque and non-opaque nouns are expected in left temporal and frontal areas (Hernandez et al., 2004; Miceli et al., 2002; Padovani et al., 2005; Quiñones et al., 2018). Besides, the interplay between the transparency of gender suffixes and the size and consistency of inflectional classes should modulate the cortical activity in specific temporal and frontal regions involved in both noun processing (Finocchiaro et al., 2010; Shapiro et al., 2005) and attentional control and error monitoring (Marangolo et al., 2006, 2003; Piras and Marangolo, 2007; Pliatsikas et al., 2014).

3.2 Materials and Methods

3.2.1 Participants

Fifty healthy participants (mean age 23.94 ± 3.66 years old, 28 females and 22 males) took part in this study. They all had no known history of neurological and language disorders, were Italian native speakers, right-handed, and had a normal or corrected-to-normal vision.

The experiment conformed to the principles of the Declaration of Helsinki and was approved by the local ethics committee. All participants signed a concise, transparent, intelligible and easily accessible written informed consent where it was clearly stated all the characteristics of the experimental procedure and that the collected data would have been used for research purposes.

3.2.2 Stimuli

Seventy-six Italian nouns were selected from the inflectional classes 2, 3, and 4 (Table 3.1) and were used as experimental stimuli in two inflectional tasks: S2P and P2S (Table 3.3).

S2P task	Gender					
	Masculine			Feminine		
Citation form suffix	Input Output	Transparency	Size and consistency	Input Output	Transparency	Size and consistency
a	Panorama Panorami	non-opaque	Small inconsistent	Pagina Pagine	non-opaque	Large, consistent
e	Paese Paesi	opaque	Medium, Partially consistent	Palude Paludi	opaque	Medium, inconsistent
P2S task	Gender					
	Masculine			Feminine		
Citation form suffix	Input Output	Transparency	Size and consistency	Input Output	Transparency	Size and consistency
a	Panorami Panorama	non-opaque	Small inconsistent	Pagine Pagina	non-opaque	Large, consistent
e	Paesi Paese	opaque	Medium, Partially consistent	Paludi Palude	opaque	Medium, incosistent

Table 3.3: Scheme of the experimental design

The stimuli were arranged into 4 experimental conditions:

1. **masculine non-opaque** (or masculine irregular): singular form ending in –a; plural form ending in –i, panorama/panorami, landscape/landscapes;
2. **masculine opaque**: singular form ending in –e; plural form ending in –i, pedale/pedali, pedal/pedals;
3. **feminine non-opaque** (or feminine transparent): singular form ending in –a; plural form ending in –e, pagina/pagine, page/pages;
4. **feminine opaque**: singular form ending in –e; plural form ending in –i, parete/pareti, wall/walls.

The experimental conditions were matched for the main psycholinguist variables as reported in Table 3.4. Values for written frequency were extracted from the CoLFIS database (Bertinetto et al., 2005), whereas values for spoken frequency were taken from the VoLIP database (De Mauro et al., 1993). Imageability values were obtained by asking an additional group of 20 participants (not enrolled in the fMRI study) to judge the imageability of the experimental stimuli using a 7-points Likert scale. The initial phoneme, the phonological complexity, and the stress pattern were kept under control across all the experimental conditions.

Grammatical Gender	Masculine				Feminine			
Citation Form Suffix	-a		-e		-a		-e	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
#letters	6.7	1.5	6.6	1.3	6.8	1.7	6.7	1.9
#syllables	3.0	0.7	3.0	0.6	3.2	1.0	2.9	0.8
#phonemes	6.4	1.5	6.4	1.3	6.4	1.7	6.4	1.9
n-count	2.3	2.0	3.0	1.8	2.5	2.0	2.9	2.3
#consonant clusters	0.6	0.6	0.5	0.5	0.5	0.6	0.5	0.5
#geminates	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4

stress on antepenultimate syllable	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2
singular form written frequency (log per million)	0.9	0.8	1.1	0.6	1.3	0.7	1.3	0.7
plural form written frequency (log per million)	0.6	0.7	0.8	0.7	1.0	0.7	0.7	0.6
singular form spoken frequency (log per million)	0.8	0.8	0.7	0.8	1.1	0.8	0.9	0.8
plural form spoken frequency (log per million)	0.3	0.7	0.5	0.7	0.7	0.7	0.5	0.6
imageability	4.5	1.0	4.9	1.0	4.7	1.1	4.7	1.0

Table 3.4: Mean and standard deviation values for the psycholinguistic parameters controlled in the experiment.

The selected stimuli allowed the comparison of masculine and feminine nouns sharing the gender suffixes, "-a" and "-e", as input (S2P task) or output forms (P2S task). These suffixes are differently associated with grammatical gender across the Italian nominal inflectional classes (see Table 3.2). Stimuli in conditions (1) and (3) are respectively masculine and feminine non-opaque nouns that share the inflectional suffix in the singular form (–a) but belong to different inflectional classes. Stimuli in condition (3) are feminine nouns showing the mapping pattern "-a/-e" between singular and plural that is a highly consistent pattern mostly associated with the feminine gender. Thereby, these noun forms have highly reliable (transparent) gender suffixes both in the singular and in the plural form (see Table 3.2). On the other hand, condition (1) contains stimuli with the mapping "-a/-i" between the singular and the plural form. This pattern is highly infrequent and violates the widespread inflectional pattern of singular nouns ending in "-a" that is strongly biased in favor of feminine nouns with the "-a/-e" mapping. Therefore, the suffix of

nouns in this condition is irregular in the singular but completely transparent in the plural form (-i is the suffix for all Italian masculine plural nouns regardless of their inflectional class, see Table 3.2). Conditions (2) and (4) contain respectively masculine and feminine nouns that share the inflectional class. Inflectional suffixes for the singular and the plural form within this class show some peculiarities: the suffix "-e" in the singular is equally associated with masculine and feminine nouns, whereas the suffix "-i" in the plural is biased in favor of masculine plurals. Therefore, the plural of feminine opaque nouns violates a prevailing association between the suffix "-i" and the masculine gender (see Table 3.2).

Additional 76 nouns were extracted from the remaining inflectional classes (classes 1 and 5, Table 3.1) and were added as filler stimuli to avoid effects of list composition. A whole list of 152 stimuli was administered to the participants.

3.2.3 Experimental procedure

Participants were presented with the singular (or plural) noun form and were requested to produce overtly the plural (or singular) form. Thereby, each participant was assigned either to the S2P task group or the P2S task group in a pseudo-random order and the stimulus set was chosen accordingly. The S2P group was composed of 10 males and 15 females (mean age 24.16 ± 3.27 years old) whereas the P2S group was composed of 12 males and 13 females (mean age 23.79 ± 4.23 years old).

During the fMRI scan, 152 nouns were serially displayed to the subject. In the S2P (P2S) experiment nouns were presented in their singular (plural) form in random order on a video display unit (connected to a back-projection screen in the MRI room) controlled by a personal computer (synchronized with the fMRI acquisition). The experimental protocol was written in Python 2.7 using PsychoPy2 module (Peirce, 2008, 2007). For off-line accuracy assessment (error counting), participants' oral responses were recorded via an MRI compatible voice microphone laying on the mouth of the subjects (Serene Sound, Resonance Technology, USA) and then transcribed by one of the experimenters. For online monitoring of the experiments,

the responses were also reproduced in the MRI console via loudspeakers connected to the stimulation computer.

Each experimental trial started with a row of white fixation crosses (+++++++) that was presented on the center of the screen against a black background. Using a rapid event-related paradigm (ISI 7.5s, jitter 4.5s), stimuli (white lower-case letters, Arial font, size 60) replaced the fixation crosses for 800 milliseconds. Participants were then asked to covertly inflect the displayed noun from the singular (plural) to the plural (singular). A 1200 ms period then followed, during which the row of fixation crosses (+++++++) appeared again. When the crosses were replaced by hashtags (#######), for 1000 ms, participants were required to produce aloud the plural (singular) form of the noun. It was decided to have participants to produce their responses only after 1200 ms after the stimulus had disappeared from the screen to minimize the impact of possible motion artefacts associated with small head movements induced by word production and to reduce as much as possible the inter-trial and inter-subject variability in the expected onset of the neural responses. Moreover, a row of fixation crosses/hashtags, rather than a simple fixation cross/hashtag, was used to avoid spurious activations in visual brain areas.

Before entering the scanner, participants received instructions about the experimental session and underwent a brief simulation of the task where they were instructed to either perform the S2P task or the P2S task using a set of 10 stimuli not included in the experimental list.

3.2.4 Image acquisition

MRI was performed on a 3 Tesla scanner (Magnetom Skyra, Siemens Healthcare, Germany) equipped with a 20-channel parallel head coil. The fMRI scan consisted of 1200 volumes of a multi-band (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013) repeated gradient-echo echo-planar imaging (EPI) sequence (repetition time (TR) = 1000 ms, echo time (TE) = 30 ms, number of axial slices = 60, matrix = 96 x 96, field of view (FOV) = 240 mm, thickness= 2.5 mm, interslice gap = 0 mm, multi-band factor = 4). Two more short multi-band gradient-echo EPI sequences were acquired. The first was

identical to the sequence of 1200 volumes, whereas the second was acquired with an opposite phase encoding. Three-dimensional T1w Magnetization Prepared Rapid Gradient-Echo (MPRAGE) sequence (TR =2400 ms, TE= 2.26 ms, TI = 950 ms, flip angle = 8°, slice thickness = 1.0 mm, matrix size = 256x256, number of slices = 192 and voxel size = 1.0 x 1.0 x 1.0 mm^3) was acquired in the same session in order to have a high resolution anatomical reference for registration and normalization of the functional images.

3.2.5 Behavioral data analysis

For both tasks, incorrect responses and non-responses were scored as errors and were used in both qualitative and quantitative analyses. Errors were analyzed through a generalized linear mixed model fit by maximum likelihood (Laplace Approximation) (Baayen et al., 2008; Jaeger, 2008) in R with the lme4 package (Bates et al., 2015). The analysis included as fixed factors the two within-subject factors (grammatical gender and citation form suffix) and one between-subject factor (the inflectional task), whereas subjects and items were considered as random factors. Frequency values for both written inputs (Bertinetto et al., 2005) and spoken outputs (De Mauro et al., 1993) were added the model as predictors and their interactions with the other variables were taken into account. To evaluate the collinearity of the predictor matrix both the degree of correlation between input and output frequency and the variance inflation factor (VIF) were assessed. A qualitative analysis of the morphological errors was performed to investigate if they are caused by the interference from highly frequent inflectional patterns or by the degree of the gender suffix transparency in the different inflectional variants of the nouns.

3.2.6 FMRI data analysis

First, slice scan time correction followed by motion correction was applied to the fMRI time-series using BrainVoyager QX (Version 2.8, Brain Innovation, The Netherlands, www.brainvoyager.com). Then, the data were converted in NIfTI format and the task sequence (1200

volumes) was corrected for distortions with FSL TOPUP tool (Andersson et al., 2003; Jenkinson et al., 2012; Smith et al., 2004) using the two short sequences acquired in opposite phase encoding. The corrected fMRI series was imported back in BrainVoyager for high-pass filtering (cut-off to 0.008 Hz) and spatial smoothing (Gaussian kernel with 6mm full width half maximum). This fMRI series was registered to the MPRAGE images, and the resulting realigned data were then transformed into Talairach space before the computation of a random-effects multi-subject general linear model (GLM).

For the first-level analysis, a single study deconvolution-based general linear model (GLM) was applied to the volume time-courses with "stick" predictors defined over each interval of 20 s from the time of visual presentation of the stimulus in its input form (singular in S2P, plural in P2S). In this way, we defined 20 predictors of interest for each stimulus. The six motion parameters were added in the GLM model as confound predictors.

For the second-level analysis, a random-effects (RFX) GLM was applied and the main effects of all four predictors corresponding to the five delays between 3 s and 7 s (i.e. around the expected peak of the BOLD response) were estimated. A three-way ANOVA (2 within factor and 1 between factor) was performed, considering grammatical gender and the citation form gender suffix as within-subject factors with two levels (gender: masculine vs. feminine; citation form suffix: opaque vs. non-opaque) and inflectional task as a between-subject factor with two levels (S2P vs. P2S). The F statistics of the main effects of each factor and all interaction terms were calculated at each voxel. The resulting statistical maps were overlaid in pseudo-color on the average Talairach-normalized anatomical scan after correction for multiple comparisons using cluster-level thresholding (Forman et al., 1995; Goebel et al., 2006). More specifically, maps were initially thresholded at a voxel-wise p-value of 0.001 (uncorrected) and then given in input to a whole-brain (no mask) correction procedure based on the estimate of the intrinsic spatial smoothness of the map and on 1000 iterations of a Monte Carlo simulation, to determine the

minimum cluster size threshold ensuring a corrected p-value of 0.05 (cluster-level corrected) at each voxel.

To disclose the direction of the interaction effects, post-hoc paired t-tests comparing the different categories were performed on the average GLM betas from all clusters with significant effects from the voxel-based analysis. The results of the post-hoc contrasts were considered statistically significant at p = 0.05 after correction for multiple comparisons using the Bonferroni criterion.

3.3 Results

Data from 6 participants were excluded from both the error and the fMRI data analyses due to excessive head movement (>3mm and/or >3°, 4 participants), technical reasons (incomplete acquisition, 1 participant) and error rate exceeding 2 standard deviations the group mean (1 participant). Therefore, data from 44 subjects (19 males and 23 females) were included in the error and fMRI data analyses.

Responses to 3 stimuli (cute, skin, duce, leader and duca, duke) were removed from the fMRI dataset of both S2P and P2S task, whereas responses to 2 stimuli (eritema, erythema, and comma, subsection) were removed only from the S2P dataset since they resulted as incorrect responses for more than 10% of participants. These responses were considered in a separate predictor of no interest in the first-level fMRI analysis.

3.3.1 Behavioral Performances

Error rates, means, and corresponding standard deviations are reported in Table 3.5. Errors constituted 3.5% of the whole response dataset and were distributed in 7 categories (Table 3.6). Only 1.4% of the overall response dataset (41% of errors) resulted in morphological errors due to incorrect computation of inflection. The remaining 59% of errors were due to other factors.

No significant correlation (r=0.2, p =0.25) was observed between input and output frequency. VIF of all variables was less than 5, thus supporting the absence of collinearity among predictors. A significant 3-factors interaction between gender, citation form suffix and

inflectional task was found (Wald chi-square [df = 1] = 4.07, p = 0.05). In particular, in the comparison between S2P and P2S tasks, non-opaque masculine nouns elicited more errors than feminine non-opaque nouns, whereas the opaque nouns showed the opposite pattern with feminine nouns eliciting more errors than masculine nouns.

S2P Inflectional Task				
	masculine non-opaque	masculine opaque	feminine non-opaque	feminine opaque
Mean ± SD	1.57 ± 0.95 (8.2%)	0.57 ± 0.59 (3%)	0.43 ± 0.66 (2.30%)	0.83 ± 1.03 (4.30%)
P2S Inflectional Task				
	masculine non-opaque	masculine opaque	feminine non-opaque	feminine opaque
Mean ± SD	0.76 ± 0.7 (4%)	0.43 ± 0.68 (2.3%)	0.19 ± 0.51 (1%)	0.43 ± 0.68 (2.3%)

Table 3.5: Error rates distribution across experimental categories and inflectional tasks (Mean, standard deviations and percentages (in parentheses) values).

Error description	Error type rate
Uninflected Stimuli	34%
Missing Responses	28%
Mixed Errors (responses related to the intended target for more than one feature)	17%
Visual Errors (response that can be ascribed to a misreading of the input stimulus)	10%
Inflectional Errors (responses where the input stimulus was not correctly inflected)	8%
Hesitations (pause made during response articulation)	2%
Other (responses not included in the described categories)	1%

Table 3.6: Qualitative analysis of incorrect responses

3.3.2 Functional MRI Results

Significant main effects of gender and significant 2- and 3-way interaction effects (gender by citation form suffix and gender by citation form suffix by inflectional task) were observed.
The significant effects are reported in Table 3.7 and are described in the following paragraphs.

Gender-by-Suffix-by-Inflectional Task					
Region	**Talairach coordinates (x,y,z)**			**size (mm³)**	**P-value**
R-SFG	33	11	46	259	0.000167
L-pCC	-6	-31	31	1794	0.000003
L-MTG	-60	-34	10	246	0.000072
Gender-by-Opacity					
Region	**Talairach coordinates (x,y,z)**			**size (mm³)**	**P-value**
L-IFG	-45	11	22	441	0.000004
L-pCC	-9	-31	34	377	0.000004
L-cerebellum	-24	-64	-26	802	0.000029
R-MTG	60	-22	-20	401	0.000024

Table 3.7: Clusters of statistically significant fMRI changes ($p<0.05$, cluster level corrected, cluster forming threshold $p=0.001$).

3.3.2.1 Main effects of Gender

Significant main effects of Gender ($p<0.05$, cluster level corrected, cluster forming threshold $p=0.001$) were found in the cerebellum (bilaterally) and the R-MTG (Figure 3.1 and Figure 3.2). Post-hoc t-tests revealed that in all these regions there was an increased activation for feminine nouns when compared to masculine nouns ($p<0.05$, corrected for multiple comparisons).
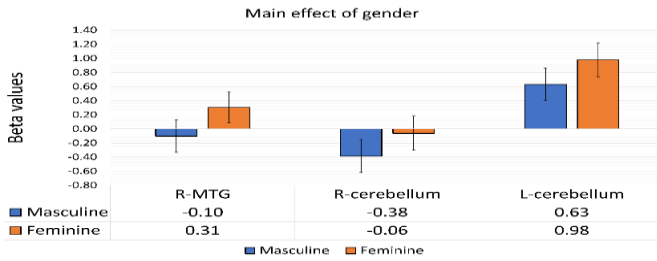
Figure 3.1: Bar plots of BOLD signal changes for masculine and feminine nouns in the R-MTG and the Cerebellum.
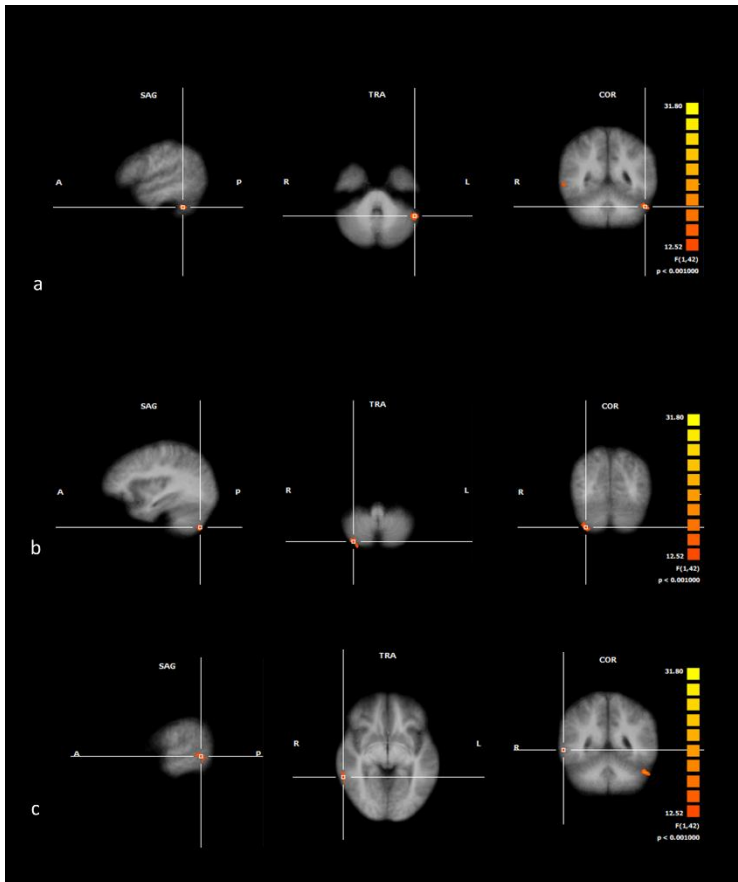
Figure 3.2: Activation pattern elicited by the "gender" factor. Clusters with statistical significant effect (overlaid on a tri-planar view of average normalized anatomical scan) are shown bilaterally in the cerebellum (b= right cerebellum; c= left cerebellum) and the right middle temporal gyrus (R-MTG, a) ($p<0.05$, cluster level corrected, cluster forming threshold p=0.001).

3.3.2.2 Gender by citation form suffix interaction

Significant gender by citation form suffix interaction effects ($p<0.05$, cluster level corrected, cluster-forming threshold p=0.001) were found in the left Inferior Frontal Gyrus (L-IFG, Figure 3.4a), in the left posterior Cingulate Cortex (L-pCC, Figure 3.4b), in the left cerebellum (Figure 3.4c) and the Right Middle Temporal Gyrus (R-MTG, Figure 3.4d).

The following contrasts reached the statistical significance in the post hoc t-tests (Figure 3.3). Increased activation was observed for masculine non-opaque (irregular) nouns when compared to masculine opaque nouns in all the significant clusters ($p<0.05$, corrected for multiple comparisons).

Feminine opaque nouns yielded a significantly higher activation in the L-IFG and the left cerebellum when compared to feminine transparent nouns ($p <0.05$, corrected for multiple comparisons).

A significant difference between masculine and feminine nouns was detected among both opaque nouns and non-opaque nouns: masculine non-opaque (irregular) nouns produced a significantly higher activation when compared to feminine non-opaque (transparent nouns) in the L-IFG, L-pCC, and R-MTG ($p<0.05$, corrected for multiple comparisons). Second, a significantly higher activation for feminine opaque nouns compared to masculine opaque nouns was found in the L-pCC, in the left cerebellum, and the R-MTG ($p<0.05$, corrected for multiple comparisons).

Gender by Citation Form Suffix Interaction

	L-IFG	L-PCC	L-CEREBELLUM	R-MTG
masculine opaque	1.29	-0.11	0.79	-0.19
masculine non-opaque	1.67	0.31	1.22	0.39
feminine opaque	1.57	0.25	1.31	0.17
feminine non-opaque	1.09	-0.03	1.02	-0.02
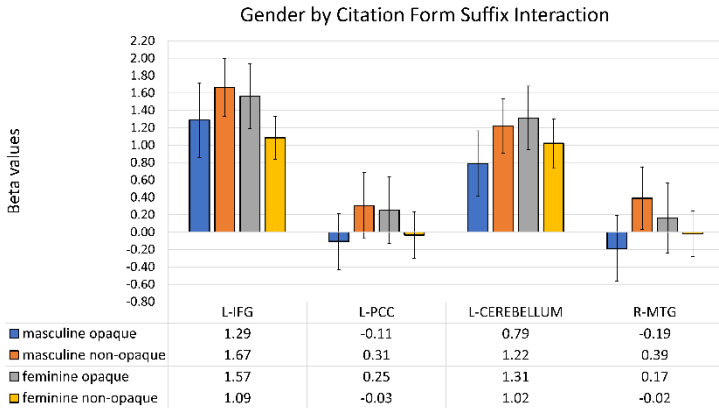
Figure 3.3: Bar plots of BOLD signal changes detected in L-IFG, L-PCC, L-Cerebellum, and R-MTG for masculine and feminine nouns with opaque vs. non-opaque citation form suffix.
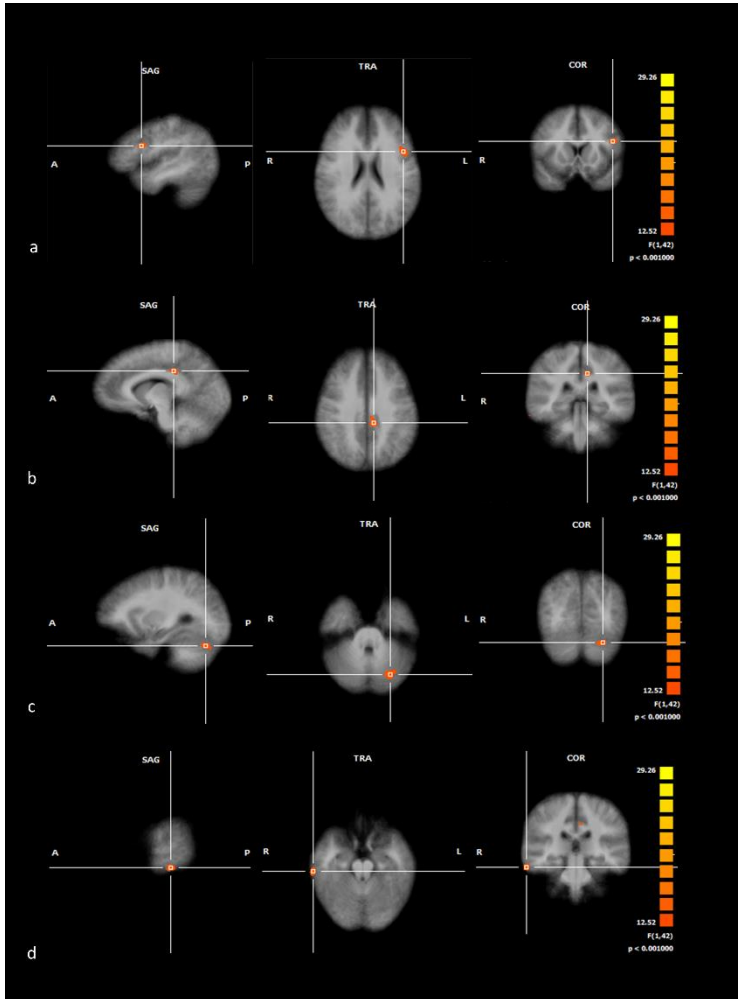
Figure 3.4: Activation pattern elicited by the interaction of two factors: gender and citation form suffix. Clusters with statistical significant effect (overlaid on a tri-planar view of average normalized anatomical scan) are shown in the left inferior frontal gyrus (a, L-IFG), left posterior cingulate cortex (b, L-pCC), left Cerebellum (c) and right middle temporal gyrus (d, R-MTG) ($p<0.05$, cluster level corrected, cluster forming threshold $p=0.001$)

3.3.2.3 Gender by citation form suffix by inflectional task interaction

Significant gender by citation form suffix by Inflectional Task interaction effects ($p < 0.05$, cluster level corrected, cluster-forming threshold $p=0.001$) were detected in the left Middle Temporal Gyrus (L-MTG, Figure 3.6a), in the L-pCC (Figure 3.6b) and the right Superior Frontal Gyrus (R-SFG, Figure 3.6c).

The following contrasts reached the statistical significance in the post hoc t-tests (Figure 3.5). In the S2P task, feminine opaque nouns exhibited significantly higher activation than masculine opaque nouns (L-MTG and L-pCC ($p<0.05$, corrected for multiple comparisons)), feminine transparent nouns (L-pCC ($p<0.05$, corrected for multiple comparisons) and masculine non-opaque nouns (L-MTG ($p<0.05$, corrected for multiple comparisons)).

Masculine irregular nouns yielded higher activation compared to both feminine transparent and masculine opaque nouns in the L-PCC ($p<0.05$, corrected for multiple comparisons).

Considering the P2S task, the R-SFG was found more activated by feminine non-opaque nouns than masculine non-opaque nouns ($p<0.05$, corrected for multiple comparisons).

