SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of

Psychology

PhD course in Psychology, Linguistics, and Cognitive Neuroscience - Cycle XXXII

Curriculum: Mind, Brain, and Behavior

Semantic and syntactic modifications in schizophrenia spectrum disorders

PhD candidate: Barattieri di San Pietro Chiara

Registration number: 823151

Tutor: Marco Marelli

Co-tutor: Claudio Giuseppe Luzzatti

Coordinatore: Marco Perugini

ACADEMIC YEAR 2019/2020

List of contents

List of abbreviations 3

Summary 8

1 Introduction 9

1.1 Schizophrenia Spectrum Disorders: the psychopathological and neuropsychological frameworks 9

1.1.1 The psychopathology of SSD 9

1.1.2 The neuropsychology of SSD 11

1.1.4 Integrity of the semantic store in SSD 16

1.1.5 Syntactic processing in SSD 18

1.1.6 Integration of semantic and syntactic information in SSD 19

1.1.7 Neural correlates of language impairments in SSD 20

1.1.8 Some methodological consideration: the assessment of language in schizophrenia using verbal fluency 21

1.2 The interpretation of language impairments in SSD: theoretical frameworks 22

1.2.1 Two fundamental psycholinguistic models 23

1.2.2 The linguistic framework 24

1.2.3 The computational cognitive architecture 27

2 Distributional semantic models applied to a semantic verbal fluency task in people with Schizophrenia Spectrum Disorders 33

2.1 Introduction 33

2.2 Materials and methods 37

2.3 Results 45

2.4 Discussion 56

3 Distributional semantic models applied to a generative associative naming task in people with Schizophrenia Spectrum Disorders 63

3.1 Introduction 63

3.2 Materials and methods 65

3.3 Results 68

3.4 Discussion 80

4 Processing of Argument Structure and Syntactic Complexity in People with SSD 87

4.1 Introduction 87

4.2 Materials and methods 92

4.3 Results 95

4.4 Discussion 105

5 Effects of disrupted lexical representations of verbs on anomalous-sentence processing in Schizophrenia Spectrum Disorders 109

5.1 Introduction 109

5.2 Materials and methods 111
5.3 Results 115
5.4 Discussion 123
6 General Discussion 128
References 136
Annex 1 Additional materials for Chapter 2 and 3 155
Annex 2 Additional materials for Chapter 5 160
Acknowledgments 162

List of abbreviations

ASCH	Argument Structure Complexity Hypothesis
BPRS	Brief Psychiatric Rating Scale
DSM	Diagnostic and Statistical Manual of Mental Disorders
FFD	First Fixation Duration
FTD	Formal Thought Disorder
GD	Gaze Duration
HP	Healthy Participants
NLP	Natural Language Processing
PPMI	Positive Pointwise Mutual Information
SD	Self-Disorder
SPQ	Schizotypal Personality Questionnaire
SSD	Schizophrenia Spectrum Disorders
SVD	Singular Value Decomposition
TFD	Total Fixation Duration
VAS	Verb Argument Structure
VF	Verbal fluency
WAT	Word Association Task
WHO	World Health Organization

Abstract

Title

Semantic and syntactic modifications in Schizophrenia Spectrum Disorders

Although the presence of language disturbances in people with Schizophrenia Spectrum Disorders (SSD) is well established (American Psychiatric Association, 2013), a full characterization of the phenomenon is still missing. The hypothesis of “schizophrenia as the price we pay for language” (Crow, 1997) opens new perspectives on the problem at stake, and suggests the need for a combined approach aiming at integrating the clinical tools nowadays employed to assess language abilities in SSD.

The overall objective of the present work is to advance the understanding of language disturbances in this population by adopting an interdisciplinary approach embracing neuropsychology, psycholinguistics, and computational linguistics. In particular, the present work is focused on: i) the differential contribution of semantic storage and executive functions to verbal fluency; ii) the production and comprehension of verbs argument structure and syntactic complexity, and; iii) the sensitivity to semantics violation on different Thematic Roles.

Forty-three persons with SSD were recruited at the IRCCS Fatebenefratelli of Brescia. Participants’ linguistic processes were investigated by means of: i) two verbal fluency tasks for the evaluation of semantic store integrity and executive function performance, both computed manually and derived from Natural Language Processing (NLP) methodologies; ii) the Northwestern Assessment of Verb Argument Structure (NAVS – Cho-Reyes & Thompson, 2012; Barbieri et al., 2019); iii) an eye-tracking study on semantic violations. For comparison, the same battery was administered to a sample of healthy control subjects matched by age and gender.

In the fluency tasks significant differences in the mean size of semantic clusters, number of switches, as well as measure of coherence were observed between groups, highlighting the differential and non-mutually exclusive contribution of the semantic store integrity and the executive functions to verbal fluency. Moreover, NLP-derived algorithms showed a high-level performance in classifying subjects with and without SSD. A specific difficulty with complex verb argument structure, as well as with non-canonical word order of sentences, both in production and comprehension, was identified in the SSD population. These results are compatible with the Argument Structure Complexity Hypothesis (ASCH – Thompson, 2003) and the presence of an underlying syntactic movement in non-canonical sentences (Chomsky, 1981). Finally, an impaired sensitivity to semantic violations on the “Agent” was observed in the eye-tracking study, compatible with the presence of a “disorder of the self” (Henriksen & Noordgard, 2014) in this population.

In summary, our results underline the presence of specific semantic and syntactic impairments in SSD as seen in language production and comprehension. Moreover, our result support the application of a multi-disciplinary approach to the issue at stake. Our study shows how the added value of fluency measures derived by a computational linguistic approach paired with a fine-grained characterization of receptive and productive language in SSD by means of the tools and theoretical frameworks derived from psycholinguistics can contribute to further characterize language modifications in SSD beyond the current knowledge.

Keywords

Schizophrenia, Verbal Fluency, NLP, Verb Argument Structure, Eye-tracking

Abstract – italiano

Titolo

Modificazioni semantiche e sintattiche nei disturbi dello spettro schizofrenico

Nonostante la comprovata presenza di alterazioni a carico del linguaggio nei disturbi dello spettro schizofrenico (DSS - American Psychiatric Association, 2013), la piena caratterizzazione del fenomeno non è ancora stata raggiunta. L'ipotesi della "schizofrenia come prezzo da pagare per il linguaggio" (Crow, 1997) apre nuove prospettive al problema sotto osservazione e suggerisce la necessità di un approccio volto a integrare gli strumenti attualmente utilizzati per la valutazione clinica delle abilità linguistiche nei DSS.

L'obiettivo generale del presente lavoro è l'avanzamento nella comprensione dei disturbi del linguaggio in questa popolazione. A tal fine è stato adottato un approccio interdisciplinare a cavallo tra la neuropsicologia, la psicolinguistica e la linguistica computazione. In particolare, il presente lavoro si concentra: i) sullo studio dell'interazione tra il magazzino semantico e le funzioni esecutive ai compiti di fluenza verbale; ii) sulla produzione e la comprensione della struttura argomentale del verbo e della complessità sintattica; e iii) sulla sensibilità alle violazioni semantiche a carico di diversi ruoli tematici.

Quarantatré persone con DSS sono state reclutate presso l'IRCCS Fatebenefratelli di Brescia. I processi di elaborazione linguistica dei partecipanti sono stati studiati attraverso: i) due fluenze verbali analizzate adottando sia punteggi tradizioni, sia algoritmi di Natural Language Processing (NLP); ii) la Northwestern Assessment of Verb Argument Structure (NAVS – Cho-Reyes & Thompson, 2012; Barbieri et al., 2019); iii) uno studio sui movimenti oculari durante un compito di lettura di frasi con violazioni semantiche. Come confronto, lo stesso protocollo sperimentale è stato somministrato a un campione di soggetti sani di controllo appaiati per sesso ed età.

Nei compiti di fluenza verbale si sono osservate differenze significative tra i due gruppi relativamente alla dimensione media dei cluster semantici, al numero di switch tra cluster, nonché alle misure di coerenza semantica, mettendo in evidenza il contributo differenziale e non mutualmente esclusivo dell'integrità del magazzino semantico e delle funzioni esecutive alla fluenza verbale. Inoltre, gli algoritmi sperimentali NLP basati sulle misure adottate hanno dimostrato un'elevata prestazione nella classificazione dei soggetti con o senza DSS. Dai risultati della batteria NAVS è stato inoltre possibile identificare una difficoltà specifica dei pazienti rispetto a strutture argomentali complesse, nonché rispetto a frasi aventi un ordine delle parole non canonico, sia in produzione sia

in comprensione, compatibili con la Argument Structure Complexiy Hypothesis (ASCH – Thompson, 2003) e con un deficit del movimento sintattico che soggiace alle frasi non canoniche (Chomsky, 1981). Infine, tramite lo studio dei movimenti oculari è stato possibile osservare una diminuita sensibilità alle violazioni sul ruolo tematico di “Agente”, compatibile con la presenza di un “disturbo del Sé” (Henriksen & Noordgard, 2014) nei DSS.

In conclusione, i nostri risultati confermano la presenza di deficit specifici a carico delle abilità semantiche e sintattiche osservabili in produzione e in comprensione nei DSS. Inoltre, l’esito dello studio supporta l’applicazione di un approccio multidisciplinare al problema in oggetto. Il presente studio dimostra come misure di fluenza verbale derivate da un approccio linguistico-computazionale associate a una dettagliata caratterizzazione del linguaggio recettivo e produttivo nei DSS grazie agli strumenti e alle cornici teoriche della psicolinguistica possono contribuire all’avanzamento nella caratterizzazione delle modificazioni del linguaggio nei DSS oltre allo stato dell’arte.

Parole chiave

Schizofrenia, Fluenze verbali, NLP, Struttura argomentale, Movimenti oculari

Summary

This work starts with an overview of the psychopathological and (neuro)cognitive aspects of the disorders, with a detailed focus on language impairments in SSD. It then presents some specific language phenomena at the interface between semantics and syntax, the relevance of adopting tools from the neurolinguistics field, and the potentialities of Natural Language Processing tools to the investigation of verbal production (Chapter 1).

In Chapter 2 and 3, I present the results of two experiments on two verbal fluency tasks (semantic and generative associative naming) with a comparison between traditional scoring methods and scoring derived from NLP tools. In Chapter 4, I present a study on Verb Argument Structure (VAS) and syntactic complexity adopting tools developed for the aphasic population. In Chapter 5, I report the results of an eye-tracking experiment on the effect of a semantic violation on the animacy feature of the grammatical subject.

The overall results of the work are commented in the General Discussion (Chapter 6) and demonstrate the added value of adopting tools developed outside the field of psychiatry for a better characterization of language disturbances in psychiatric conditions.

1 Introduction

In order to appreciate the modifications of language occurring in schizophrenia, it is essential to both understand what we mean by the term "schizophrenia", and which aspects of that complex phenomenon we call "language" we are dealing with. I will thus briefly introduce this spectrum of disorders as clinically interpreted today, with a particular emphasis on the pathological signs affecting language; then, I will describe the theoretical framework adopted in the present work, which may help interpreting those aspects of language I think is worth exploring in this condition.

1.1 Schizophrenia Spectrum Disorders: the psychopathological and neuropsychological frameworks

1.1.1 The psychopathology of SSD

Although the earliest descriptions of its symptoms can be traced back to the second millennium B.C. (Okasha & Okasha, 2000), schizophrenia has been object of systematic medical investigations only since the seminal work on "dementia praecox" by Emil Kraepelin (1886). During this time frame, the definition of the condition evolved, with different aspects of the disorder being differently emphasized over time. Today, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) collectively categorizes schizophrenia, other psychotic disorders, and schizotypal personality disorder under the label of "Schizophrenia Spectrum Disorders" as:

> "*abnormalities in one or more of the following five domains: delusions, hallucinations, disorganized thinking (speech), grossly disorganized or abnormal motor behavior (including catatonia), and negative symptoms*"

with a diagnosis issuable if at least two diagnostic criteria are met, with the first being one of the three core symptoms (delusions, hallucinations, or disorganized speech), and the second either disorganized motor behavior or negative symptoms.

People with SSD may present *delusions*, strongly held beliefs that cannot be changed even considering contradictory evidence. Themes of delusion can be persecutory, referential, somatic, religious, or grandiose, and examples include the belief that one's internal organs have been removed and replaced with some else's organs without leaving a scar (bizarre delusion) or being under police surveillance despite lack of evidence (non-bizarre delusion). Thought withdrawal (the belief that one's thoughts have been removed) and thought insertion (alien's thoughts have been put in one's mind), or delusions of control (one's body or action are controlled by some outside force) are also observed. *Hallucinations* are perception-like experience that occur without an external stimulus and are not under voluntary control. All sensory modalities can be affected, but in SSD auditory hallucinations, experienced as voices that are distinct from the individual's own thoughts, are the most common. Individual with SSD may also manifest an *abnormal motor behavior* (including catatonia), which goes from excessive agitation, stereotyped movements, and echoing of speech, to a complete lack of verbal and motor responses. Problems are reflected on any goal-directed behavior, which translates to difficulties in performing daily living activities.

The distinction between positive and negative symptom is now recognized as fundamental to understand functional limitations among individuals with schizophrenia (Andreasen & Olsen, 1982; Crow, 1980). Positive symptoms are represented by excess or distorted normal behavior or cognition (e.g., hallucinations and delusions); on the other hand, negative symptoms are distinguished by the absence of some normal capabilities (e.g., blunted affect, emotional withdrawal, or lack of spontaneity and flow of conversation). Also, while positive symptoms are often episodic, negative symptoms are typically stable (Rollins et al., 2010).

From the observation of the individual's speech it is possible to infer the presence of the third core symptom, *disorganized thinking* (or disorganized speech, also called formal thought disorder, or FTD). People with SSD can present derailment or loosening of association (i.e., switching from one topic to another), as well as tangentiality, whereby answers to questions may be obliquely related or completely unrelated to them. Rarely, speech may be severely impaired so as to be nearly incomprehensible, resembling receptive aphasia in its linguistic disorganization (incoherence or "word salad"). Symptoms can be severe enough to impair effective communication, although less severe presentations may occur during the prodromal and residual periods of the disorder. Among the negative symptoms expressed in SSD (avolition, anhedonia, asociality), diminished emotional expression and alogia can affect speech: individuals can present flat intonation of speech (prosody) and reduced emotional emphasis to speech, while alogia presents as a diminished speech output. I will later discuss language disturbances in schizophrenia more in details.