BOLD signal changes in the S2P task

	L-MTG	L-pCC	R-SFG
■ Masculine "-e"	0.64	-0.32	-0.09
■ Masculine "-a"	0.83	0.28	0.36
■ Feminine "-e"	1.31	0.26	0.44
■ Feminine "-a"	0.80	-0.20	-0.05
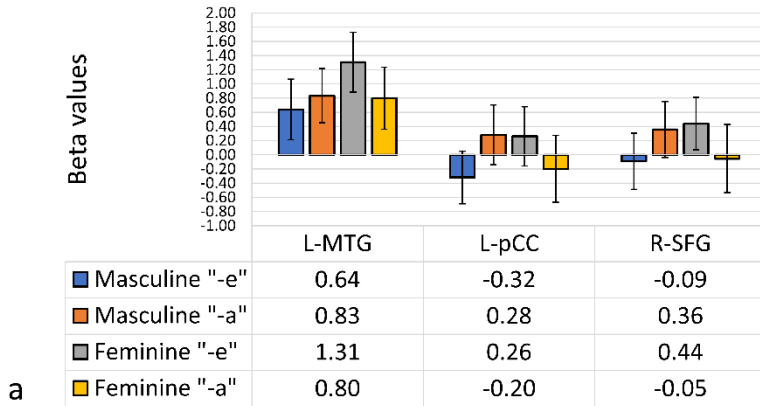
a

BOLD signal changes in the P2S task

	L-MTG	L-pCC	R-SFG
■ Masculine "-e"	1.16	0.00	-0.72
■ Masculine "-a"	0.93	-0.04	-0.78
■ Feminine "-e"	1.11	-0.07	-0.79
■ Feminine "-a"	1.49	0.27	-0.28
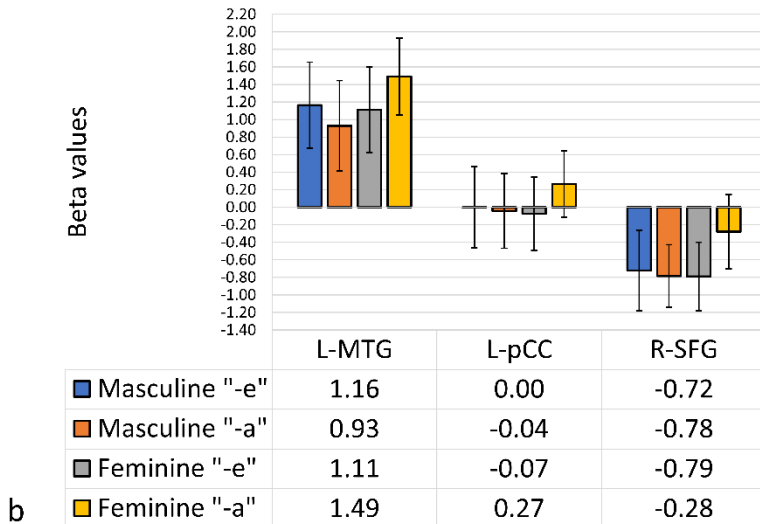
b

Figure 3.5: Bar plots of BOLD signal changes detected in R-SFG, L-MTG and L-PCC for masculine and feminine nouns with opaque vs. no opaque citation form suffix in the S2P (a) and P2S (b) inflectional task.
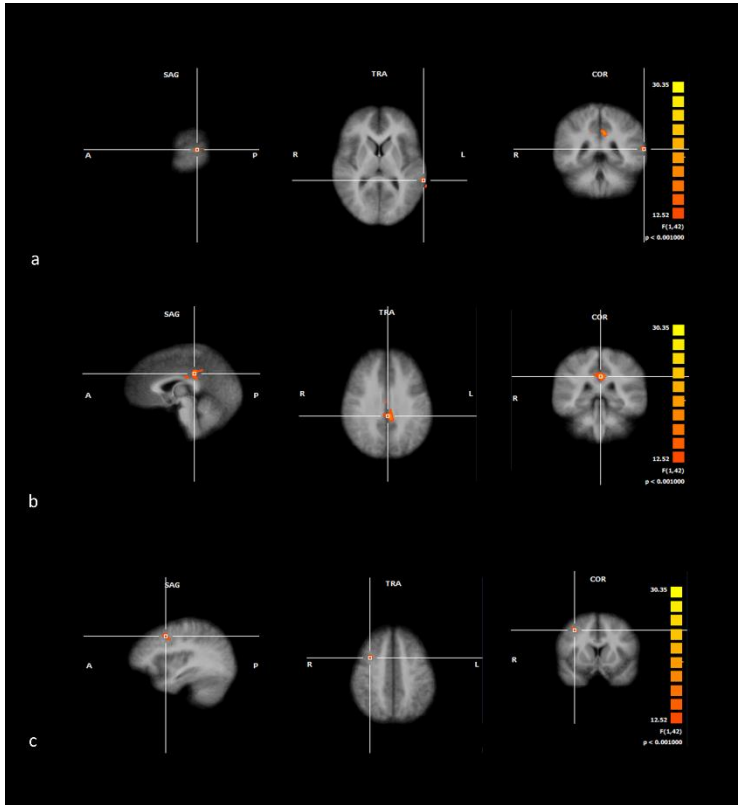
Figure 3.6: Activation pattern elicited the interactions of three factors: gender, citation form suffix and inflectional task. Clusters with statistical significant effect (overlaid on a tri-planar view of average normalized anatomical scan) are shown in the left middle temporal gyrus (a, L-MTG), left posterior cingulate cortex (b, L-pCC) and right superior frontal gyrus (c, R-SFG) (p<0.05, cluster level corrected, cluster forming threshold p=0.001).

3.4 Discussions

This work investigated the neural correlates of nominal inflection by exploring the effects of the gender transparency and the distributional

properties of inflectional classes (size and consistency) by manipulating the grammatical gender, the gender suffixes, and the inflectional classes in a factorial analysis of both behavioral (error) and neural (fMRI) data. FMRI data showed that nominal inflectional operations activate an extensive bilateral cortical network involving L-pCC, frontal areas (L-IFG, R-SFG), temporal areas (R- and L-MTG) and the cerebellum, that is in line with previous functional neuroimaging findings (Benetello et al., 2016; Finocchiaro et al., 2010; Marangolo et al., 2006, 2003; Miceli et al., 2002; Padovani et al., 2005; Piras and Marangolo, 2007; Pliatsikas et al., 2014; Quiñones et al., 2018; Shapiro et al., 2005).

Moreover, for the first time, our data support that nominal inflectional processes are sensitive to the size and consistency of the inflectional classes rather than be influenced only by the variable transparency of gender morphemes. In fact, smaller and/or less consistent inflectional classes yielded higher cortical activity, suggesting a higher cognitive and attentional demand and confirming our hypothesis that inflectional distributional factors modulate cognitive processes and neural mechanisms underlying lexical access (Baayen et al., 2011; De Martino et al., 2019, 2020; Hendrix, 2016; Milin et al., 2009; Mirković et al., 2011; Nevat et al., 2017; Zwitserlood et al., 2000).

It is worth noting that, despite the focus on a language-specific phenomenon (Italian nominal inflection), our data support the more general claim that the functioning of the mental lexicon is sensitive to the existence of highly frequent phenomena in a linguistic environment whether they are formal, grammatical or semantic (Cibelli et al., 2015; Pylkkänen et al., 2004). The lower cortical activity elicited by the inflection of noun belonging to highly frequent and consistent inflectional classes suggests that speakers benefit from the presence of words in the mental lexicon that are inflectionally consistent with the input. On the contrary, the inflection of nouns belonging to less consistent inflectional classes needs extra cognitive processing load, attentional control and error monitoring.

In the following subsections, the specific findings from behavioral and neural data will be discussed.

3.4.1 Behavioral performance

Statistical analyses performed on incorrect responses revealed an interaction between inflectional task, grammatical gender and citation form suffix showed that in the comparison between the S2P task and the P2S task. In particular, non-opaque masculine nouns (panorama, landscape) elicited more errors than feminine non-opaque nouns (pagina), whereas the opaque nouns showed the opposite pattern, with feminine nouns (parete) eliciting more errors than masculine nouns (pedale). The finding suggests that when noun forms have an unreliable gender suffix and belong to small and inconsistent inflectional classes they can be disadvantaged during inflectional operations.

3.4.2 Main effect of gender

The inflectional processing of feminine, compared to masculine, nouns were associated with significantly higher activation bilaterally in the cerebellum and the R-MTG. This finding suggests the presence of a higher attentional resource demand induced by the asymmetry of the gender-to-ending association between masculine and feminine nouns stimuli used in the present study. In our experiments, masculine nouns have a transparent (or at least a highly biased) suffix, whereas feminine nouns have a transparent suffix only when they are the input forms in the S2P task. Therefore, the inflectional suffixes of feminine nouns turned out to be, more ambiguous and trickier, thus requiring higher neural activity in regions devoted to error monitoring and attentional control. Particularly, the increase of the cerebellar activity for feminine nouns is likely to be an epiphenomenon of the executive attentional control required by the lower degree of gender transparency of feminine suffixes and it is supported by a growing number of studies suggesting a crucial role for this region in high-order cognitive functions related to language processing (Kellett et al., 2012; King et al., 2019; Marek et al., 2018; Mariën et al., 2014; Pleger and Timmann, 2018; Pliatsikas et al., 2014; Ullman, 2004). Similarly, the activation observed in the R-MTG is in line with previous lesion-

studies that proposed the involvement of right-hemisphere areas in linguistic tasks when a choice between alternatives is necessary (Marangolo et al., 2006, 2003; Piras and Marangolo, 2007).

These findings show that the noun inflectional processing elicits neural responses in regions involved in control activities related to linguistic tasks and that it is modulated by the distributional properties of inflectional morphemes.

3.4.3 Gender by citation form suffix interaction

The gender by citation form suffix interaction elicited significant activations in the L-pCC, the L-IFG, the left cerebellum and the R-MTG. In all these regions, increased neural activation was observed for masculine non-opaque nouns when compared to masculine opaque nouns. A less extended pattern was observed among feminine nouns: in the L-IFG and left cerebellum feminine opaque nouns elicited higher activations than feminine non-opaque nouns. Significantly higher activation for the masculine non-opaque compared to the feminine non-opaque nouns were observed in the L-IFG, the L-pCC and the R-MTG. Feminine opaque nouns elicited higher activation compared to masculine opaque nouns in the L-pCC, the left cerebellum and the R-MTG.

All these effects cannot be explained by the transparency of gender suffixes per se thus supporting the hypothesized role of size and consistency of the inflectional class (Baayen et al., 2011; De Martino et al., 2020; Hendrix, 2016; Milin et al., 2009; Mirković et al., 2011; Nevat et al., 2017; Zwitserlood et al., 2000). In fact, as these findings come from the pooled data of both inflectional tasks, the observed neural response is likely to be ascribed to the interplay of the nominal inflection features rather than the transparency of each specific form. Specifically, the L-IFG showed increased activation for nouns belonging to scarcely consistent inflectional classes (i.e. feminine opaque and masculine non-opaque) both compared to nouns from a highly consistent class (i.e. feminine non-opaque). These findings are supported by previous observations about the involvement of the L-IFG (BA 44, Broca's area) in the processing of gender-related inflectional affixes (Heim, 2008; Miceli et al., 2002; Padovani et al.,

2005) and the inflectional morphology of verbs (De Martino et al., 2020). The similar effects observed in the R-MTG are also coherent with this explanation and are in line with previous research (Padovani et al., 2005; Quiñones et al., 2018).

The L-pCC showed higher activation for the feminine opaque and masculine non-opaque nouns compared respectively to masculine opaque and feminine non-opaque in the L-pCC. In line with the previous observations, also this finding can be ascribed to higher attentional control and/or self-monitoring of the ongoing performance required by nouns from less consistent inflectional patterns. The L-pCC is part of a so-called "task-negative" network (Fox et al., 2005) and its activation in language tasks has been associated with the difficulty of the linguistic task or to its metalinguistic nature (Miceli et al., 2002). Moreover, a recent study documented anatomical changes in L-pCC volume after language therapy sessions based on a combination of anodal transcranial direct current stimulation (tDCS) and written naming/spelling therapy in primary progressive aphasia (de Aguiar et al., 2020).

Finally, the differences observed between masculine non-opaque and masculine opaque nouns with the same gender but belonging to different inflectional classes in the left cerebellum confirm that inconsistent inflectional paradigms require increased executive control in the linguistic task (De Smet et al., 2007; Guell et al., 2018; Kellett et al., 2012; King et al., 2019; Mariën et al., 2014; Pleger and Timmann, 2018).

3.4.3 Gender by citation form by inflectional class interaction

The L-MTG, the L-pCC and the R-SFG, were found to respond significantly to the interaction between gender, citation form suffix and inflectional task. Specifically, during the S2P task, L-pCC was significantly more activated for feminine opaque, compared to feminine non-opaque nouns and for masculine non-opaque, compared to both feminine non-opaque and masculine opaque nouns. In both cases the experimental category that showed higher activation is characterized by an inflectional morpheme that is distributionally biased towards another experimental category, thus increasing the

difficulty of the S2P inflectional task and the need of language-related attentional control and error monitoring that has been already observed in this brain region (Miceli et al., 2002).

During the S2P task, feminine opaque nouns also elicited significantly higher activation than masculine opaque nouns, in the L-MTG and the L-pCC. In this case, the effect cannot be accounted by the (suffix) transparency per se, as both the suffix –e and the S2P inflectional mapping -e/-i are equally distributed between the two genders. However, the observed effect can be ascribed to the interplay between the degree of transparency of the input suffix and the consistency of the inflectional pattern. In fact, the suffix –i is present in all masculine plural forms but only in 26% feminine plural nouns, therefore generating an asymmetry that might increase the cognitive demand for the selection of the appropriate gender for feminine plural nouns ending in -i. Moreover, L-MTG and L-pCC, have been associated respectively to lexical storage and retrieval (Binder, 2015; Heim, 2008) and increased attentional control in linguistic tasks (Miceli et al., 2002). Crucially, this finding suggests that, rather than being a fixed independent cue to grammatical gender, the transparency of gender suffixes is likely to interact with the size and consistency of the gender-to-suffix association within the noun paradigm.
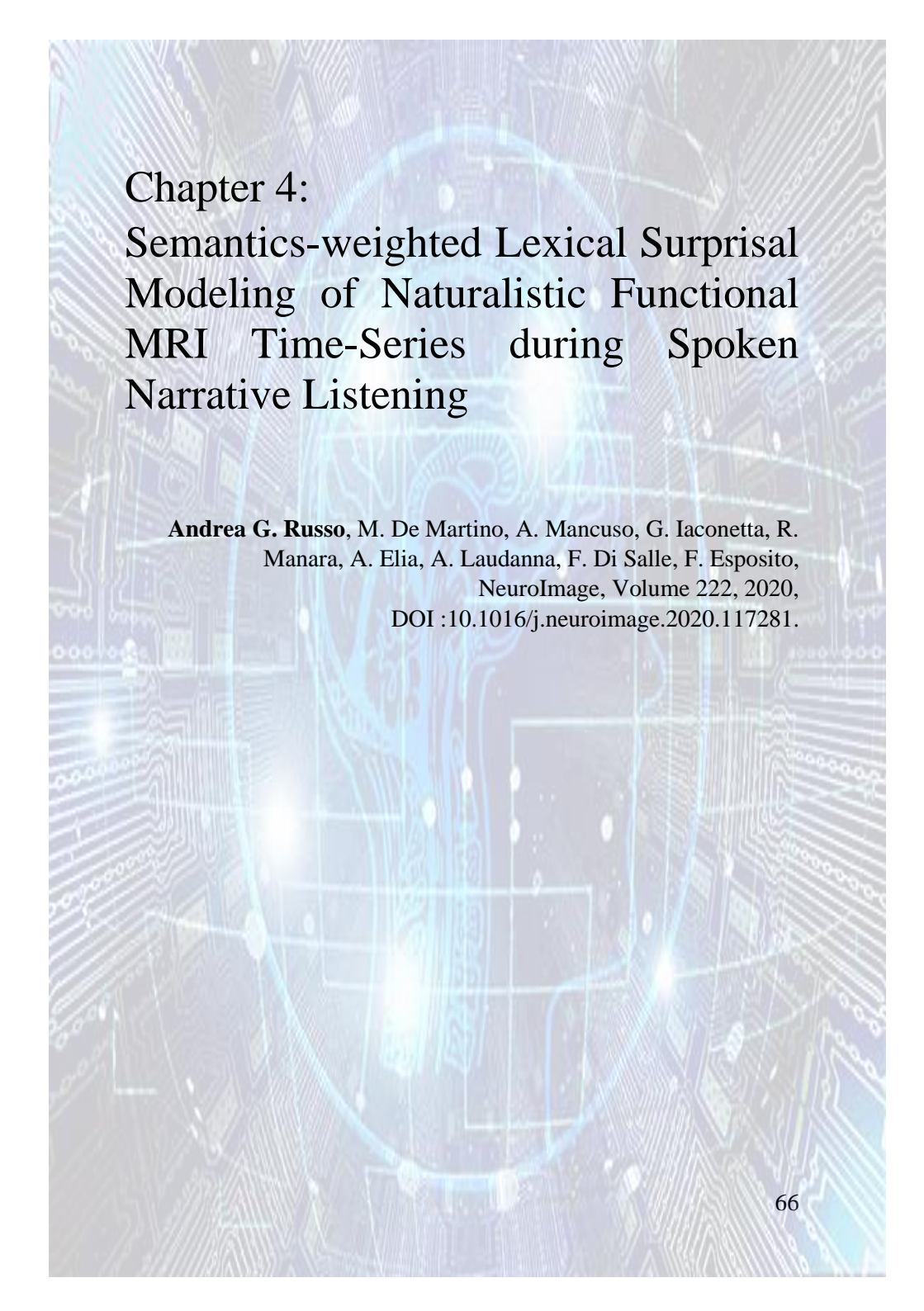
When considering the P2S task, the R-SFG was found significantly more activated for feminine non-opaque nouns when compared to masculine non-opaque nouns. This effect can be explained considering specific task-related requirements as, in the input forms of the P2S task, the suffix –i occurs more frequently than the suffix –e. In fact, it is associated to three out of four experimental categories, therefore, reproducing the distribution in the natural context (64% of noun plurals ending in -i vs. 36% of noun plurals ending in -e (De Martino et al., 2019, 2018)). Thus, the less frequent feminine non-opaque plurals ending in -e might require higher attentional control and error monitoring compared to all other plurals ending in -i and involve the intervention of an error detection system supported by the R-SFG (Dapretto and Bookheimer, 1999; Hagoort and Brown, 1999; Indefrey and Levelt, 2004; Padovani et al., 2005).

3.5 Conclusions

Our findings suggest that the neural correlates of nominal inflection are supported by an extensive bilateral cortical network involving L-pCC, frontal areas (L-IFG, R-SFG), temporal areas (R- and L-MTG) and the cerebellum. The rich and complex Italian inflectional system provided a valid testing ground to disclose the interplay between morphological (e.g. the transparency of gender suffixes) and distributional (i.e. the size and consistency of inflectional classes) factors in language processing, that have not been fully considered in the available cognitive and neuroanatomical models.

Moreover, our findings suggest that the access to the mental lexicon benefit of highly frequent phenomena in a linguistic environment, whether they are formal, grammatical or semantic, therefore supporting the general notion that the statistical structure of a language affects the functioning of the mental lexicon and its neural correlates. Future investigations are needed to further analyze the role of qualitative and quantitative properties of words from different inflectional classes in cross-linguistical, naturalistic and clinical experimental framework to provide additional insights on how the brain processes crucial morphological operations.

This research also suggests that current standards in implementing protocols for language evaluation, treatment and mapping in brain-damaged patients can be improved when inflectional morphology processing is considered, especially for highly inflected language.

Chapter 4:
Semantics-weighted Lexical Surprisal Modeling of Naturalistic Functional MRI Time-Series during Spoken Narrative Listening

Andrea G. Russo, M. De Martino, A. Mancuso, G. Iaconetta, R. Manara, A. Elia, A. Laudanna, F. Di Salle, F. Esposito,

4.1 Introduction

Probabilistic language models are increasingly used in combination with naturalistic fMRI to provide neural representations of linguistic processes beyond the perceptual experience by generating informative neural predictors from (series of) spoken or written words, which can be used to map the neural correlates of single words (Huth et al., 2016; Pereira et al., 2018) or encode language-related sequential processing occurring in the human brain (Armeni et al., 2017; Brennan et al., 2016; Lopopolo et al., 2017; Willems et al., 2016).

It is thought that, during normal functioning, the human brain is constantly engaged in predicting what (input) is coming next on the base of the available information (Clark, 2013; Friston, 2005). Similar to other cognitive domains, natural language comprehension relies on specific neural processes based on the expectation or prediction of words, letters or other linguistic elements, when these are embedded in naturalistic streams (Armeni et al., 2017; Frank and Willems, 2017; Kuperberg and Jaeger, 2016). For the naturalistic analysis of linguistic features, mismatches between the actual input and the expected one can be quantified using an information-theoretic measure known as surprisal (Hale, 2001; Levy, 2008). The latter can be estimated with the use of any probabilistic language model assigning conditional probabilities to linguistic units such as words, letters, phonemes or part-of-speech tags (Armeni et al., 2017). Therefore, a word surprisal indicates its unexpectedness (i.e. the surprise) given the previous words, and it has been parametrically linked to the language-related cognitive effort or difficulty (Demberg and Keller, 2008; Hale, 2016, 2001; Levy, 2008; Smith and Levy, 2013).

In previous fMRI studies, different linguistic predictability measures have been correlated with the blood oxygen level dependent (BOLD) signal measured during both active (reading) and passive (listening) language comprehension tasks (Carter et al., 2019; Frank et al., 2015; Frank and Willems, 2017; Henderson et al., 2016; Lopopolo et al., 2017; Shain et al., 2020; Willems et al., 2016). Among these, two previous studies employed word-level conditional probabilities in a listening task: in the first, using multiple excerpts from an audiobook, Willems et al. reported significant effects of a lexical surprisal

predictor within a broad network including the left inferior temporal sulcus, the superior temporal gyrus and the temporal poles bilaterally (Willems et al., 2016). In the second, a re-analysis of the same data set was proposed by Lopopolo et al. (2017), using the so-called perplexity (an exponential transformation of surprisal), which was developed in three different versions to address the lexical, syntactic and phonological information. The lexical perplexity was found to be more selectively correlated with the BOLD activity in the left inferior temporal gyrus and the superior temporal gyrus bilaterally (Lopopolo et al., 2017). Furthermore, Shain et al. (2020) recently showed that linguistic prediction is specifically supported by the language network and that these predictions are sensitive both to local word co-occurrence patterns and to the hierarchical structure of sentences (Shain et al., 2020).

Recent studies also showed the flexibility and the power of using distributed word embedding models to address the neurological bases of semantics in different naturalistic experimental settings (Nishida and Nishimoto, 2017; Pereira et al., 2018; Wang et al., 2018). However, to the best of our knowledge, distributed word vectors have never been used to account explicitly for the semantic dimension in linguistic prediction models like the surprisal.

The aim of this work was to provide additional insights about the neural underpinnings of linguistic prediction during spoken narrative listening by analyzing naturalistic fMRI data with an augmented surprisal model explicitly accounting for the semantic dimension. Therefore, a naturalistic fMRI experiment was designed in which healthy Italian participants were asked to listen to a 12-minute narrative in Italian while their BOLD signal was recorded and, later, to respond to a post hoc questionnaire to assess their comprehension. To model continuous fMRI time-series, two word-level surprisal predictors were derived from a large written Italian corpus (Lyding et al., 2014), respectively based on purely lexical information (Lopopolo et al., 2017; Willems et al., 2016) and on a combination of lexical and semantic features of each word (Mitchell, 2010; Mitchell and Lapata, 2009; Sayeed et al., 2015). In light of previous findings (Willems et al., 2016), we expected that the model associated with pure lexical

information would activate secondary auditory areas in the temporal cortex, whereas the augmented semantics-weighted surprisal would possibly increase the correlation between the surprisal predictor and the fMRI signal in higher-order areas such as the left inferior frontal gyrus, where the new upcoming information is supposed to be integrated in the context (Binder et al., 2009; Hagoort, 2013, 2005; Zhu et al., 2012). Moreover, as the surprisal measure has been proven to be a good proxy of the linguistic cognitive effort (Demberg and Keller, 2008; Hale, 2016, 2001; Levy, 2008; Smith and Levy, 2013), a possible correlation between the participants' elicited BOLD response (taken as a neural index of cognitive effort) and the participants' scores to the post hoc questionnaire (taken as a behavioral index of their comprehension of the story) was also theorized.

4.2 Material and Methods

4.2.1 Participants

Thirty-one healthy volunteers (Italian native proficient speakers, 23 females, mean age 24.2 ± 4.4 years old) without known psychiatric or neurological problems, with normal or corrected-to-normal vision and without hearing, developmental and language-related problems, were enrolled in the experiment. All participants were right-handed by self-report, and all participants were naive concerning the purpose of the experiment. Written informed consent was obtained following the Declaration of Helsinki, and the study was approved by the local ethics committee.

4.2.2 Stimuli

The stimulus used in the present work has been selected from the website of "Progetto Babele Rivista Letteraria" (http://www.progettobabele.it) where several written and spoken Italian narratives of semi-professional and amateur writers are available. The excerpt used in this work is "Storia di Gianna e delle

sue chiavi" ("Story of Gianna and her keys") written by Carlo Santulli and read by Silvia Cecchini (https://www.progettobabele.it/AUDIOFILES/ascolta.php?ID=841). The excerpt was spoken at an average rate of 156 words per minute that is in line with the recommended word per minute rate for audiobooks (Williams, 1998). Stimulus duration was 11:50 min (1878 written words).

The bigrams with the presence of an apostrophe were entered in the model first as two separate words, then only the surprisal relative to the second word was considered as it includes the whole bigram and it was listened by the subjects as a single word. Therefore, the final number of words used in the subsequent analysis was 1856.

Reversed speech version of the stimulus was obtained with Audacity 2.03 (https://www.audacityteam.org/) by importing the original audio track and then applying a transformation so that the end of the audio was heard first and the beginning last. This technique has been used in previous studies (Lopopolo et al., 2017; Willems et al., 2016) as the reversed speech is comparable to forward speech in terms of auditory characteristics while omitting the linguistic components (but see also Brown et al., 2012). To avoid and reduce biases due to previous knowledge of the story, in this study a short narrative of an amateur writer was chosen as our stimulus. Indeed, all subjects declared that they did not have any prior knowledge of the story.

4.2.3 Definition of Surprisal

A human speaker, using specific rules (syntax) sets the order of a list of words to convey a message. As a consequence, it can be postulated that the language system, after the processing of the first $t-1$ words $(W_1,..., W_{t-1})$ of a stream will be in a state that implicitly assigns a conditional probability: $P(W_t|W_1,...,W_{t-1})$ to each potentially upcoming word W_t (Kuperberg and Jaeger, 2016).

The surprisal associated with an upcoming new word appearing at a time (or position) t is defined as the negative logarithm of its conditional occurrence probability:

$$\text{surprisal (t)} = -\log_{10}(P((W_t \mid W_1...., W_{t-1})) \quad (1)$$

If the conditional probability of the observed word is one it means that given the left-side context of the word there are no other possible outcomes, thereby the surprise in observing the word is null. On the other hand, the occurrence of a word that was not among the candidate outcomes, i.e. an event with a null probability, corresponds to infinite surprisal value.

The conditional probabilities essential for the estimation of the surprisal values can be calculated by any probabilistic language model (Henderson et al., 2016; Willems et al., 2016).

4.2.4 Definition of semantic vector space

In 1954, Harris postulated the distributional hypothesis of language affirming that the meaning of a word can be inferred from the contexts in which it is used (Harris, 1954). Using word co-occurrences in a large corpus, it is possible to observe that, for example, the contexts in which the word "merchant" is used are similar to those in which the word "dealer" occurs, whereas the context of occurrence of "dog" and "pillars" are essentially different (Tripodi and Pira, 2017). In the natural language processing field, this hypothesis supported the construction of several algorithms to create vector space models where each word is represented by an N-dimensional vector (Dumais, 2004; Mikolov et al., 2013; Pennington et al., 2014; Sayeed et al., 2015; Turney and Pantel, 2010), where N is the number of the so-called context words that are the most common in the language corpus (excluding the determiners and conjunctions).

The i-th component of the word vector w (c_i) is estimated as the ratio between the conditional probability of the context word c_i given the word w and the (unconditional) probability of the context word c_i:

$$v_i = p\,(c_i|\,w)\,/\,p(c_i) \quad (2)$$

The numerator in (2) can be obtained by iterating through the corpus and counting how many times word w appears together with a context word ci within a fixed window of words. A window size of 10 was selected to weaken possible syntactic effects and the 1000 most common words (excluding function words such as determiners) of the

corpus were chosen as context words (Huth et al., 2016; Sayeed et al., 2015). Vector representations of unknown words (i.e. words do not present in the corpus) were artificially created using a vector of ones to guarantee the mathematical structure of the model and to not bias these unknown words towards a specific context word.

This representation allows estimating the semantic similarity between two words using the cosine of the angle between the word vectors $\hat{w}$ and $\hat{h}$

$$\cos(\varphi) = (\hat{w} * \hat{h}) /(|\hat{w}|*|\hat{h}|) \qquad (3)$$

In this work to prepare the corpus and to estimate the vector semantic space a combination of custom scripts in Python (Software Foundation. Python Language Reference, version 3.5. available at http://www.python.org) was used based on the package Natural Language Toolkit (NLTK) (Bird et al., 2009) and TreeTagger (Schmid, 1994). Additional custom scripts in R (Tierney, 2012) and bash scripts were also used.

4.2.5 Estimation of the surprisal models

Two different types of surprisal were estimated. The first is based on the word co-occurrences of the observed lexical form given the preceding local context estimated on a large corpus, therefore it can be referred to as a lexical surprisal (LS). In the second type, a multiplicative factor expressing the semantic similarity of the current word with a preceding broader context is introduced to modulate the co-occurrences of the current lexical form with the left-side local context (independent from the local one), using the distributed representations of the words in a vector space model, therefore it can be called semantics-weighted surprisal (SwS).

The word co-occurrences used to build both the semantic vector space model and the probability for the LS need to be estimated on a large linguistic corpus. In this study, the PAISÀ corpus that is a Creative Commons licensed large web corpus of contemporary Italian was used. The latter corpus contains approximately 388,000 documents

from 1,067 different websites, for a total of about 250M tokens (Lyding et al., 2014).

Word co-occurrences were estimated after a series of pre-processing steps on the corpus that was necessary to fulfil the requirements of the used software (see next sections), i.e. to reduce the sparsity of the word vector and the computational cost. Thereby, residuals and spurious HTML tags, words with a total frequency of less than fifty (in most cases they represented typos or obsolete words) were removed from the corpus. Finally, the corpus was split into sentences (one sentence per line), all words were lowered, and the punctuation marks were removed.

4.2.5.1 Estimation of the lexical surprisal

The LS was based only on the co-occurrences of the observed lexical form given its context in a large corpus. In this work, a stable and widely used stochastic model was used: the second-order Markov model, also known as trigram model (Armeni et al., 2017; Willems et al., 2016). It is based on the idea that, in a given sentence formed by words $\{W_1, W_2...., W_N\}$ the probability of observing the word W_t given the whole left-side context $P(W_t|W_1,..., W_{t-1})$ can be approximated to $P(W_t|W_{t-2}, W_{t-1})$ (i.e. the probability of observing the word Wt given only the two preceding words). Surprisal values estimated by trigram models have been used successfully in recent neuroimaging studies (Frank and Willems, 2017; Lopopolo et al., 2017; Willems et al., 2016) and in many psycholinguistic EEG studies that demonstrated that trigram-based surprisal correlates positively with the N400 (Frank et al., 2015). The LS values were estimated using the software SRILM and Kneser-Ney smoothing was used to control for possible unknown words, i.e. words that are not present in both the raw and the preprocessed version of the corpus (Stolcke et al., 2011). Unknown words were either atypical inflected forms of an Italian word or obsolete words. The absence in the corpus was either due to the corpus preprocessing (e.g. the removal of words with a frequency lower than 50) or to the fact that these words were already absent in the web documents that were searched during the creation of the corpus (Lyding et al., 2014). In total, after corpus preprocessing,

3.7% of the words in the corpus were labelled as unknown. For example, an unknown (obsolete) word was the word "vapoforno" (literally steam-oven) that refers to a specific oven used in the past to bake the bread, whereas a word with a frequency of 1, and thus lower than 50, was "sbattei" (I slammed).

4.2.5.2 Estimation of the semantics-weighted surprisal

The SwS was estimated following the semantic surprisal model originally presented in (Mitchell and Lapata, 2009) and recently implemented and used by Sayeed and colleagues in a psycholinguistic study (Sayeed et al., 2015). In this study, it has been shown that the SwS can successfully predict spoken word durations in naturalistic scenarios such as workgroup meetings (Sayeed et al., 2015). Briefly, this model integrates the probabilities obtained by the trigram model and the semantic similarity calculated from the vector space model between the considered word and its preceding words (i.e. the history of the word). However, a distinction between the "content" words (i.e. words that carry relevant semantic meaning such as "cat" or "dog") and "function" words (i.e. words that do not have a clear semantic meaning such as the determiners) is made. For content words, the trigram probability is scaled by a positive factor depending on the semantic similarity of the current word with its recent history (i.e. the words constituting the preceding context) (Mitchell and Lapata, 2009). The theoretical range of this factor is between zero and infinite and therefore it can either result in a downscaling (if it is lower than 1) or an upscaling (if it is higher than 1) of the trigram probability. For function words, the semantic scaling factor is simply set to 1. Therefore, assuming that for the function word the SwS is $P(W_t|W_{t-2}, W_{t-1})$ (i.e. it equals LS), for the content words the SwS is estimated as follows:

$$SwS = p(W_t|W_{t-1}, W_{t-2}) \, \Sigma_i w_i h_i p(c_i) \ (4)$$

where w_i is the i-th component of the semantic vector associated with the word W_t, h_i is the i-th component of the vector H that is the result of the sum of the semantic vectors associated to $\{W_1, W_2, W_3, ..., W_{t-}$

$_3$} (the words W_{t-1} and W_{t-2} are not considered to separate semantic and lexical contributes). This formulation originates from two modifications in the original definition of the semantic similarity (scalar product in eq. 3), to properly account for the influence of the frequency of the target word: (i) the terms inside the scalar product are multiplied by the unconditional probabilities of the context words (i.e. the factor $p(c_i)$) and (ii) the probability of the word to be predicted (i.e. the factor $p(w)$) is replaced by the trigram probability (Mitchell and Lapata, 2009; Sayeed et al., 2015). In our implementation of the model, the vector history H is obtained by multiplication of the semantic vectors of the four content words preceding the lower bound of the trigram (i.e. considering the current word as W_0 its history vector is composed by the words ranging from W_{-8} to W_{-3}). The four-word length has been chosen to both have a suitable amount of contextual information and to reduce as much as possible any syntactic relationship between the context and the word under examination.

However, as pointed out by Mitchell (2009) the SwS calculated using the formula (4) is no longer a probability, so a normalization step is required (Mitchell and Lapata, 2009; Sayeed et al., 2015). The normalization factor ensures (i) that the semantic similarity factor assigned to the content word depends on the semantic similarity assigned to all other words and (ii) that only the trigram probability factor is re-distributed across the other words (Mitchell and Lapata, 2009; Sayeed et al., 2015):

$$\Sigma_{Wc}\, p(W_c|W_{t-1}, W_{t-2})/\, \Sigma_{Wc}\, p(W_c) \quad (5)$$

Therefore, combining (4) and (5) the SwS of the content word can be estimated as follows:

$$SwS = p(W_t|W_{t-1}, W_{t-2}) * \Sigma_i w_i h_i p(c_i) *(\, \Sigma_{Wc}\, p(W_c|W_{t-1},W_{t-2})/\, \Sigma_{Wc}$$
$$p(W_c)) \quad (6)$$

In conclusion, given a sequence of words {W_1, W_2, W_3, ..., W_N} in a sentence, the SwS is:

- Content words = $p(W_t|W_{t-1}, W_{t-2}) * \Sigma_i w_i h_i p(c_i) *(\Sigma_{Wc}$

  $p(W_c|W_{t-1},W_{t-2})/ \Sigma_{Wc} p(W_c))$ (7)
- Function words = $p(W_t|W_{t-1}, W_{t-2})$ (8)

For a more detailed description of the semantic surprisal model and its linguistic background see (Mitchell and Lapata, 2009; Sayeed et al., 2015). For a more detailed empirical analysis of the relation between the two surprisal measures (i.e. LS and SwS), with a specific focus on the numerical impact of the semantic weighting (scaling factor) on the corresponding LS values, as well as on the specific neural effects of the semantic component isolated from the SwS model, see section 2 of the Appendix.

4.2.6 Experimental procedure

Participants listened to a story, as well as to its reversed version, while in the MRI scanner. Half of the participants started with the non-reversed stimulus and half with the reversed speech stimulus. Before entering the scanner, the participant was instructed to listen as carefully as possible. A short break separated the two versions of the stimulus.
Stimuli were presented using a custom script written in Python 2.7 with the use of the PsychoPy2 module (Peirce, 2008, 2007) via MRI compatible earphones (Serene Sound, Resonance Technology, USA). For online monitoring of the experiments, the stimuli were also reproduced in the MRI console via loudspeakers connected to the stimulation computer. Before the beginning of the acquisition, a volume test was performed: a fragment from another story with comparable voice and sound quality was presented while the scanner was collecting images. The volume of the audio was adjusted to the optimal level based on feedback from the participant.

4.2.7 Post-hoc questionnaire

After the acquisition, participants underwent a questionnaire to test their comprehension of the non-reversed story. The questionnaire contained 8 multiple choice questions regarding the non-reversed story with 4 answer options to each question. Questions were about general content, and correct answers were summed to have an overall level of understanding and attention of each participant.

4.2.8 Image acquisition

MRI was performed on a 3 Tesla scanner (Magnetom Skyra, Siemens Healthcare, Germany) equipped with a 20-channel parallel head coil. The fMRI scan consisted of 750 volumes of a multi-band (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013) repeated gradient-echo echo-planar imaging (EPI) sequence (repetition time (TR) = 1000 ms, echo time (TE) = 30 ms, number of axial slices = 60, matrix = 96 x 96, field of view (FOV) = 240 mm, thickness= 2.5 mm, interslice gap = 0 mm, multi-band factor = 4). Two more multi-band gradient-echo EPI sequences (5 volumes each) were acquired. The first was identical to the sequence of 750 volumes, whereas the second was acquired with an opposite phase encoding. Three-dimensional T1w Magnetization Prepared Rapid Gradient-Echo (MPRAGE) sequence (TR = 2400 ms, TE = 2.26 ms, TI = 950 ms, flip angle = 8°, slice thickness = 1.0 mm, matrix size = 256x256, number of slices = 192 and voxel size = 1.0 x 1.0 x 1.0 mm3) was acquired in the same session in order to have a high resolution anatomical reference for registration and normalization of the functional images.

4.2.9. Prediction performances

When comparing the prediction performances between a pure lexical and a semantics-weighted surprisal model, it is important to consider that (i) the SwS model integrates information of different nature (i.e. word co-occurrences and semantic vectors) and (ii) the words included in the conditioning context are not the same in the two models.
Regarding the first point, it is worth noting that the SwS is a weighted version of the LS where the semantic of the word is explicitly modelled, whereas, in the LS model, the semantic of the word is still

present in the lexical form but it is not explicitly taken into account during the surprisal estimation.

Regarding the second point, Frank and Willems, (2017) introduced a modified lexical surprisal model, called "skip-bigram", to overcome this disparity in the conditioning context. In brief, the "skip-bigram" probability (P_{SB}) is estimated from the co-occurrence of the pair (bigram) formed by the considered word and each word outside the context used for the trigram model, that has been used to estimate the semantic distance. The estimated probabilities are then interpolated with the lexical probability (see (Frank and Willems, 2017) for more details). According to this study, an adapted version of this model could reduce the influence of the length of the conditioning domains (see section 2.5.2) in the comparison as it takes into account the same context of the semantic distance that is based (by definition) on a higher number of context words (four context words vs. two context words). Therefore, as the difference between the LS and the SwS is in how they model the word meanings, the LS interpolated with the P_{SB} allows us to investigate and to isolate the "semantics-weighting" as it made the two conditioning domains between the pure lexical and the semantics-weighted model perfectly equitable.

The "skip-bigram" factor for each word was estimated by considering the same content words used for the estimation of the SwS semantic component. Thus, a modified version of the LS, called "skip-bigram" LS (LS_{SB}), was estimated by linearly interpolating the P_{SB} with the LS:

$$LS_{SB} = -\log_{10}(\lambda\ p(W_t|W_{t-1},\ W_{t-2})\ + (1-\lambda)P_{SB}) \quad (9)$$

where $\lambda \in [0,1]$ indicates the contribution of the P_{SB} to the surprisal. The lower the value in this modified surprisal, the better is the prediction (Frank and Willems, 2017). Therefore, the λ factor that minimized the LS_{SB} average was chosen.

An estimate of the average surprisal was computed for (and compared between) the three models (SwS, LS and LS_{SB}), by averaging the surprisal values across all words in the narrative text.

4.2.10 Functional MRI Data Analysis

MRI data pre-processing was performed using BrainVoyager (Brain Innovation, The Netherlands, www.brainvoyager.com) (Goebel, 2012), SPM12 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm/), MATLAB R2018b (The MathWorks, Inc., Natick, MA, www.mathworks.com) and FSL (Jenkinson et al., 2012). FMRI statistical analysis was performed in BrainVoyager.

4.2.10.1 FMRI Data Preprocessing

The raw slice time-series of each scan (DICOM series) were first imported in BrainVoyager and carefully reviewed with the time-course movie tool to promptly detect the occurrence of spikes in the images or other acquisition-related technical problems. In BrainVoyager, slice timing correction, followed by motion correction and high-pass filtering (cut-off to 0.008 Hz) was applied to all fMRI time-series. Moreover, starting from the motion parameters, the maximum absolute translation/rotation values and the mean framewise displacement (FD) were estimated. No de-spiking was applied to the fMRI time-series, which were exported to NIfTI format and corrected for EPI geometrical distortions with the FSL tool TOPUP (Andersson et al., 2003; Jenkinson et al., 2012; Smith et al., 2004) using the two additional short fMRI time-series acquired with opposite phase encoding to estimate the warps. The corrected series were then further pre-processed using the Data Processing Assistant for Resting-State fMRI toolkit (DPARSF; http://www.rfmri.org) (Chao-Gan and Yu-Feng, 2010). In particular, the anatomical and functional scans were spatially normalized to the standard Montreal Neurological Institute (MNI) template, functional images were resampled to 2x2x2 mm voxel sizes and then, white matter (WM), cerebrospinal fluid (CSF) signals and the Friston-24 movement parameters (Friston et al., 1996) were regressed out from fMRI time series. Finally, functional images were spatially smoothed using an isotropic 4-mm full-width at half-maximum Gaussian kernel.

After these steps, all individual fMRI series were imported back in BrainVoyager and further transformed into the Talairach space. This

step allowed the statistical analysis and presentation of group-level activation maps in the surface space on the cortical surface meshes.

4.2.10.2 FMRI Data Modeling and Statistical Analysis

For the whole-brain voxel-based statistical analysis of the fMRI time-courses, a two-level (mixed-effects) general linear model (GLM) was performed in BrainVoyager. In the first-level GLM, the correlation between surprisal estimates and fMRI time-courses was estimated as a fixed effect in every single subject in both narrative conditions (original and reversed speech). In the second-level GLM, the inter-subject variability of these effects was assessed by treating subjects as random observations.

The first-level general linear model (GLM) was applied to the volume time-courses of each subject (Friston et al., 1995) alternatively accounting for the LS or the SwS of the words. In addition to the predictor of interest, three predictors of no interest (confounds) accounting for word duration (WD), lexical frequency (LF) and root mean squared (RMS) amplitude of the word sound (WS) were added to the design matrix. The inclusion of the LF ensures that no extra effects result from the correlation of fMRI responses with LS and SwS predictors because of the absolute (context-independent) rarity of a word in a language-representative corpus, whereas the inclusion of the WD and the RMS of the WS in two additional parametric nuisance predictors ensures that no extra effects result from purely acoustic sources of variance, such as the (changing) volume of the sound, or the variable prosody, intonation and tone of the speaker's voice. To justify the use of separate fMRI data models for the two surprisal predictors (LS, SwS), as well as the inclusion of the LF as confound predictor in both models, some additional analyses were performed to explore in greater details the multicollinearity of all three predictors (LS, SwS, LF) as well as the neural effects associated with the LF variable. The description and the results of these analyses are provided in section 1 of the Appendix.

For every single word, starting from the original speech, LS and SwS values were estimated using a custom Python script. The LS and SwS series were reversed for the reversed speech. Word onsets

(timestamps) and durations were separately calculated for the original and reversed speech using Speechmatics (https://www.speechmatics.com/). The (log-transformed) LF of each word was obtained from the PAISÀ corpus (Lyding et al., 2014). As the single word represents the linguistic unit of interest for our model, all parameters (SwS, LS, WD, LF and RMS) were related to the single words, thereby a parametric design was used to build the predictors in the GLM. To this purpose, word onsets and durations were used to create a box-car predictor, i.e. a function that is one for the whole duration of a word and zero otherwise, at the time resolution of 0.01s (which is the time resolution of the word timestamp). To generate the LS, SwS and LF predictors, the boxcar predictor was amplitude-modulated at each word by the value of the parameter at that word. To generate the WS predictor, the RMS value of the audio signal was estimated over the time interval between the onset and the offset of each word. Before modulation, the series of each parameter were converted to z scores. The unmodulated (box-car) predictor was also taken to account for the variable durations of each word. All four predictors were then convolved with the hemodynamic response (to account for the hemodynamic delay) at the sampling rate of 100Hz. Finally, the convolved predictors were down-sampled at the time resolution of fMRI (1 Hz). This procedure was applied identically for the original and reversed speech stimuli. As the two surprisal models were applied separately, two separate GLM were performed where the predictor of interest was either the LS or the SwS predictor. To decorrelate the other (confound) predictors, the Gram-Schmidt procedure was implemented according to its hierarchical formulation, whereby the first predictor is simply unaffected by the procedure, the second predictor is orthogonal to the first, the third predictor is orthogonal to the subspace spanned by the first two predictors, and so on. This is equivalent to replacing each confounding variable by its residuals from a least-squares regression on the previous variables. Because each subject performed two runs (in random order), the modelled time courses from both runs (original and reversed speech) were concatenated along time and the two design matrices were both concatenated along the time dimension (rows) and duplicated (with

zero-padding) along the predictor dimension (columns) to keep the original and reverse speech conditions separate.

For the second (group) level analysis, two separate random-effects (RFX) GLM analysis were performed: one with the LS and one with the SwS, as predictor of interest in the first-level GLM. In both cases, the contrast between the "original" and the "reversed" narrative stimulus condition was evaluated, resulting in two statistical maps (t statistics). These maps were overlaid in pseudo-color on, and, for visualization purposes, projected on an inflated cortical mesh of a Talairach-normalized T1-weighted scan of a single subject. To correct for the multiple voxel-wise comparisons, a cluster-level thresholding approach was applied (Forman et al., 1995; Goebel et al., 2006), thereby, the t maps were initially thresholded at a maximum voxel-wise p-value of 0.001 (uncorrected) and then given in input to a whole-brain (no mask) correction procedure based on the estimate of the intrinsic spatial smoothness of the map and on 1000 iterations of a Monte Carlo simulation, to determine the minimum cluster size threshold ensuring a corrected p-value of 0.05 (cluster-level corrected) at each voxel.

4.2.10.3 Likelihood-based model comparison and model fitting

The clusters where the surprisal models elicited significant activations in the original speech compared to the reversed speech were further investigated to understand how well the two surprisal models explain the variance of the BOLD signal. In particular, the average BOLD time-courses for both the original and the reversed story were extracted from these clusters in each subject's data set and the resulting inter-subject data matrix (N subjects by 750x2 volumes) was used in a likelihood-based model comparison test (Brennan et al., 2016) where the surprisal, the experiment-related and stimulus-related (confounds) predictors were treated as fixed effects and the subjects as random effects. First, a specific base model containing the stimulus-related confounds predictors (LF, WD and RMS) and the experiment-related predictor (speech direction) was created for each surprisal model (two different models to take into account the different effects of the Gram-Schmidt procedure applied). Second, the corresponding

surprisal predictors (i.e. one for the "real" and one for the "reversed" speech) were added to each base model to create a specific augmented surprisal model.

To compare the two augmented statistical models (i.e. the ones that include the surprisal as first predictor) in their ability to explain the variance of the neural data (i.e. the BOLD signal from each ROI), a simulated likelihood ratio test (LRT) (Royston and Thompson, 1995) was performed.

The LRT is used to compare (and choose between) two competing statistical models. For example, within the current literature of natural language processing applied to fMRI, Brennan et al. (2016) have shown how LRTs can be repeatedly applied to hierarchically compare different syntactic models, with the purpose to evaluate the unique contribution made by each type of syntactic structure to the BOLD signal across ROIs (Brennan et al., 2016). Within the context of generalized linear models, the simulated LRT can be seen as an extension of the LRT enabling the comparison of two competing statistical models that are not nested into one another (see, e.g. Lewis et al., 2011). In the present study, the two competing models used in the fMRI data analysis contain the same number of predictors, and therefore they are not nested.

In short, the simulated LRT is performed by first creating a reference (null) distribution for the LRT statistic using one of the two models (chosen as the reference model). Then, a statistical significance for the other (alternative) model (given the null hypothesis that the two models are not different) is obtained non-parametrically by comparing the observed LRT statistic with the reference distribution (see, e.g. Lewis et al., 2011). In our case, the SwS was naturally chosen as the alternative model, as it was originally introduced as a special version of the LS, which was, instead, chosen as the reference model. Thus, for each iteration of the simulation process, a new simulated data set was created from the LS model with the parameters set to the maximum likelihood estimates (as resulting from the fitting to the observed data). Then, both models (LS and SwS) were fitted to the simulated data set, and the result of the LRT was evaluated to the reference distribution (Lewis et al., 2011). After 10000 simulations,

the estimated p-value of the model comparison was obtained as the ratio of the number of times the simulated LRT statistic is equal to or exceeds, the observed value plus one, divided by the total number of iterations plus one (Manly, 2007).

4.2.10.4 Correlation between behavioral and neural responses

The regression coefficients for the surprisal predictors from the first-level GLM fit of the regional BOLD signals were extracted for each participant and each region of significant activations in the contrast "real vs. reversed speech". These values and the percentage of correct answers of the participants in the post-hoc questionnaire were analyzed using a generalized linear model, where the regression coefficients were considered as fixed factors. The statistical significance of the effect of the regression coefficients was evaluated with an F test.

All the regional analyses were performed in MATLAB.

4.3 Results

Participants correctly answered at least five out of the eight questions (mean ± std 86% ± 10%) presented in the questionnaire, indicating a satisfactory level of comprehension of the story ($p < 0.05$, binomial distribution, chance level: 0.25). Three subjects were excluded from the fMRI data analysis due to excessive motion or incomplete data acquisition; one subject was excluded due to a technical problem which caused the loss of the data set, resulting in a group of 27 participants (21 females).

Across included participants, translation and rotation movements were all below the threshold of 3mm and 3 degrees (N=27, translation: mean ± std 0.34 ± 0.40 mm; rotation: mean ± std 0.32 ± 0.37 degree). The mean FD of each participant was below 0.5 mm (Power et al., 2012) and the grand mean FD of the analyzed cohort (N=27, mean ± std 0.11 ± 0.10 mm) was below the grand mean FD of an exemplary "low-motion" cohort (N=43, mean ± std 0.29 ± 0.12 mm) analyzed in (Patel et al., 2014).

In the surprisal analysis of the narrative text, the LS model yielded a higher average surprisal compared to the SwS (2.74 vs. 2.85) with the SwS model almost systematically reducing the actual probability space for each upcoming word in comparison with the LS model. To test if this difference was only due to the different word contexts (and more specifically to the higher number of words) used in the estimation of the SwS model, the LS was linearly interpolated with the word conditional probabilities estimated from the same context words used in the SwS estimation by creating the LS_{SB} model. In this case, the lowest average surprisal on our narrative stimulus (i.e. the best model) was obtained by choosing an interpolation factor of 0.98, indicating that the integration of the LS with further conditional probabilities estimated on the additional context used in the SwS calculation had a negligible impact on the model (Figure 4.1). Therefore, as the improvement over a purely Markov model (i.e. the LS) is very small (LS average surprisal 2.86 ± 1.60, LS_{SB} average surprisal 2.86 ± 1.59), with changes in the fMRI predictor on the second decimal digit, only the LS model was used in the fMRI analysis.
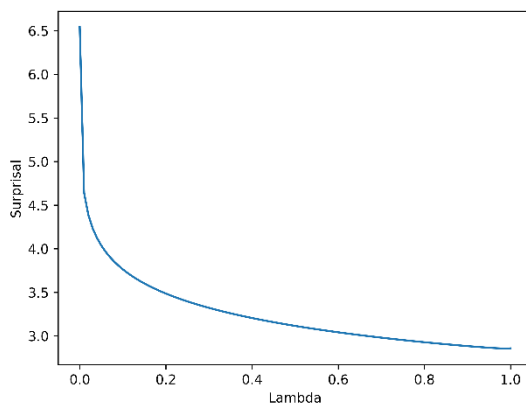
Figure 4.1: Skip-bigram interpolation effect. The average surprisal value of the narrative stimulus estimated for different interpolation factors (lambda) with the model that integrates lexical surprisal and the skip-bigram factor.

In the whole-brain voxel-based fMRI data analysis, LS and SwS predictors produced both overlapping and distinct activations in the contrast between original and reversed story conditions. In all significant clusters, the fMRI time-courses were found to be positively modulated by word surprisal (i.e. higher surprisal levels were associated with higher activation levels and vice versa) and the size of this effect was significantly higher when listening to the original, compared to the reversed, speech condition.

The LS predictor elicited significant activations (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) in both hemispheres. In the left hemisphere, activation was observed in a compact cluster extending from the posterior to the anterior portion of the superior temporal gyrus and the middle temporal gyrus (L-STG/MTG). In the right hemisphere, significant activations were detected in the superior temporal gyrus (R-STG), in the anterior temporal lobe (BA 38) (R-ATL) and the right cerebellar lobule VIIb (R-cerebellum) (Figure 4.2, Table 4.1).
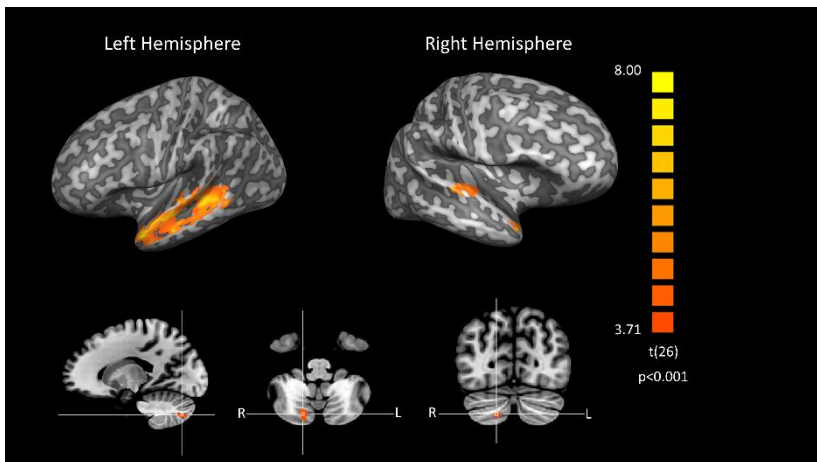
Figure 4.2: *Lexical surprisal brain response.* Activation pattern elicited by the LS model in the real speech compared to the reversed speech. Clusters with statistically significant effect (overlaid on an inflated cortical mesh from a Talairach-normalized anatomical scan) are shown in the left and right superior (L-, R-STG), in the right anterior temporal lobe (R-ATL) and in the right cerebellum (R-cerebellum).

LS			
	Talairach coord. (x,y,z)	t-value	mm^3
R-STG	57, -11, 8	6.09	717
R-ATL	45, 19, -20	7.58	1173
R-cerebellum	15, -73, -40	5.08	294
L-STG/MTG	-49, 11, -18	8.78	11831
SwS			
	Talairach coord. (x,y,z)	t-value	mm^3
R-ATL	45, 19, -20	7.17	1356
L-STG/MTG	-49, 11, -18	9.40	11770
L-IFG	-55, 19, 20	6.57	869

Table 4.1. Clusters of statistically significant fMRI activations (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) for both surprisal predictors. Activated regions had a higher positive correlation with the surprisal of the narrative during actual compared to reversed speech. The description of the area, the Talairach coordinates of the peak voxel, the cluster extent of the cluster, and the t-value of the peak voxel in the cluster are displayed in the table.

The SwS predictor activated a more left-lateralized pattern in the original, compared to the reversed, speech condition. Similar to the LS, a significant activation (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) was observed in a cluster encompassing the L-STG and L-MTG and in a cluster in the R-ATL. These two clusters overlapped by 90% (10681 mm3) and 85% (1080 mm3) respectively between the LS and SwS models. Differently from the LS, an additional cluster was detected in the left inferior frontal gyrus (L-IFG), encompassing (part of) left pars-triangularis (BA 45) and the left pars-opercularis (BA 44) (Figure 4.3, Table 4.1).
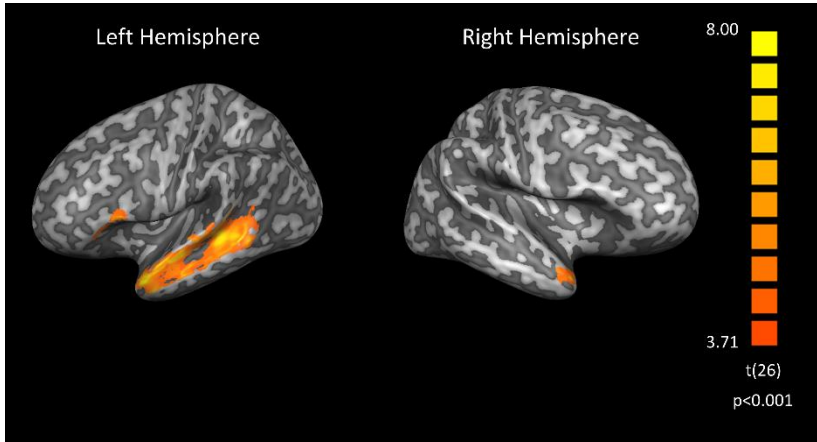
Figure 4.3: *Semantics-weighted surprisal brain response.* Activation pattern elicited by the Semantic-weighted Surprisal (SwS) in the real speech compared to the reversed speech. Clusters with statistically significant effect (overlaid on an inflated cortical mesh from a Talairach-normalized anatomical scan) in the left superior temporal gyrus (L-STG), in the right anterior temporal pole (R-ATL) and the left inferior frontal gyrus (L-IFG).

The subject-specific time-courses extracted from the above-mentioned clusters were used in a likelihood-based model comparison to evaluate differences between models in explaining the BOLD signal and to evaluate the prediction of behavioral performances from the individual model coefficients. In all regions where the LS (or SwS) predictor elicited significantly higher activation in the real, compared to the reversed, speech condition, both surprisal models significantly improved the fitting compared to the base models without surprisal predictor. Most notably, the SwS yielded better fitting compared to the LS model in the L-STG/MTG, R-ATL and L-IFG ($p<0.05$) (Table 4.2).

Likelihood-based model comparisons	
Area	**LS vs SwS**
R-STG	0.99

R-ATL	0.13
R-Cerebellum	0.87
L-STG/MTG	<0.05*
Likelihood-based model comparisons	
Area	**LS vs SwS**
R-ATL	<0.05*
L-STG/MTG	<0.05*
L-IFG	<0.05*

Table 4.2. Summary of the results of the likelihood-based model comparisons considering the neural data of all the subjects. The SwS model yielded significant better fitting compared to the LS model in the L-STG/MTG, considering the areas where the LS showed higher activation in the original compared to the reversed speech, and in all the regions (R-ATL, L-STG/MTG and L-IFG) where the SwS yielded higher activations in the original compared to the reversed speech.