1.1.2 The neuropsychology of SSD

In the last decades, it has been hypothesized that schizophrenia may represent a possible long-term consequence of early neural development anomalies (the so-called "neurodevelopmental model of schizophrenia"). The large body of research carried out over the last 50 years clearly demonstrates the presence of a significant cognitive impairment in patients with schizophrenia (Gold et al., 2002), with cognitive disorders affecting 75% of people with SSD (Harvey et al., 2001). Although not considered diagnostic criteria, cognitive impairments in SSD are broadly described in the literature, being so general and diffuse among this population (Fioravanti et al., 2012) that they may be considered endophenotypic traits of the disorder (Haug et al., 2012). The raising consensus is that the study of neurobiological mechanisms of cognition may provide a means to better understand the symptoms of schizophrenia. The most consistent impairments are in the domains of (i) attention, (ii) memory, and (iii) executive functions, which are typically found before the onset of clinical symptoms (Saykin et al., 1994), and persist throughout the course of the illness (Nieuwenstein et al., 2001).

i. Impairments in sustained attention and slow post-perceptual processing have been found in this population. The hypothesis offered to explain such findings are multiple and range from a defective filtering or screening of incoming information; the loss of information in the short term memory; an impaired control and maintenance of selective processing strategies; a reduced processing capacity; as well as an impaired effort or controlled processing (Callaway & Naghdi, 1982). Cohen and Servan-Schreiber (1992) proposed that a disturbance in the internal representation of contextual information, corresponding to the neuromodulatory effect of dopamine in the prefrontal cortex, could account for schizophrenic deficits in attention.
ii. Performance at tests assessing both long- and short-term memory shows significant difference between people with and without SSD (Fioravanti et al., 2005; Fioravanti et al., 2012). Impairment in encoding and retrieval processes have been attributed to the poor memory performance of these patients (Gold & Harvey, 1993). Impaired verbal memory is well documented in adults with schizophrenia, over and above normal general intellectual functioning, also in subjects who have never been exposed to neuroleptics (Landrø & Ueland, 2008). In patients

with early phase schizophrenia, impaired verbal memory has been associated with high level of self-disorder (Haug et al., 2012) (see later);

iii. Consistent with the presence of a frontal dysfunction, impairment of executive functions (Everett et al., 2001), including attentional control, inhibition, and cognitive flexibility, as well as higher order executive functions, including action planning (Zalla et al., 2001), have been repeatedly reported in schizophrenia, leading to a dysexecutive model of thought disorder (Stirling et al, 2006). In particular, the ability to adjust one's behavior according to a changing environment, i.e., the so called “cognitive flexibility” (Scott, 1962), has long been studied in people with SSD, with considerable evidence pointing to the presence of deficits in attentional set-shifting and task switching, similar to those observed in individuals with frontal lobe lesions (Elliott & Sahakian, 1995). In fact, early studies on these latter patients lesion using the Wisconsin Card Sort Test (WCST) showed a characteristic difficulty in shifting from one sorting criterion to another, in the face of negative feedback (Milner, 1963). These kinds of errors were called “perseverative errors”, to distinguish them from other types of incorrect responses on the test, and these perseverative errors became the foremost example of stuck-in-set behavior. The WCST has been used in schizophrenia, with evidence indicating that SSD participants make a significantly higher number of perseverative responses than normal control subjects do (Everett et al., 2001).

1.1.3 Language impairments in SSD - disorder of thought or disorder of language?

The bizarre aspects of positive symptoms observable in schizophrenic speech is traditionally referred to as “disorder of thought”. Rather than a primary disorder of language, as in aphasia, speech disturbances in SSD have long been regarded as reflective of an underlying disorder of thinking. However, given that “thinking” cannot be probed directly, thought disorder can only be deduced from speech, making this assumption critically tautologic (Rochester & Martin, 1979). Such definition, in fact, implies that what the person with SSD says is the direct consequence of what he or she thinks, and that the ability to convey such thoughts by means of language is intact. Patients often express abnormal thoughts with normal language, and psychiatrists thus distinguish between disturbances of “thought content” from “thought form”. The former relates to the actual thoughts described, and can report delusions (false beliefs), depersonalisation and derealisation, phobias, preoccupations and obsessions. The latter, on the contrary, refers to how the person's thoughts are expressed in speech and the way ideas are linked (logical, goal-directed, loose associations): it can range from an easily understandable, coherent speech, to an incomprehensible “word salad”.

Although there is a fundamental difference between language and thought, this difference has been limitedly investigated in schizophrenia, despite an impaired verbal communication in the form of "*disorganized speech*" and of "*breaks or interpolations in the train of thoughts, resulting in incoherence or irrelevant speech, or neologisms*" is one of the core diagnostic feature of schizophrenia as indicated both in the DSM-5 (APA, 2013), and in the ICD-10 (1992).

In fact, language deficits are pervasive among subjects with a diagnosis of a disorder in the schizophrenic spectrum (DeLisi, 2001) and there is now evidence that all levels of language are disrupted (Bellani et al, 2009).

Early descriptive works of spontaneous speech in schizophrenic patients have tried to formalize the linguistic abnormalities observed in this population. The scale for the assessment of "Thought, Language and Communication" (TLC - Andreasen, 1979) was developed with the intent of formalizing thought disorder in absence of an agreement concerning its definition. Table 1.1 summarizes the most common features identified by Andreasen, with some explicatory examples (reported from Andreasen 1979 if not otherwise specified).

Type	Definition	Example
Derailment	Unclear or confusing connections between larger units, such as sentences or ideas.	"Why do you think people believe in God?" "Urn, because making a do in life. Isn't none of that stuff about evolution guiding isn't true anymore now. It all happened a long time ago. It happened in eons and eons and stuff they wouldn't believe in him."
Loss of goal	Failure to follow a chain of thought through to its natural conclusion.	"I always liked geography. My last teacher in that subject was Professor August A. He was a man with black eyes. I also like black eyes. There are also blue and grey eyes and other sorts, too..." (Bleuler, 1950)
Poverty of content of speech	Although replies are long enough so that speech is adequate in amount, it conveys little information. Language tends to be vague, often over-abstract or over-concrete, repetitive, and stereotyped.	"Tell me what you are like" "Ah one hell of an odd thing to say perhaps in these particular circumstances, I happen to be quite pleased with who I am or how I am and many of the problems that I have and have been working on I have are difficult for me to handle or to work on because I am not aware of them as problems which upset me personally."
Tangentiality	Replying to a question in an oblique, tangential, or even irrelevant manner.	"What city are you from?" "Well that's a hard question to answer because my parents. ... I was born in Iowa, but I know that I'm white instead of black so apparently, I came from the North somewhere and I don't know where, you know, I really don't know where my ancestors came from."
Poverty of speech	Monosyllabic answers	"Yes", "No", "Maybe", "Don't know", "Last week".
Illogicality	A pattern of speech in which conclusions are reached that do not follow logically. This may take the form of *non sequiturs*.	"Parents are the people that raise you. Anything that raises you can be a parent. Parents can be anything, material, vegetable, or mineral, that has taught you something."
Perseveration	Persistent repetition of words, ideas, or	"I think I'll put on my hat, my hat, my hat, my hat."

	subjects.	
Incoherence	Word Salad, jargon aphasia, schizophasia, paragrammatism	"Where did all this start could it possibly have started the possibility operates some of the time having the same decision as you and possibility that I must now reflect or wash out any doubts that that's bothering me and one instant what's bothering me in the whole thing must stop immediately otherwise that is damned if that is damned that's no use and if I don't tell the truth that's bothering me an awful lot in my wisdom the truth is I've got the truth to tell you with mine signing here and as I am as God made me and understand my position and you'll listen with intelligence your intelligence works lit again and is recorded in my head" (Wykes & Leff, 1982)
Self-reference	The patient repeatedly refers the subject under discussion back to him or herself.	"What time is it?" "Seven o'clock. That's my problem. I never know what time it is. Maybe I should try to keep better track of the time."

Table 1.1 Types of communication anomalies in people with SSD.

In the next decades, these findings have been extensively replicated (for example, see Elvevåg et al., 2007), although with slightly different terminologies derived from new clinical instruments (the Scale for the Assessment of Thought, Language and Communication; TLC - Andreasen, 1986). Spontaneous speech in SSD was found to be characterized by reduced syntactic complexity (DeLisi, 2001), as well as reduced verbal and speech fluency (Covington et al., 2005).

At the pragmatic level, in people with SSD it has been observed an impaired comprehension of proverbs, metaphors, idioms, as well as of non-literal expression comprehension, irony and sarcasm (Mitchell & Crow, 2005; Tavano et al., 2008), indicative of a predisposition to concretism (Kircher et al., 2007). In terms of conversational skills, people with SSD tend to disrespect turns in conversation. Intonation has been reported to be flattened, with linguistic prosody difficulty in differentiating declarative or interrogative sentences by prosodic sentence contour (Covington et al., 2005).

The link between schizophrenia and language has been further strengthened with Crow's epigenetic theory of schizophrenia (1997a; 1997b; 2008), in which schizophrenia has been defined as "*the price we pay for language*", hypothesizing that language-related nuclear symptoms might be due to a failure in establishing left hemisphere dominance for language (the prevalence of left-handedness in psychotic disorders is 40%, much higher than the rate of the general population, which is around 10%). Specific morphological changes reported in the schizophrenic population is the reduction or absence of cerebral asymmetry. Crow's theory stems from the observation of the differences between the forms of psychotic illness, going from maniac-depressive to schizophrenic psychoses. Such heterogeneity and lack of clear boundaries between diagnoses would be, according to the Author, suggestive of the "continuum" nature of the disorder. In this sense, normal variations within the spectrum would be present in the normal population, but would become visible, pathological, and hence diagnosable, only in its extreme variations. To further support the epigenetic origin of schizophrenia, Crow report the results of the WHO Ten Country Study of Incidence (Jablensky et al., 1992): if we reduce the definition of schizophrenia to its core symptoms (also called first rank or nuclear symptoms, which includes *Gedankenlautwerden*[1], thought insertion and removal), the disorder have a constant incidence worldwide: in other words, its incidence is constant and irrespective of social, geographic, and cultural variation. Such constant incidence indicates that the mutation that originates the disorder must have preceded the separation of the populations in which it is now present. Crow traces this mutation back to a genetic variation emerging during the transition from a precursor hominid toward the population of *Homo sapiens* (between 137.000 and 250.000 years ago). In order to be kept in the genetic endowment of the species, given the clear disadvantage for the single person who would show that mutation, this variation must have brought, or be associated to, an evolutionary advantage for the species. This is the core of Crow's theory: if we assume that these two cognitive phenomena, language and schizophrenia, share the same genetic basis, the latter can be considered "the price" that humans, as a species, had to pay in order to have the former, despite the clear disadvantage it brings for the individual. The evolutionary advantage might have well been the capacity of language, which like schizophrenia, appears to be intrinsic to *Homo sapiens* (Chomsky, 1965), and relatively constant across populations. The two abilities may reflect different aspects of a single genetic change, a "speciation event" that introduced an innovation in the functional organization of the brain. The salient fact about the neural basis of language is that it is lateralized, and such specialization is accompanied by a preference to use the right hand for tasks requiring fine motor skill. Dominance for language and handedness is reflected in the anatomical asymmetry in the brain, derived by the so-called "cerebral torque", a neuro-developmental re-organization of the brain. Remarkably, people with schizophrenia show a reliable loss of the asymmetry (to the point of symmetry) of the planum temporale (Shapleske et al., 1999), visible also in the high incidence of left-handedness. According to the Author, both language anomalies in schizophrenia and nuclear symptoms would derive from anomalies of the intra-hemispheric functioning. In the last decades, the importance of both environmental and genetic factors in the development of the disorder has been demonstrated. Factors associated with increased risk of developing schizophrenia are: growing up in a urban environment, immigration, cannabis usage, male gender and perinatal events (van Os & Kapur, 2009). Of all the known risk factors for schizophrenia, genetics is the single most important one, but the precise mode of inheritance is still unknown: a meta-analysis of twin studies suggests a strong genetic component, estimating the heritability to be

[1] The German term *Gedankenlautwerden* can be roughly translated as "thought sonorization". This is a hallucination where a person hears voices which anticipate what he or she is about to think, or which state what the person is thinking as he thinks it. Together with thought insertion and thought removal, they represent a loss of the boundary between the self and the outside world and crucially constitute a pathology of the relationship between language and thought.

81% (95% confidence interval, 73% to 90% - Sullivan et al., 2003). Crow's theory is, in this sense, fundamentally epigenetic, in the sense that it suggests the involvement of phenotype changes in addition to the traditional genetic basis of inheritance. The large amount of disease variation we observe in the clinical population is unlikely to be caused by a single DNA sequence variation. More probably, a set of epigenetic processes are likely to act dynamically and independently in the control of gene expression, regulating neurobiological and cognitive processes (Dempster et al., 2013).

Given that a complete review on language impairments in SSD would be out of the scope of the present work (especially works on phonology and morphology, but see Bellani et al., 2009 for a comprehensive review), I will focus the following section on studies investigating the integrity of the semantic store (verbal fluencies in general, and differential abilities in verb and noun naming), syntax processing (in production and comprehension), as well as integrative processes involving the two levels in this population.

1.1.4 Integrity of the semantic store in SSD

In neuropsychology, the integrity of the semantic store is traditionally assessed with verbal fluency tasks. A verbal fluency tasks requires the subject to generate as many words as possible in a fixed time period. In the present work we will only consider semantic fluency, i.e., test requiring the participant to produce words from a certain semantic category, for example "animals"[2].

The two versions that are most frequently used in neuropsychological research and the clinical practice are phonemic fluency (producing words that start with a certain letter: this test is considered to reflect the role of executive functions, and it is not considered a semantic task. We will not consider this in the text) and semantic fluency (producing words from a certain semantic category, for example "animals").

In SSD, an increased number of perseverations (inappropriate repetitions of earlier responses) and intrusions (the unintentional recollection of inappropriate information) in verbal fluency tasks are observed. These phenomena have been interpreted as due to an increased susceptibility to interferences in recall strategies (Galaverna et al., 2016) tapping on an impaired executive functionality. However, the reduced performance in fluency tasks where the executive demand is low (such as semantic associations and sorting tests) observed in this population (Doughty & Done, 2009) points to some other causing factors: these finding have been attributed to a substantial impairment of the semantic network (Paulsen et al., 1996), at least in the non-paranoid subtype of schizophrenia.

[2] In the present work we will not consider phonemic fluency, i.e., test requiring participants to produce words starting with a certain letter. This test is considered to reflect the role of executive functions, and it is not considered a lexical-semantic task.

Deficit in verbal fluency appear to be particularly associated with transition to psychosis in people with clinical high risk (Fusar-Pioli et al., 2012) and poor performance in semantic verbal fluency tasks remain stable in individuals with schizophrenia during the time course of the disease (Robert et al., 1998).

Verbal fluency scoring is usually calculated according to the total number of coherent items produced within a given time frame. However, as semantic verbal fluency is thought to involve "clustering" (Gruenewald & Lockhead, 1980), that is, a cognitive process of forming clusters of related words within a certain category, a more fine-grained qualitative description of these tasks must include the analysis of the number and size of semantic clusters, as well as the switches (transitions) between them. However, such studies are not conclusive[3]. Early studies found that, in comparison to healthy controls, schizophrenic patients showed a significant impairment of both switching and clustering in a semantic fluency task (Robert et al., 1998). However, later evidences showed that people with SSD spend more time switching to words within clusters and to different clusters than controls; nonetheless, group differences in number of switches disappear when the total number of words produced was covaried, suggesting that, despite a general slowness attributable to difficulties in finding new words, patients are similar to controls with respect to number of concepts in their semantic network (Elvevåg et al., 2002). Similarly, Bozikas and colleagues (2005) found that the differences in terms of number of clusters and number of switches disappeared after controlling for total output: this finding has been interpreted hypothesizing that patients would in fact adopt the same cognitive strategies as healthy controls, but less effectively (since, in the end, they produced fewer words than healthy individuals). On the other hand, no significant differences were found in the number of switches and in the mean size of clusters in a semantic verbal fluency task (category "animal") in a group of adolescents with SSD compared to a healthy control group (Landrø & Ueland, 2008). Later, it has been shown that people with SSD generate fewer words than healthy controls and present intact clustering, but decreased switching (Okruszek et al., 2013). In summary, evidences from the previous literature on clustering and switching seem to point toward the impact of a disrupted executive functioning in this population on the performance in semantic verbal fluency tasks, but the role of the integrity of the semantic store has not been clarified yet.