No significant effects were observed for the regression coefficients of the fMRI analysis, extracted from each ROI, in the analysis of the participants' answers to the post-hoc questionnaire ($p > 0.05$).

4.4 Discussions

In this study, the neural correlates of linguistic predictions have been investigated using naturalistic fMRI (natural listening of full-spoken narratives) and a novel semantics-weighted word-level surprisal model of continuous fMRI responses. The main objective was to augment a pure lexical surprisal model of fMRI responses by integrating lexical and semantic information within the same probabilistic language framework (Mitchell and Lapata, 2009; Sayeed et al., 2015) in the express attempt to more selectively capture the neural responses associated with word (un)predictability in a natural listening scenario. Notably, this model has been shown to have lower perplexity on a held-out data set (Mitchell and Lapata, 2009) and to elicit better performances in predicting word pronunciation duration compared to the trigram model (Sayeed et al., 2015), although it was

not previously employed for studying continuously spoken language with fMRI.

4.4.1 Surprisal modelling issues

The basic assumption of surprisal modelling is that a more surprising (i.e. less predictable) word stimulates greater prediction errors which may, in turn, result in more effective modulations of neural responses in those brain regions that handle such errors and contribute to successful speech comprehension (Friston et al., 2010; Tuennerhoff and Noppeney, 2016). On these premises, the neural correlates of two (LS, SwS) word-level surprisal predictors were comparatively analyzed within a mixed-effects GLM of continuous fMRI recordings. As the SwS model was expressly formulated as a weighted version of the LS model (assigning the same LS values to the function words and resulting in a modulation with a semantic factor of the LS values for the content words), the two models were not combined in the first-level GLM and two separate whole-brain statistical analyses were performed. Indeed, as expected, for the narrative text used here, the SwS and LS predictors were highly (positively) correlated between each other ($r = +0.77$, see section 1 of the Appendix for more details), thereby the SwS predictor was strictly intended to possibly replace, and not be added to, the LS predictor, in the fMRI data model. In contrast, three stimulus-related confounds were added to the first-level GLM to properly account for the variability in linguistic (lexical frequency) and non-linguistic (duration, acoustic energy) word features, that should in principle not affect the context-specific predictability of a word when this is embedded in a spoken narrative. However, because infrequent words are in general less predictable (i.e. more surprising), among context-independent linguistic features, the neural effects of the lexical frequency were also considered in a separate "context-free" whole-brain analysis using LF as predictor of interest and WD and acoustic energy as confounds (see section 1 of the Appendix).

The SwS model showed a lower average surprisal value for the narrative stimulus compared to the LS, suggesting that, on average, the semantic information mostly reduces the conditional probability of

a word in a context. This is apparently in contrast from what could be expected based on a previous work suggesting that the integration of semantic distance with the word predictability may not be necessarily effective to improve the prediction performances (Frank and Willems, 2017). However, in our study, the context words used for estimating the lexical and the semantic components of the SwS model were purposely kept different and independent (Mitchell and Lapata, 2009; Sayeed et al., 2015) and were integrated using a scaling approach instead of a linear interpolation (Mitchell, 2010).

When comparing the raw numeric values of the LS and SwS models on content words in the narrative, there were some visible extrema in the SwS values (Figure S2, see section 2 of the Appendix for more details). However, these only occurred for 47 (out of a total of 893) content words (i.e. ~6% of the content words and ~3% of all the words in the narrative text), mostly in the lower quartiles of LS values, and no particular patterns of heteroscedasticity between the two measures were noted across LS quartiles (see Figure S2). Moreover, when performing the same comparative analysis on the two resulting fMRI predictors (i.e. the series of values across time points of the fMRI series, not across words of the narrative, after preprocessing and hemodynamic convolution), the resulting scatter plot showed no extrema (and no heteroscedasticity patterns) for the SwS predictor (see Figure S3).

Finally, to test if the observed difference between the two models was just due to the inclusion of a larger word context in the SwS, a different lexical surprisal model was estimated by following the workflow described in Frank and Willems (2017). In the present study, the further interpolation of the LS with a factor that takes into account the conditional probabilities derived from the broader (non-local) word context used to estimate the semantic component of the SwS model, had very little impact on the LS model, and this observation is in agreement with the analysis performed by Frank and Willems (2017) on the English language (Frank and Willems, 2017).

4.4.2 Surprisal-related neural effects

Both the LS and SwS predictors elicited a significant activation in the L-STG/MTG and in the R-ATL in the fMRI contrast between the actual speech (real story) and the reversed speech (reversed story). However, while the LS also activated the R-STG and the R-cerebellum, the SwS selectively activated a cluster in the L-IFG. The activation of this network of areas is perfectly in line with a recent study by Shain et al. (2020) which also showed how the cognitive processes underlying linguistic prediction, as indexed by word co-occurrence models such as the surprisal models, are fully supported by the language network and not by more general domain areas (Shain et al., 2020).

Notably, the more extended activation pattern obtained via the SwS model, encompassing both the whole L-STG/MTG and the L-IFG, could not be equally obtained by the sole semantic component (i.e. the semantic scaling factor isolated from the SwS formulation). Indeed, a separate whole-brain fMRI analysis based on the semantic component as the predictor of interest (see section 2 of the Appendix for more details) revealed that, when isolating the semantic scaling factor from the SwS model, and using it to modulate an independent word-level measure of semantic relatedness of each content word with its previous context, the significant effects of the resulting fMRI contrast within the language network were substantially reduced and mainly confined within the secondary auditory cortex (see Figure S5). Thus, considering also that the semantics weighting mainly produced a linear (R2=0.78), rather than exponential (R2=0.08), transform of the semantic similarity values across the content words (i.e. changes in semantic similarity and semantic weighting, after the transformation and normalization, were mainly proportional, see Figure S4), it is plausible that the increased sensitivity of the SwS model within the language network may truly reflect the constructive integration of a context-dependent semantic similarity with the conditional probabilities.

On the other hand, a context-free whole-brain analysis (i.e. with the LF as predictor of interest) also revealed significant positive effects (i.e. more infrequent words were associated with higher positive BOLD signal changes) within the same language network, albeit the

shape and size of the cortical patches were different from the corresponding ones highlighted by the LS and SwS predictors (Figure S1, see section 1 of the Appendix for more details). Thus, similar to previous studies (Levy, 2008; Shain et al., 2020; Staub, 2015), we also observed that both LF and (two) surprisal measures can capture a significant portion of the BOLD signal variance within the language network. These observations further support the notions that lower-frequency words are also (generally speaking) more unexpected than higher-frequency words (Staub, 2015) and that, during natural listening, word frequency effects are strictly linked with word predictability effects at the neural level (Levy, 2008; Shain et al., 2020). It should be also noted, however, that, the sign of neural effects for frequency was not always consistent among previous studies with the (most obvious) expectation that more infrequent words trigger higher (positive) BOLD signal changes (implying that the processing of these words should be carried out with greater neural processing cost). Actually, while the sign of frequency-related neural effects was positive in our naturalistic fMRI data, in full agreement with the previous work by Staub (Staub, 2015), the recent study by Shain et al. (2020) found exactly the opposite, i.e. more infrequent words were associated with lower BOLD signal changes. Thus, we could speculate that the impact of the LF measure, albeit potentially highly significant, could be in principle less specific in the prediction of the BOLD signal changes within the language network, possibly due to its strict dependency on the specific corpus with no link to the local context of the narrative being listened during measurements. However, future studies are needed to clarify the neural interactions between lexical surprisal and lexical frequency within the language network, as also suggested by Shain et al. (2020).

4.4.3 Whole-brain analysis of surprisal-related fMRI activity

The STG and the MTG are two essential cortical structures for both speech comprehension (Binder et al., 2009, 2008; Friederici et al., 2003; Lopopolo et al., 2017) and voice recognition (Belin et al., 2011; Pernet et al., 2015; Watson et al., 2014).

The bilateral involvement of the STG from the LS predictor applied to an Italian story is perfectly in line with a previous study using an identical implementation of the lexical surprisal model applied to Dutch stories. Thus, this finding also proposes the postulated prediction mechanism whereby the STG activity is modulated by the prediction (error) of the upcoming lexical form of a word (Willems et al., 2016). Contrariwise, the lack of activation of the R-STG for the SwS model might be due to the fact that the lexical predictability would involve the right hemisphere more heavily (Bonhage et al., 2015; Carter et al., 2019) whereas the semantic prediction (that is strictly related to the semantic system (Carter et al., 2019)) should be more left-lateralized (Binder et al., 2009; Carter et al., 2019).

For what concerns the involvement of the L-MTG, it has been previously suggested that this region is deeply involved in speech processing, and, more specifically, in triggering the retrieval of lexical and syntactic features of an incoming word from the long-term memory (Dronkers et al., 2004; Lopopolo et al., 2017). However, it also plays a fundamental role in facilitating the collection and storage of semantic knowledge related to the words (Binder et al., 2009). Therefore, the activation of this cortical region for both LS and SwS predictors (but also the better fitting found for the SwS predictor) would be in line with the interpretations provided in two previous studies using models of linguistic prediction and reporting the link between the L-MTG activation and both lexical (Lopopolo et al., 2017) and semantic (Weber et al., 2016) predictability of the stimulus. The activation detected in the right posterior lobule of the cerebellum corroborates the hypothesis, supported by previous findings, that this area has a role in the linguistic expectancy (Argyropoulos, 2016; Moberget and Ivry, 2016; Pleger and Timmann, 2018; Sokolov et al., 2017). Particularly, the cerebellum has been found responsive to the predictability of the upcoming word (Lesage et al., 2017), to the prediction of semantic (Carter et al., 2019; D'Mello et al., 2017) and syntactic (Carter et al., 2019) features, and, more in general, during several language processing tasks (Mariën et al., 2014).

The bilateral activation of the ATL has been observed by Willems et al. (2016) using an identical implementation of the LS on a different

language (Dutch) (Ferstl et al., 2008; Willems et al., 2016). Our finding replicates the same mechanism in Italian and therefore both strengthens the notion of this region implicated in language processing in general and, more specifically, in the integration of linguistic information within a stream of words (Carter et al., 2019), thereby confirming the sensitivity of the surprisal measure in capturing the modulation of neural responses underlying the comprehension of the text.

The left IFG has been extensively associated with language-related processing tasks such as syntax computation, semantic selection and phonological demanding tasks (Ardila et al., 2016; Binder et al., 2008; Carreiras et al., 2012; Clos et al., 2014). However, studies focusing on linguistic prediction showed mixed findings (Henderson et al., 2016; Lopopolo et al., 2017; Willems et al., 2016). In our study, the SwS elicited a significant activation in the L-IFG encompassing both the pars opercularis and the pars triangularis, which seems in line with the putative role of the L-IFG in the semantic prediction as a higher-level cognitive control processing unit (Binder et al., 2009; Ferstl et al., 2008; Hagoort, 2013, 2005; Thompson-schill et al., 2009; Zhu et al., 2012). The activation of the L-IFG for the SwS predictor and the better fitting showed by the model, therefore, suggest that the semantic integration gathers a better (more sensitive) measure of the ongoing high-level cognitive engagement required by the linguistic prediction than the LS predictor (Mitchell and Lapata, 2009; Sayeed et al., 2015). The lack of activation in the left fusiform gyrus in both LS and SwS patterns is partly surprising, as its activation has been previously reported using two different measures of linguistic prediction, albeit on the same data set (Lopopolo et al., 2017; Willems et al., 2016). However, two different explanations were provided in those studies: Willems et al. (2016) pointed out the emergence of priming effects in the fusiform gyrus which were successfully detected by the model because of the prediction (error) in the word form (Willems et al., 2016). Instead, Lopopolo et al. ascribed this effect to the continuous access to the lexico-semantic information of the word via this region (Lopopolo et al., 2017). In our case, we might speculate that the activation of this region did not reach the statistical significance for

both the LS and the SwS models most likely because, compared to Dutch and English, Italian has higher orthographic transparency and lower syllabic complexity (Borleffs et al., 2017; Seymour et al., 2003) whereby the postulated continuous access to the word form might be less necessary in general and consequently the surprisal modulation would be less effective.

4.4.4 Region-based comparison of surprisal models of fMRI activity

The fMRI time-courses from the above clusters were also used to compare the ability of the surprisal predictors to fit these neural data. Particularly, the use of a simulated LRT allowed a rigorous comparison of the two statistical models that, although related to one another, could not be nested as they included the same number of predictors and the first predictor was an alternative version of the surprisal model.

The comparisons of the two surprisal models with their corresponding base models confirmed that the use of a surprisal predictor may help in explaining more variance of the neural response as measured with fMRI during natural listening. Moreover, in comparison with the LS, the SwS model showed a significantly better performance in fitting the neural data in the L-STG/MTG, R-ATL and L-IFG. Altogether these findings support, first, the idea of using surprisal models to capture more variance in the BOLD response during language comprehension and, second, the hypothesis that the integration of the semantic information into the surprisal probabilistic framework may further improve the ability of the model in fitting the neural responses.

4.4.5 Behavioral correlates of surprisal-related fMRI activity

The regression coefficients from the ROI analysis of fMRI time-courses in the above clusters were also used in the analysis of the participants' answers to the post hoc questionnaire to test the possible correlation between the magnitude of the BOLD response and the participants' level of comprehension. In fact, as the surprisal measure has been proven to be a good proxy of the cognitive effort related to

linguistic processing (Demberg and Keller, 2008; Hale, 2016, 2001; Levy, 2008; Smith and Levy, 2013) it was possible that the higher (lower) the cognitive effort required for integrating the incoming input with the current information, the more positively (negatively) the BOLD signal had to be expected to change within the activated regions. As a consequence, the variation of the BOLD response of a participant could signal a higher (lower) cognitive engagement in the comprehension of the entire sentence albeit this does not necessarily imply that the same participant is ultimately understanding more (less) of its meaning.

The analysis of the participants' answers did not yield a significant effect of the regression coefficients of the group fMRI analysis on the participants' scores, therefore we cannot provide supporting evidence to the possibility of using fMRI measures as a possible index of comprehension. On the other hand, as all participants were healthy, fully compliant and highly educated young subjects, most of them responded correctly to the majority of the questions, and therefore it is likely that the observed variability in the participants' scores was far too small to reveal possible correlations with the BOLD responses. Thus, we cannot rule out (and actually would rather still expect) that, in future studies, the use of more complex stories and/or the submission of far more detailed questionnaires (e.g. more questions, more choices per question, etc.) or the enrollment of subjects overtly experiencing more difficulties in understanding the narrative (e.g. hypoacusic or mentally retarded patients) could gather more inter-subject variability in the participants' levels of understanding that can be better correlated with the size of surprisal-related BOLD effects.

4.4.6 Future directions

In general, the proposed experimental paradigm has several advantages that could make it useful or convenient for certain clinical applications. The use of a naturalistic stimulus (i) enables the possibility to study language-related brain mechanisms without involving the subjects in an explicit cognitive performance and (ii) may improve the overall quality of the data as naturalistic fMRI measurements are more accurate and reproducible than resting-state

fMRI measurements and also less vulnerable to the major confounds of head motion and physiological artefacts (Chen et al., 2019; Hasson et al., 2010). Particularly, the possibility to have informative neural predictors that target one or more linguistic aspects from the BOLD response changes during listening to a spoken narrative could be clinically relevant for studying subjects affected by receptive language disorder (i.e. a syndrome characterized by problems in language understanding but not production, mostly present in the children and adolescents (Goldstein and Naglieri, 2011)), as it can be used to obtain a fixed response model for both a healthy and a pathological group without the need of co-varying the response with the behavioral performance. For instance, in a cohort of participants affected by language comprehension deficits, it is likely to expect different neural behavior in key areas of the language network compared to the pattern elicited in a population with normal language abilities. However, as this approach has never been used (to the best of our knowledge) in this scenario we cannot yet formulate clear expectations about the direction of these changes.

Finally, the existence of natural language processing models based on recurrent neural networks, that appear to learn both the semantic features and the position of words within the streams (Bengio et al., 2003; De Mulder et al., 2015), suggests that future investigations and developments of the presented surprisal models, and in particular the SwS model, should possibly attempt to take the position of the words in the context into account.

4.5 Conclusions

A novel variant of the surprisal model for word prediction incorporating both lexical and semantic information has been presented and used to map the putative neural correlates of story comprehension in a naturalistic fMRI experiment. Compared to a pure lexical surprisal, the semantics-weighted surprisal provides a better predictor of the fMRI activity in left STG and MTG and is critically more sensitive in detecting linguistic processes in the left IFG. While future studies are certainly warranted, as far as the amount of surprisal

modulation detectable from continuous fMRI recordings can be interpreted as an index of the cognitive engagement during language comprehension in subjects listening to spoken narratives, the proposed approach may gather a possible neurological marker of (in)sufficient understanding of a story. In this case, analyzing different groups of people (e.g. from healthy and clinical populations) listening to the same story (or to different stories of varying language and complexity), eventually under (physiologically or pharmacologically) modified conditions, will possibly expand or challenge the validity of the proposed models. Finally, the use of highly naturalistic stimuli such as full-spoken narratives could be a crucial feature for studying linguistic processes in poorly compliant patients such as children or elderly patients.

Chapter 5:
Towards Semantic fMRI Neurofeedback: Navigating among Mental States using Real-time Representational Similarity Analysis

A. G. Russo, Michael Lührs, Francesco Di Salle, Fabrizio Esposito, Rainer Goebel (under review, preprint https://doi.org/10.1101/2020.11.09.374397)

5.1 Introduction

Real-time functional magnetic resonance imaging neurofeedback (rt-fMRI-NF) is a psychophysiological approach in which the online measured blood oxygen level dependent (BOLD) signal is provided to the subject as visual (Cohen Kadosh et al., 2016; Yamin et al., 2017), auditory (Ramot et al., 2016), haptic (Buch et al., 2008) or electrical (Young et al., 2015) feedback to allow the self-regulation of his/her neural activity towards target levels (Sitaram et al., 2017). The increasing performance in this task has been previously associated with measurable improvements in specific neurological functions and/or positive changes in behaviors (Watanabe et al., 2017), thereby rt-fMRI-NF has been successfully applied in a great variety of domains such as motor function (Scharnowski and Weiskopf, 2015; Sitaram et al., 2011), emotion regulation (Herwig et al., 2019; Linhartová et al., 2019), prosody (Rota et al., 2009) and visual task performance (Shibata et al., 2011). Furthermore, rt-fMRI-NF has found applications in the treatment of different neuropsychiatric disorders (Linden et al., 2012; Mehler et al., 2018; Orlov et al., 2018; Young et al., 2014; Zweerings et al., 2019).

The flexibility of fMRI as a functional neuroimaging tool and its successful integration with advanced computational and real-time analysis methodologies has contributed to the development of various experimental rt-fMRI-NF frameworks (Paret et al., 2019). However, there is an ongoing debate on whether or not participants in rt-fMRI-NF sessions need to be provided with an explicit strategy to regulate their brain activity. Indeed, both approaches have their strengths and weaknesses and in most cases, the choice of one over the other depends on the research question (Paret et al., 2019). Thus, there are studies that provide the subjects with a clear strategy (Kober et al., 2013; Scharnowski and Weiskopf, 2015; Zilverstand et al., 2015) as well as studies in which the subject is left free to choose the task (Amano et al., 2016; Ramot et al., 2016; Shibata et al., 2019, 2011; Watanabe et al., 2017). Furthermore, there are various choices for the source of the NF signal and how it can be calculated. For example, some studies have used the variation in the average fMRI signal from a single region of interest (ROI) compared to a baseline (deCharms et

al., 2004; Weiskopf et al., 2003; Young et al., 2014), whereas others rely on (a measure of) the functional connectivity within a network of two or more ROIs (Koush et al., 2013) to target more complex cognitive functions (Cao et al., 2014; Fair et al., 2007; Watanabe et al., 2017). Estimating the likelihood of a decoded neural activity pattern starting from a previously recorded brain pattern, is also possible (Shibata et al., 2019, 2011; Watanabe et al., 2017). Finally, there is a wide spectrum of viable options also concerning the feedback modalities (e.g. visual, auditory, haptic and electrical) and how to present the NF signal to the subject. For instance, visual feedback can be presented using simple shapes (e.g. a vertical bar or a circle) (Krause et al., 2017) as well as more immersive virtual reality interfaces (Yamin et al., 2017).

However, although different NF approaches have been successfully applied in several domains (Linden et al., 2012; Linhartová et al., 2019; Mehler et al., 2018; Paret et al., 2019; Watanabe et al., 2017), the possibility to provide more semantically-driven feedback, that, by some means, encodes the current mental state of the subject from high-level brain representations, as well as its differences from previous and other mental states, could be critically useful in some challenging scenarios such as emotion regulation (Linhartová et al., 2019). Along these lines, a semantic representation of the stimulus estimated with the application of a multidimensional approach to the current participant's brain activity may provide the extra degrees of freedom to self-modulate a mental state.

The aim of this work is to propose a novel rt-fMRI-NF paradigm based on a real-time incremental version of the representational similarity analysis (RSA) (Kriegeskorte et al., 2008a), namely real-time RSA (rt-RSA). RSA allows an abstract representation of the brain activity, in terms of concepts and associated mental states modulated by a given task. Thus, the core of this approach is a visual feedback that is semantically related to the participant's brain activity and this is done via the generation of a multi-dimensional NF signal from a high order brain region. In its real-time implementation, this feedback represents the subject's current neural activity as a movable point in a plane where a set of possible

target neural states are displayed as fixed points. The feasibility and validity of the proposed approach have been tested on real and artificial data.

5.2 Material and Methods

5.2.1 Representational Similarity Analysis

Representational theory (Dennett, 1987) can help neuroscientists in interpreting the spatial distribution of brain activity or pattern across several neuro-physical units (e.g. neurons or voxels). It is assumed that, in a given time window of observation, the spatial pattern encodes the neural representation of contents of, e.g., an image, a sound, or a motor action. On these premises, the combination of different patterns of neuronal activity defines a multi-dimensional space where the dimensions are the neuro-physical units and a single point in this space is a pattern (Kriegeskorte and Kievit, 2013). The geometrical properties of this space can be characterized and analyzed using RSA (Kriegeskorte and Kievit, 2013).

For example, considering a set of experimental conditions and their corresponding brain-activity patterns from an ROI, it is possible to estimate their mutual (dis)similarities and to encode these in a representational dissimilarity matrix (RDM) (Kriegeskorte et al., 2008a; Kriegeskorte and Kievit, 2013; Nili et al., 2014). The RDM operationally defines the geometrical space of the representations of the selected region and can be used to compare different representations (from different brain regions or conditions) via a signed-rank test (Kriegeskorte et al., 2008a; Nili et al., 2014; Wilcoxon, 1945).

The most common measure to compute the dissimilarities between two brain patterns is the correlation distance (i.e. one minus the linear correlation between spatial patterns). Statistically, this measure normalizes for both the mean and the standard deviation of the spatially variable activity. Geometrically, it is related to the angle between two high dimensional vectors (i.e. the brain patterns) and

ranges from a minimum of 0 to a maximum of 2: if the two patterns are highly correlated, their dissimilarity goes to 0, whereas if they are anti-correlated, this value goes to 2. Other possible measures are the Euclidean distance, the Mahalanobis distance, the absolute activation differences and a computational measure that is based on the pairwise classification accuracies of discriminative models (e.g. via linear discriminant analysis) (Kriegeskorte et al., 2008a; Nili et al., 2014; Walther et al., 2016).

Using a dimensionality reduction approach like multidimensional scaling (MDS) (Borg and Groenen, 2005; Kruskal and Wish, 1978; Wang, 2012), the RDM can be visualized in a two or three-dimensional scatter plot where the mutual distances among points reflect the dissimilarities among the response patterns (Kriegeskorte et al., 2008a; Nili et al., 2014).

RSA has gathered important insights in various domains, including vision (Brouwer and Heeger, 2009; Op de Beeck et al., 2008), audition (Giordano et al., 2013), categorical perception (Kriegeskorte et al., 2008b; Naselaris et al., 2012), memory (Polyn et al., 2005) and motor control (Wiestler et al., 2011). For a more exhaustive review see (Kriegeskorte and Kievit, 2013).

5.2.2 Real-time RSA

The rt-RSA approach integrates RSA into an NF paradigm enabling the comparison of multiple neural patterns (representing a given set of experimental conditions) between each other and intuitively summarizing their differences (Kriegeskorte and Kievit, 2013), in real-time. In its simplest version, for a given ROI, a visual feedback is provided to the subject consisting of a constellation of points anchored on a plane, called "representational space" (RS), corresponding to the target neural representations of a set of base stimuli, and a moving point corresponding to a variable neural representation. The coordinates of these points on the plane are estimated (and eventually those of the moving point updated) from multi-voxel patterns of regional BOLD activation which have been previously shown to identify (or encode the dynamic variability of) mental (e.g. cognitive

or emotional) states. Thereby, following the initial estimation of the RS from an fMRI localizer session, the participant's ROI activity pattern will be estimated in real-time and provided as a novel movable point in this space. The aim is to let the subject move this point towards a selected target point in the RS by learning to self-modulate his/her multi-voxel activation pattern in such a way to engage in a specific mental state. To this purpose, the subjects are provided intermittently with a visual feedback showing the position of their current brain state within the RS with respect to the base stimuli.

5.2.3 Mathematical description of the method

The possibility to project a newly calculated brain activity pattern on the existing (previously estimated) RS is based on the application of a linear solution to the distance-based triangulation problem, introduced in the field of MDS by de Silva and Tenenbaum (2004). The solution was part of a different multidimensional scaling approach originally proposed to overcome the limitation of classical multidimensional scaling (cMDS) (Kruskal and Wish, 1978; Torgerson, 1958), also known as Principal Coordinate Analysis (PCoA), when the number of entries of the input matrix is very large compared to the intrinsic dimensionality of the data (for a complete description of the method, please see de Silva and Tenenbaum, 2004). The cMDS belongs to the family of MDS methods, as it projects elements from a high-dimensional space (e.g. a brain pattern) to a lower-dimensional space while preserving the geometry as faithfully as possible (de Silva and Tenenbaum, 2004).

The solution for the distance-based triangulation problem provides a convenient mathematical framework to project a new data point (i.e. a new brain pattern) onto an existing RS by simply applying a previously estimated linear transformation. Specifically, given a set of base stimuli and their evoked activity patterns in an ROI, an RDM is initially calculated based on their pairwise dissimilarities. In our case, the metric chosen for this calculation is the correlation distance. Then, the RS corresponding to the estimated RDM is obtained by applying the cMDS to the RDM itself. Finally, using a fixed linear

transformation function, the coordinate vector $\underline{x}$ representing the position of the pattern of a new stimulus in the estimated RS is obtained according to the following formula:

$$\underline{x} = -\frac{1}{2} * L * (\underline{\delta_a} - \underline{\delta_\mu}) \qquad (1),$$

where

$$L = eigenvectors(RDM)^T / \sqrt{(eigenvalues(RDM)} \qquad (2),$$

$\underline{\delta_a}$ is the squared vector of the dissimilarities between the new stimulus and the base stimuli and $\underline{\delta_\mu}$ is the sum of the squared columns of the RDM divided by its number of entries (i.e. the number of base stimuli).

5.2.4 General experimental framework

An rt-fMRI-NF experiment with the use of rt-RSA requires at least two separate sessions: a localizer session and an NF session. To increase the statistical power of the estimated brain activity patterns, the localizer session may include multiple runs where each run can include several trials of one base stimulus (one run per base stimulus) or interleaved trials of different base stimuli. In each localizer session, a series of N trials (corresponding to different repetitions of each base stimulus) are delivered to the subject who performs an active (e.g. button pressing) or passive (e.g. image viewing) task. The ROI is ultimately selected at the end of the localizer eventually using a combination of functional and anatomical criteria based on the measured brain activation and a priori regional hypotheses, if available (Walther et al., 2016).

A single activation pattern (for a given stimulus) can be obtained from the ROI map of regression coefficients (as effect size estimates, see, e.g., Walther et al., 2016) or statistical parameters (e.g. t scores, signal-to-noise estimates, see, e.g., Walther et al., 2016) via the general linear model (GLM) analysis of the fMRI responses at each voxel. The GLM

estimation can be either performed online, thereby the signal (and noise) incremental estimates are recursively updated over successive trials, or offline, by eventually concatenating multiple runs using the most complete basis set of stimulus and confound predictors, to maximize the accuracy and power of all base stimulus patterns (to be used for the base RS).

During the NF session, the multi-voxel pattern of the selected ROI is updated via online GLM and its position within the previously estimated RS is periodically displayed as a visual stimulus to the subject. The participants are thus periodically informed about the position of their current mental representation (as obtained from their current brain activity) with respect to the mental representations associated with all base stimuli (as obtained from the brain activity measured during the localizer runs). In this way, the subjects are stimulated to self-modulate their brain activity to change its position within the RS towards the target position. Operationally, at the end of each task block, the participant's multi-voxel pattern is extracted from the predefined ROI and the dissimilarities of this pattern vs. the base stimulus patterns are calculated and given as input to (1) to update the corresponding coordinates in the RS. Moreover, while collecting the series of new patterns (i.e. mental states) over successive NF trials, the trajectory of the current mental state is also displayed to the subject, thereby the history of the modulation is kept visible to the subject to possibly incentive (or disincentive) an undertaken strategy towards reaching the target state.

A graphical description of the procedure is provided in Figure 5.1.

Localizer session

Neurofeedback session
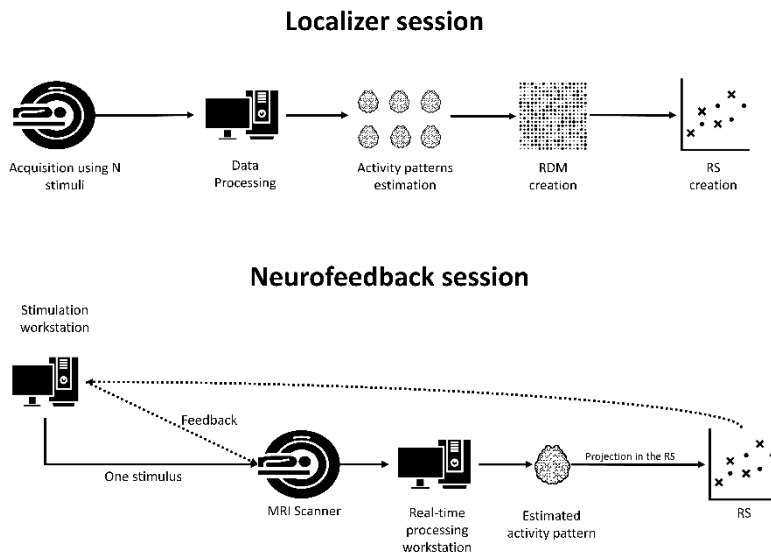
Figure 5.1 General workflow of an fMRI-NF experiment with the use of the rt-RSA. In the localizer session, the subjects perform a task (e.g. passive viewing) and are exposed to a set of experimental conditions. The brain-activity related to each experimental condition is extracted from a region-of-interest (ROI) and used to estimate a representational dissimilarity analysis (RDM). Then, the latter is projected on a plane that is the representational space (RS) of the ROI.

In the neurofeedback session, the subjects perform the same task but are provided with one stimulus. The fMRI data are processed in real-time and the brain activity of the ROI (used in the localizer session) are periodically extracted and projected in the RS. The position of this projection in the RS is provided intermittently to the subject as visual feedback.

5.2.5 Method implementation

The proposed method was implemented in Python (Python Software Foundation. Python Language Reference, version 3.7. Available at http://www.python.org)(van Rossum, 1995). The fMRI data were processed online and offline using Turbo-BrainVoyager version 4.0 (TBV) (Brain Innovation B.V., Maastricht, The Netherlands) and

imported in Python using the TBV Network Access Plugin (Network Access Plugin) and the corresponding interface (GitHub - expyriment/expyriment-stash) as part of the open-source library Expyriment (Krause and Lindemann, 2014).

All the scripts and the data used in this work are available at https://github.com/andreagrusso/rtRSA

5.2.6 fMRI Neurofeedback experiment

5.2.6.1 Participant

Functional and anatomical data of one healthy left-handed volunteer (male, age 27), with normal vision and without known neurological or psychiatric disorders, were obtained using a 7 Tesla scanner (Siemens Healthcare, Germany). Informed written consent was obtained before the study and the experimental procedure was approved by the local Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

5.2.6.2 Experimental design

The complete scanning session was divided into two main parts: an offline training session (outside the MR scanner) and a scanning session including the localizer and the NF experiment. In both experiments (localizer and NF) the participant performed an imagery task upon the delivery of an auditory cue.

In the training session, the subject was asked to familiarize with the stimuli by visually inspecting and memorizing the images of selected objects as well as listening to their accompanying cues. Two animate (cat, dog) and two inanimate (chair, hammer) objects were chosen as base stimuli. The training continued inside the MR scanner during the acquisition of the anatomical data.

The localizer experiment was composed of four consecutive functional runs, one for each object. A single run was experimentally designed as a series of ten tasks and eleven rest blocks of 20 seconds

in which the participant alternately imagined the object corresponding to the delivered auditory cue (e.g. "cat" or "chair"). At the end of each task-block, the word "stop" was delivered to the participant to announce the beginning of the resting period. The subject was requested to focus the gaze on a white fixation cross centered on a dark grey screen during the whole acquisition. The four images were selected from a database of naturalistic objects (Hebart et al., 2019) while the corresponding auditory cues were generated using https://soundoftext.com/ that creates audio files from text using the text to speech engine of Google Translate. Both the auditory cues and the visual feedback were delivered to the subject using a custom made Python script using PsychoPy3 module (Peirce et al., 2019).

The NF run was designed according to a similar experimental paradigm. Specifically, the subject was requested to imagine for a period of 20 seconds the object upon delivery of the corresponding auditory cue. At the end of the task-block, the subject heard the word "stop" that indicated the beginning of a 20 second period of rest. However, differently from the training and localizer runs, the rest-period was followed by an extra-period of 5 seconds during which the visual feedback was displayed. This consisted in the RS display where the positions of the base stimuli (i.e. the activity patterns acquired in the localizer session) were displayed as yellow points with a label tag indicating the object name. The position of the current activity pattern, as estimated from the last time window of measurement (40 s) ending right before the feedback-block, was displayed as a red star. In addition, the positions of the current pattern from all previous task-blocks were also displayed as light-grey stars and linked to one another with a green line. Thereby, the subject was provided, not only with the position in the RS of the current brain pattern, but also with the trajectory performed across the previous task-blocks. The experimental procedure is summarized in Figure 5.2.
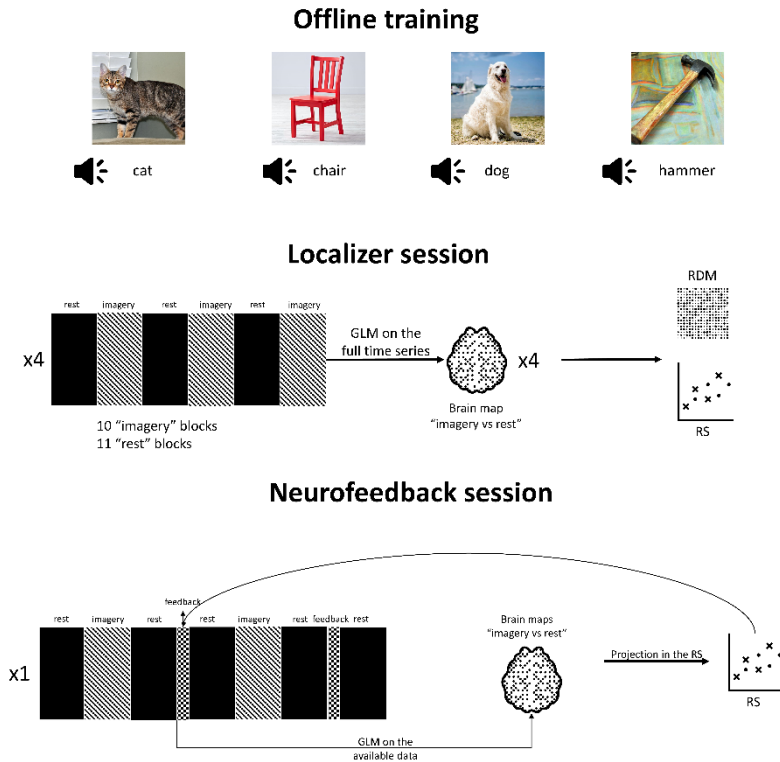
Figure 5.2 Workflow of the fMRI-NF experiment performed.

In the offline training session, the subject was requested to memorize four images and their corresponding auditory cues. In the localizer session, the subject underwent four block-design fMRI acquisitions, one for each stimulus, in which he was requested to imagine as clear as possible the image corresponding to the provided auditory cue. The imagery blocks were interleaved by periods of rest. At the end of each functional acquisition the statistical values relative to the contrast "imagery vs rest" was extracted for all the voxels of a defined ROI. The dissimilarities among these activation patterns were estimated and first, encoded in a RDM and projected onto a plane to create the RS of the ROI. In the neurofeedback session the subject was provided with only one auditory cue and requested to imagine the corresponding image. The imagery blocks were interleaved by periods of rest and small period of feedback. At the beginning of the feedback period the activity pattern of the defined

111

ROI relative to the contrast "imagery vs. rest" is extracted and projected in the RS space. The RS with the new point is given as feedback to the subject.

5.2.6.3 Data acquisition

FMRI data were obtained using a 7 Tesla whole-body scanner (Magnetom; Siemens AG, Erlangen, Germany). The participant was placed comfortably in the MRI scanner with foam padding next to the head to minimize spontaneous or task-related motion.

For the functional acquisition, multi-band (Feinberg et al., 2010; Moeller et al., 2010; Xu et al., 2013) repeated gradient-echo echo-planar imaging (EPI) sequences were used. Except for the number of time points (localizer session: 430 volumes; NF session: 680 volumes), identical scanning parameters were used for all functional measurements (repetition time (TR) = 1000 ms, echo time (TE) = 21 ms, number of axial slices = 60, matrix = 112 x 112, field of view (FOV) = 224 mm, thickness = 2 mm, interslice gap = 0 mm, multi-band factor = 4). In the NF session, functional images were reconstructed and exported using a direct TCP/IP connection from the image reconstruction computer to the real-time analysis computer and stored on the hard drive. The real-time data analysis software (TBV) running on the real-time analysis computer was able to read and process the exported images in real-time.

For the acquisition of the structural data, a high-resolution T1-weighted anatomical scan was acquired using a three-dimensional (3D) magnetization prepared rapid-acquisition gradient-echo (MP2RAGE) sequence (192 slices, 0.9 mm iso voxels, no gap, TR = 4500 ms, TE = 2.39 ms, TI1 = 900ms, TI2 = 2750, FA = 5, FOV = 230 x 230mm2, matrix size = 256 x 256, total scan time = 8 min and 34 s.

5.2.7 Online data processing

The functional runs of the localizer were processed with TBV to optimally define the target ROI to be used for the extraction of the brain activity patterns elicited by the base stimuli and for the subsequent NF runs. During the localizer session, the online data

processing allowed the real-time monitoring of the subject's head motion and the general quality of the data. More, specifically, motion correction with sinc interpolation, linear trend removal, temporal high-pass filtering and spatial smoothing with an isotropic 4-mm full-width at half-maximum (FWHM) Gaussian kernel were applied online to the functional time-series. The online voxel-wise GLM was computed incrementally with one predictor time-course for the imagery task (derived from an incremental boxcar function convolved with a standard hemodynamic response function (Boynton et al., 1996)) and six motion parameters (incrementally derived from the motion correction procedure) as confound predictors. The online (incremental) GLM was fitted at each voxel using a recursive least squares estimation of the regression coefficients. At the end of the session, pre-processed functional data were reloaded in TBV and the offline GLM analysis was also applied to the complete pre-processed fMRI time series. In both cases, the t-contrast "imagery vs. rest" was calculated.

Starting from offline GLM results, an ROI was defined in the inferior temporal cortex (ITC) as this region is well known to encode high-level (semantic) representations of natural objects at the interface between vision and semantics (Haxby et al., 2001; Kriegeskorte et al., 2008b; Mur et al., 2013). For this study, the ROI definition was performed using a combined anatomical and functional approach. Namely, a whole-brain probabilistic functional map in MNI space was generated from the Neurosynth database (Yarkoni et al., 2011) using the keyword "object". The statistical threshold was set to q=0.1 using the False Discovery Rate (Genovese et al., 2002). This map was imported in BrainVoyager 21.4 (Brain Innovation, Maastricht, The Netherlands) and an ROI was initially defined by selecting a cluster of activation encompassing the left ITC. Before the NF run, the ROI definition was further adapted to the estimated brain activity from the offline GLM of localizer runs. Namely, the extracted ROI was imported in TBV where it was transformed to the native space of EPI images and used as a guide to manually define a functional ROI in the ITC from the offline GLM contrast "imagery vs. rest" (main effects of all stimuli, p<0.001). At the end of each functional run, for each voxel

of the selected ROI, the t-values relative to the contrast "imagery vs. rest" were extracted and stored to local disk. Upon completion of all the functional runs of the localizer sessions, the stored data were assembled in a matrix, whose dimensions were the number of stimuli and the number of voxels of the ROI, and which was used to estimate the RDM and the corresponding RS (with its transformation) to be used in the NF session.

The data of the NF were processed in real-time with TBV with the same preprocessing steps used in the localizer session. The ROI activity pattern related to the contrast "imagery vs. rest" was extracted for the incremental time window ending one time point before the beginning of the feedback block, thereby encompassing the whole time series up to this point in time. The incremental GLM included the estimated motion predictors as confounds. To generate the feedback stimulus at the beginning of the feedback block, the extracted t-values from the ITC ROI were used to estimate the position of the current brain state in the RS estimated from the localizer data.

5.2.8 Offline data analysis

The full time-series data acquired during both the localizer and the NF session were reloaded in TBV and used for further analyses. Namely, the data from the localizer session were used to assess the computational and statistical performances of the rt-RSA approach on both real and artificial fMRI time-series, whereas the data from the NF session were used to analyze the experimental performance of the subject. For the validation, using the same GLM predictor as ideal response time-course, an artificial dataset with simulated brain activity was created to analyze the performances of the rt-RSA approach under different signal-to-noise conditions. In this way, the same incremental GLM analysis, as implemented in TBV, was used for the simulations.

5.2.8.1 Stability of the rt-RSA

The data of the functional runs of the localizer session were reloaded in TBV. Starting from this data set, four NF experiments were

simulated to investigate the stability of the rt-RSA approach. Namely, the t-maps from the incremental GLM were first extracted at the end of each task-block and used to simulate the dynamically updated brain pattern as the input of a virtual NF display. This pattern was projected as an additional point onto the RS estimated from the complete time-series of all functional runs. The distances, and the trajectory, from the input to the target are estimated at each block. As both online (incremental) and offline GLM analyses are conducted on the same localizer data, this scenario simulates the ideal successful NF outcome whereby the input NF pattern (from the online GLM) overlaps perfectly to the pattern associated with one of the base stimuli. Thereby, the stability of the rt-RSA can be evaluated.

5.2.8.2 Performance evaluation

To evaluate the performance of the subject, both the localizer and the NF data were analyzed.

Using the localizer data, an RDM was calculated at the end of each task-block via the incremental GLM. The last RDM (i.e. the one estimated from the full time-series) was chosen as reference RDM and the monotonic correlation between the vectorized upper triangular part of the RDM at each block and the vectorized upper triangular part of the reference RDM was statistically evaluated with a signed-rank test (Kriegeskorte et al., 2008a; Nili et al., 2014; Wilcoxon, 1945). The analysis of the RDM correlation series allowed us to have a possible estimate of the minimum number of blocks needed for the subject to incrementally generate an RDM which is not significantly different from the reference RDM. Moreover, it also allows evaluating the ability of the subject to maintain this similarity, consistently over time, across the subsequent blocks. The idea behind this analysis is that with an increase in the number of time points and, as a consequence, in the number of task blocks, it is possible to understand if the brain activity has become more stable and if the subject has been able to modulate consistently his/her brain activity. Therefore, this analysis could help us to evaluate prospectively, before the NF session, the quality of the localizer data as well as the stability of the base RS.

Using the data of the NF session, the performance of the subject was evaluated by measuring the distances in the RS of the projections of current brain patterns to the corresponding target. This analysis allowed us to have a measure of the subject's performance in modulating his brain state in relation, not only to the target stimulus but also the other base stimuli whose dissimilarities generate the RS. In practice, the fewer the steps (i.e. the number of task blocks) to reach the target, the faster the dissimilarity with the target brain state decreases and the dissimilarities with the other base stimuli replicate the original ones. Furthermore, estimating the distance between the projection of the last block and the target provides a measure of success for the NF outcome, as the lower this distance, the better the subject has likely fulfilled the task of engaging in that specific mental representation.

5.2.8.3 Simulations

An ensemble of artificial datasets was created to simulate the outcome of the rt-RSA analysis under different noise conditions.
For each simulation, five different artificial time-series of 430 images (the same number of time points of the real localizer data) were simulated in an ROI of 100 voxels. All voxel time-series of all data sets were initialized with random Gaussian noise with zero mean and different variances ($\sigma = 0.5, 1, 2, 3$). In each data set, an ideal activation time-course was injected into a variable percentage of voxels, whereas the rest of the voxels only contained noise. To simulate five different patterns (associated with five simulated stimuli), the subsets of active voxels were varied across the five artificial data sets. In particular, assigning a numerical index to the five patterns from 0 to 4, the percentage of active voxels for the even patterns was randomly extracted from a uniform distribution of integers ranging from 30 to 60, whereas the percentage of active voxels for the odd patterns was randomly extracted from a uniform distribution of integers ranging from 20 to 50. The activation time course was generated by convolving the canonical hemodynamic response function with a box-car function according to the same

paradigm of the real localizer experiment. However, the signal amplitude of each block was scaled with a random factor from a uniform distribution ranging from -1 to 3, to simulate variable modulation performances of the subject across task blocks.

The procedure described in the previous paragraph was repeated 1000 times, and for each simulation, the distances of the projected brain activity in the RS from the target, as well as the correlation of the RDM calculated at the end of each task block with the reference RDM, were calculated. These performances were thus reported on the average of 1000 simulations. The whole analysis was repeated four times after increasing level of noise variance ($\sigma = 0.5$, 1, 2, 3) to evaluate the performance of the rt-RSA under various noise conditions.

5.3 Results

The localizer data were preliminary used to simulate a different NF experiment for each stimulus, by extracting the brain pattern of the ROI at the end of each task block and calculating its position in the RS. The estimated distance between each projection (as obtained via incremental online GLM) and its corresponding target (as obtained via full offline GLM) showed, for all the stimuli, a decreasing trend over task blocks and a value of zero at the end of the session. In particular, the stimulus "cat" showed a maximum distance from the target of 0.6 at the first task block and a distance lower than 0.2 from the 6th task block on. The stimulus "dog" showed a maximum distance from the target of 0.5 at the first task block and a distance lower than 0.2 from the 8th task-block on. Similarly, the stimulus "hammer" showed a maximum distance from the target of 0.65 at the 2nd task block and a distance lower than 0.2 from the 8th task-block on. The stimulus "chair" showed more variability with a maximum distance of 1.2 at the 3rd task-block and a distance lower than 0.2 only at the 10th task block (Figure 5.3).
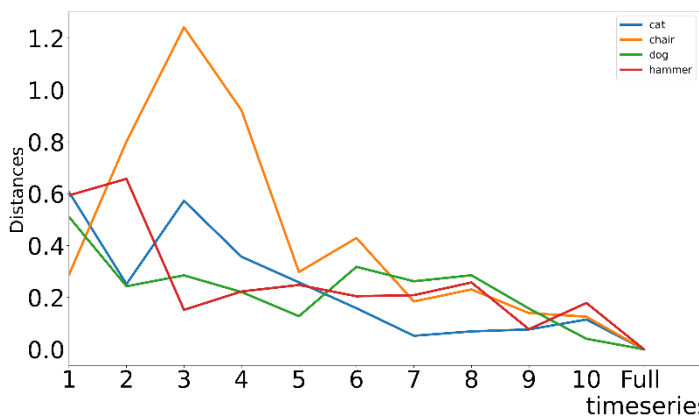
Figure 5.3 Distance of the projections from the target in the RS over the task-blocks, evaluated on the localizer data. Four neurofeedback experiments were simulated sampling incrementally the fMRI data of one of the stimuli and projecting these values in the defined RS. The estimated distances have a decreasing trend for all the stimuli and a zero value in the last task-block.

The trajectories of the projections in the RS showed, for all the stimuli, that, at the end of the time series, the positions of the estimated brain patterns coincide with the positions of the corresponding targets. Besides, the visual inspection of the trajectories, suggested how each stimulus follows a different path towards its target (Figure 5.4).
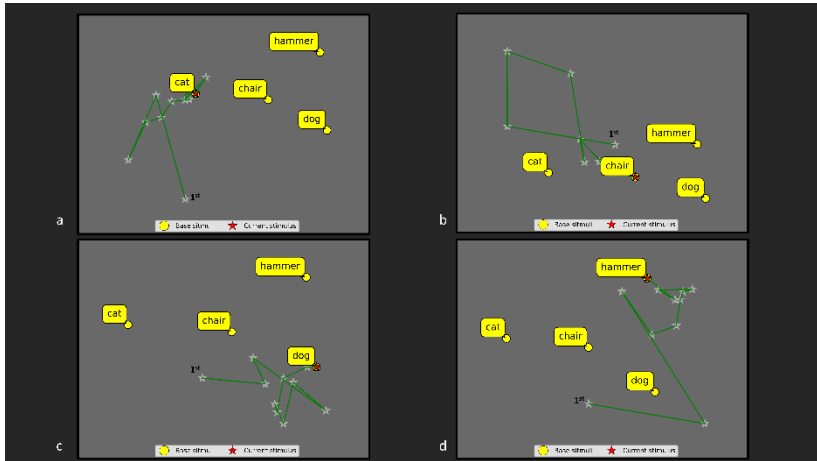
Figure 5.4 Trajectories of the projections on the RS over the task-blocks, evaluated on the localizer data. Four neurofeedback experiments were simulated by sampling incrementally the fMRI data of one of the stimuli and projecting these values in the defined RS. The estimated projections for all the stimuli move towards (and at the end perfectly overlap with) the corresponding targets.

The performances of the subject during the real NF experiment were evaluated using the data from both the localizer and the NF session.

The Spearman correlation between RDMs, estimated at each task block, and the reference RDM, estimated at the end of the localizer session with a full GLM, showed an overall increasing trend across the task blocks (Figure 5.5a). The correlation coefficient remained negative from the 1st to the 4th task block, before turning positive at the 5th task block, reaching a maximum of 0.8 at the 7th task block, and, finally, falling to lower values (~0.4) in the last two task blocks (Figure 5.5a). The estimated distances from the target in the RS showed an overall decreasing trend across the task-blocks. In the first four task blocks, the distances remained higher than 0.4 with a maximum of 0.82 at the 3rd task block, but then the distance from the target decreased towards a minimum of 0.12 at the last time-point (Figure 5.5b).
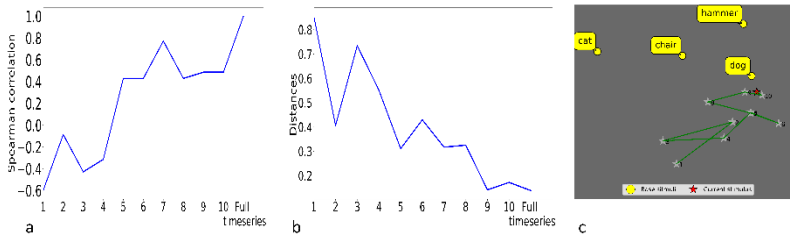
Figure 5.5 Evaluation of the subject's performance: correlation between RDMs in the localizer data (a), distances between the projection and the target in the RS (b) and trajectory of the projection in the RS (c). The Spearman's correlation between the RDMs estimated at the end of each task-block (i.e. a subset of the time series) and the reference RDM estimated at the end of the time series, show an increasing trend suggesting a stabilization of the brain patterns in the localizer data (a). The estimated distances between the projection in the RS and the target show a decreasing trend supporting the idea that, in the neurofeedback session, the subject learned to modulate his brain pattern to engage in the target mental state (b). The trajectories of the projections in the RS confirm that the subject managed to move the point corresponding to his brain activity towards the target (i.e. "dog") (c).

The analyses of the artificial datasets were useful to evaluate the stability and accuracy of the rt-RSA under different noise conditions by estimating the distances from the projection of the current brain activity (at each task block) to the target in the RS and the correlation between the RDMs at each task block and the reference RDM. After 1000 simulations, the average distance exhibited a monotonic decreasing trend over time for different noise conditions (Figure 5.6) with a minimum of ~0.05 for $\sigma = 0.5$ and a maximum of ~0.2 for $\sigma = 3$.
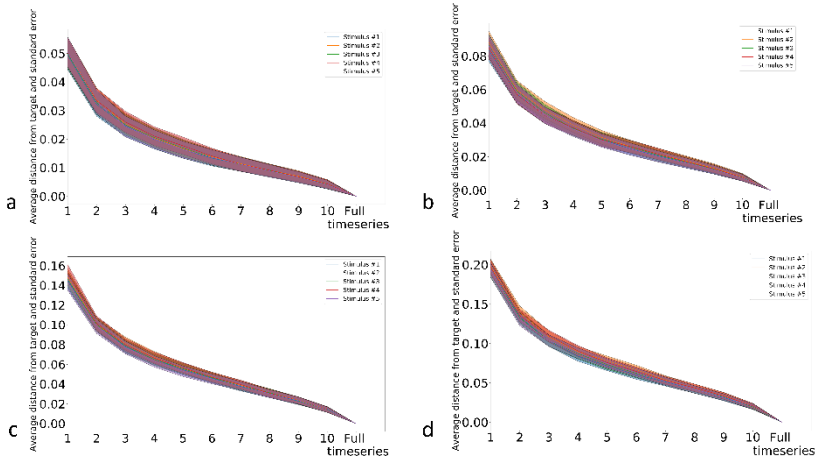
Figure 5.6 Distances from the target in the RS on the average on 1000 simulated datasets in various noise conditions: data with gaussian noise $\mu = 0$, $\sigma = 0.5$ (a), data with gaussian noise $\mu = 0$, $\sigma = 1$ (b),data with gaussian noise $\mu = 0$, $\sigma = 2$ (c) and data with gaussian noise $\mu = 0$, $\sigma = 3$ (d). In all the cases, the distances from the target have a decreasing trend towards zero.

An opposite trend was observed in the analysis of the RDM correlations. Namely, the average Spearman correlation coefficient exhibited a monotonic increasing trend over time for each noise level (Figure 5.7). The exponential nature of the curve was more evident at lower noise levels ($\sigma = 0.5$, 1). At the first block, the starting point of the curve showed lower values at higher noise levels, ranging from a maximum of 0.825 ($\sigma = 0.5$) to a minimum of about 0.2 ($\sigma = 3$).
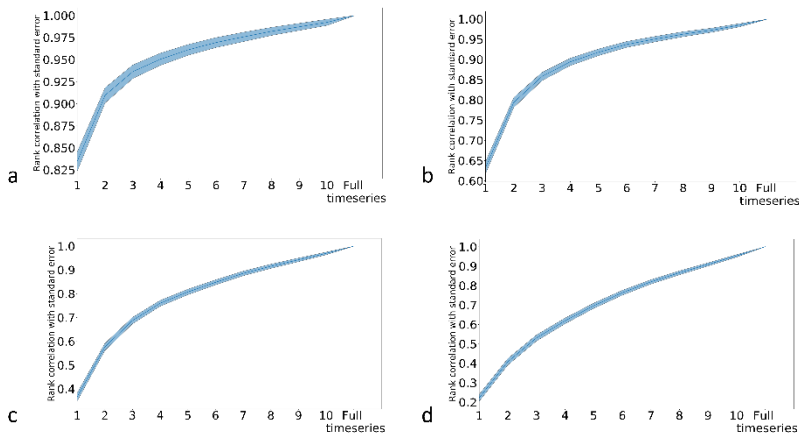
Figure 5.7 Spearman's correlation between RDMs estimated at the end of each task-block and the reference RDM on the average on 1000 simulated datasets in various noise conditions: data with Gaussian noise $\mu = 0$, $\sigma = 0.5$ (a), data with Gaussian noise $\mu = 0$, $\sigma = 1$ (b), data with Gaussian noise $\mu = 0$, $\sigma = 2$ (c) and data with Gaussian noise $\mu = 0$, $\sigma = 3$ (d). In all the cases, the correlations with the reference RDM have an increasing exponential trend towards one.

5.4 Discussions

In this work, rt-RSA, a method based on the integration of RSA within an rt-fMRI-NF paradigm, has been introduced. The online implementation of RSA (in a real-time incremental procedure) enabled a novel type of multi-dimensional feedback of the participant's brain activity which can be semantically related to an internal stimulus representation (thereby reflecting the actual mental state of the subject) via multi-voxel pattern analysis. Using rt-RSA, the online estimated neural pattern is displayed as a movable point on a plane where two or more target points are also displayed at fixed positions in such a way to approximate the mutual (dis)similarity between the current and some candidate target neural representations of (sets of) base stimuli which were previously presented to the subject during a localizer session. In this way, the subject is requested to change the position of the movable point towards one of the target

points in a fixed RS space by self-modulating his/her brain activity until reaching a desired mental state.

There are (at least) two potential advantages of this framework over classical NF paradigms: first, the NF display encodes (and therefore the self-modulation operates on) multi-voxel (i.e. spatially distributed) patterns of brain activity within one or more selected ROIs, which are (known or found to be) critically involved in the semantic processing of the stimulus, rather than one-dimensional (spatially averaged) ROI signals (deCharms et al., 2004; Weiskopf et al., 2003; Young et al., 2014) or pairwise connectivity estimates (Koush et al., 2013); second, this (locally) distributed pattern of brain activity is dynamically compared, not just to one, but many, base or reference patterns, thus providing additional degrees of freedom, both to the experimenter (in the preparation and set-up of the NF materials) and the subject (in the choice of the mental strategy), to more efficiently (self-) modulate a mental representation along multiple dimensions.