Despite various abnormalities in action processing have been reported in SSD (poverty of action, disorganized behavior, and stereotyped actions) (APA, 2013), studies on differential abilities in naming nouns and verbs in people with SSD are limited, being the majority of studies on verbal fluency focused on object words (nouns). Nonetheless, findings so far seem to converge towards a significant action fluency deficit in patients with schizophrenia or SSD. In a verb generation task,

[3] See the section "Some methodological consideration: the assessment of language in schizophrenia using tests of verbal fluency" below for a detailed description of scoring systems based on number and size of clusters.

requiring subject to describe "*what an object does or something you can do with an object*", Marvel and colleagues (2004) found that schizophrenic patients produced a high number of errors, irrespective of the strength of the association between the noun and the target verb (i.e., noun in the low selection condition were expected to trigger only one verb – indicating a strong response strength – while nouns in the high selection condition could trigger more than one response, a weak response strength). Consistently with these findings, Woods and colleagues (2007) found that persons with schizophrenia performed approximately one standard deviation below non schizophrenic subjects on a test of action (verb) fluency. In a study investigating verb/noun differences in adults with schizophrenia, Kambanaros and colleagues (2010) found that action names were significantly more difficult to retrieve than object names in schizophrenic patients, pointing to a grammatical class effect. Similarly, in a test requiring subject to generate action verbs, Badcock and colleagues (2011) found a significant action (verb) fluency deficit in a sizeable proportion of individuals with chronic schizophrenia (66%), with performance over 1 standard deviation below that of healthy controls; within the schizophrenia sample, poor fluency for actions, tools and musical instruments (but not fruit or phonological fluency) was significantly associated with increased scores to the "Odd Speech" subscale of the Schizotypal Personality Questionnaire (SPQ; Raine, 1991), suggesting a specific role of action-based language production deficit in thought disorder. In the attempt to further characterize this selective deficit for verbs, Smirnova and colleagues (2017) compared people with and without schizophrenia on a task requiring subject to generate action or mental state verbs. The results indicated that the percentage of action verbs produced was significantly lower in patients than controls, whilst the percentage of mental state verbs produced did not differ. Moreover, patients' produced verbs were significantly less concrete, positively correlated with memory and intelligence, and negatively correlated with interpersonal symptoms.

1.1.5 Syntactic processing in SSD

Several studies have found an impaired processing of complex syntax in SSD patients and, in general, relative to healthy controls, people with SSD have deficits with both comprehension and production of syntactically complex sentences.

Results of an early work on language production in schizophrenic subjects pointed to a reduced syntactic complexity (examined using the Brief Syntactic Analysis by Sacks et al., 1974) in schizophrenic speech whose negative symptoms were prominent (Thomas, 1996), and in schizophrenic patients in the early stage of the condition (Thomas et al., 1996). Similarly, further research found that patients with and without formal thought disorder produced significantly more syntactic errors than healthy controls (Oh et al., 2002). Moreover, Lelekov and colleagues (2000) tested a group of schizophrenic patients using a task developed for the aphasic clinical population, requiring participants to identify (by pointing to a set of photographs) the agent, object, and recipient of sentences presented auditorily. The experimental set of sentences contained canonical (syntactically simple, such as active and subject cleft sentences like "*It was the elephant that hit the monkey*") sentence, involving no syntactic movement, as well as non-canonical (syntactically complex, such as passive, and subject- object-relative sentences like "*The elephant that the monkey hit hugged the rabbit*") involving syntactic movement. The performance of patients to the test was highly correlated to that of a non-linguistic sequence processing task, indicating a significant effect of complexity, independent of the nature of the task. Authors interpreted these results as suggestive of a dysfunction of cognitive sequence processing mechanisms that would affect both linguistic and non-linguistic sequences. Indeed, these findings point out to a specific difficulty of SSD patients when processing complex sentences involving syntactic movement, like that observed in aphasic subjects with left hemisphere lesions in the peri-sylvian cortex (Caplan et al., 1985; Caplan et al., 1996). More recently, Bagner and colleagues (Bagner et al., 2003) tested the correlation between language comprehension, working memory, and thought disorder and hallucinations in schizophrenic patients compared to healthy controls, with a task where sentences differing in length (simple short and long sentences) and syntactic complexity (center-embedded subject- and object-relative sentences) were presented auditorily followed by comprehension questions: results indicated that patients showed significantly more severe comprehension deficits (measured in terms of response accuracy) than controls, with a significant effect of sentence type.

1.1.6 Integration of semantic and syntactic information in SSD

Kuperberg and colleagues (1998) investigated the performance of people with and without schizophrenia to an on-line monitoring task. In this experiment, linguistic stimuli were presented as pre-recorded sentences and participants were instructed to press a response button when they hear a target word. The target word was always the direct object of the verb, and sentences were rendered anomalous by pragmatic, semantic or syntactic violations. Results indicated that healthy controls and schizophrenic patients without thought disorder took longer to recognize words preceded by linguistic anomalies compared with words in normal sentences. On the contrary, schizophrenic thought-disordered patients showed significantly smaller differences in reaction times, suggestive of a relative insensitivity to linguistic violations. These findings have been interpreted in the light of the theory that schizophrenic thought disorder arises from a deficit in the use of linguistic context to process and produce speech. To test the hypothesis of the presence, in schizophrenia, of an impairments in building up sentence context due to abnormalities in combining semantic and syntactic information, Kuperberg and colleagues (2006) administered a self-paced reading task to a sample of people with and without schizophrenia paired with acceptability judgments. Target stimuli were sentences with semantic violation created by assigning an inanimate subject noun to verbs that assign the role of Agent (e.g., in a sentence such as "*For breakfast the eggs would only eat toast and jam*" the verb "to eat" requires an animate Agent as grammatical subject: the semantic violation is given by replacing the it with an inanimate noun. i.e., "eggs"). This condition was compared to non-violated, pragmatically-violated, and morpho-syntactically-violated sentences. Results showed that at sentence-final words and decisions, patients showed smaller reactions time differences than healthy participants (HPs) to animacy-violated sentences or morpho-syntactically violated sentences compared to pragmatically- or non-violated sentences. Such relative insensitivity to these violations was interpreted as suggestive of a specific difficulty, in schizophrenic patients, in combining semantic and syntactic information to build up sentence context. Electrophysiological evidences reviewed by Kuperberg and colleagues (2010) further suggest that, when studying the processes underlying the building up of sentence and discourse structure, it is not possible to separate, on one hand, the structure and function of semantic memory and, on the other hand, the ability to combine and integrate words together. Rather, according to the Authors, the language impairment in schizophrenia would result from a dysfunctional interaction between these systems in an effort to build up higher-order meaning.

1.1.7 Neural correlates of language impairments in SSD

In the last decades, a high number of neuropathological and neuroimaging studies have been carried out in order to better understand cognitive impairments in SSD: overall, these findings have confirmed the theoretical model that interprets schizophrenia as a pathology of the neocortex (Gold et al., 2002), and it is known that the pathophysiology of this disorder involves the abnormal development of language-related regions (Nicolson et al., 2000). A complete review of these findings would be out of the scope of the present work: I will thus present here briefly only those works that are relevant for the study of language impairments in SSD.

Neuroimaging studies found a reduced hemispheric dominance for language processing in this population, affecting the right homologous of the Broca's area (Spaniel et al., 2007), regardless of gender (Sommer et al., 2003), also when experiencing psychotic symptoms (Weiss et al., 2006) in tasks of verbal fluency, verb generation, and reading. Patients with schizophrenia show a reliable loss of the asymmetry (to the point of symmetry) of the planum temporale, due to an increase of the size of the right homologous (Shapleske et al., 1999). Functional anomalies such as smaller activations in left hemisphere regions usually engaged in word generation, with compensatory functional pattern in the right hemisphere (Artiges et al., 2000) were also found. Abnormalities in the frontal cortex, containing Broca's area (Shenton et al., 2001), as well as abnormalities of the corpus callosum (Brambilla et al., 2005) have also been reported in schizophrenia.

1.1.8 Some methodological consideration: the assessment of language in schizophrenia using verbal fluency

In the clinical setting, the subjective evaluation by the treating clinician of the patient's self-verbal presentation (how the person speaks and interacts with others) is an essential diagnostic tool. Psychopathological scales such as the Brief Psychopathological Rating Scale (BPRS) (Overall & Pfefferbaum, 1962) provide a 0 to 7 qualitative scale to assess, among others, the presence of unusual thought content and of conceptual disorganization.

Standardized neuro-psychological batteries and tests are also available. Among these, verbal fluency tasks are the most widely used. Verbal fluency (VF), also known as list generation task, is generally used to get a measure of the subject's cognitive flexibility and of the ability to retrieve words from the lexical store system (Spinnler & Tognoni, 1987). VF tasks have several variations (oral or written, for example) and require subjects to list as many items as possible in a given time frame following a categorical aspect, which can be either semantic, phonemic, or even syntactic. Another possible task is the free association task, requiring the subject to name either the first related word that come to her/his mind (Jenkins & Palermo, 1965) or, in a multiple-answer task, as in Spinnler and Tognoni (1987) as many words as possible whose meaning are in relationship with the probe item. Performing VF tasks requires mental flexibility, efficient retrieval and recall of words, as well as inhibition of dominant responses (Henry & Crawford, 2010). VF can either be considered i) a proxy measure of executive functions, ii) a linguistic measure of integrity and organization of the mental lexicon, or iii) a combination of the two. Greater involvement of the executive functions or the integrity of the semantic store is required if the task at hand is semantic or phonemic. Executive functions, such as strategic research and inhibitory controls, are deemed more involved in phonological fluency tasks than in semantic VF tasks: in the former, subjects are required to held a phonological cue in the working memory, and to scan and retrieve from the lexical store only those words starting by that specific cue, an effort which is considered more cognitively heavy than navigating through a pre-set semantic category. In fact, fluency tests requiring participants to name items in a category (e.g., animals) provide a "conceptual" structure (i.e., the category itself) which is missing in phonemic VF tests. Thus, VF tasks requiring word generation according to an initial letter give the greatest scope to subjects seeking a strategy for guiding the search for words and are most difficult for subjects who cannot develop strategies on their own. Impaired VF is also associated with frontal lobe damage (Janowsky et al., 1989). However, when fluency is tested during PET/MRI scanning, the metabolic pattern suggests that rather than a predominantly frontal involvement in this task, both temporal and frontal regions participate bilaterally: of interest was the finding that normal subjects with the lowest fluency scores had the highest metabolic rates, suggesting that poorer performers must invest more effort in the task (Parks et al., 1988).

An initial methodology to score semantic VF tasks in terms of number of clusters and switches between them was developed by Troyer and colleagues (Troyer et al., 1997) and successively improved by Koren and colleagues (Koren et al., 2005). The formers suggested that performance at a VF task implies the deployment of at least two components: i) clustering, the production of words within a semantic or phonemic subcategory; and ii) switching, the ability to shift efficiently to a new category. Whereas the first component would rely on processes such as verbal memory and word store, pertaining to functions of the temporal lobe, the second would be funded on frontal processes such as strategic search and cognitive flexibility. On semantic fluency, clusters are defined as groups of words generated subsequently and belonging to the same subcategory. For example, within the category "animals", several clusters such as "farms animals", "birds", and "zoo animals" can be formed. The categories (and subcategories) proposed in Troyer's original work included geographical living environment, type of human use, and zoological categories, and were derived by the actual patterns of words generated by the participants of the study. Switches are the number of transitions between clusters, including single words. Troyer considers errors and repetitions in the calculation of cluster size. The original definition of these indexes, as proposed by Troyer and colleagues (1997) included:

1. *Cluster Size*: the mean number of words in a semantic cluster, minus 1 (errors and repetition included). Clusters consist of successively generated words belonging to the same taxonomic subcategory (living environment, human use, etc.), as specified by the Authors.
2. *Number of Switches*: total number of transitions between clusters, including single words.

As later pointed out (Abwender et al., 2001), this system considers single, non-clustered words as clusters of size 0. Thus, a production such as (1) and (2)

1. <u>gatto, cane</u>
2. balena, <u>leone, tigre, leopardo, ghepardo</u>, <u>aquila, falco</u>, <u>cane, lupo</u>, canguro, <u>rana, rospo</u>

would both yield a mean cluster size of 1, although they reflect different performance levels. More precisely, (1) presents one single cluster of two words, which, according to Troyer's method, should be scored as a mean cluster of size equal to 1; at the same time, (2) would score 6 clusters, whose sizes would be 0, 3, 1, 1, 0, and 1 respectively. The resulting cluster mean size for (2) would hence be (0+3+1+1+0+1)/6 = 1. To overcome this problem, Koren and colleagues (2005) proposed to define clusters as 2 or more related words, leaving single words aside.

1.2 The interpretation of language impairments in SSD: theoretical frameworks

As outlined above, language impairments in SSD affect all levels of the linguistic processing, both in production and in comprehension. A proper interpretation of these phenomena requires a well-structured theoretical framework. In this work, I propose that the combination of three perspectives – two classical psycholinguistic models, a functional linguistic framework, and a promising model of cognitive architecture derived from computational linguistics – might help unveiling some specific aspects of language production and comprehension in SSD.

1.2.1 Two fundamental psycholinguistic models

Spreading activation – Collin and Loftus 1975

According to Collins and Lofts (1975), the organization of conceptual knowledge in the long-term memory is made up of interconnected units of information, or concepts in a semantic network (Figure 1.1). Here, associations are based on personal experience and are not necessarily logical. Links between such units, which varies in length/strength, are at the basis of processes entailing associations and information retrieval: the search process is initiated by the "activation" of a node, which is propagated or "spread" to other nodes. Advantages of this model are that it can explain the priming effect, familiarity effect, the typicality effect, and direct concept-property associations.

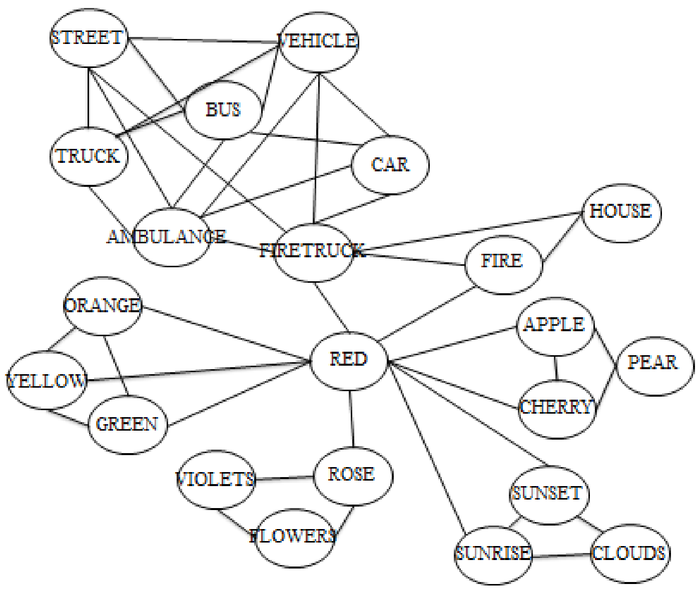


Fig.1.1 Collins and Loftus' (1975) representation of concepts relatedness

Lexical access – Bock and Levelt 1994

According to Bock and Levelt (Bock & Levelt, 1994), speech production entails a sequence of independent, sequential stages. In its latest version (Levelt et al., 1999), after a first step of conceptual preparation, word generation proceeds through lexical selection, morphological and phonological encoding, phonetic encoding and articulation (Figure 1.2). Each stage produces its own output representations: lexical concepts, lemmas, morphemes, phonological words, and phonetic gestural scores (which are executed during articulation). The conceptual preparation stage entails the so-called "verbalization problem": here, both pragmatic (context-dependent) and semantic (network-dependent) causes of activation are taken into consideration for the speaker to get from the notion to be expressed to a message of lexical concepts. To do so, he/she selects a lemma from the mental lexicon. An active lexical concept spreads some of its activation to "its" lemma node[4]: in the case of verbs, these are specified by the verb argument structure and its thematic grid. Upon selection of a lemma, its syntax becomes available for further grammatical encoding, that is, creating the appropriate syntactic environment for the word. After that, the speaker goes from the conceptual/syntactic domain to the phonological/articulatory domain, where the task is to prepare the articulatory gestures: the first step is to retrieve the word's phonological shape from the mental lexicon. Morphological encoding and syllabification take place, and the output morpheme to be produced is ready to be processed phonetically for articulation.

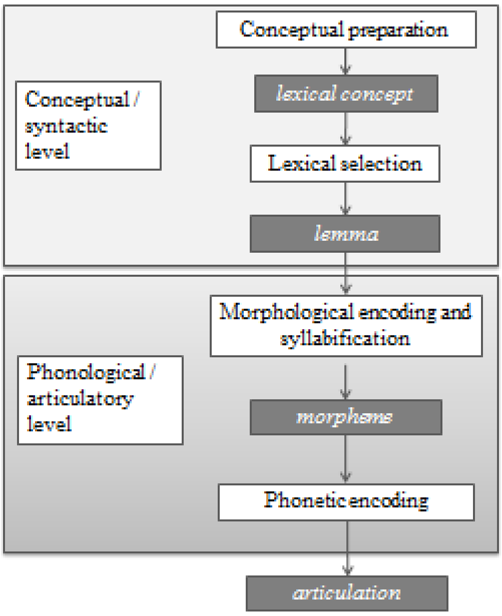


Figure 1.2. The (adapted) model of lexical production proposed by Bock and Levelt (1999).

1.2.2 The linguistic framework

Accessing the lexical form of verbs entails, according to the Bock and Levelt's model of lexical production (1994), retrieving (i) the verb argument structure and (ii) its thematic grid. The notion of argument structure was first adopted by researchers working in the government-binding framework:

[4] According to the Authors, lemma selection is a statistical mechanism, which favors the selection of the highest activated lemma, while the selection of function words takes place on purely syntactic grounds.

according to the generative theoretical framework of Chomsky's Government and Binding Theory (Chomsky, 1981), verb arguments are obligatory elements required by the verb, which specify the number and type of participants in the event described by the verb itself. The minimum number of arguments required is one, as in the case of the grammatical subject of intransitive verbs (1), while transitive verbs require at least two arguments, one subject and a direct object (2). There are also verbs requiring obligatorily three arguments (3), although this categorization is far from straightforward, as in the case of optionally transitive verbs.

1. L'uomo ride

    [$_{NP}$The man [$_{VP}$laughs]]

2. L'uomo versa il succo

    [$_{NP}$The man [$_{VP}$pours $_{NP}$the juice]]

3. La ragazza appoggia un libro sullo scaffale

    [$_{NP}$The girl [$_{VP}$puts $_{NP}$the book $_{PP}$on the shelf]]

Arguments can be either generated outside the verbal phrase (VP), as in the case of the grammatical subject, or inside the VP, as the direct object of transitive verbs. For instance, in the transitive construction (4) the noun phrase (NP) acting as grammatical subject ("the dog") is external, while the NP in the object position ("the cat") is internal.

4. Il cane rincorre il gatto

    [$_{IP}$The dog [$_{VP}$chases $_{NP}$the cat]]

The identification of the direct object (and thus of the transitive subclass of verbs) as a fundamental grammatical category is justified by the realization of passive construction as in (5).

5. Il gatto è rincorso dal cane

    [$_{IP}$The cat [$_{VP}$is chased $_{PP}$by the dog]]

According to the Argument Structure Complexity Hypothesis (ASCH), postulated by Thompson (2003), the more complex is the argument structure of a verb (in terms of number of arguments and the presence of a syntactic movement), the more difficult is producing it. This complexity is reflected in the number of arguments taken by the verb: compared to transitive verbs, which require at least two arguments, intransitive verbs requiring only one argument (as in the case of unergatives, see further) are deemed easier.

Intransitive verbs can be further subcategorized into unergatives and unaccusative, each of which has a distinct underlying syntactic representation (Perlmutter, 1978). According to Chomsky

(Chomsky, 1981), the application of transformational rules to the deep structure (d-structure) of a subject-verb-object (SVO) sentence results in the sentence as it appears to the speaker (surface structure, or s-structure). Active, SVO sentences (e.g. “The dog chases the cat”) are considered as having a canonical word order, whereas for examples passives and object relative clauses show a non-canonical word order. In the case of sentences with a non-canonical word order, the object, generated inside the VP, is moved across the verb and the subject, and surfaces in the clause-initial position via syntactic operation (Fig. 1.3c). According to Chomsky, unergative verbs take a d-structure subject and no object (6), whereas an unaccusative verb takes a d-structure object and no subject (7), which is later moved to the pre-verbal position (see further for a description of canonical and non-canonical word orders). In this latter case, the Unaccusative Hypothesis (Burzio, 1986; Perlmutter, 1978) states that the themes of unaccusative verbs are indeed generated in the direct object position, and are then moved outside the VP to the position of the grammatical subject, implying a syntactic movement similar to that of passive sentences. Unaccusative verbs assign the thematic role in the d-structure, whereas the case is assigned only after the NP syntactic movement to the pre-verbal position in the s-structure. Indeed, in both unaccusative and passive sentences, the NP-theme of the verb moves out of the VP leaving a trace (t) behind. On the contrary, unergative verbs assign the Agent thematic role in the d-structure to the NP in the subject position and, being the nominative case assigned at the same time, there is no need for a further movement.

6. Unergative: [$_{\text{NP}}$The man$_{\text{AGENT}}$ [$_{\text{VP}}$laughs]]

7. Unaccusative: ____ [$_{\text{VP}}$falls $_{\text{NP}}$the baby$_{\text{THEME}}$]]

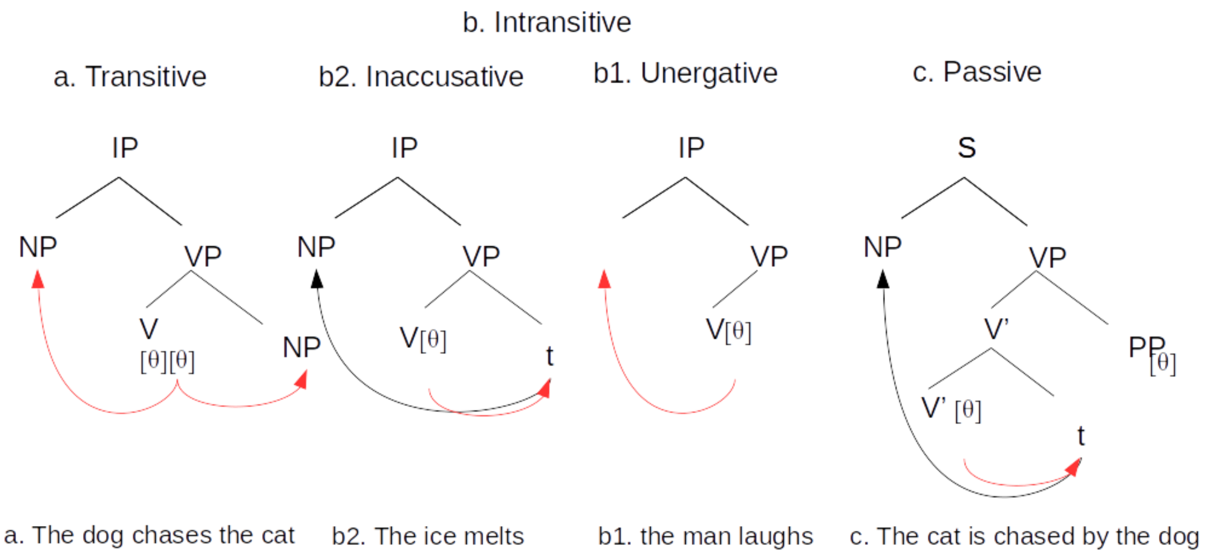


Figure 1.3 Syntactic structures for transitive (a) and intransitive (b) (unergative and unaccusative) verbs, and passive structure (c). Black lines represent NP movement to subject position (leaving a trace behind, "t"); Red lines illustrate theta-marking.

For this reason, an additional measure of complexity postulated by the ASCH is given by the possible presence of a syntactic movement, as in the case of unaccusative verbs.

Argument structure is supposedly stored in the lemma, the cognitive representation of words dealing with its morpho-syntactic features, other than the lexeme, which carries its phonological representation. Arguments are assigned to a specific Thematic Role, which identifies the function each participant plays in the action described by the verb. The Thematic Role knowledge is the understanding of "*who is doing what to whom*" in an action event, and it is typically determined in sentences by the argument structure organized by verbs (Wu et al., 2007). Although a standard definition of the number and nature of Thematic Roles is still to come, according to Dowty (1991), two fundamental Thematic Roles can be identified: "Proto-Agent" and "Proto-Patient", whereby Agent is the participant in an event that causes things to happen, while the Patient is the entity passively involved in the event expressed in the predicate (in this latter category, the thematic role of "Theme" is usually integrated). In active transitive sentences as (4) the NP-subject "the dog" receives the role of Agent since it performs the action; the NP-object "the cat" receives instead the role of Theme, which identifies the participant who undergoes the action. On the other hand, both in passive and unaccusative sentences, the main verb assigns the role of Theme to the NP-subject. A relevant feature of unaccusative verbs relies in their semantic underpinnings, in the sense that unaccusative verbs require a non-agentive argument (Dowty, 1991; van Valin, 1997).

1.2.3 The computational cognitive architecture

In the last few decades, corpus-based models of semantic representations have caught the attention of both computational linguists and cognitive scientists. These models are based on the so-called "distributional hypothesis" (DH) of language, which have been dated back to Wittgenstein's (1953) claim that "*Die Bedeutung eines Wortes liegt in seinem Gebrauch*", or that the meaning of a word is in its use. Advocating for a distributional methodology for the study of language, Harris (1954) contributed to further develop the DH by pointing out the inherent differential nature of the linguistic meaning: from this point of view, difference in meaning shall correlate with difference in distribution. As pointed out by Sahlgren (2008), these approaches are fundamentally descriptive in nature, in the sense that they acquire meaning entirely on the basis of the available linguistic data, and they do not imply any aprioristic assumption on language. If we assume the existence of a formal cognitive mechanism able to learn semantics by exposure to language data in the surrounding environment, DH is not simply an analytical method, but rather a theoretical framework for a computational model of semantic memory.

Stemming from these premises, several Distributional Semantic Models have been successfully applied to different tasks of semantic relationships. Corpus-based semantic representations of words (also called word vectors) exploit statistical properties of textual structure. All these methods rely on the idea that words with similar meanings tend to occur in similar contexts, which is in line with the DH exposed above. These models start from a text corpus and, applying different analytical techniques, end up representing words in a multi-dimensional space as numerical vectors derived from a numerical matrix. In this space, terms with similar meanings tend to be located close to each other. Thanks to the contemporary advancement of computing technologies, it is today possible to extract lexical co-occurrences from (very) large text corpora. Turney and Pantel (2010) offered a thorough survey about the use of vector space models for semantic processing.

The creation of a semantic space entails, in short, two levels of processing: a linguistic processing and a mathematical processing. As the linguistic (pre-)processing of the raw corpus is a mandatory step before applying any kinds of further transformation (which, as we shall see, are peculiar to each different model), I will discuss it here briefly, and it has to be considered a common first preliminary step for different mathematical models.

Linguistic processing involves basically three steps: tokenization, normalization and, possibly, annotation of the corpus. With *tokenization* we define what constitutes a term: for example, we may want to consider a term any strings of characters separated by spaces. *Normalization* involves converting similar strings (such as “Gatto” and “gatto”) to the same form (“gatto”). If relevant to the task at hand, we may eventually want to *annotate* the resulting text, that is, to label different strings (e.g., “andare” and “andato”) in a similar manner (“andare-V”).

Following Baroni and colleagues (Baroni et al., 2014a), I will distinguish two types of models: those that are based on actual counts of word co-occurrences (*count-models*) and models based on predictive neural networks (*predict-models*). For count-models, the mathematical process consists of three steps: generating a matrix of frequencies, adjusting its weights, and reducing its dimensionality. Predict-based models are neural networks that work adjusting weights of the node in the network to predict a word from its context or the context from the word (as the CBOW and skip-gram models of Word2vec – see below).

Creation of a semantic space with a count model

The method used to learn the distributional representation of words also affects the resulting semantic space. Matrix models such as Latent Semantic Analysis (LSA, Landauer & Dumais, 1997), hyperspace analogue of language (HAL, Burgess, 1998), and dependency vectors (DV, Padó &

Lapata, 2007) entails essentially three steps: i) the extraction and counting of co-occurrences between lexical items and the selected contexts (either words, or region of text) from the corpus; ii) the arrangement of such co-occurrences in a matrix, whereas each unique word in the text is a row, and each text passage a column; and iii) the (mathematical) transformation of raw frequencies by means of a weighting function.

A practical example will help illustrate the process. Suppose we extracted the following co-occurrences of targets ("*violino*", "*bicicletta*", "*gatto*", and "*tigre*") with the context lexemes of interest ("*guidare*", "*suonare*", "*mangiare*", "*dormire*", "*prendere*", and "*morire*") from a corpus:

$$
\begin{array}{c}
\\ \textit{violino} \\ \textit{bicicletta} \\ \textit{gatto} \\ \textit{tigre}
\end{array}
\begin{pmatrix}
\textit{guidare} & \textit{suonare} & \textit{mangiare} & \textit{dormire} & \textit{prendere} & \textit{morire} \\
0 & 15 & 0 & 0 & 5 & 0 \\
8 & 2 & 0 & 0 & 8 & 0 \\
0 & 0 & 15 & 15 & 10 & 9 \\
0 & 0 & 13 & 18 & 7 & 12
\end{pmatrix}
$$

The entries of the matrix will be numbers representing the frequency of the word in that context. The resulting matrix will be very *sparse*, in the sense that most of its entries will be zeros, since most words will not appear in a given sample of text.

As the association of some words might be not that informative (think for example at articles preceding nouns), weighting the resulting frequencies aims at enhancing word associations that are particularly informative. For a term-document matrix, such as LSA (Landauer & Dumais, 1997), a popular measure of association is the Positive Pointwise Mutual Information (PPMI) (Niwa & Nitta, 1995), which has been proven effective for the creation of semantic spaces for semantic tasks (Bullinaria & Levy, 2012). PPMI is the log transformation of the ratio between the joint and the individual probability of two words to occur in the text, eventually replacing negative values with 0.

$$PPMI_{(w1,w2)} = max\left(log_2 \frac{P_{(w1,w2)}}{P_{(w1)}P_{(w2)}}, 0\right)$$

In a Matrix M where W is the number of words (rows) and C the number of contexts (columns), we define f_{ij} the number of times a word w_i occurs in the context c_j, hence the formula to compute the joint and individual probability (p_{ij}) values of the matrix will be:

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{W}\sum_{j=1}^{C} f_{ij}}$$

The resulting joint and individual probability matrix of our example above will be:

	guidare	*suonare*	*mangiare*	*dormire*	*prendere*	*morire*	
violino	.00	.12	.00	.00	.10	.00	.15
bicicletta	.07	.02	.00	.00	.17	.00	.13
gatto	.00	.00	.15	.21	.26	.43	.36
tigre	.00	.00	.26	.49	.37	1.00	.36
	.06	.12	.20	.24	.22	.15	

Once all frequency values are re-weighted, a term-frequency matrix is turned into a term-weight matrix, using the logarithmic formula above. Going back to our example matrix, the resulting weighted matrix would be:

	guidare	*suonare*	*mangiare*	*dormire*	*prendere*	*morire*
violino	-	2.68	-	-	1.62	-
bicicletta	3.16	.17	-	-	2.60	-
gatto	-	-	1.05	1.29	1.75	2.97
tigre	-	-	1.80	2.47	2.20	4.16

The next step is the reduction of the space dimensionality. A typical approach in this respect is Singular Value Decomposition (SVD), which reduce statistical noise without losing the geometrical metrics of the matrix. The matrix is decomposed into its three singular values or *eigenvalues* as follows:

$$M_w = U \sum V^T$$

Where U is the orthogonal matrix (word vectors), V^T is the conjugate transposed matrix (document vectors), and $\sum$ the diagonal matrix. The number of dimensions k can be chosen arbitrarily, but it has been shown that choosing about 300 to 1,000 dimensions results in a good performance of LSA (Landauer & Dumais, 1997).