Rt-RSA was successfully tested in a real rt-fMRI-NF experiment at 7 Tesla where the subject performed a visual imagery task and the brain activity from an ROI in the left ITC was recorded. The imagery task was chosen for a proof-of-concept study, as it enabled the possibility to modulate the brain activity related to semantic features of the stimuli in a region known to encode the high-level representation of natural objects (Haxby et al., 2001; Kriegeskorte et al., 2008b; Mur et al., 2013). The functional data from a localizer session were used to both generate the RS for a set of four base stimuli (to be used in the NF session) and to simulate NF experiments. While the actual NF performances of the subject were evaluated on the real NF data from the NF session, the computational feasibility and the statistical accuracy of the rt-RSA approach under simulated ideal conditions of successful modulation were evaluated on the same localizer data. Besides, 1000 simulations of multi-voxel ROI patterns were performed to evaluate the rt-RSA approach under different signal-to-noise conditions.

The analysis of the simulated NF experiments with real fMRI data from the localizer session demonstrated that it is possible to implement a visual two-dimensional NF by estimating in real-time the current

brain pattern by dynamically updating (with negligible delays) the position of the associated current mental representation in a previously defined (and fixed) RS. Indeed, the resulting trajectory of a point on a two-dimensional plane (which is obtained by analyzing incrementally the fMRI time-series for just one target stimulus at the time) and the final collapsing of this point to the target position (when the full time-series is used in the GLM) demonstrated that the provided visual feedback could be in principle correct and stable. More specifically, the results of the simulated NF experiments clearly showed that if the participant is in principle able to become engaged in the same mental state which causes exactly the same distribution of brain activity elicited by the target stimulus in the target ROI, the resulting visual feedback could in principle guide this process towards the perfect match between the positions of the current and target brain patterns.

The results from the rt-fMRI-NF experiment demonstrate that the rt-RSA approach can be applied in a real-world scenario, with little additional computational steps and no additional hardware requirements. Thus, our proof-of-concept study suggests that it would be possible to integrate an rt-RSA based procedure within several experiments using existing NF paradigms.

The integration of a multi-dimensional NF display did not introduce additional difficulties for the subject to understand the task, at least according to the report from the single subject examined in this study. Indeed, while the task progress is perceived by the subject like "a journey in a geographical map", the investigator can still rely on the distance from the target as a simple (one-dimensional) index of success to evaluate the subject's performance (Paret et al., 2019). Moreover, the offline analyses performed on the localizer data showed how it is possible to obtain a preliminary assessment of the stability of the RS as initially created from the base stimuli, similarly as it happens in a one-dimensional NF where it is possible to evaluate the variability of the BOLD percent signal changes over the trials.

The analyses of the artificial datasets allowed to assess the impact of noise on the trajectories under ideal conditions of successful NF. These gathered two main observations: First, the projection of the current neural representation, and its final convergence to the target

points, remains stable also when the fMRI signal is affected by a relatively higher amount of noise. However, higher noise levels critically affect the position of the current point at the beginning of the experiment, thereby potentially increasing the length of the trajectory in the RS, and therefore the minimum number of steps required, to reach the target. Second, the amount of noise has an impact on the reference RDM itself and, therefore, on the RS, thereby the monotonic correlations between current and reference RDMs may be reduced.

Taken together, these results may suggest that, in an ideal scenario, the presence of higher noise levels simply causes longer paths to the target state, i.e. noise itself does not necessarily undermine or disrupt the stability of the rt-RSA under the assumption that the subject has successfully learnt to regulate his/her mental states with the decided strategy. However, it should be also pointed out that the stabilization of the statistical values does not represent per se an indication that the subject has successfully consolidated an optimal mental strategy. Actually, if the subject chooses a wrong strategy, there is no guarantee of successful convergence, even at very low noise levels. Nonetheless, as far as the noise is assumed constant across the blocks, the final distance of the current, from the target representation may still provide a useful indication of the overall performance in modulating the mental state according to that strategy.

The proposed method appears to be a promising tool for the self-regulation of the brain signals as it is flexible and versatile. The visual feedback modality provides a straightforward solution to implement, and easy to understand for the subject, and can therefore be considered the default solution. However, there are in principle no limitations about the feedback modality as different physical dimensions can be used as different feedback channels. The choice of the mental task, for which no detailed instructions are needed, is completely up to the subjects and their own representations, the only requirement being that a robust activation pattern is elicited (and verified) during the localizer session. As a consequence, besides integrating a visual feedback during object imagery, as shown in the present study, rt-RSA could be in principle used in more complex scenarios such as those employed in NF-based emotion regulation

(Linhartová et al., 2019). In fact, basic (positive or negative) emotions have been previously associated with specific neural signatures within different brain regions (Saarimäki et al., 2016), including, e.g., the amygdala (Sergerie et al., 2008). Therefore, it could be possible to project (and target) patterns with different emotional valence, on the same RS from one (or more) of these brain regions, after training the subject to engage in several different mental states that become associated with different local patterns (Linhartová et al., 2019) during the localizer session. Technically, it is only essential to keep the same number and order of voxels in the ROI(s) to correctly fit the original dimensions of the RS (Kriegeskorte and Kievit, 2013).

Finally, towards a proper generalization of this approach, and to counteract the inter-subject variability in local patterns, as resulting from the localizing/training phase with more complex stimuli, it could be possible to force higher levels of smoothing to the functional data and, in principle, derive one unique RDM from the group-level analysis of several subjects (e.g. from a healthy or control cohort). Along these lines, it would be possible to create one "external" RS to be used in the NF session of a single individual (e.g. a patient) by aligning the estimated neural pattern from the individual to a common space (Frost and Goebel, 2013; Haxby et al., 2011). As a consequence, the participants could be asked to self-modulate their own mental state by navigating through predefined mental states associated, not (only) with their neural representations, but rather with some "control" representations from different people. However, this approach will need proper design and careful testing to be validated before any clinical applications.

5.5 Conclusions

In conclusion, a new method for rt-fMRI-NF has been introduced which promises to go beyond the classical approach of fMRI signal self-modulation. The presented simulation and the preliminary results from a real rt-fMRI-NF demonstrate that rt-RSA gathers the possibility to use a fixed RS for a given ROI enabling a semantic feedback to the subject for the self-regulation of mental states in a

multi-dimensional space. This approach has been shown to be computationally and statistically feasible and its application to a real NF experiment based on a simple imagery task at 7 Tesla has yielded encouraging results. Nonetheless, future studies are warranted to increase the number of subjects and to assess performances in a multi-subject study, eventually at lower (i.e. clinical) magnetic fields (e.g. 1.5 or 3 Tesla).

Chapter 6:
General discussions and conclusions

The contributions of this thesis to the neuroscience community involved two main aspects: (i) the investigation of the influence that the linguistic statistical factors have in language production and comprehension and (ii) the proposal of a novel multi-dimensional experimental method for the rapidly evolving field of the rt-fMRI-NF.

The first aspect was addressed with two rapid event-related fMRI experiments on the inflection of the Italian verbs and nouns and with a naturalistic fMRI experiment dedicated to the neural mechanisms underlying the linguistic prediction in a more ecologically valid scenario, such as the listening to an audiobook.

All these experiments support the idea that our brain exploits the statistical features of a language both during production and comprehension and suggest the necessity of considering these aspects in both theoretical frameworks and practical applications.

The first two studies show that the inflection of both nouns and verbs are supported by a bilateral cortical network involving frontal and temporal areas. Their findings corroborate the hypothesis that the statistical features of a language, whether they are formal, grammatical or semantic, influence the functioning of the mental lexicon and its neural correlates for language production. Moreover, the presented results challenge the current theories on the lexical access and support the necessity of considering, at least for highly inflected languages, these statistical properties in both physiological and clinical studies. In particular, the findings of the study on the verb inflection demonstrates that in a highly inflected language, like Italian, this phenomenon reflects the distributional features of the three conjugations (i.e. inflectional class), such as the size, the productivity and the ortho-phonological consistency. Besides, the fMRI study on the nominal inflection shows that this process is modulated by the interplay between morphological (e.g. the transparency of gender suffixes) and distributional (i.e. the size and consistency of inflectional classes) factors in language processing, that have not been fully considered in the available cognitive and neuroanatomical models.

The hypothesis about the influence of the distributional factors on language processing has been strengthened by the findings of the third study, as they show that our brain exploits the recurrence of statistical patterns in the lexicon to infer the incoming word and maintain the language comprehension. The cortical responses observed are in line with previous similar studies showing that the surprisal measure is a good proxy of the linguistic prediction during language comprehension and that the linguistic prediction process is mainly hosted in the temporal poles. Furthermore, the results demonstrate that the integration of information from multiple linguistic levels increase the possibility to explain better the BOLD variance as the novel variant of the surprisal model, that integrates both lexical and semantic information, provides a better predictor of the fMRI activity in left STG and MTG and is critically more sensitive in detecting linguistic processes in the left IFG, compared to a pure lexical surprisal. Finally, although future investigations are definitely needed, the proposed experimental approach may find possible applications in studying language comprehension in poorly compliant subjects and exploring possible differences between, e.g., a healthy and a clinical cohort listening to the same story or different stories of varying complexity.

The second aspect entailed the development of a novel experimental procedure that integrates the ability of the RSA in summarizing the neural representations of a set of stimuli with the possibility of the rt-fMRI-NF of guiding the subject towards the self-modulation of the measured brain response. The novel method promises to go beyond the classical NF approach by enabling the possibility to provide a semantic feedback to the subject for the self-modulation of a mental state in a multi-dimensional space. The analyses of the pilot data set acquired at 7 Tesla show that the method is computationally and statistically feasible. Moreover, the preliminary encouraging results, coming from a simple imagery task, demonstrate that implemented procedure can be easily integrated into an rt-fMRI-NF as it does not require any special equipment or effort of the subjects and the researchers. Nonetheless, future investigations

are warranted to test the performances both in a multi-subjects study and in scanning session at lower magnetic fields.

References

Allen, M., Badecker, W., 2002. Inflectional Regularity: Probing the Nature of Lexical Representation in a Cross-Modal Priming Task. J. Mem. Lang. 46, 705–722. https://doi.org/10.1006/jmla.2001.2831

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., Watanabe, T., 2016. Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. Curr. Biol. CB 26, 1861–1866. https://doi.org/10.1016/j.cub.2016.05.014

Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. NeuroImage 20, 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Ardila, A., Bernal, B., 2016. Neuroimaging of Language: The Contribution of fMRI 10.

Ardila, A., Bernal, B., Rosselli, M., 2016. How Localized are Language Brain Areas? A Review of Brodmann Areas Involvement in Oral Language. Arch. Clin. Neuropsychol. 31, 112–122. https://doi.org/10.1093/arclin/acv081

Argyropoulos, G.P.D., 2016. The cerebellum, internal models and prediction in 'non-motor' aspects of language: A critical review. Brain Lang., Contributions of the Cerebellum to Language Functions 161, 4–17. https://doi.org/10.1016/j.bandl.2015.08.003

Armeni, K., Willems, R.M., Frank, S.L., 2017. Probabilistic language models in cognitive neuroscience: Promises and pitfalls. Neurosci. Biobehav. Rev. 83, 579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001

Aronoff, M., 1994. Morphology by Itself: Stems and Inflectional Classes. MIT Press.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang., Special Issue: Emerging Data Analysis 59, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R.H., Milin, P., Đurđević, D.F., Hendrix, P., Marelli, M., 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. Psychol. Rev. 118, 438.

Badre, D., Wagner, A.D., 2002. Semantic Retrieval, Mnemonic Control, and Prefrontal Cortex. Behav. Cogn. Neurosci. Rev. 1, 206–218. https://doi.org/10.1177/1534582302001003002

Baggio, G., Hagoort, P., 2011. The balance between memory and unification in semantics: A dynamic account of the N400. Lang. Cogn. Process. 26, 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Bates, D.M., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Linear Mixed-Effects Models using "Eigen" and S4.

Bates, E., Goodman, J.C., 1997. On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. Lang. Cogn. Process. 12, 507–584. https://doi.org/10.1080/016909697386628

Belin, P., Bestelmeyer, P.E.G., Latinus, M., Watson, R., 2011. Understanding Voice Perception. Br. J. Psychol. 102, 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Benetello, A., Finocchiaro, C., Capasso, R., Capitani, E., Laiacona, M., Magon, S., Miceli, G., 2016. The dissociability of lexical retrieval and morphosyntactic processes for nouns and verbs: A functional and anatomoclinical study. Brain Lang. 159, 11–22. https://doi.org/10.1016/j.bandl.2016.05.005

Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C., 2003. A Neural Probabilistic Language Model. J. Mach. Learn. Res. 3, 1137–1155.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thornton, A.M., 2005. CoLFIS (Corpus e Lessico di Frequenza dell'Italiano Scritto). Available Httpwww Istc Cnr Itmaterialdatabase.

Binder, J.R., 2015. The Wernicke area Modern evidence and a reinterpretation. Neurology 85, 2170–2175. https://doi.org/10.1212/WNL.0000000000002219

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19, 2767–2796. https://doi.org/10.1093/cercor/bhp055

Binder, J.R., Swanson, S.J., Hammeke, T.A., Sabsevitz, D.S., 2008. A comparison of five fMRI protocols for mapping speech comprehension systems. Epilepsia 49, 1980–1997. https://doi.org/10.1111/j.1528-1167.2008.01683.x

Bird, S., Lopez, E., Klein, E., 2009. Elaborazione del Linguaggio Naturale con Python.

Blumstein, S.E., Baker, E., Goodglass, H., 1977a. Phonological factors in auditory comprehension in aphasia. Neuropsychologia 15, 19–30. https://doi.org/10.1016/0028-3932(77)90111-7

Blumstein, S.E., Cooper, W.E., Zurif, E.B., Caramazza, A., 1977b. The perception and production of Voice-Onset Time in aphasia. Neuropsychologia 15, 371–383. https://doi.org/10.1016/0028-3932(77)90089-6

Bonhage, C.E., Mueller, J.L., Friederici, A.D., Fiebach, C.J., 2015. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. Cortex, Special issue: Prediction in speech and language processing 68, 33–47. https://doi.org/10.1016/j.cortex.2015.04.011

Bordag, D., Pechmann, T., 2009. Externality, internality, and (in)dispensability of grammatical features in speech production: evidence from Czech declension and conjugation. J. Exp. Psychol. Learn. Mem. Cogn. 35, 446–465. https://doi.org/10.1037/a0014874

Borg, I., Groenen, P.J.F., 2005. Modern Multidimensional Scaling: Theory and Applications, 2nd ed, Springer Series in Statistics. Springer-Verlag, New York. https://doi.org/10.1007/0-387-28981-X

Borleffs, E., Maassen, B.A.M., Lyytinen, H., Zwarts, F., 2017. Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative

review. Read. Writ. https://doi.org/10.1007/s11145-017-9741-5

Boroditsky, L., 2011. How Language Shapes Thought. Sci. Am. 304, 62–65. https://doi.org/10.1038/scientificamerican0211-62

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. J. Neurosci. 16, 4207–4221. https://doi.org/10.1523/JNEUROSCI.16-13-04207.1996

Bozic, M., Marslen-Wilson, W., 2010. Neurocognitive Contexts for Morphological Complexity: Dissociating Inflection and Derivation. Lang. Linguist. Compass 4, 1063–1073. https://doi.org/10.1111/j.1749-818X.2010.00254.x

Brennan, J.R., Stabler, E.P., Van Wagenen, S.E., Luh, W.M., Hale, J.T., 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. Brain Lang. 157–158, 81–94. https://doi.org/10.1016/j.bandl.2016.04.008

Broca, P., 1861. Sur la faculté du langage articulé.

Brouwer, G.J., Heeger, D.J., 2009. Decoding and Reconstructing Color from Responses in Human Visual Cortex. J. Neurosci. 29, 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Brown, E.C., Muzik, O., Rothermel, R., Matsuzaki, N., Juhász, C., Shah, A.K., Atkinson, M.D., Fuerst, D., Mittal, S., Sood, S., Diwadkar, V.A., Asano, E., 2012. Evaluating reverse speech as a control task with language-related gamma activity on electrocorticography. NeuroImage 60, 2335–2345. https://doi.org/10.1016/j.neuroimage.2012.02.040

Buch, E., Weber, C., Cohen, L.G., Braun, C., Dimyan, M.A., Ard, T., Mellinger, J., Caria, A., Soekadar, S., Fourkas, A., Birbaumer, N., 2008. Think to Move: a Neuromagnetic Brain-Computer Interface (BCI) System for Chronic Stroke. Stroke 39, 910–917. https://doi.org/10.1161/STROKEAHA.107.505313

Bunger, A., Skordos, D., Trueswell, J.C., Papafragou, A., 2016. How Children and Adults Encode Causative Events Cross-

Linguistically: Implications for Language Production and Attention. Lang. Cogn. Neurosci. 31, 1015–1037. https://doi.org/10.1080/23273798.2016.1175649

Burzio, L., 1998. Multiple correspondence. Lingua 104, 79–109. https://doi.org/10.1016/S0024-3841(97)00025-9

Bybee, J.L., 1985. Morphology: A Study of the Relation Between Meaning and Form. John Benjamins Publishing.

Caffarra, S., Barber, H.A., 2015. Does the ending matter? The role of gender-to-ending consistency in sentence reading. Brain Res. 1605, 83–92. https://doi.org/10.1016/j.brainres.2015.02.018

Caffarra, S., Janssen, N., Barber, H.A., 2014. Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. Neuropsychologia 63, 124–134. https://doi.org/10.1016/j.neuropsychologia.2014.08.016

Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., Song, X.-W., Xia, M.-R., Shu, N., Dong, Q., Milham, M.P., Castellanos, F.X., Zuo, X.-N., He, Y., 2014. Topological organization of the human brain functional connectome across the lifespan. Dev. Cogn. Neurosci. 7, 76–93. https://doi.org/10.1016/j.dcn.2013.11.004

Carota, F., Bozic, M., Marslen-Wilson, W., 2016. Decompositional Representation of Morphological Complexity: Multivariate fMRI Evidence from Italian. J. Cogn. Neurosci. 28, 1878–1896. https://doi.org/10.1162/jocn_a_01009

Carreiras, M., Carr, L., Barber, H.A., Hernandez, A., 2010. Where syntax meets math: Right Intraparietal Sulcus activation in response to grammatical number agreement violations. NeuroImage 49, 1741–1749. https://doi.org/10.1016/j.neuroimage.2009.09.058

Carreiras, M., Pattamadilok, C., Meseguer, E., Barber, H., Devlin, J.T., 2012. Broca's area plays a causal role in morphosyntactic processing. Neuropsychologia 50, 816–820. https://doi.org/10.1016/j.neuropsychologia.2012.01.016

Carstairs-McCarthy, A., 1994. Inflection Classes, Gender, and the Principle of Contrast. Language 70, 737–788. https://doi.org/10.2307/416326

Carter, B.T., Foster, B., Muncy, N., Luke, S.G., 2019. Linguistic networks associated with lexical, semantic and syntactic predictability in reading: A fixation-related fMRI study. NeuroImage 189, 224–240. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.018

Carter, C.S., van Veen, V., 2007. Anterior cingulate cortex and conflict detection: An update of theory and data. Cogn. Affect. Behav. Neurosci. 7, 367–379. https://doi.org/10.3758/CABN.7.4.367

Chao-Gan, Y., Yu-Feng, Z., 2010. DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front. Syst. Neurosci. 4. https://doi.org/10.3389/fnsys.2010.00013

Chen, G., Taylor, P.A., Qu, X., Molfese, P.J., Bandettini, P.A., Cox, R.W., Finn, E.S., 2019. Untangling the relatedness among correlations, part III: Inter-subject correlation analysis through Bayesian multilevel modeling for naturalistic scanning. NeuroImage 116474. https://doi.org/10.1016/j.neuroimage.2019.116474

Cibelli, E.S., Leonard, M.K., Johnson, K., Chang, E.F., 2015. The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. Brain Lang. 147, 66–75. https://doi.org/10.1016/j.bandl.2015.05.005

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. https://doi.org/10.1017/S0140525X12000477

Clos, M., Langner, R., Meyer, M., Oechslin, M.S., Zilles, K., Eickhoff, S.B., 2014. Effects of prior information on decoding degraded speech: An fMRI study. Hum. Brain Mapp. 35, 61–74. https://doi.org/10.1002/hbm.22151

Cohen Kadosh, K., Luo, Q., de Burca, C., Sokunbi, M.O., Feng, J., Linden, D.E.J., Lau, J.Y.F., 2016. Using real-time fMRI to influence effective connectivity in the developing emotion regulation network. Neuroimage 125, 616–626. https://doi.org/10.1016/j.neuroimage.2015.09.070

Colombo, L., Laudanna, A., De Martino, M., Brivio, C., 2004. Regularity and/or consistency in the production of the past

participle? Brain Lang., Third International Conference on the Mental Lexicon 90, 128–142. https://doi.org/10.1016/j.bandl.2003.12.009

Colombo, L., Stoianov, I., Pasini, M., Zorzi, M., 2006. The role of phonology in the inflection of Italian verbs: A connectionist investigation. Ment. Lex. 1, 147–181. https://doi.org/10.1075/ml.1.1.09col

Corbett, G.G., 1991. Gender. Cambridge University Press.

Crepaldi, D., Berlingeri, M., Paulesu, E., Luzzatti, C., 2011. A place for nouns and a place for verbs? A critical review of neurocognitive data on grammatical-class effects. Brain Lang. 116, 33–49. https://doi.org/10.1016/j.bandl.2010.09.005

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., Job, R., 2005. Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. J. Mem. Lang. 53, 42–59. https://doi.org/10.1016/j.jml.2005.02.007

Dahl, Ö., 2008. Tense and Aspect in the Languages of Europe, Eurotyp. De Gruyter Mouton.

Damasio, H., Damasio, A.R., 1980. THE ANATOMICAL BASIS OF CONDUCTION APHASIA. Brain 103, 337–350. https://doi.org/10.1093/brain/103.2.337

Danziger, S., Ward, R., 2010. Language Changes Implicit Associations Between Ethnic Groups and Evaluation in Bilinguals. Psychol. Sci. 21, 799–800. https://doi.org/10.1177/0956797610371344

Dapretto, M., Bookheimer, S.Y., 1999. Form and Content : Dissociating Syntax and Semantics in Sentence Comprehension 24, 427–432.

de Aguiar, V., Zhao, Y., Faria, A., Ficek, B., Webster, K.T., Wendt, H., Wang, Z., Hillis, A.E., Onyike, C.U., Frangakis, C., Caffo, B., Tsapkini, K., 2020. Brain volumes as predictors of tDCS effects in primary progressive aphasia. Brain Lang. 200, 104707. https://doi.org/10.1016/j.bandl.2019.104707

de Diego Balaguer, R., Rodríguez-Fornells, A., Rotte, M., Bahlmann, J., Heinze, H.-J., Münte, T.F., 2006. Neural circuits subserving the retrieval of stems and grammatical features in regular and irregular verbs. Hum. Brain Mapp. 27, 874–888. https://doi.org/10.1002/hbm.20228

De Martino, M., Bracco, G., Laudanna, A., 2011. The activation of grammatical gender information in processing Italian nouns. Lang. Cogn. Process. 26, 745–776. https://doi.org/10.1080/01690965.2010.491977

De Martino, M., Bracco, G., Postiglione, F., Laudanna, A., 2018. Declinazione e genere dei nomi italiani: DeGNI, una banca dati lessicale. [WWW Document]. URL http://www.lapsus.unisa.it/ricerca/index

De Martino, M., Bracco, G., Postiglione, F., Laudanna, A., 2017. The influence of grammatical gender and suffix transparency in processing Italian written nouns. Ment. Lex. 12, 107–128. https://doi.org/10.1075/ml.12.1.05dem

De Martino, M., Mancuso, A., Russo, A.G., Elia, A., Salle, F.D., Saponiero, R., Vietri, S., Esposito, F., Laudanna, A., 2020. Producing regularly and irregularly inflected verb forms: behavioural and neuroimaging data from the three Italian conjugations. Lang. Cogn. Neurosci. 35, 420–439. https://doi.org/10.1080/23273798.2019.1668953

De Martino, Postiglione, F., Bracco, G., Laudanna, A., 2019. Declinazione e genere dei nomi italiani: DeGNI, una banca dati lessicale. G. Ital. Psicol. https://doi.org/10.1421/93789

De Mauro, T., Mancini, F., Vedovelli, M., Voghera, M., 1993. Lessico di frequenza dell'italiano parlato. Etaslibri, Milano.

De Mulder, W., Bethard, S., Moens, M.-F., 2015. A survey on the application of recurrent neural networks to statistical language modeling. Comput. Speech Lang. 30, 61–98. https://doi.org/10.1016/j.csl.2014.09.005

de Silva, V., Tenenbaum, J.B., 2004. Sparse multidimensional scaling using landmark points 1–41.

De Smet, H.J., Baillieux, H., De Deyn, P.P., Mariën, P., Paquier, P., 2007. The cerebellum and language: The story so far, in:

Folia Phoniatrica et Logopaedica. pp. 165–170. https://doi.org/10.1159/000102927

deCharms, R.C., Christoff, K., Glover, G.H., Pauly, J.M., Whitfield, S., Gabrieli, J.D.E., 2004. Learned regulation of spatially localized brain activation using real-time fMRI. NeuroImage 21, 436–443. https://doi.org/10.1016/j.neuroimage.2003.08.041

Demberg, V., Keller, F., 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition 109, 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Dennett, D.C., 1987. The Intentional Stance. The MIT Press.

Di Tella, S., Baglio, F., Cabinio, M., Nemni, R., Traficante, D., Silveri, M.C., 2018. Selection processing in noun and verb production in left- and right-sided Parkinson's disease patients. Front. Psychol. 9, 1–13. https://doi.org/10.3389/fpsyg.2018.01241

Dikker, S., Pylkkänen, L., 2013. Predicting language: MEG evidence for lexical preactivation. Brain Lang. 127, 55–64. https://doi.org/10.1016/j.bandl.2012.08.004

Dikker, S., Silbert, L.J., Hasson, U., Zevin, J.D., 2014. On the Same Wavelength: Predictable Language Enhances Speaker–Listener Brain-to-Brain Synchrony in Posterior Superior Temporal Gyrus. J. Neurosci. 34, 6267–6272. https://doi.org/10.1523/JNEUROSCI.3796-13.2014

D'Mello, A.M., Turkeltaub, P.E., Stoodley, C.J., 2017. Cerebellar tDCS Modulates Neural Circuits during Semantic Prediction: A Combined tDCS-fMRI Study. J. Neurosci. 37, 1604–1613. https://doi.org/10.1523/JNEUROSCI.2818-16.2017

Dronkers, N.F., Wilkins, D.P., Van Valin, R.D., Redfern, B.B., Jaeger, J.J., 2004. Lesion analysis of the brain areas involved in language comprehension. Cognition, Towards a New Functional Anatomy of Language 92, 145–177. https://doi.org/10.1016/j.cognition.2003.11.002

Dumais, S.T., 2004. Latent semantic analysis. Annu. Rev. Inf. Sci. Technol. 38, 188–230. https://doi.org/10.1002/aris.1440380105

Eddington, D., 2002. Dissociation in Italian Conjugations: A Single-Route Account. Brain Lang. 81, 291–302. https://doi.org/10.1006/brln.2001.2525

Fair, D.A., Dosenbach, N.U.F., Church, J.A., Cohen, A.L., Brahmbhatt, S., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2007. Development of distinct control networks through segregation and integration. Proc. Natl. Acad. Sci. U. S. A. 104, 13507–13512. https://doi.org/10.1073/pnas.0705843104

Fausey, C.M., Boroditsky, L., 2011. Who dunnit? Cross-linguistic differences in eye-witness memory. Psychon. Bull. Rev. 18, 150–157. https://doi.org/10.3758/s13423-010-0021-5

Fedorenko, E., Nieto-Castañón, A., Kanwisher, N., 2012. Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. Neuropsychologia 50, 499–513. https://doi.org/10.1016/j.neuropsychologia.2011.09.014

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. PLoS ONE 5. https://doi.org/10.1371/journal.pone.0015710

Ferstl, E.C., Neumann, J., Bogler, C., von Cramon, D.Y., 2008. The Extended Language Network: A Meta-Analysis of Neuroimaging Studies on Text Comprehension. Hum. Brain Mapp. 29, 581–593. https://doi.org/10.1002/hbm.20422

Finocchiaro, C., Basso, G., Giovenzana, A., Caramazza, A., 2010. Morphological complexity reveals verb-specific prefrontal engagement. J. Neurolinguistics 23, 553–563. https://doi.org/10.1016/j.jneuroling.2010.04.004

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (f

MRI): Use of a Cluster-Size Threshold. Magn. Reson. Med. 33, 636–647. https://doi.org/doi:10.1002/mrm.1910330508

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Essen, D.C.V., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic , anticorrelated functional networks.

Frank, S.L., Otten, L.J., Galli, G., Vigliocco, G., 2015. The ERP response to the amount of information conveyed by words in sentences. Brain Lang. 140, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frank, S.L., Willems, R.M., 2017. Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. Lang. Cogn. Neurosci. 32, 1192–1203. https://doi.org/10.1080/23273798.2017.1323109

Franzon, F., Peressotti, F., Arcara, G., Semenza, C., 2014. Semantic interpretability speeds up the processing of morphological features. A psycholinguistic experiment on gender agreement. Stem- Spraak- En Taalpathologie 19, 189–191.