Despite the computations behind count models are well understood, how these can be performed by a human cognitive system is less clear. In LSA models, the reduction step has been considered its defining component: it is this dimensionality reduction step that seems to make LSA able to capture deeper semantic and associative structures (Landauer & Dumais, 1997). In other words, the basic assumption of count models such as LSA is that the mathematical analysis implied in its creation is able to extract deep ("latent") relations among language items. From this process would descend the fitness of its similarity estimates as predictors of human judgments and performance in tasks on meaning of words. Indeed, in their original work, Landauer and Dumais do not claim that the brain, nor the mind, are supposed to compute an exact SVD of a complete sparse matrix. Instead, the Authors suggest that the mind-brain store reprocess the input in a way that has roughly the same effect. It has been argued that counting and weighting in an LSA model might approximate the processes of conditioning or association, and that at the neural level, there might take place a similar dimensionality reduction process (Landauer & Dumais, 1997; Mandera et al., 2017). Nonetheless, LSA models have been proven successfully both in approximating several cognitive phenomena and processes and in providing a theoretical principle of semantic organization (Foltz et al., 1998)

Creation of a semantic space with a predict model (word2vec)

Recent developments in the field of computer science and NLP moved from techniques based on co-occurrence counts to predictive systems based on neural networks. Models based on neural networks are considered more reliable in terms of predictive power (Baroni et al., 2014), but also more sound from a cognitive perspective (Hollis & Westbury, 2016; Mandera et al., 2017). Of relevance is also the fact that they are computationally more effective, being able to process large corpora of text in just a few hours, in contrast to previous models that may need weeks to train (Berardi et al., 2015).

The elementary unit of a neural network is the so-called "artificial neurons". An artificial neuron is basically a mathematical function that receives some input values and produces an output value. Each input is differentially weighted, and the sum is forwarded to a function (f) before being outputted. A neural network is hence a network of artificial neurons (also called nodes). One of the most popular neural networks for natural language processing is word2vec (Mikolov et al., 2013). Basically, this kind of neural networks consists of an input, a hidden, and an output layer: each word of the corpus corresponds to a node in the input layer, while the number of hidden layers is a parameter that must be set arbitrarily. Word representations obtained with predict models are generally low-dimension dense vectors, having a dimensionality that ranges typically from 20 to 500 (Günther et al., 2015).

Word2vec implements two different prediction models: the continuous bag-of-words, or CBOW model, and the skip-gram model (Fig. 1.4).

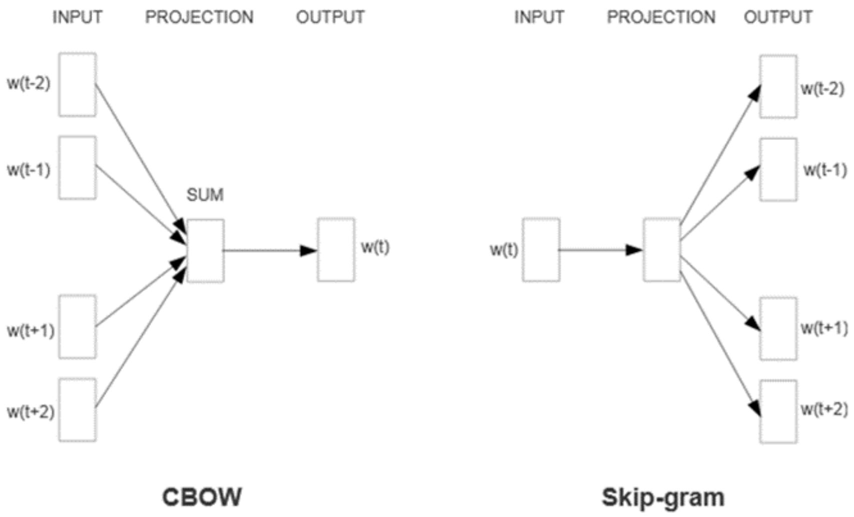


Figure 1.4 Word2vec's architecture. The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word.

In both cases the algorithm uses a window of predefined length that slides along the text corpus[5]. By changing the weights within the network, the model learns to predict the word in the center of the window from the surrounding words (CBOW model), or to predict the context based on the central word (skip-gram model). The mathematical function (*f*) implemented in word2vec is the hierarchical softmax function, optimized by stochastic gradient descent and back-propagated to all the nodes of the network (see Rong, 2016 for a detailed explanation of word2vec learning parameters). Once the training is completed, the weights between the input layer and the hidden nodes are exported as the resulting word vectors (or word-embeddings). The number of dimensions of the resulting semantic space is related to the number of nodes considered in the hidden layer.

Compared to count models, the psychological grounding of predict models appear to be sounder. In fact, predict models operate following an implicit learning principle, and the learning process takes place incrementally over a number of subsequent attempts (i.e., the neural network does not need to have all the information available before applying the transformation): in fact, these models learn to associate a cue to an outcome and, if the predicted outcome (i.e., the outcome based on the current data) does not match the desired one, to update the weight. As such, predict models have been linked to the Rescorla-Wagner model of learning (1972): formally, this latter corresponds to the delta rule that describes the stochastic gradient descend in a neural network with a single layer, which has been later expanded to multiple-layer networks with nonlinear back-propagation functions.

[5] The term "bag of words" refers to the notion that the CBOW algorithm uses a continuous distribution representation of the context, in the sense that the order of previous words does not influence the projection (i.e. the non-linear hidden layer is removed and the projection layer is shared for all words, thus averaging their vectors).

2 Distributional semantic models applied to a semantic verbal fluency task in people with Schizophrenia Spectrum Disorders

2.1 Introduction

The occurrence of impairment in semantic VF tasks in SSD patients has been repeatedly reported in the literature.

As with other neuropsychological test, semantic VF tests rely on multiple cognitive functions. At least two main interacting neural systems are supposed to be involved: the semantic store system, supported by the left temporal lobe, and the executive control system, supported by pre-frontal regions. The "total word" measure (i.e., considering only the total number of words produced according to the instructions given) do not distinguish to what degree the two systems are compromised. As proposed by Troyer and colleagues (1997), a valuable measures of semantic store sanity is the size of semantic clusters (consecutive words arranged in groups that share similar properties), while a measure of the executive system performance is the number of switches (total number of transitions between clusters) (see Chapter 1).

Previous studies have shown that people with SSD generate fewer words than HPs in VF tasks (Bokat & Goldberg, 2003); however, studies applying scoring methods taking into consideration number and size of clusters are inconclusive (see Chapter 1). Analyzing patterns of word meaning produced during SVF test, especially in conditions characterized by language disturbances such as SSD, by means of automated procedures can offer a valuable diagnostic (de Boer et al., 2018) and prognostic (Corcoran et al., 2018) tool for researchers and clinicians.

However, analyses of semantic patterns and production coherence are rarely performed in clinical settings. Moreover, manual scoring following the approach proposed by Troyer and colleagues (1997) bears some important limitation, i.e., the rigidity of the taxonomic approach paired with an intrinsic subjectivity of the scoring methods, and the time constraint factor. Recent advances of NLP techniques have already provided researchers with a wide range of powerful tools that may come into help to overcome the actual limitations. Below, I explore each of these limitations in details, and I will suggest how NLP may come to help in overcoming them. I will also present the application of a measure of semantic coherence (Elvevåg et al., 2007) to compute coherence in SVF.

Perhaps the most important limitation of Troyer's approach is that it is taxonomic in nature (Paula et al., 2018). This means that it requires the application of well-defined criteria to recognize a cluster: the human rater reads the words one by one and compares them to the content of pre-determined categories (i.e., living environment, human use, and zoological categories) and their subcategories. If, reading through the list, the rater encounters a word not belonging to the category of the previous ones, she/he marks for a switch and re-starts the process. Two scores are eventually computed: the total number of switches and the mean size of clusters (number of words within all clusters divided by the overall number of clusters)[6]. This approach, in fact, excludes all other kinds of associations, in that subjects might want to use a different strategy to group words: for example, "cat" and "mouse" are by no mean to be considered as forming a cluster by Troyer's instructions, as they belong to two distinct subcategories (namely "pets" for "human use" and "rodent" in "zoological category", respectively). This, however, leaves out a very simple connection between these two words (for example, they are two very popular fictional cartoon characters, who often appears together; also, they often appear together in multi-word expressions and idioms). Strictly related to this, another limitation is that, in case of ambiguity (i.e., when one word can fit into two subcategories), it's up to the rater to decide either to include it as coherent or not coherent with the category under investigation. This limitation has already been identified by previous studies (Pauselli et al., 2018). In this sense, the application of measures of semantic relatedness, derived by calculating proximity between word vectors of LSA semantic spaces (see Chapter 1), have already proven their usefulness in clinical

[6] Troyer proposed the "number of clusters" as an additional measure of mental flexibility: however, this measure is linearly correlated with the number of switches (Farzanfar et al., 2018), and for this reason we have decided to not consider it in the present study.

studies, for example on Parkinson's disease (Farzanfar et al., 2018), formal thought disorder in schizophrenia (Holshausen et al., 2014), Alzheimer's Disease (Pakhomov & Hemmy, 2014), as well as to identify subtle brain damages in healthy subject (Ryan et al., 2013). Finally, a major setback of Troyer's method is that a human rater must carry it out manually. This approach is time consuming, making it unlikely to be deployed in everyday clinical settings outside of controlled research studies, where clinical burden is already heavy. Scoring a verbal fluency test requires few minutes for each subject, while, on the contrary, an automated procedure would be incomparably faster, requiring just a few seconds to compute the results of tens of subjects.

As application of NLP techniques to clinical data is still at a relatively early stage, there lacks a standard approach to automatic analysis of SVF for clinical use in psychiatry and neurology. A recent systematic review and meta-analysis by de Boer and colleagues (de Boer et al., 2018) pointed out that the handful of studies published so far adopted different solutions, mostly pragmatic in nature, with the ultimate aim of identifying the best approach to get the designated objective (e.g., which type of processing works best to categorize subjects). Furthermore, the terminology applied across studies is still far from homogeneous: terms such as "similarity", "coherence", and "relatedness" have been used to define different and sometimes overlapping word-similarity measures (Goodkind et al., 2018; Pakhomov et al., 2012; Pauselli et al., 2018) applied to different types of linguistic data, ranging from single-word association tasks (i.e., verbal fluency task and word association tasks) to narratives.

2.1.1 Objectives and hypothesis

The objective of the present study is threefold: i) develop different automatic procedures to analyze verbal fluency data able to compute semantic clusters and to analyze patterns of semantic relatedness; ii) apply these algorithms to a SVF task administered to a sample of people with SSD and a matched group of people without psychiatric disorders; iii) compare the ability of the automatic procedures to categorize participants, both comparing each automatic procedures against each other, as well as against human-derived measures.

According to previous results, showing the superiority of prediction-based models over count-based models (see Chapter 1) in psycholinguistics tasks (Baroni et al., 2014), we hypothesize that categorization of participants using automatic measures of semantic relatedness extracted from the computational models obtained with prediction-based models will outperform categorization based on values obtained from count-based models: more specifically, given that different hyperparameters applied to the predict model are known to affect the performance of the model (Baroni et al., 2014), we foresee that a larger vector dimensionality will perform better at predicting spontaneous word association than a narrower vector dimensionality, and that both models will outperform the count-based model. Such performance will contribute to a fit categorization of the two experimental groups of participants.

Given the results of previous studies pointing out lower coherence in the production of people with SSD (Elvevåg et al., 2007), we expect healthy participants to produce words that are on average more correlated between each other, than people with SSD. In other words, we expect to find lower mean value of coherence between adjacent words in the verbal production of people with SSD compared to the production of healthy participants. Again, we expect measures of semantic relatedness derived from predict-models to outperform those derived from count-based models in the categorization of subjects based on such values of coherence.

Also, given the limitations of taxonomy-based approaches outlined above, we predict that computational-based scores will outperform the traditional manual approach in categorizing participants in the two Groups.

Used individually and in combination, these measures might help us grasping different aspects of "coherence" in schizophrenia (and possibly, providing an approach to other disorders in which there are language deviances). Figure 2.1 depicts the flowchart of the study, which will serve as reference in the following discussion.

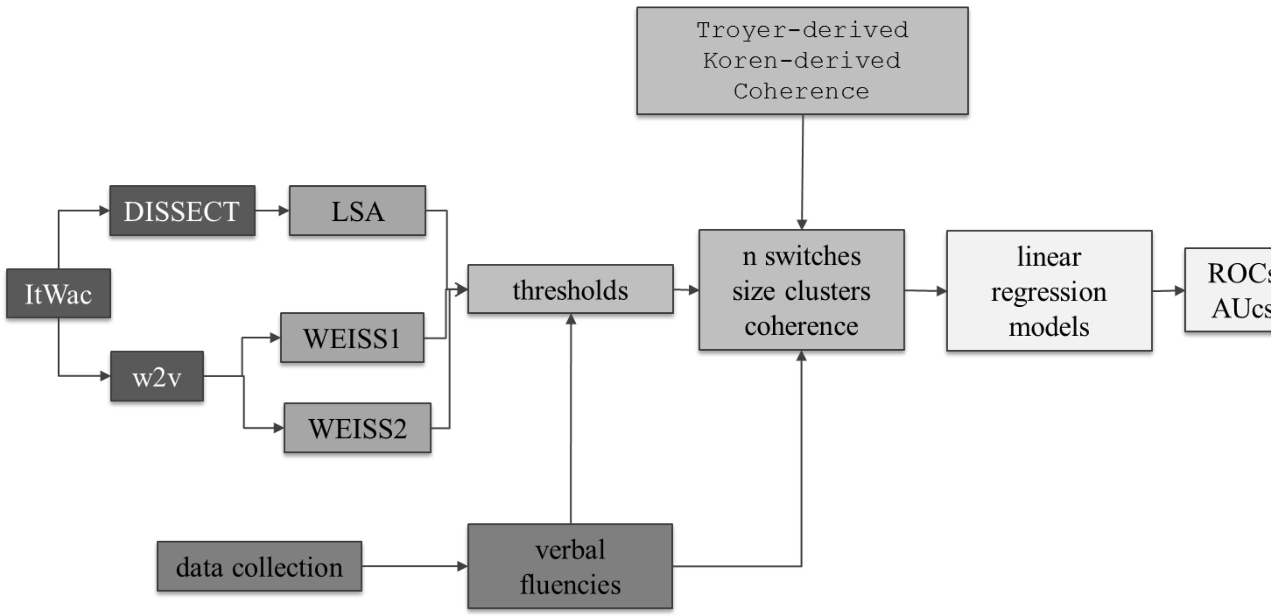


Figure 2.1. Flowchart of the study

2.2 Materials and methods

2.2.1 Participants

Thirty-seven persons with a diagnosis of SSD according to DSM-5 (APA, 2013) were recruited. Participants were recruited from the outpatients' service and the residential facilities of the IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, between February 2018 and April 2019. Diagnoses were made by the treating clinicians (staff psychiatrists). Participants were aged 18-65, able to give informed consent, right-handed, and had normal or correct-to-normal visual acuity. The diagnosis on Axis I had to be unique, but co-morbidities on Axis II were admitted. Exclusion criteria were: additional neurological disorders, head trauma with cognitive sequalae, mental retardation, substance abuse in the 3 months preceding the enrollment. The final sample of SSD participants (24 males and 9 females) had a mean age of 49 years (SD = 9.24). At the time of recruitment, SSD participants had been on treatment with at least one anti-psychotic medication (M = 2.49, SD = 1.57) for at least the previous 6 months. The mean length of illness in the SSD group was 23.58 (SD = 11.85, N = 36 – for one SSD participant, it was not possible to date the age of onset).