Friederici, A.D., Meyer, M., von Cramon, D.Y., 2000. Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. Brain Lang. 74, 289–300. https://doi.org/10.1006/brln.2000.2313

Friederici, A.D., Rueschemeyer, S.-A., Hahne, A., Fiebach, C., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension. Cereb. Cortex 13, 170–177.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. 360, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K.J., Daunizeau, J., Kilner, J., Kiebel, S.J., 2010. Action and behavior: a free-energy formulation. Biol. Cybern. 102, 227–260. https://doi.org/10.1007/s00422-010-0364-z

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Hum. Brain Mapp 2, 189–210. https://doi.org/10.1002/hbm.460020402

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S.J., Turner, R., 1996. Movement-Related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355. https://doi.org/10.1002/mrm.1910350312

Frost, M.A., Goebel, R., 2013. Functionally informed cortex based alignment: An integrated approach for whole-cortex macro-anatomical and ROI-based functional alignment. NeuroImage 83, 1002–1010. https://doi.org/10.1016/j.neuroimage.2013.07.056

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878. https://doi.org/10.1006/nimg.2001.1037

Geschwind, N., 1965. DISCONNEXION SYNDROMES IN ANIMALS AND MAN. Brain 88, 237–237. https://doi.org/10.1093/brain/88.2.237

Giordano, B.L., McAdams, S., Zatorre, R.J., Kriegeskorte, N., Belin, P., 2013. Abstract Encoding of Auditory Objects in Cortical Activity Patterns. Cereb. Cortex 23, 2025–2037. https://doi.org/10.1093/cercor/bhs162

Gleitman, L.R., January, D., Nappa, R., Trueswell, J.C., 2007. On the give and take between event apprehension and utterance formulation. J. Mem. Lang. 57, 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Goebel, R., 2012. BrainVoyager — Past, present, future. NeuroImage, 20 YEARS OF fMRI 62, 748–756. https://doi.org/10.1016/j.neuroimage.2012.01.083

Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of Functional Image Analysis Contest (FIAC) data with BrainVoyager QX: From single-subject to cortically aligned group General Linear Model analysis and self-organizing group Independent Component Analysis. Hum. Brain Mapp. 27, 392–401. https://doi.org/10.1002/hbm.20249

Goldstein, S., Naglieri, J.A. (Eds.), 2011. Encyclopedia of Child Behavior and Development. Springer US, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9

Guell, X., Gabrieli, J.D.E., Schmahmann, J.D., 2018. Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort. NeuroImage 172, 437–449. https://doi.org/10.1016/j.neuroimage.2018.01.082

Hagoort, P., 2019. The neurobiology of language beyond single-word processing. Science 366, 55–58. https://doi.org/10.1126/science.aax0289

Hagoort, P., 2013. MUC (Memory, Unification, Control) and beyond. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00416

Hagoort, P., 2005. On Broca, brain, and binding: a new framework. Trends Cogn. Sci. 9, 416–423. https://doi.org/10.1016/j.tics.2005.07.004

Hagoort, P., Brown, C.M., 1999. The neurocgnition of syntactic processing, in: Neurocognition of Language. https://doi.org/10.1038/nrn3566

Hale, J., 2016. Information-theoretical Complexity Metrics. Linguist. Lang. Compass 10, 397–412. https://doi.org/10.1111/lnc3.12196

Hale, J., 2001. A probabilistic earley parser as a psycholinguistic model. pp. 1–8. https://doi.org/10.3115/1073336.1073357

Harris, Z.S., 1954. Distributional Structure. WORD 10, 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14, 40. https://doi.org/10.1016/j.tics.2009.10.011

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293, 2425–2430. https://doi.org/10.1126/science.1063736

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the

representational space in human ventral temporal cortex. Neuron 72, 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Hebart, M.N., Dickter, A.H., Kidder, A., Kwok, W.Y., Corriveau, A., Van Wicklin, C., Baker, C.I., 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLOS ONE 14, e0223792. https://doi.org/10.1371/journal.pone.0223792

Heim, S., 2008. Syntactic gender processing in the human brain: A review and a model. Brain Lang. 106, 55–64. https://doi.org/10.1016/j.bandl.2007.12.006

Heim, S., Alter, K., Friederici, A.D., 2005. A dual-route account for access to grammatical gender: Evidence from functional MRI. Anat. Embryol. (Berl.) 210, 473–483. https://doi.org/10.1007/s00429-005-0032-6

Henderson, J.M., Choi, W., Lowder, M.W., Ferreira, F., 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. NeuroImage 132, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050

Hendrix, P., 2016. Experimental explorations of a discrimination learning approach to language processing. https://doi.org/10.15496/publikation-9333

Hernandez, A.E., Kotz, S.A., Hofmann, J., Valentin, V.V., Dapretto, M., Bookheimer, S.Y., 2004. The neural correlates of grammatical gender decisions in Spanish. NeuroReport 15, 863–866. https://doi.org/10.1097/00001756-200404090-00026

Herwig, U., Lutz, J., Scherpiet, S., Scheerer, H., Kohlberg, J., Opialla, S., Preuss, A., Steiger, V.R., Sulzer, J., Weidt, S., Stämpfli, P., Rufer, M., Seifritz, E., Jäncke, L., Brühl, A.B., 2019. Training emotion regulation through real-time fMRI neurofeedback of amygdala activity. NeuroImage 184, 687–696. https://doi.org/10.1016/j.neuroimage.2018.09.068

Hickok, G., Poeppel, D., 2007. The Cortical Organisation fo Speech Processing. Nature 8, 393–402. https://doi.org/10.7554/eLife.14521

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. https://doi.org/10.1038/nature17637

Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jaeger, T.F., 2008. Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. J. Mem. Lang. 59, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Joanisse, M.F., Seidenberg, M.S., 2005. Imaging the past: Neural activation in frontal and temporal regions during regular and irregular past-tense processing. Cogn. Affect. Behav. Neurosci. 5, 282–296. https://doi.org/10.3758/CABN.5.3.282

Justus, T., Larsen, J., Yang, J., Davies, P. de M., Dronkers, N., Swick, D., 2011. The role of Broca's area in regular past-tense morphology: An event-related potential study. Neuropsychologia 49, 1–18. https://doi.org/10.1016/j.neuropsychologia.2010.10.027

Kandylaki, K.D., Bornkessel-Schlesewsky, I., 2019. From story comprehension to the neurobiology of language. Lang. Cogn. Neurosci. 34, 405–410. https://doi.org/10.1080/23273798.2019.1584679

Kellett, K.A., Stevenson, J.L., Gernsbacher, M.A., 2012. What Role does the Cerebellum Play in Language Processing? Handb. Neuropsychol. Lang. 1, 294–316. https://doi.org/10.1002/9781118432501.ch15

Kielar, A., Joanisse, M.F., 2009. Graded Effects of Regularity in Language Revealed by N400 Indices of Morphological Priming. J. Cogn. Neurosci. 22, 1373–1398. https://doi.org/10.1162/jocn.2009.21353

Kielar, A., Milman, L., Bonakdarpour, B., Thompson, C.K., 2011. Neural correlates of covert and overt production of tense and agreement morphology: Evidence from fMRI. J. Neurolinguistics, Clinical Speech and Language Studies in Honour of Susan Edwards 24, 183–201. https://doi.org/10.1016/j.jneuroling.2010.02.008

King, M., Hernandez-Castillo, C.R., Poldrack, R.A., Ivry, R.B., Diedrichsen, J., 2019. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. Nat. Neurosci. 22, 1371–1378. https://doi.org/10.1038/s41593-019-0436-x

Kober, S.E., Witte, M., Ninaus, M., Neuper, C., Wood, G., 2013. Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00695

Koush, Y., Rosa, M.J., Robineau, F., Heinen, K., W. Rieger, S., Weiskopf, N., Vuilleumier, P., Van De Ville, D., Scharnowski, F., 2013. Connectivity-based neurofeedback: Dynamic causal modeling for real-time fMRI. Neuroimage 81, 422–430. https://doi.org/10.1016/j.neuroimage.2013.05.010

Krause, F., Benjamins, C., Lührs, M., Eck, J., Noirhomme, Q., Rosenke, M., Brunheim, S., Sorger, B., Goebel, R., 2017. Real-time fMRI-based self-regulation of brain activation across different visual feedback presentations. Brain-Comput. Interfaces 4, 87–101. https://doi.org/10.1080/2326263X.2017.1307096

Krause, F., Lindemann, O., 2014. Expyriment: A Python library for cognitive and neuroscientific experiments. Behav. Res. Methods 46, 416–428. https://doi.org/10.3758/s13428-013-0390-6

Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: Integrating cognition, computation, and the brain. Trends Cogn. Sci. 17, 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational
similarity analysis - connecting the branches of systems
neuroscience. Front. Syst. Neurosci. 2.
https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J.,
Esteky, H., Tanaka, K., Bandettini, P.A., 2008b. Matching
Categorical Object Representations in Inferior Temporal
Cortex of Man and Monkey. Neuron 60, 1126–1141.
https://doi.org/10.1016/j.neuron.2008.10.043

Kruskal, J.B., Wish, M., 1978. Multidimensional Scaling. SAGE.

Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction
in language comprehension? Lang. Cogn. Neurosci. 31, 32–
59. https://doi.org/10.1080/23273798.2015.1102299

Lau, E.F., Weber, K., Gramfort, A., Hämäläinen, M.S., Kuperberg,
G.R., 2016. Spatiotemporal Signatures of Lexical–Semantic
Prediction. Cereb. Cortex N. Y. NY 26, 1377–1387.
https://doi.org/10.1093/cercor/bhu219

Laudanna, A., 2007. Representation and Processing of Regular and
Sub-Regular Verbal Forms in Italian. Lingue E Linguaggio
227–246. https://doi.org/10.1418/25652

Laudanna, A., 1999. Regular versus irregular inflection: A question
of levels. Behav. Brain Sci. 22, 1029–1030.
https://doi.org/10.1017/S0140525X99382224

Laudanna, A., Gazzellini, S., Martino, M.D., 2004. Representation of
grammatical properties of Italian verbs in the mental lexicon.
Brain Lang., Third International Conference on the Mental
Lexicon 90, 95–105. https://doi.org/10.1016/S0093-
934X(03)00423-1

Leminen, A., Smolka, E., Duñabeitia, J.A., Pliatsikas, C., 2019.
Morphological processing in the brain: The good
(inflection), the bad (derivation) and the ugly
(compounding). Cortex, Structure in words: the present and
future of morphological processing in a multidisciplinary
perspective 116, 4–44.
https://doi.org/10.1016/j.cortex.2018.08.016

Levy, R., 2008. Expectation-based syntactic comprehension. Cognition 106, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, F., Butler, A., Gilbert, L., 2011. A unified approach to model selection using the likelihood ratio test. Methods Ecol. Evol. 2, 155–162. https://doi.org/10.1111/j.2041-210X.2010.00063.x

Linden, D.E.J., Habes, I., Johnston, S.J., Linden, S., Tatineni, R., Subramanian, L., Sorger, B., Healy, D., Goebel, R., 2012. Real-Time Self-Regulation of Emotion Networks in Patients with Depression. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0038115

Linhartová, P., Látalová, A., Kóša, B., Kašpárek, T., Schmahl, C., Paret, C., 2019. fMRI neurofeedback in emotion regulation: A literature review. NeuroImage 193, 75–92. https://doi.org/10.1016/j.neuroimage.2019.03.011

Lopopolo, A., Frank, S.L., Van Den Bosch, A., Willems, R.M., 2017. Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain. PLoS ONE 12, 1–18. https://doi.org/10.1371/journal.pone.0177794

Luzzatti, C., De Bleser, R., 1996. Morphological Processing in Italian Agrammatic Speakers: Eight Experiments in Lexical Morphology. Brain Lang. 54, 26–74. https://doi.org/10.1006/brln.1996.0060

Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Orletta, F.D., Dittmann, H., Lenci, A., Pirrelli, V., 2014. The PAISÀ Corpus of Italian Web Texts. Proc. 9th Web Corpus Workshop WaC-9 36–43. https://doi.org/10.1109/ICIEV.2016.7760164

Manly, B.F.J., 2007. Randomization Tests, 4th Edition by Eugene S. Edgington, Patrick Onghena. Int. Stat. Rev.

Marangolo, P., Incoccia, C., Pizzamiglio, L., Sabatini, U., Castriota-Scanderbeg, A., Burani, C., 2003. The right hemisphere involvement in the processing of morphologically derived

words. J. Cogn. Neurosci. 15, 364–371.
https://doi.org/10.1162/089892903321593090

Marangolo, P., Piras, F., Galati, G., Burani, C., 2006. Functional
anatomy of derivational morphology. Cortex 42, 1093–1106.
https://doi.org/10.1016/S0010-9452(08)70221-1

Marek, S., Siegel, J.S., Gordon, E.M., Raut, R.V., Gratton, C.,
Newbold, D.J., Ortega, M., Laumann, T.O., Miller, D.B.,
Zheng, A., Lopez, K.C., Berg, J.J., Coalson, R.S., Nguyen,
A.L., Dierker, D., Van, A.N., Hoyt, C.R., McDermott, K.B.,
Norris, S.A., Shimony, J.S., Snyder, A.Z., Nelson, S.M.,
Barch, D.M., Schlaggar, B.L., Raichle, M.E., Petersen, S.E.,
Greene, D.J., Dosenbach, N.U.F., 2018. Spatial and
Temporal Organization of the Individual Human
Cerebellum. SSRN Electron. J. 1–17.
https://doi.org/10.2139/ssrn.3188429

Mariën, P., Ackermann, H., Adamaszek, M., Barwood, C.H.S.,
Beaton, A., Desmond, J., De Witte, E., Fawcett, A.J.,
Hertrich, I., Küper, M., Leggio, M., Marvel, C., Molinari,
M., Murdoch, B.E., Nicolson, R.I., Schmahmann, J.D.,
Stoodley, C.J., Thürling, M., Timmann, D., Wouters, E.,
Ziegler, W., 2014. Consensus paper: Language and the
cerebellum: An ongoing enigma. Cerebellum 13, 386–410.
https://doi.org/10.1007/s12311-013-0540-5

Marslen-Wilson, W.D., Hare, M., Older, L., 1995. Priming and
blocking in the mental lexicon: The English past tense
[WWW Document]. URL
https://www.semanticscholar.org/paper/Priming-and-
blocking-in-the-mental-lexicon%3A-The-Marslen-Wilson-
Hare/89e5e4a4205c9dee2daecaa4368d5c80b2816df8
(accessed 5.5.20).

Marslen-Wilson, W.D., Tyler, L.K., 2007. Morphology, language
and the brain: the decompositional substrate for language
comprehension. Philos. Trans. R. Soc. B Biol. Sci. 362, 823–
836. https://doi.org/10.1098/rstb.2007.2091

McClelland, J.L., Patterson, K., 2002. Rules or connections in past-
tense inflections: what does the evidence rule out? Trends

Cogn. Sci. 6, 465–472. https://doi.org/10.1016/S1364-6613(02)01993-9

Mehler, D.M.A., Sokunbi, M.O., Habes, I., Barawi, K., Subramanian, L., Range, M., Evans, J., Hood, K., Lührs, M., Keedwell, P., Goebel, R., Linden, D.E.J., 2018. Targeting the affective brain—a randomized controlled trial of real-time fMRI neurofeedback in patients with depression. Neuropsychopharmacology 43, 2578–2585. https://doi.org/10.1038/s41386-018-0126-5

Meunier, F., Marslen-Wilson, W., 2004. Regularity and irregularity in Frenchverbal inflection. Lang. Cogn. Process. 19, 561–580. https://doi.org/10.1080/01690960344000279

Miceli, G., Mazzucchi, A., Menn, L., Goodglass, H., 1983. Contrasting cases of Italian agrammatic aphasia without comprehension disorder. Brain Lang. 19, 65–97. https://doi.org/10.1016/0093-934X(83)90056-1

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., Caramazza, A., 2002. The neural correlates of grammatical gender: an fMRI investigation. J. Cogn. Neurosci. 14, 618–628. https://doi.org/10.1162/08989290260045855

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. ArXiv13013781 Cs.

Milin, P., Kuperman, V., Kostić, A., Baayen, R.H., 2009. Paradigms bit by bit: an information-theoretic approachto the processing of paradigmatic structure in inflectionand derivation. Analogy Gramm. Form Acquis. 381, 214–252. https://doi.org/9780199547548.003.0010

Miozzo, M., Fischer-Baum, S., Postman, J., 2010. A selective deficit for inflection production. Neuropsychologia 48, 2427–2436. https://doi.org/10.1016/j.neuropsychologia.2010.04.001

Mirković, J., Seidenberg, M.S., Joanisse, M.F., 2011. Rules vs. Statistics: Insights from a Highly Inflected Language. Cogn. Sci. 35, 638–681. https://doi.org/10.1111/j.1551-6709.2011.01174.x

Mitchell, J., 2010. Composition in Distributional Models of Semantics. University of Edinburgh.

Mitchell, J., Lapata, M., 2009. Language Models Based on Semantic Composition, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Presented at the EMNLP 2009, Association for Computational Linguistics, Singapore, pp. 430–439.

Moberget, T., Ivry, R.B., 2016. Cerebellar contributions to motor control and language comprehension: searching for common computational principles. Ann. N. Y. Acad. Sci. 1369, 154–171. https://doi.org/10.1111/nyas.13094

Moeller, S., Yacoub, E., Olman, C. a, Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband Multislice GE-EPI at 7 Tesla, With 16-Fold Acceleration Using Partial Parallel Imaging With Application to High Spatial and Temporal Whole-Brain FMRI. Magn. Reson. Med. 63, 1144–1153. https://doi.org/10.1002/mrm.22361.Multiband

Molinaro, N., Barber, H.A., Caffarra, S., Carreiras, M., 2015. On the left anterior negativity (LAN): The case of morphosyntactic agreement: A Reply to Tanner et al. Cortex 66, 156–159. https://doi.org/10.1016/j.cortex.2014.06.009

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P.A., Kriegeskorte, N., 2013. Human Object-Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00128

Naselaris, T., Stansbury, D.E., Gallant, J.L., 2012. Cortical representation of animate and inanimate objects in complex natural scenes. J. Physiol.-Paris, New trends in neurogeometrical approaches to the brain and mind problem 106, 239–249. https://doi.org/10.1016/j.jphysparis.2012.02.001

Nevat, M., Ullman, M.T., Eviatar, Z., Bitan, T., 2017. The neural bases of the learning and generalization of morphological inflection. Neuropsychologia 98, 139–155. https://doi.org/10.1016/j.neuropsychologia.2016.08.026

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A Toolbox for Representational Similarity Analysis. PLoS Comput. Biol. 10. https://doi.org/10.1371/journal.pcbi.1003553

Nishida, S., Nishimoto, S., 2017. Decoding naturalistic experiences from human brain activity via distributed representations of words. NeuroImage 1–11. https://doi.org/10.1016/j.neuroimage.2017.08.017

Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. 87, 9868–9872.

Op de Beeck, H.P., Torfs, K., Wagemans, J., 2008. Perceived Shape Similarity among Unfamiliar Objects and the Organization of the Human Object Vision Pathway. J. Neurosci. 28, 10111–10123. https://doi.org/10.1523/JNEUROSCI.2511-08.2008

Orlov, N.D., Giampietro, V., O'Daly, O., Lam, S.-L., Barker, G.J., Rubia, K., McGuire, P., Shergill, S.S., Allen, P., 2018. Real-time fMRI neurofeedback to down-regulate superior temporal gyrus activity in patients with schizophrenia and auditory hallucinations: a proof-of-concept study. Transl. Psychiatry 8, 1–10. https://doi.org/10.1038/s41398-017-0067-5

Padovani, R., Calandra-Buonaura, G., Cacciari, C., Benuzzi, F., Nichelli, P., 2005. Grammatical gender in the brain: Evidence from an fMRI study on Italian. Brain Res. Bull. 65, 301–308. https://doi.org/10.1016/j.brainresbull.2004.11.025

Papafragou, A., Hulbert, J., Trueswell, J., 2008. Does language guide event perception? Evidence from eye movements. Cognition 108, 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Paret, C., Goldway, N., Zich, C., Keynan, J.N., Hendler, T., Linden, D., Cohen Kadosh, K., 2019. Current progress in real-time functional magnetic resonance-based neurofeedback: Methodological challenges and achievements. NeuroImage

202, 116107.
https://doi.org/10.1016/j.neuroimage.2019.116107

Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vértes, P.E., Ersche, K.D., Suckling, J., Bullmore, E.T., 2014. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. Neuroimage 95, 287–304. https://doi.org/10.1016/j.neuroimage.2014.03.012

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., 2019. PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods 51, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peirce, J.W., 2008. Generating stimuli for neuroscience using PsychoPy. Front. Neuroinformatics 2, 1–8. https://doi.org/10.3389/neuro.11.010.2008

Peirce, J.W., 2007. PsychoPy-Psychophysics software in Python. J. Neurosci. Methods 162, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Penke, M., Krause, M., 2002. German Noun Plurals: A Challenge to the Dual-Mechanism Model. Brain Lang. 81, 303–311. https://doi.org/10.1006/brln.2001.2526

Pennington, J., Socher, R., Manning, C.D., 2014. GloVe : Global Vectors for Word Representation. https://doi.org/10.3115/v1/D14-1162

Perani, D., Schnur, T., Tettamanti, M., Italy, Cappa, S.F., Fazio, F., 1999. Word and picture matching: a PET study of semantic category effects. Neuropsychologia 37, 293–306. https://doi.org/10.1016/S0028-3932(98)00073-6

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S.J., Kanwisher, N., Botvinick, M., Fedorenko, E., 2018. Toward a universal decoder of linguistic meaning from brain activation. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-03068-4

Perlovsky, L., Sakai, K.L., 2014. Language and Cognition. Front. Behav. Neurosci. 8. https://doi.org/10.3389/fnbeh.2014.00436

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Pinker, S., 1991. Rules of language. Science 253, 530–535. https://doi.org/10.1126/science.1857983

Pinker, S., Ullman, M.T., 2002. The past and future of the past tense. Trends Cogn. Sci. 6, 456–463. https://doi.org/10.1016/S1364-6613(02)01990-3

Piras, F., Marangolo, P., 2007. Noun–verb naming in aphasia: a voxel-based lesion-symptom mapping study. NeuroReport 18, 1455–1458. https://doi.org/10.1097/WNR.0b013e3282ef6fc9

Pleger, B., Timmann, D., 2018. The role of the human cerebellum in linguistic prediction, word generation and verbal working memory: evidence from brain imaging, non-invasive cerebellar stimulation and lesion studies. Neuropsychologia 115, 204–210. https://doi.org/10.1016/j.neuropsychologia.2018.03.012

Pliatsikas, C., Johnstone, T., Marinis, T., 2014. Grey Matter Volume in the Cerebellum is Related to the Processing of Grammatical Rules in a Second Language: A Structural Voxel-based Morphometry Study. Cerebellum Lond. Engl. 13, 55–63. https://doi.org/10.1007/s12311-013-0515-6

Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-Specific Cortical Activity Precedes Retrieval During Memory Search. Science 310, 1963–1966. https://doi.org/10.1126/science.1117645

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Pylkkänen, L., Feintuch, S., Hopkins, E., Marantz, A., 2004. Neural correlates of the effects of morphological family frequency and family size: an MEG study. Cognition 91, B35–B45. https://doi.org/10.1016/j.cognition.2003.09.008

Quené, H., van den Bergh, H., 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. J. Mem. Lang., Special Issue: Emerging Data Analysis 59, 413–425. https://doi.org/10.1016/j.jml.2008.02.002

Quiñones, I., Molinaro, N., Mancini, S., Hernández-Cabrera, J.A., Barber, H., Carreiras, M., 2018. Tracing the interplay between syntactic and lexical features: fMRI evidence from agreement comprehension. NeuroImage 175, 259–271. https://doi.org/10.1016/j.neuroimage.2018.03.069

Ramot, M., Grossman, S., Friedman, D., Malach, R., 2016. Covert neurofeedback without awareness shapes cortical network spontaneous connectivity. Proc. Natl. Acad. Sci. 113, E2413–E2420. https://doi.org/10.1073/pnas.1516857113

Romberg, A.R., Saffran, J.R., 2010. Statistical learning and language acquisition. Wiley Interdiscip. Rev. Cogn. Sci. 1, 906–914. https://doi.org/10.1002/wcs.78

Rota, G., Sitaram, R., Veit, R., Erb, M., Weiskopf, N., Dogil, G., Birbaumer, N., 2009. Self-regulation of regional cortical activity using real-time fmri: the right inferior frontal gyrus and linguistic processing. Hum. Brain Mapp. 30, 1605–1614. https://doi.org/10.1002/hbm.20621

Royston, P., Thompson, S.G., 1995. Comparing Non-Nested Regression Models. Biometrics 51, 114–127. https://doi.org/10.2307/2533319

Rumelhart, D.E., McClelland, J.L., 1986. Parallel Distributed Processing. MIT Press.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete Neural Signatures of Basic Emotions. Cereb. Cortex 26, 2563–2573. https://doi.org/10.1093/cercor/bhv086

Sach, M., Seitz, R.J., Indefrey, P., 2004. Unified inflectional processing of regular and irregular verbs: a PET study. NeuroReport 15, 533–537.

Sahin, N.T., Pinker, S., Halgren, E., 2006. Abstract grammatical processing of nouns and verbs in Broca's area: Evidence from fMRI. Cortex 42, 540–562. https://doi.org/10.1016/S0010-9452(08)70394-0

Salvi, G., Vanelli, L., 2004. Nuova grammatica italiana. Il Mulino.

Say, T., Clahsen, H., 2002. Words, Rules and Stems in the Italian Mental Lexicon, in: Nooteboom, S., Weerman, F., Wijnen, F. (Eds.), Storage and Computation in the Language Faculty, Studies in Theoretical Psycholinguistics. Springer Netherlands, Dordrecht, pp. 93–129. https://doi.org/10.1007/978-94-010-0355-1_4

Sayeed, A., Fischer, S., Demberg, V., 2015. Vector-space calculation of semantic surprisal for predicting word pronunciation duration. Proc. 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. Vol. 1 Long Pap. 763–773.

Scharnowski, F., Weiskopf, N., 2015. Cognitive enhancement through real-time fMRI neurofeedback. Curr. Opin. Behav. Sci. 4, 122–127. https://doi.org/10.1016/j.cobeha.2015.05.001

Schmid, H., 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees, in: Proceedings of the International Conference on New Methods in Language Processing.

Schreuder, R., Jong, N. de, Krott, A., Baayen, H., 1999. Rules and rote: Beyond the linguistic either-or fallacy. Behav. Brain Sci. 22, 1038–1039. https://doi.org/10.1017/S0140525X9947222X

Sergerie, K., Chochol, C., Armony, J.L., 2008. The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. 32, 811–830. https://doi.org/10.1016/j.neubiorev.2007.12.002

Serianni, L., 1988. Grammatica italiana. Suoni, forme, costrutti - Google Scholar. Utet.

Seymour, P.H.K., Aro, M., Erskine, J.M., 2003. Foundation literacy acquisition in European orthographies. Br. J. Psychol. 94, 143–174. https://doi.org/10.1348/000712603321661859

Shain, C., Blank, I.A., van Schijndel, M., Schuler, W., Fedorenko, E., 2020. fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. Neuropsychologia 138, 107307. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/doi: 10.1002/j.1538-7305.1948.tb01338.x

Shapiro, K.A., Moo, L.R., Caramazza, A., 2006. Cortical signatures of noun and verb production. Proc. Natl. Acad. Sci. 103, 1644–1649. https://doi.org/10.1073/pnas.0504142103

Shapiro, K.A., Mottaghy, F.M., Schiller, N.O., Poeppel, T.D., Flüß, M.O., Müller, H.-W., Caramazza, A., Krause, B.J., 2005. Dissociating neural correlates for nouns and verbs. NeuroImage 24, 1058–1067. https://doi.org/10.1016/j.neuroimage.2004.10.015

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., Kawato, M., 2019. Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. NeuroImage 188, 539–556. https://doi.org/10.1016/j.neuroimage.2018.12.022

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M., 2011. Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. Science 334, 1413–1415. https://doi.org/10.1126/science.1212003

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M.L., Rana, M., Oblak, E., Birbaumer, N., Sulzer, J., 2017. Closed-loop brain training: the science of neurofeedback. Nat. Rev. Neurosci. 18, 86–100. https://doi.org/10.1038/nrn.2016.164

Sitaram, R., Veit, R., Stevens, B., Caria, A., Gerloff, C., Birbaumer, N., Hummel, F., 2011. Acquired Control of Ventral Premotor Cortex Activity by Feedback Training: An Exploratory Real-Time fMRI and TMS Study. Neurorehabil. Neural Repair. https://doi.org/10.1177/1545968311418345

Slioussar, N., Kireev, M.V., Chernigovskaya, T.V., Kataeva, G.V., Korotkov, A.D., Medvedev, S.V., 2014. An ER-fMRI study of Russian inflectional morphology. Brain Lang. 130, 33–41. https://doi.org/10.1016/j.bandl.2014.01.006

Smith, N.J., Levy, R., 2013. The effect of word predictability on reading time is logarithmic. Cognition 128, 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Smolka, E., Khader, P.H., Wiese, R., Zwitserlood, P., Rösler, F., 2013. Electrophysiological Evidence for the Continuous Processing of Linguistic Categories of Regular and Irregular Verb Inflection in German. J. Cogn. Neurosci. 25, 1284–1304. https://doi.org/10.1162/jocn_a_00384

Sokolov, A.A., Miall, R.C., Ivry, R.B., 2017. The Cerebellum: Adaptive Prediction for Movement and Cognition. Trends Cogn. Sci. 21, 313–332. https://doi.org/10.1016/j.tics.2017.02.005

Staub, A., 2015. The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. Lang. Linguist. Compass 9, 311–327. https://doi.org/10.1111/lnc3.12151

Stolcke, A., Zheng, J., Wang, W., Abrash, V., 2011. SRILM at Sixteen : Update and Outlook. Proc. IEEE Autom. Speech Recognit. Underst. Workshop ASRU 5–9.