Thirty-seven healthy volunteers, matched by age-class (18-30, 31-40, 41-50, and 51-65) and gender, were recruited among the hospital staff and through public announcements also in Brescia. Level of education was not matched at recruitment between the two groups (W = 949.5, $p < .01$). The exclusion criteria for the control group were: any documented psychiatric disorders or being first-degree relative of a patient with diagnosis of SSD.

All subjects were native speakers of Italian.

After a complete description of the study, informed consent to participation was obtained from all subjects. In case of patients with support administration, the participation to the study was first discussed with the patient; then, written consent was obtained both from the patient and the appointed administrator. The study was approved by the IRCCS Ethical Committee (Opinion 61/2017) and followed the principles of the Helsinki Declaration.

2.2.2 Tasks

As part of a wider assessment battery, the semantic VF task contained in the BACS battery (Keefe et al., 2008), adapted for the Italian population (Anselmetti et al., 2008), was administered to all participants. During this task, participants were given 60 seconds to produce as many words as possible belonging to the given category ("animals"). All sessions were audio recorded and later transcribed by a training psychologist. The transcripts were converted to plain ASCII text to make them machine-readable, and hand-edited to enforce standard spelling.

2.2.3 Description of the experimental (independent) variables

Manual scoring - baseline

The SVF test was manually scored by a training psychologist not blind to the experimental design, following both Troyer et al. (1997) scoring approach and the revised approach proposed by Koren and colleagues (2005) for assessing the mean size of clusters and number of switches.

Numerical representation of word meaning (word vectors)

Automatizing a method to compute measure of semantic relatedness in the context of a verbal fluency test means to devise a computer algorithm (a set of instructions in the form of a machine-readable code) able to "understand" language and to recognize words that have similar meaning. In order to do so, two basic issues need to be solved: i) provide the algorithm with a numerical representation of word meaning; and ii) provide the algorithm with an unambiguous rule that defines occurrence of a semantic shift (that is, a topic shifts). I will present both issues in details in the "Materials and Methods" section.

In this study, the numerical representations of word meanings, needed to compute relatedness measures, were derived from both predict and count models. Two of them (Word-Embeddings Italian Semantic Space 1 and 2 - later labeled as "WEISS1" and "WEISS2") were created by Marelli (2017) with the CBOW method from word2vec (Mikolov et al., 2013) and one (hence "LSA") was created ad-hoc for this study applying the DISSECT toolkit (Dinu et al., 2013). All three spaces were based on the itWac corpus (Baroni et al., 2009), the largest available Italian text corpus (http://wacky.sslmit.unibo.it), built through web crawling and consisting of about 1.9 billion tokens. The raw (untagged) itWac corpus was converted to a standard coding system (UTF8), tokenized and eventually converted to lower case. All special characters (except for vowels with orthographically marked stresses - à, è, é, ì, ò, and ù) were previously removed from the corpus. The output files were saved as a plain text file and queried with R on a local machine.

Word vectors from a predict model

WEISS1 (http://meshugga.ugent.be/snaut-italian), is based on a CBOW model with 400 dimensions and a 9-word window. WEISS2 (http://meshugga.ugent.be/snaut-italian-2/) is based on a CBOW model with 200 dimensions and a 5-word window. Both consider words with a minimum frequency of 100.

Word vectors from a count model

The third set of word vectors were derived from an LSA space. To do so, we followed a three-step pipeline: i) extraction of occurrence counts from the corpus; ii) build-up of a word-by-document matrix; and iii) transformation of the raw counts.

The occurrence counts were extracted from the itWac corpus and outputted as .xml file containing two strings (the target and the document in which the word appears) and a number (the corresponding count) on each line, following the procedure proposed by Günther and colleagues (2015). However, the heavy computational load required for the dimensionality reduction step would have prevented its application on the entire itWac corpus (1.9 billion token). For this reason, considering that most psycholinguistic studies for the English language made use of far smaller corpora, such as the TASA corpus[7] (http://lsa.colorado.edu), we extracted a subset of the itWac corpus so to match these settings: an untagged set of 91,058 documents, comprising 180,080 words (selected from the word2vec vocabulary list), was randomly extracted from itWac. This was done, on the one hand, to ensure comparability with the previous literature (mostly based on the TASA corpus, including 12,190,931 tokens) and, on the other hand, for practical reasons, since computing a LSA space on a large corpus requires a substantial amount of computational resources, which many research facilities can not necessarily afford nor maintain.

Step two and three were carried out by means of the DISSECT toolkit (Dinu et al., 2013). A Positive Pointwise Mutual Information (PMI) weighting scheme was applied. Dimensionality reduction was obtained by means of Singular Value Decomposition (SVD). The number of dimensions was chosen to be 300, following the original paper of Landauer and Dumais (1997), which indicates a reduction ranging from 1,000 to 300 dimensions for good performance.

How to recognize a semantic switch

Instructing an algorithm to recognize a semantic switch means, in short, to implement a binary function that operates on the sequence of words (w_1, w_2, …, w_n) produced by a subject: there is a switch when the similarity between two words (or a cluster and a word) falls below a specific threshold. But, how to choose such threshold value? We decided to consider the distribution of semantic relatedness values in our sample and to compute the values of cosine proximity between all adjacent words (N = 1,049) produced by the study cohort (N = 74) and to select the 10th, 30th, 50th, 70th, and 90th quantiles of that distribution as thresholds. For each semantic space, five thresholds were established. Switches were identified applying only thresholds specific for that semantic space (numerical values are reported in Table 2.1).

[7] The TASA corpus is made up of 44,486 documents with 98,646 unique terms (approximately corresponding to the amount and type of general reading that an average US student would be exposed to by the first year of university).

	Thresholds				
Semantic space	10th	30th	50th	70th	90th
LSA	.25	.31	.34	.37	.43
WEISS1	.20	.21	.22	.23	.26
WEISS2	.32	.34	.36	.37	.40

Table 2.1. Cosine values adopted as thresholds for the three semantic spaces to be applied at the semantic VF task.

Compute clusters: "Troyer-derived" and "Koren-derived" functions

We developed two R-based functions to compute number of switches and mean size of clusters from VF tasks following Troyer's and Koren's proposed methodologies. The annotated scripts of the "Troyer-derived" and "Koren-derived" functions are reported in Annex 1.

Both functions take three arguments: a vector containing the list of words produced by the subject, a numerical value indicating the threshold, and a matrix of word vectors. The only feature distinguishing the functions is the minimum cluster size, which is equal to 1 in the first case (as in Troyer's method) and to 2 in the second case (following Koren's approach).

In short, each function retrieves from the semantic space the word vector corresponding to the first word produced by the subject and calculates a measure of semantic relatedness (cosine proximity) between it (w_n) and the following word (w_{n+1}). The function then compares this value to the pre-specified threshold and, if found equal or above such value, the two words are considered as part of a cluster. In this case, a mean vector is created by averaging the two vectors, representing the cluster meaning. The function then compares the next word in the list against this mean vector, and repeats the last two steps (computing mean vector and comparing it to the next word) until the cosine proximity to the next word eventually falls below the threshold: in this case, the function registers a switch, and re-starts the process with a new cluster. As a result, the function outputs two variables of interest: number of switches, and mean size of clusters.

Let's make a practical example. Suppose we have an SVF output as follows: "elefante, picchio, giraffa, rinoceronte, scimpanzè, aquila, gallina, gatto, leone"[8] and we want to analyze it with our function using WEISS1 as semantic space of reference and setting a threshold equal to the 30th quantile (.21) applying Troyer's methodology (i.e., considering single words as a cluster, that is, having size = 1). The algorithm starts computing the cosine similarity between "elefante" and "picchio", giving back a cosine similarity of .12. As this value is under the specified threshold, the function registers a switch and moves on to compare the similarity between "picchio" and "giraffa".

[8]Elephant, woodpecker, giraffe, rhino, chimpanzee, eagle, hen, cat, lion.

The outputted value is “.14” - again under the threshold: the function registers a second switch and moves on to compare “giraffa” with “rinoceronte”. Now the cosine similarity is = .29 – above the threshold: this means that we are within a cluster. In this case, the function carries on and computes the mean vector of these two words; then it moves on to compare this “cluster vector” to the next word vector, which is “scimpanzé”. The result is again above the threshold (.32): we are still within the same cluster, so no switch is registered; rather, a new cluster vector is calculated, which will now include “giraffa”, “rinoceronte”, and “scimpanzè”. Our algorithm will now move on to the next word, “aquila”. The cosine value between the mean vector of the previous cluster of words (comprising “giraffa”, “rinoceronte”, and “scimpanzé”) and “aquila” now drops at = .19, below the defined threshold. The algorithm “closes” the current cluster, registers a shift, and starts considering a new cluster from “aquila” onwards. “Aquila” is now the first word of a possible new cluster: the algorithm computes the cosine values with the following word (“gallina”) and outputs = .19, which is under the threshold – a new shift is registered. The algorithm moves on. The cosine proximity between this word (“gallina”) and the next one (“gatto”) is computed: the output is .19, again under the threshold, and hence an additional shift is added to the counting. Eventually, the algorithm computes the cosine proximity between “gatto” and the last word (“leone”): the result is again under the threshold (= .13) and an additional shift is added to the overall count. Figure 2.2 illustrate these passages.

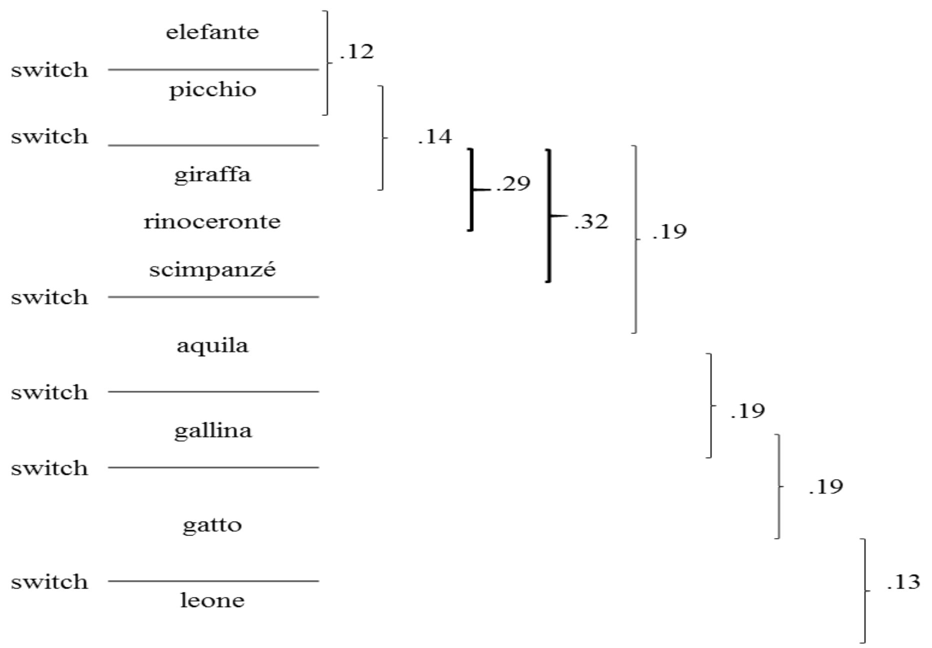


Figure 2.2 The “Troyer-derived” algorithm computes the cosine values of adjacent words in order to identify semantic clusters and switches, according to pre-set values.

To summarize, the final output of the function for this example will be: six switches (i.e., seven clusters) and mean size of cluster = (1+1+3+1+1+1+1)/7 = 1.286.

Differently, in the computation of mean size of clusters, the Koren-based algorithm would only consider those cluster having at least two words in it. In this case, we would have 1 cluster with size 3, and hence = 3/1 = 3.

The adopted definition of a semantic cluster (i.e., defined by the mean vector of all its words) does in fact preserve local associations. Let's make another practical example to clarify this point. Suppose we have two persons producing the same set of words, but in different order, such as in (1-a) and (1-b):

(1-a) cane gatto agnello avvoltoio abbaiare

(1-b) avvoltoio agnello gatto cane abbaiare

By using the same setting of the previous example, the "Troyer-derived" algorithm would identify in 1-a 3 switches occurring between 4 semantic clusters (namely: "cane + gatto", "agnello", "avvoltoio", "abbaiare"): the resulting mean cluster size would then be = (2+1+1+1)/4 = 1.25. In 1-b, the same algorithm would identify 2 switches occurring between 3 clusters (namely: "avvoltoio", "agnello", "gatto + cane + abbaiare"), with mean cluster size equal to (1+1+3)/3 = 1.67. Hence, the local association of "cane" with "abbaiare" is correctly preserved in 1-b.

Measure of overall coherence

As a measure of overall coherence, we calculated a measure of semantic relatedness between words having 1-, 3-, 5-, and 7-word in-list distance between each other (Fig. 2.3). In order to do so, we developed another R function (the "Coherence" function, see Appendix 1 for the annotated script), which takes three arguments: a vector containing the list of words produced by the subject, a numeric value, and a matrix of word vectors. In short, the function retrieves from the semantic space the word vectors corresponding to the words produced by the subject and calculates a measure of semantic relatedness (cosine similarity) between them. Let suppose that the list of responses is made of 10 words (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}): according to the specified target, the function will calculate the cosine between w_1 and w_2 (target = 1), between w_1 and w_3 (target = 3), between w_1 and w_5 (target = 5), and between w_1 and w_7 (target = 7), iterating the computation for each succeeding words in the list. As a result, the function outputs the mean cosine similarity between all comparisons that it was possible to compute (the limit being given by the actual length of the list).

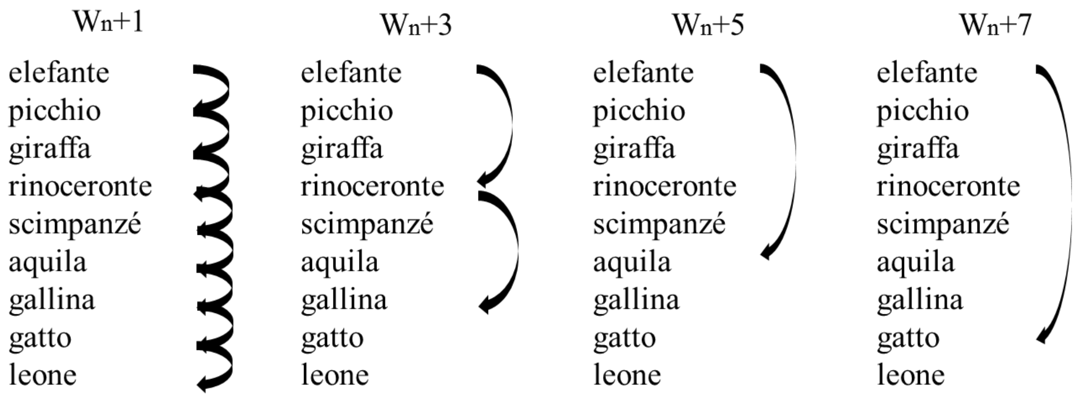


Fig.2.3 The "coherence" function computes cosine similarity measures of words at different distances in the list of words produced by the participant. The figure illustrates which words are considered for the four possible different distances considered (1, 3, 5, and 7 items, respectively).

Standard quantitative measures

As a quality check of the data collected, we adopted a set of standard quantitative measures, and namely: (i) number of words; (ii) number of repetitions; (iii) number of unique words (equal to the total number of items produced, minus repetitions and non-words; words not included in the semantic spaces were also removed); (iv) mean length of words; and (v) mean frequency of the words produced. Frequencies of words were taken from Subtlex-IT (http://crr.ugent.be/subtlex-it/; Crepaldi et al. 2013), a set of frequency norms for the Italian language based on movie subtitles. The corpus size is around 130 million words, with a frequency range of .01 to 17854 per million.

Traditional scoring

The "traditional" VF scoring was calculated according to the total number of coherent items produced within a given time frame. Following Troyer's methodology (1997), "mean cluster size" was calculated considering the mean number of words in a semantic cluster, minus 1 (errors and repetition included). Clusters were created taking as reference the categories (and subcategories) proposed in Troyer's original. Moreover, "number of switches" was computed as the total number of transitions between clusters, including single words. Furthermore, following Koren's revised procedure (2005), additional scores for "number of switches" and "size of clusters" were counted considering clusters as 2 or more related words, leaving single words aside.

Automatic scoring

The proposed algorithms output:

1. Number of switches and mean size of clusters computed by the algorithm and based on measures of semantic relatedness obtained from an LSA space;
2. Number of switches and mean size of clusters computed by the algorithm and based on

measures of semantic relatedness obtained from two semantic spaces created with a predict-based model (WEISS1 and WEISS2), differing in co-occurrence window and vector dimensionality;

3. Mean cosine similarity between words at 1, 3, 5, and 7 of word-distance.

These indexes will be used as independent variables to predict the two Groups (SSD participants and HPs).

2.2.4 Data analyses

First, results of both the manual and automated procedures (based on WEISS1, WEISS2, and LSA) used to compute mean number of switches and mean cluster size, as well as coherence measures, were entered as continuous dependent variables in generalized linear regression models, having group membership (pre-identified by the clinical staff) as categorical predictor. Moreover, albeit known to have minimal to small effect sizes as a predictor of the mean size of clusters in animal fluency tests (Troyer, 2003), number of years of formal education was included as covariate in the regression models, given the proven effect of education on verbal fluency tasks (Novelli et al., 1986; Spinnler & Tognoni, 1987). Given that the results of the regression models indicated that *Education* was not a significant predictor for any (N = 96) of the verbal fluency indexes under investigation (except "Means size of cluster" calculated according to Koren's method and having set the threshold at the 30th quantile within the WEISS2 semantic space: in this case, Education: t = 2.181, p <.05), and that we do not hypothesize a differential effect of this variable in the two Groups, it was excluded from the ROC analyses.

Moreover, we created an additional set of classifiers by applying a logistic regression model having group membership (pre-identified by the clinical staff) as categorical dependent variable and fluency indexes (number of switches*means size of clusters, and the best predictor of the first step with cosine similarity between words at different distances) calculated manually and based on the three semantic spaces (WEISS1, WEISS2, and LSA) as continuous independent variables.

Second, to test the predictive ability of the different models (manual scoring, WEISS1, WEISS2, and LSA) to categorize subjects, we compared Area Under the Curve (AUC) values derived from Receiver Operating Characteristics (ROC) graphs. ROC graph is "*a technique for visualizing, organizing and selecting classifiers based on their performance*" (Fawcett, 2006). As predicting a categorical response (group membership) from an observation (verbal fluency value) can be considered a classification process (James et al., 2013), our logistic regression can be considered, in fact, classifiers. On a ROC graph, the rate of true positive is plotted on the Y axis, while the rate of false positive on the X axis: the resulting curve is the relative tradeoff between the cost of the classifier (any time a negative instance is classified as positive, or the rate of false positive) and its benefit (any time a positive instance is classified as such, or true positive). In this sense, a ROC graph depicts the ability of a classifier to rank the positive instances relative to the negative ones, whereby the diagonal line is the strategy of randomly guessing the class. The AUC is the portion of the area of the unit square of the ROC graph, and hence its value will be 0 < AUC < 1. However, to exclude unrealistic classifiers, AUC < .5 are generally disregarded. In short, the greater the area, the better is the performance of the classifiers. ROC curves were estimated for fitted values of the logistic regression against pre-identified group memberships, for all the considered thresholds and semantic spaces. We then compared our models' performances, with AUC = 1 indicating perfect accuracy. All data were analyzed using the R software V3.6 (R CoreTeam, 2019).

2.3 Results

2.3.1 Verbal fluency data

The final dataset used for the analysis consists of 1,049 words (284 unique words), as produced by the cohort of 74 subjects described above.

Standard Quantitative measures

Table 2.2 reports mean values, standard deviations, as well as estimated partial coefficients for *Group* of the SVF standard quantitative values.

	HPs		SSD Participants						
index	**M**	**SD**	**M**	**SD**	**estimate**	**Std.Error**	**t value**	**p value**	**sign**
Number of items	22.11	4.7	17.03	4.37	-4.49	1.11	-4.04	<.001	***
Unique words	21.86	4.56	16.43	4.29	-4.84	1.08	-4.47	<.001	***
Repetitions	.24	.55	.59	.93	.35	.19	1.84	.070	
Mean word-length	6.47	.29	6.39	.44	-.13	.09	-1.41	.16	
Mean frequency	1,586.72	314.92	1,927.17	681.98	328.06	132.25	2.48	.015	*
Mean frequency (log-transformed)	6.14	.36	6.43	.43	.28	.10	2.79	<.01	**

Table 2.2 Mean, SD, and estimated coefficients for *Group* of the SVF standard quantitative measures.

Once having partialled out the effect of *Education* (which was not significant for any of the variable under investigation), the results show that, on average, people with SSD produced significantly (t = 4.04, $p < .001$) fewer words (M = 17.03, SD = 4.37) than HPs (M = 22.11, SD = 4.7). After subtracting the number of repetitions, non-words, and words outside the SS, people with

SSD still produced significantly less (t = 4.47, p < .001) unique words (M = 16.43, SD = 4.29) than HPs (M = 21.86, SD = 4.56). With respect to the mean frequency of words produced, people with SSD used words that are significantly (t = 2.79, p <.01) more frequent (M = 1,927.17; SD = 681.98) than HPs (M = 1,586.72; SD = 314.92). No significant differences were observed in the mean length of words (p = .16), and number of repetitions (p = .070).

Results of the manual scoring procedure

Table 2.3 reports mean values, standard deviations, as well as estimated partial coefficients for *Group* of the manual scoring assessment of the SVF test.

	HPs		SSD Participants						
Index	**Mean**	**SD**	**Mean**	**SD**	**Estimate**	**Std.error**	**T value**	**p.value**	**Sign.**
Troyer									
Number of switches	9.84	2.54	8.5	2.82	-1.1	.66	-1.66	.102	
Size of clusters	1.18	.72	1	.44	-.17	.15	-1.14	.259	
Koren									
Number of clusters	5.65	1.69	4.77	1.47	-.78	.39	-1.99	.051	
Size of clusters	2.16	.9	1.9	.7	-.26	.2	-1.27	.207	

Table 2.3 Mean, SD, and estimated coefficients for Group of the SVF manual scoring.

Once having partialled out the effect of *Education* (which was not significant for any of the variable under investigation), no significant differences were found in the number of switches and mean cluster size neither according to the Troyer's method (p = .102 and p = .259, respectively) nor following Koren's approach (p = .051 and p = .207, respectively).

Results of the automated scoring procedure

Table 2.4 and 2.5 report means, standard deviations, and estimated coefficients for *Group* for the two indexes of verbal fluency (number of switches, and mean size of cluster) calculated by means of the Troyer-derived procedure, applied to the SVF task.

Troyer – Number of switches

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	3.51	3.44	2.7	3.01	-.73	.81	-.9	.37	
30th	4.19	3.78	3.32	3.38	-.87	.89	-.97	.333	
50th	5.7	3.76	4.22	3.84	-1.37	.95	-1.45	.153	
70th	8.03	3.88	5.89	4.07	-1.94	.99	-1.96	.054	
90th	11.16	4.22	7.84	3.66	-3	.98	-3.07	.003	**
WEISS2									
10th	3.65	3.7	2.14	2.34	-1.25	.76	-1.63	.107	
30th	4.49	3.79	3.08	2.77	-1.18	.82	-1.43	.157	
50th	5.24	4.17	3.7	3.35	-1.09	.93	-1.17	.244	
70th	6.3	4.36	4.22	3.61	-1.94	1	-1.94	.055	
90th	9.54	4.51	6.89	3.61	-2.4	1.01	-2.37	.020	*
LSA									
10th	6.27	3.63	3.76	3.21	-2.2	.85	-2.6	.012	*
30th	7.7	4.05	5.24	3.48	-2.15	.94	-2.3	.024	*
50th	8.84	3.75	6.27	3.88	-2.43	.95	-2.56	.012	*
70th	10.11	3.86	7.35	4.13	-2.4	.99	-2.43	.018	*
90th	11.7	4.01	8.89	4.42	-2.46	1.04	-2.35	.021	*

Table 2.4. Mean, SD, and estimated coefficients for *Group* of number of switches computed by the Troyer-derived computational algorithm.

Once having partialled out the effect of *Education* (which was not significant for any of the variable under investigation), the results of regression model on the number of switches calculated using WEISS1 indicated significant differences at the 90th quintile, whereby people with SSD produced on average less switches than healthy volunteers (t = -3.07, p <.01).

Using WEISS2 and having set a threshold at the 90th quintile, we found a significant difference (t = -2.37, p <.05) between people with and without SSD, whereby healthy control participants produced on average more switches than SSD participants.

Using the LSA-derived semantic space, significant differences in the mean number of produced switches were found for all the thresholds (10th, 30th, 50th, 70th, and 90th quintiles), with people with SSD producing always fewer switches than healthy control participants (p <.05).

Troyer – Size of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	9.73	8.34	7.95	6.23	-.77	1.8	-.43	.669	
30th	8.3	7.53	7.32	6.4	-.23	1.72	-.13	.895	
50th	5.33	5.21	6.11	6.14	1.48	1.4	1.06	.293	
70th	3.19	2.48	4.08	3.73	.96	.79	1.21	.228	
90th	2.03	.78	2.45	2	.48	.38	1.27	.207	
WEISS2									
10th	9.72	8.31	8.13	4.73	-1.83	1.68	-1.09	.279	
30th	7.56	7.13	6.67	4.77	-1.12	1.51	-.75	.458	
50th	6.69	6.48	6.2	4.9	-1.09	1.42	-.77	.443	
70th	5.4	5.69	5.43	4.36	-.1	1.26	-.08	.934	
90th	2.63	1.8	2.94	2.89	.22	.6	.38	.708	
LSA									
10th	4.67	4.81	6.11	5.16	.69	1.21	.57	.572	
30th	3.69	3.89	4	3.12	-.08	.87	-.09	.927	
50th	2.58	1.11	3.32	2.67	.71	.51	1.4	.167	
70th	2.15	.62	2.55	1.52	.36	.29	1.26	.212	
90th	1.82	.38	1.98	.93	.19	.18	1.1	.275	

Table 2.5 Mean, SD, and estimated coefficients for Group of mean cluster size computed by the Troyer-derived computational algorithm.

Once having partialled out the effect of *Education* (which was not significant for any of the variable under investigation), no significant difference was found in the mean size of clusters of the two Groups calculated by Troyer-derived and using any of the three semantic spaces.

Table 2.6 and 2.7 report the estimated fitted values for *Group* for the two indexes of verbal fluency (number of switch/cluster, and mean size of cluster) calculated by means of the Koren-derived procedure, applied to the SVF task.

Koren – Number of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	1.81	1.65	0.86	1.18	-.93	.36	-2.61	.011	*
30th	2.38	1.91	1.08	1.21	-1.28	.4	-3.21	.002	**
50th	2.54	1.76	1.11	1.2	-1.45	.37	-3.86	<.001	***
70th	3.16	1.46	1.54	1.28	-1.57	.34	-4.59	<.001	***
90th	3.43	1.32	2.05	1.18	-1.34	.31	-4.3	<.001	***
WEISS2									
10th	1.46	1.5	0.89	1.22	-.47	.34	-1.39	.168	
30th	1.89	1.45	1.08	1.19	-.75	.33	-2.29	.025	*
50th	2.27	1.63	1.19	1.15	-.98	.35	-2.81	.006	**
70th	2.51	1.48	1.35	1.14	-1.14	.33	-3.45	<.001	***
90th	3.3	1.35	1.92	1.04	-1.34	.3	-4.46	<.001	***
LSA									
10th	2	1.22	1.35	1.18	-.61	.3	-2.05	.044	*
30th	2.32	1.18	1.57	1.34	-.7	.31	-2.23	.029	*
50th	2.62	1.3	1.65	1.4	-.93	.34	-2.76	.007	**
70th	3.14	1.48	2.08	1.23	-.99	.34	-2.92	.005	**
90th	3.19	1.47	2.35	1.18	-.77	.33	-2.32	.023	*

Table 2.6 Mean, SD, and estimated coefficients for Group of number of clusters computed by the Koren-derived computational algorithm.

Once having partialled out the effect of *Education* (which was not significant for any of the variable under investigation), the results of the regression model on the number of clusters indicated significant differences between the two Groups when this index of verbal fluency was calculated using each and every semantic spaces, as well as considering thresholds set at any quintile (except for the threshold set at the 10th quantile of the WEISS2 space): overall, healthy volunteers produced more clusters than SSD participants, with significance ranging from p = .005 to p = .044.

Koren – Size of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	4.1	4.48	1.84	2.99	-2.6	.94	-2.76	.007	**
30th	3.77	3.45	1.76	2.07	-2.25	.7	-3.2	.002	**
50th	3.78	3.21	2	2.58	-1.95	.72	-2.69	.009	**
70th	5.33	4.87	3.04	4.7	-1.72	1.17	-1.46	.147	
90th	3.74	2.79	3.55	2.08	0.09	.61	.14	.888	
WEISS2									
10th	4.92	5.4	2.78	4.09	-1.63	1.18	-1.38	.172	
30th	5.03	4.76	2.98	3.8	-1.28	1.04	-1.23	.223	
50th	4.21	3.45	2.98	3.4	-.87	.84	-1.03	.306	
70th	4.15	3.03	3.1	2.98	-.68	.74	-.91	.363	
90th	3.77	2.44	3.69	1.87	-.04	.54	-.07	.943	
LSA									
10th	6.26	4.43	4.22	4.2	-1.74	1.07	-1.63	.108	
30th	5.26	3.3	3.79	3.41	-1.26	.83	-1.51	.134	
50th	4.77	2.5	3.45	3.27	-1.08	.72	-1.5	.137	
70th	4.07	2.48	3.77	2.49	-.44	.62	-.72	.474	
90th	3.59	1.82	3.6	2.05	-.16	.48	-.33	.745	

Table 2.7 Mean, SD, and estimated coefficients for Group of size of clusters computed by the Koren-derived computational algorithm.

The effect of *Education* was evident on for the mean size of clusters calculated according to the Koren-derived algorithm with WEISS2 at the 30th quintile (t = 2.18, p < .05), significant differences were found in the size of clusters of the two groups calculated by the Koren-derived algorithm using WEISS1: having set the threshold at the 10th, 30th, and 50th quintiles, we found that on average HPs produced clusters bigger in size than SSD participants (p <.05).

Comparison between automated and manual scoring at the categorization performance

ROC is a probability curve we can use to estimate positive and negative instances, and AUC represents the degree or measure of separability: in other words, the area under the curve tells us how much a model is capable of distinguishing between the two classes under investigation (in our case, “SSD participants” vs “HPs”).

The following graphs depict the categorization ability of the verbal indexes (number of switches and mean size of clusters) as defined by two proposed algorithms as well as by the manual annotator. In other words, we used the fitted values of a logistic regression (with the two measures of verbal fluency as interacting independent variables, and *Group* as categorical dependent variables) to compute a ROC curve (see Figure 2.4 and 2.5).