Strickland, B., 2017. Language Reflects "Core" Cognition: A New Theory About the Origin of Cross-Linguistic Regularities. Cogn. Sci. 41, 70–101. https://doi.org/10.1111/cogs.12332

Thompson-schill, S.L., Ramscar, M., Chrysikou, E.G., 2009. Current Directions in Psychological Science When a Little Frontal Lobe Goes a Long Way 18, 259–263. https://doi.org/10.1111/j.1467-8721.2009.01648.x

Thornton, A.M., Iacobini, C., Burani, C., 1997. VDVDB-Una base di dati sul Vocabolario di base della lingua italiana. Bulzoni Editore., Roma.

Tierney, L., 2012. The R Statistical Computing Environment, in: Feigelson, E.D., Babu, G.J. (Eds.), Statistical Challenges in Modern Astronomy V, Lecture Notes in Statistics. Springer, New York, NY, pp. 435–447. https://doi.org/10.1007/978-1-4614-3520-4_41

Torgerson, W.S., 1958. Theory and methods of scaling. Wiley.

Tripodi, R., Pira, S.L., 2017. Analysis of Italian Word Embeddings.

Tsigka, S., Papadelis, C., Braun, C., Miceli, G., 2014. Distinguishable neural correlates of verbs and nouns: A MEG study on homonyms. Neuropsychologia 54, 87–97. https://doi.org/10.1016/j.neuropsychologia.2013.12.018

Tuennerhoff, J., Noppeney, U., 2016. When sentences live up to your expectations. NeuroImage 124, 641–653. https://doi.org/10.1016/j.neuroimage.2015.09.004

Turney, P.D., Pantel, P., 2010. From Frequency to Meaning: Vector Space Models of Semantics Peter. J. Artif. Intell. Res. 37, 141–188. https://doi.org/10.1016/j.jaci.2011.12.1000

Tyler, L.K., Bright, P., Fletcher, P., Stamatakis, E.A., 2004. Neural processing of nouns and verbs: The role of inflectional morphology. Neuropsychologia 42, 512–523. https://doi.org/10.1016/j.neuropsychologia.2003.10.001

Ullman, M.T., 2004. Contributions of memory circuits to language: the declarative/procedural model. Cognition, Towards a New Functional Anatomy of Language 92, 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M.T., 2001. A neurocognitive perspective on language: The declarative/procedural model. Nat. Rev. Neurosci. 2, 717–726. https://doi.org/10.1038/35094573

Ullman, M.T., 1993. The computation of inflectional morphology (Thesis). Massachusetts Institute of Technology.

Ünal, E., Papafragou, A., 2018. The relation between language and mental state reasoning. Oxford University Press. https://doi.org/10.1093/oso/9780198789710.003.0008

Ünal, E., Papafragou, A., 2016. Interactions Between Language and Mental Representations. Lang. Learn. 66, 554–580. https://doi.org/10.1111/lang.12188

van der Lely, H.K.J., Ullman, M.T., 2001. Past tense morphology in specifically language impaired and normally developing children. Lang. Cogn. Process. 16, 177–217. https://doi.org/10.1080/01690960042000076

van Rossum, G., 1995. Python tutorial (No. Report CS-R9526).

Veríssimo, J., Clahsen, H., 2009. Morphological priming by itself: A study of Portuguese conjugations. Cognition 112, 187–194. https://doi.org/10.1016/j.cognition.2009.04.003

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., Cappa, S.F., 2011. Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci. Biobehav. Rev. 35, 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Vigliocco, G., Vinson, D.P., Paganelli, F., Dworzynski, K., 2005. Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. J. Exp. Psychol. Gen. 134, 501–520. https://doi.org/10.1037/0096-3445.134.4.501

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. NeuroImage 137, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wang, J., 2012. Classical Multidimensional Scaling, in: Wang, J. (Ed.), Geometric Structure of High-Dimensional Data and

Dimensionality Reduction. Springer, Berlin, Heidelberg, pp. 115–129. https://doi.org/10.1007/978-3-642-27497-8_6

Wang, M., Chen, Y., Schiller, N.O., 2018. Lexico-syntactic features are activated but not selected in bare noun production: Electrophysiological evidence from overt picture naming. Cortex. https://doi.org/10.1016/j.cortex.2018.05.014

Wang, M., Schiller, N.O., 2019. A Review on Grammatical Gender Agreement in Speech Production. Front. Psychol. 9, 1–7. https://doi.org/10.3389/fpsyg.2018.02754

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M., 2017. Advances in fMRI Real-Time Neurofeedback. https://doi.org/10.1016/j.tics.2017.09.010

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., Belin, P., 2014. Crossmodal Adaptation in Right Posterior Superior Temporal Sulcus during Face-Voice Emotional Integration. J. Neurosci. 34, 6813–6821. https://doi.org/10.1523/JNEUROSCI.4478-13.2014

Weber, K., Lau, E.F., Stillerman, B., Kuperberg, G.R., 2016. The Yin and the Yang of prediction: An fMRI study of semantic predictive processing. PLoS ONE 11, 1–25. https://doi.org/10.1371/journal.pone.0148637

Weiskopf, N., Veit, R., Erb, M., Mathiak, K., Grodd, W., Goebel, R., Birbaumer, N., 2003. Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): Methodology and exemplary data. NeuroImage 19, 577–586. https://doi.org/10.1016/S1053-8119(03)00145-9

Wernicke, C., 1874. Der aphasische Symptomenkomplex, in: Wernicke, C. (Ed.), Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis. Springer, Berlin, Heidelberg, pp. 1–70. https://doi.org/10.1007/978-3-642-65950-8_1

Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender

Agreement in Spanish Sentence Reading. J. Cogn. Neurosci. 16, 1272–1288. https://doi.org/10.1162/0898929041920487

Wiestler, T., McGonigle, D.J., Diedrichsen, J., 2011. Integration of sensory and motor representations of single fingers in the human cerebellum. J. Neurophysiol. 105, 3042–3053. https://doi.org/10.1152/jn.00106.2011

Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. Biom. Bull. 1, 80–83. https://doi.org/10.2307/3001968

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., Van Den Bosch, A., 2016. Prediction during Natural Language Comprehension. Cereb. Cortex 26, 2506–2516. https://doi.org/10.1093/cercor/bhv075

Willems, R.M., van Gerven, M.A.J., 2018. New FMRI Methods for the Study of Language, in: Rueschemeyer, S.-A., Gaskell, M.G. (Eds.), The Oxford Handbook of Psycholinguistics. Oxford University Press, pp. 974–991. https://doi.org/10.1093/oxfordhb/9780198786825.013.42

Williams, J.R., 1998. Guidelines for the use of multimedia in instruction. Proc. Hum. Factors Ergon. Soc. 2, 1447–1451. https://doi.org/10.1177/154193129804202019

Willms, J.L., Shapiro, K.A., Peelen, M.V., Pajtas, P.E., Costa, A., Moo, L.R., Caramazza, A., 2011. Language-invariant verb processing regions in Spanish–English bilinguals. NeuroImage 57, 251–261. https://doi.org/10.1016/j.neuroimage.2011.04.021

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73. https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 Tesla. NeuroImage 83, 790–795. https://doi.org/doi:10.1016/j.neuroimage.2013.07.055.

Yamin, H.G., Gazit, T., Tchemodanov, N., Raz, G., Jackont, G., Charles, F., Fried, I., Hendler, T., Cavazza, M., 2017. Depth electrode neurofeedback with a virtual reality interface. Brain-Comput. Interfaces 4, 201–213. https://doi.org/10.1080/2326263X.2017.1338008

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670. https://doi.org/10.1038/nmeth.1635

Young, B.M., Nigogosyan, Z., Walton, L.M., Remsik, A., Song, J., Nair, V.A., Tyler, M.E., Edwards, D.F., Caldera, K., Sattin, J.A., Williams, J.C., Prabhakaran, V., 2015. Dose-response relationships using brain–computer interface technology impact stroke rehabilitation. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00361

Young, K.D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2014. Real-Time fMRI Neurofeedback Training of Amygdala Activity in Patients with Major Depressive Disorder. PLoS ONE 9. https://doi.org/10.1371/journal.pone.0088785

Zhu, Z., Hagoort, P., Zhang, J.X., Feng, G., Chen, H.-C., Bastiaansen, M., Wang, S., 2012. The anterior left inferior frontal gyrus contributes to semantic unification. NeuroImage 60, 2230–2237. https://doi.org/10.1016/j.neuroimage.2012.02.036

Zilverstand, A., Sorger, B., Sarkheil, P., Goebel, R., 2015. fMRI neurofeedback facilitates anxiety regulation in females with spider phobia. Front. Behav. Neurosci. 9, 148. https://doi.org/10.3389/fnbeh.2015.00148

Zweerings, J., Hummel, B., Keller, M., Zvyagintsev, M., Schneider, F., Klasen, M., Mathiak, K., 2019. Neurofeedback of core language network nodes modulates connectivity with the default-mode network: A double-blind fMRI neurofeedback study on auditory verbal hallucinations. NeuroImage 189, 533–542. https://doi.org/10.1016/j.neuroimage.2019.01.058

Zwitserlood, P., Bölte, J., Dohmes, P., 2000. Morphological effects on speech production: Evidence from picture naming. Lang. Cogn. Process. 15, 563–591. https://doi.org/10.1080/01690960050119706

Publications

- **Russo** et al. (2020) "Semantics-weighted Lexical Surprisal Modeling of Naturalistic Functional MRI Time-Series during Spoken Narrative Listening", NeuroImage, Volume 222, 2020, 117281, ISSN 1053-8119, DOI: 10.1016/j.neuroimage.2020.117281
- **Russo** et al. (2020) "The neural substrate of noun morphological inflection: A Rapid Event-related fMRI Study in Italian", Neuropsychologia, Volume 151, 2021, 107699, ISSN 0028-3932, DOI: 10.1016/j.neuropsychologia.2020.107699.
- **Russo** et al. (under review) "Towards Semantic fMRI Neurofeedback: Navigating among Mental States using Real-time Representational Similarity Analysis" (preprint: doi: https://doi.org/10.1101/2020.11.09.374397)
- **Russo** et al., (2018) "Computer-Assisted Review of Auto-Segmented MRI Images in Surface-based Analyses", in Proceedings of the VI Conference of Gruppo Nazionale di Bioingegneria (GNB), 25th-27th June 2018, Milan, Italy
- **Russo** et al., (2020) "A Stochastic Language Model of Italian Applied to Functional MRI during Narrative Listening", in Proceedings of the VII Conference of Gruppo Nazionale di Bioingegneria (GNB) 2020, Trieste, 9th-11th June 2021, Trieste, Italy
- **Russo** & Ponticorvo et al., 2019 "No increased cerebrovascular involvement in adult beta-thalassemia by advanced MRI analyses" May 2019, Blood Cells Molecules and Diseases 78, DOI: 10.1016/j.bcmd.2019.05.001
- Canna & **Russo** et al. 2018 "Automated search of control points in surface-based morphometry", April 2018, NeuroImage, Volume 176, 2018, Pages 56-70, ISSN 1053-8119, DOI: 10.1016/j.neuroimage.2018.04.035

- De Martino et al. 2020 "Producing regularly and irregularly inflected verb forms: behavioral and neuroimaging data from the three Italian conjugations", Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2019.1668953
- Canna & Ponticorvo et al. 2018 "A group-level comparison of volumetric and combined volumetric-surface normalization for whole brain analyses of myelin and iron maps", December 2018, Magnetic Resonance Imaging, 54, 225-240
- Tartaglione et al. 2019 "No evidence of increased cerebrovascular involvement in adult neurologically-asymptomatic β-Thalassaemia. A multicenter multimodal magnetic resonance study" March 2019, British Journal of Haematology 185(4) DOI: 10.1111/bjh.15834
- Canna et al. 2019 "Intensity-related distribution of sweet and bitter taste fMRI responses in the insular cortex", May 2019, Human Brain Mapping 40(28), DOI: 10.1002/hbm.24621
- Tartaglione et al. 2019 "Brain functional impairment in beta-thalassaemia: the cognitive profile in Italian neurologically asymptomatic adult patients in comparison to the reported literature", May 2019, British Journal of Haematology, DOI: 10.1111/bjh.15959
- Cascino et al. 2020 "Cortical Thickness, Gyrification Index and Fractal Dimensionality in people with acute and recovered Anorexia Nervosa and in people with Bulimia Nervosa", March 2020, Psychiatry Research: Neuroimaging, DOI: 10.1016/j.pscychresns.2020.111069
- Manara et al. 2020 "Asymptomatic intracranial aneurysms in beta-thalassemia: a three-year follow-up report", January 2020, Orphanet Journal of Rare Diseases, Volume 15, Numero 1, Pagine 1-4
- Tartaglione et al. 2020 "Headache in beta-thalassemia: An Italian multicenter clinical, conventional MRI and MR-

angiography case-control study, Blood Cells, Molecules, and Diseases", Volume 81, 2020, 102403, ISSN 1079-9796, DOI: 10.1016/j.bcmd.2019.102403.

- Manara et al. (2020) "Visual cortex changes in children with sickle cell disease and normal visual acuity: a multimodal magnetic resonance imaging study". Br J Haematol. DOI:10.1111/bjh.17042

- Bardozzo et al., (2020) "Motor strength classification with machine learning approaches applied to anatomical neuroimages" per IEEE World Congress on Computational Intelligence (WCCI) 2020, 19 - 24th July 2020, Glasgow (UK)

Appendix
1. Lexical surprisal and frequency predictors of BOLD signals

In the main article, the effects of the lexical surprisal (LS) and semantics-weighted surprisal (SwS) on the BOLD signal were analyzed separately in two whole-brain voxel-based two-level (mixed-effects) general linear model (GLM) analyses using BrainVoyager. In the first-level GLM, the correlation between surprisal measures and fMRI time-courses was estimated as a fixed effect in every single subject in both narrative conditions (original and reversed speech). In addition to the predictor of interest (either LS or SwS), three predictors of no interest (confounds), accounting for word duration (WD), lexical frequency (LF) and root mean squared (RMS) amplitude of the word sound (WS), were added to the design matrix. To decorrelate the confound predictors from the predictor of interest, the Gram-Schmidt orthonormalization procedure was implemented according to its hierarchical formulation. In the second-level GLM, the inter-subject variability of these effects was assessed by treating subjects as random observations.

In this section, the multicollinearity of the two surprisal predictors, together with the LF predictor, and the correlation of the LF predictor with BOLD signals across the whole brain, are analyzed. To this purpose, the correlation matrix and the Variance Inflation Factor (VIF) of each predictor are computed for a design matrix composed by all three predictors (LS, SwS, LF), with no additional confounds and no orthonormalization. Also, a whole-brain analysis of the fMRI data using LF as the main predictor of interest (and WD and WS as confound predictors), is performed. In this case, LF values were represented on the surprisal scale (negative log probability), such that larger values correspond to less frequent words in the corpus, and the design matrix underwent the same (hierarchical) Gram-Schmidt orthonormalization procedure used for LS and SwS models.

The contrast between real and reversed speech conditions was considered in the group-level analysis.

The mutual correlations between SwS, LS and LF predictors revealed that the two surprisal predictors were highly (positively) correlated (r

= +0.77), whereas the correlation with the word frequency was lower (in absolute values) and negative (r = -0.26 for the SwS and r = -0.38 for the LS) (Table S1). The VIF values calculated on the matrix containing all three predictors (SwS, LS and LF) were: 4.20 for the SwS, 4.81 for the LS and 1.43 for the LF.

The whole-brain analysis with the LF as the predictor of interest revealed several clusters with significantly positive LF effects. These were localized in the superior temporal gyrus (bilaterally), in the right anterior temporal gyrus, in the left inferior frontal gyrus and the right cerebellum (p<0.05, cluster level corrected, cluster-forming threshold p=0.001). Additional clusters were also detected in the premotor cortex (Figure S1).

	SwS	LS	LF
SwS	1	0.77	-0.26
LS	0.77	1	-0.38
LF	-0.26	-0.38	1

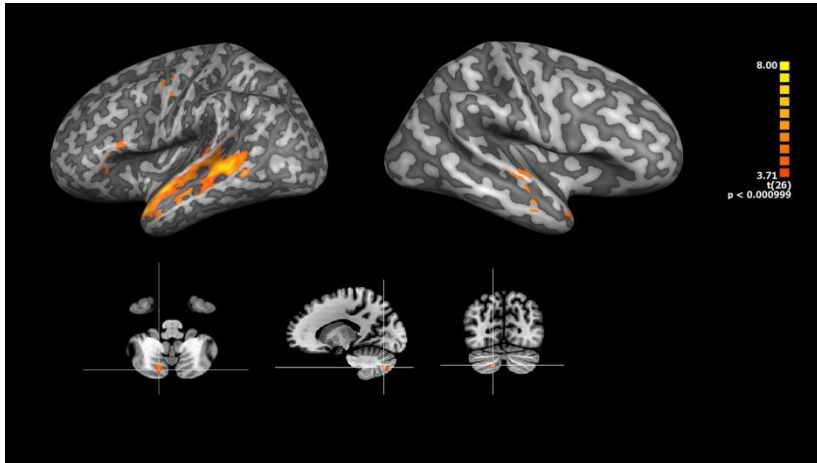**Table S1.** Correlation matrix of the SwS, LS and LF predictors.

Figure S1. *Lexical frequency brain response.* Activation pattern elicited by the (full variance) LF predictor in the real speech, compared to the reversed speech, condition. Clusters with statistically significant effects (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) are overlaid in pseudo-color on an inflated cortical mesh obtained from a Talairach-normalized anatomical scan.

2. Semantic scaling factor

In the main article, the SwS is formulated as a combined measure of the trigram probability and the semantic similarity. According to the provided formulation, the SwS modulates the co-occurrences of the current lexical form of each content word with its left-side local context via a multiplicative factor expressing its semantic similarity with a preceding broader context of words (i.e. the semantic scaling factor). For content words, the theoretical range of the semantic scaling factor is between zero and infinite and therefore it can either result in a downscaling (if it is lower than 1) or an upscaling (if it is higher than 1) of the trigram probability. The semantic scaling factor is identically set to 1 when the considered token is a function word.

To empirically explore the relation between the two surprisal measures (i.e. LS and SwS), and to statistically analyze how the weighting by semantic information specifically affects the LS

measure, a scatter plot of all LS and SwS raw values assigned by the models to the content words in the narrative text is presented, together with a box plot of LS and SwS raw values across all four LS quartiles (figure S2). There were some visible extrema in the SwS values although these were only 47 out of a total of 893 content words (i.e. ~6% of the content words and ~3% of all the words in the narrative). However, these only occurred for 47 (out of a total of 893) content words (i.e. ~6% of the content words and ~3% of all the words in the narrative text), mostly in the lower quartiles of LS values, and no particular patterns of heteroscedasticity between the two measures were noted across LS quartiles (see Figure S2). Furthermore, when performing the same analysis on the two resulting fMRI predictors (i.e. the series of values across time points of the fMRI series, not across words of the narrative, after preprocessing and hemodynamic convolution) the scatter plot showed no extrema (and no heteroscedasticity patterns) for the SwS predictor (see Figure S3).

The linear/non-linear nature of the semantic weighting was empirically addressed by fitting both a linear and an exponential line to, and by drawing these lines on a scatter plot of, all scaling factor values vs. corresponding semantic similarity values, for all content words in the corpus.

For instance, given the transformation/normalization in the SwS formula, it is possible to say that, the semantics weighting produces a linear ($R^2=0.78$), rather than exponential ($R^2=0.08$), transformation of the semantic similarity values (Figure S4).

To possibly isolate the neural effects of the semantic scaling factor, i.e. the semantic component of the SwS model, an additional whole-brain analysis of the fMRI data was performed where this factor was isolated from the formula and considered as a predictor of interest in the first-level GLM, exactly as previously done for the LS, SwS and LF predictors. In this case, however, a special regressor was created for the same set of words used in the other whole-brain analyses, where each content word was assigned with the corresponding value of the semantic scaling factor and each function word was assigned with the value of one (implying no semantic modulation for these words).

The same contrast between real and reversed speech has been evaluated in the second-level GLM analysis and the whole-brain statistical map exhibited a significantly positive activation in the superior temporal gyrus (bilaterally) and in the right cerebellum (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) (Figure S5).
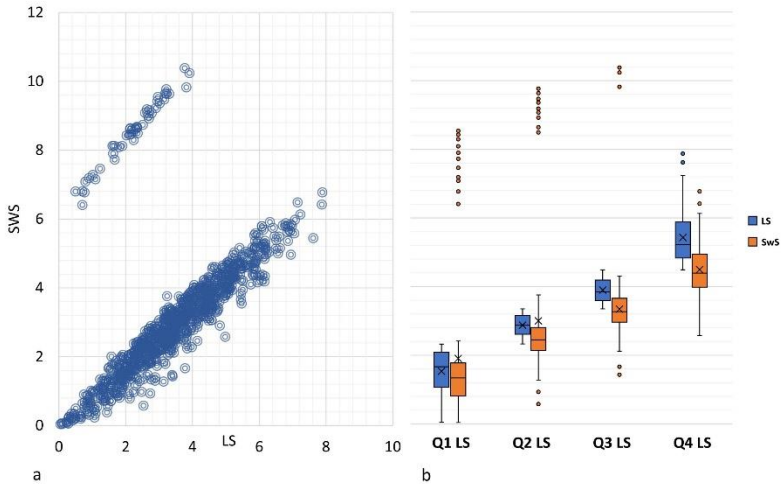

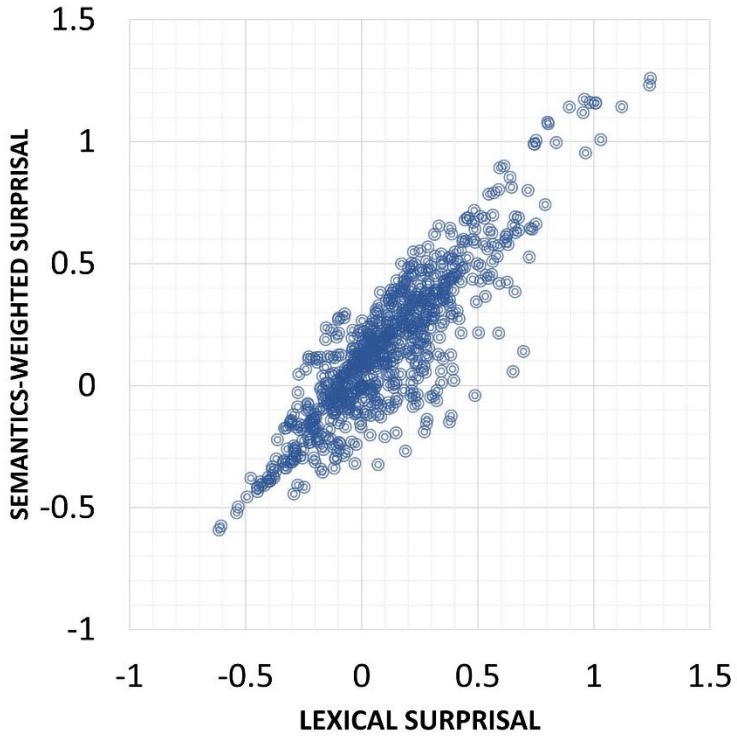**Figure S2.** Scatter plot (a) and box plot (b) of LS and SwS raw values.

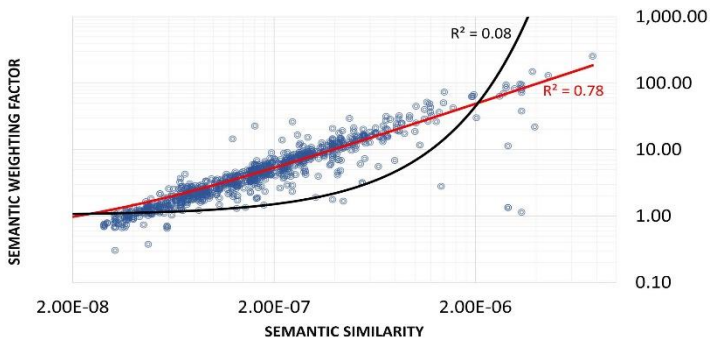**Figure S3.** Scatter plot of LS and SwS fMRI predictors (z scores).

Figure S4. Scatter plot of the semantic weighting (scaling) factor values vs. the semantic similarity values for all content words in the 1800-word corpus. Both axes are logarithmically scaled (base 10). Linear (red) and exponential (black) trend lines are drawn, and the corresponding R^2 value of the fit are displayed, on the graph.
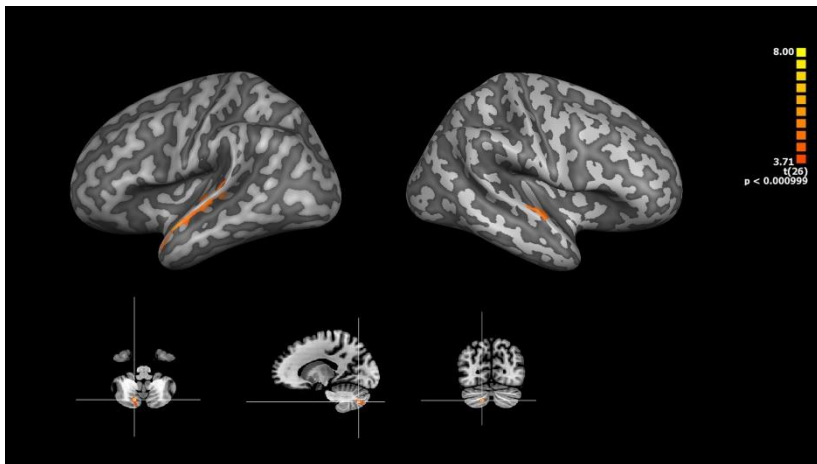

Figure S5 *Semantic factor brain response.* Activation pattern elicited by the isolated semantic scaling factor in the SwS model in real speech, compared to the reversed speech, condition. Clusters with statistically significant effects (p<0.05, cluster level corrected, cluster-forming threshold p=0.001) are overlaid in pseudo-color on an inflated cortical mesh obtained from a Talairach-normalized anatomical scan.

175

Acknowledgements

Questo lavoro di tesi è il frutto di tre anni di dottorato, che come ogni percorso che si rispetti, sono stati più simili a un sentiero accidentato di montagna che a una passeggiata in riva al mare. Se sono arrivato fin qui e non mi sono perso nei boschi è perché intorno a me ho sempre avuto persone con cui condividere ogni momento difficile e ogni piccola grande gioia.

Ringrazio mia madre per aver sempre creduto in me più di quanto io avessi mai fatto. Forse un giorno anche i neuroscienziati riusciranno a capire perché la mamma ha sempre ragione.

Ringrazio mio padre a cui sono allo stesso tempo tanto simile e tanto diverso. Grazie per avermi insegnato a modo tuo che nonostante dovessi costruirmi da solo la mia strada tu ci saresti sempre stato.

Ringrazio mia sorella per avermi sopportato in tutti questi anni di studio. Anche se abbiamo spesso idee diverse ci somigliamo più di quanto vogliamo ammettere soprattutto adesso che hai iniziato ad arrovellarti sulla statistica anche tu.

Ringrazio Daniela che ha vissuto di riflesso tutte le gioie e le difficoltà di questo dottorato. Grazie per essere stata una sorgente viva di entusiasmo e una roccia ferma a cui potermi aggrappare durante le tempeste.

Ringrazio anche le mie colleghe Antonietta e Sara. La prima per avermi (in)volontariamente attirato nel pazzo mondo della ricerca, la seconda per aver condiviso i sali-scendi di questo percorso e i dilemmi su cosa prendere ai distributori automatici.

Ringrazio il mio migliore amico Marco per aver ascoltato con pazienza tutti i miei dubbi e i miei deliri sui vettori semantici, sugli

esperimenti fMRI, sui cieli grigi di Maastricht e su quanto la Guinness sia in fondo la miglior birra del mondo.

Infine, voglio ringraziare tutti quelle persone che ho incontrato sulla mia strada in questi anni e che, anche inconsapevolmente, sono state fonte di ispirazione e conforto. Grazie per avermi teso una mano quando la corrente era più forte.

"If our brains were simple enough for us to understand them, we'd be so simple that we couldn't."

Professor **Ian Stewart**