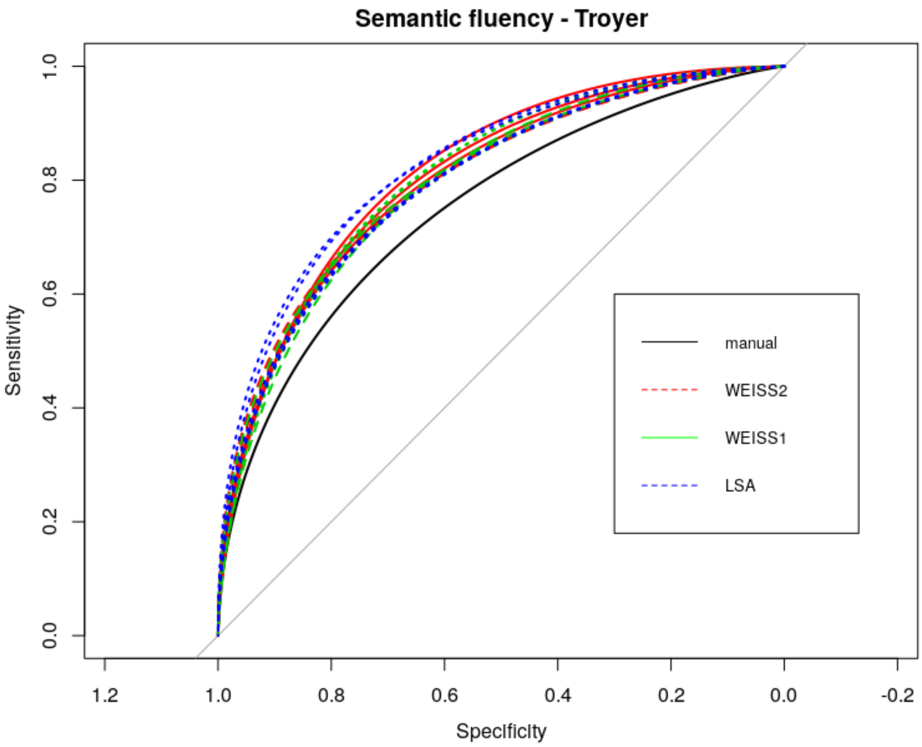


Figure 2.4 ROC curves for Troyer-derived measures of verbal fluency as estimated fitted values with Group as dependent variables

From visual inspection, Troyer-derived measures based on NLP methods (red, green, and blue lines) appear to outperform the measures obtained from the manual rater (black line) in the classification task. More specifically while WEISS2 (red line) seems to demonstrate good sensitivity, the classifiers based on LSA appear to have a higher AUC, with higher specificity. WEISS1, albeit being more performant than the manual rater, appears to be associated with a smaller AUC than the other two semantic spaces.

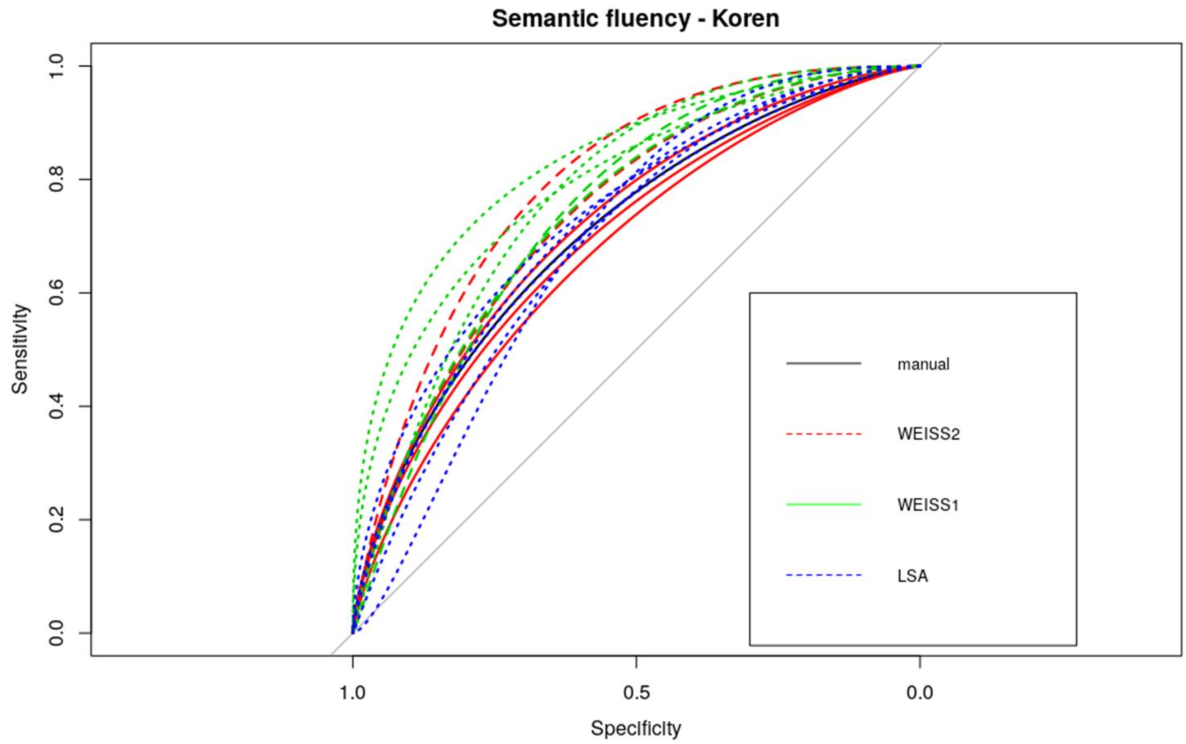
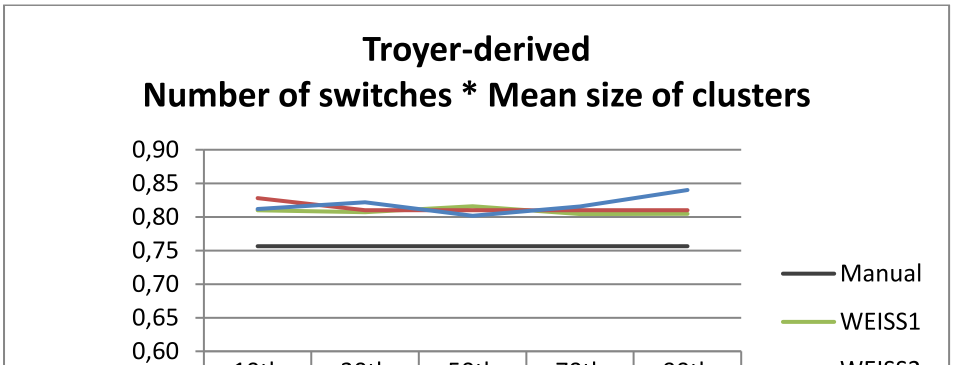


Figure 2.5 ROC curves for Koren-derived measures of verbal fluency as estimated fitted values with Group as dependent variables

From visual inspection, measures based on the Koren-derived method (red, green, and blue lines) appear, on average, to outperform the measures obtained from the manual rater (black line) in the classification task, albeit not always. In this case, WEISS1 (green line) computes the highest sensitivity, and WEISS2 (red line) the highest specificity, while LSA-derived measures seem to lag behind.

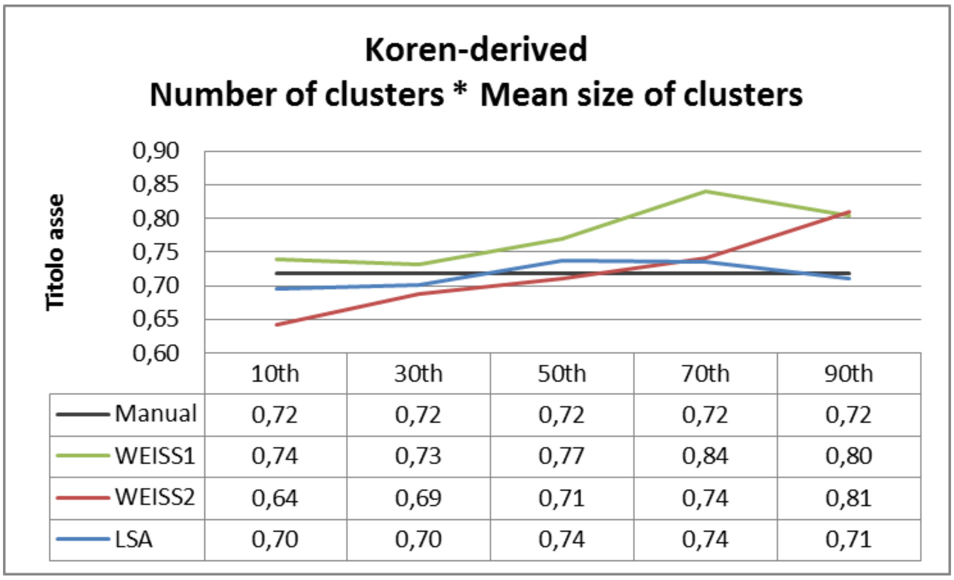
Figure 2.6 and 2.7 illustrate the comparison for all selected measures of verbal fluency.



	10th	30th	50th	70th	90th
Manual	0,76	0,76	0,76	0,76	0,76
WEISS1	0,81	0,81	0,82	0,80	0,80
WEISS2	0,83	0,81	0,81	0,81	0,81
LSA	0,81	0,82	0,80	0,82	0,84

Figure 2.6. AUC values of Troyer-derived (semantic VF task) for number of switches.

Figure 2.6 summarizes AUC values (Y-axis) of ROC curves computed based on the fitted values of the logistic regression model, with the “number of switches” and “mean size of clusters” calculated by the Troyer-derived algorithm as continuous independent variable and *Group* as dichotomous dependent variable. On the X-axis, we report the different thresholds used to define shifts. The highest value (AUC = .84) was scored by the procedure using LSA and having set a threshold at .43 (90th percentile). This system outperformed the manual method (AUC = .76), as well as the best performances of both the algorithms based on WEISS1 (AUC = .81) and WEISS2 (AUC = .83).



	10th	30th	50th	70th	90th
Manual	0,72	0,72	0,72	0,72	0,72
WEISS1	0,74	0,73	0,77	0,84	0,80
WEISS2	0,64	0,69	0,71	0,74	0,81
LSA	0,70	0,70	0,74	0,74	0,71

Figure 2.7 AUC values of koren-derived (semantic VF task) for mean size of clusters.

Figure 2.7 summarizes AUC values (Y-axis) of ROC curves computed based on the fitted values of the logistic regression model, with the “number of switches” and “mean size of clusters” calculated by the Koren-derived algorithm as continuous independent variable and *Group* as dichotomous dependent variable. On the X-axis, we report the different thresholds used to define shifts. The highest value (AUC = .84) was scored by the procedure using WEISS1 and having set a threshold at .23 (70th percentile). This system outperformed the manual method (AUC = .72), as well as the best performances of both the algorithms based on WEISS2 (AUC = .81) and LSA (AUC = .74).

Coherence analysis

Table 2.8 reports the estimated fitted values for *Group* for mean cosine values between words at 1-, 3-, 5-, and 7-word in-list distance computed by the “Coherence” function using the three semantic spaces.

	HPs		SSD Participants						
	M	**SD**	**M**	**SD**	**estimate**	**Std.Error**	**t value**	**p-value**	**sign**
WEISS1									
word N+1	.22	.02	.22	.03	0	.01	.25	.801	
word N+3	.17	.02	.17	.02	0	.01	.93	.357	
word N+5	.15	.02	.17	.03	.02	.01	2.91	.004	**
word N+7	.15	.02	.17	.05	.01	.01	1.65	.102	
WEISS2									
word N+1	.34	.03	.34	.03	0	.01	.34	.733	
word N+3	.29	.03	.29	.03	0	.01	.47	.643	
word N+5	.27	.03	.29	.04	.01	.01	1.34	.185	
word N+7	.27	.03	.29	.05	.01	.01	.8	.424	
LSA									
word N+1	.34	.04	.35	.04	0	.01	.44	.658	
word N+3	.27	.04	.29	.04	.02	.01	2.33	.022	*
word N+5	.27	.04	.29	.04	.02	.01	1.96	.054	
word N+7	.27	.04	.28	.06	.01	.01	.51	.61	

Table 2.8 Fitted values of *Group*Education* interaction on mean cosine similarity between words at different distances calculated according to the three different semantic spaces

As expected, the farther apart are two words in a list, the lower the mean cosine similarity between them, in both groups: as the participants explore the semantic store to retrieve words, there is a gradual loosening of semantic relatedness. However, Control participants show a steeper decrease: the difference between cosine similarities of adjacent words and with six interpolated items is -.07, while it ranges from -.05 to -.07 in SSD participants.

Statistical differences were found between the mean cosine values for SSD participants *versus* Control participants at a 5-word distance calculated with WEISS1. In this case, Control participants produced less related words (M = .15, SD = .02) than SSD participants (M = .17, SD = .02) (t = 2.91, p <.01). A significant difference was also found for the mean cosine values between SSD participants *vis-à-vis* Control participants with respect to the LSA cosines at 3-word distance: consistently, Control participants produced less related words (M = .27, SD = .04) than SSD Participants (M = .29, SD = .04) (t = 2.33, p < .05).

Figure 2.8 depicts the categorization ability of the estimated fitted values of the regression models having Group as dependent variable and measures of coherence as independent variable. To maximize the performance of the classifier, we selected the best predictors based on number of switches and mean size of clusters for each semantic space (baseline) from the previous section and added as predictor to compute ROC curves.

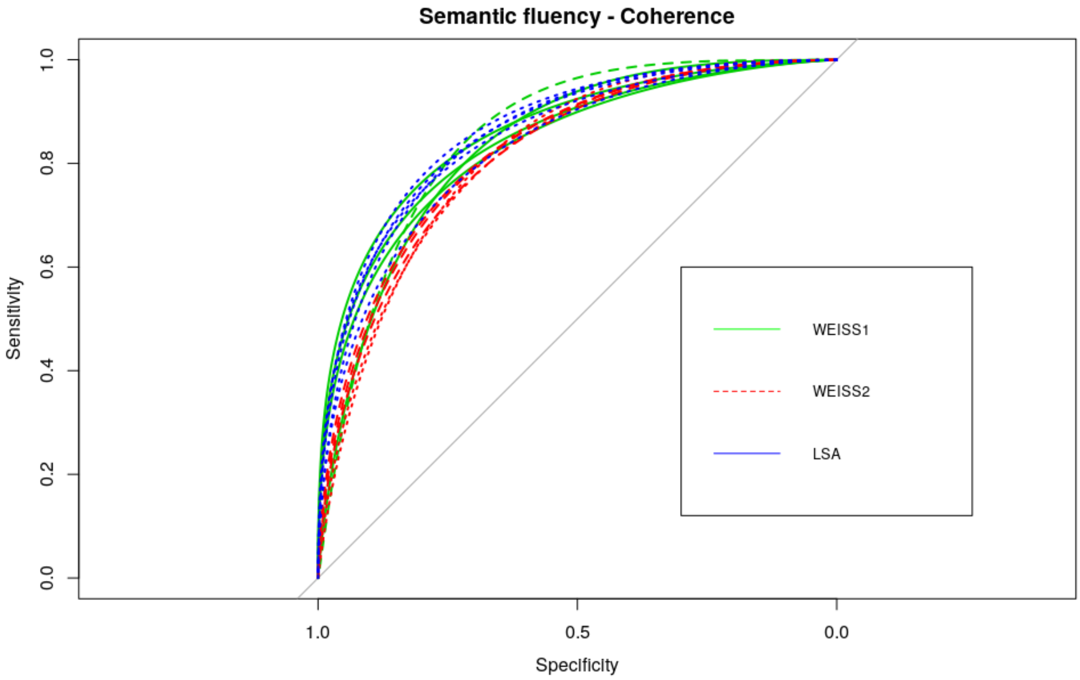
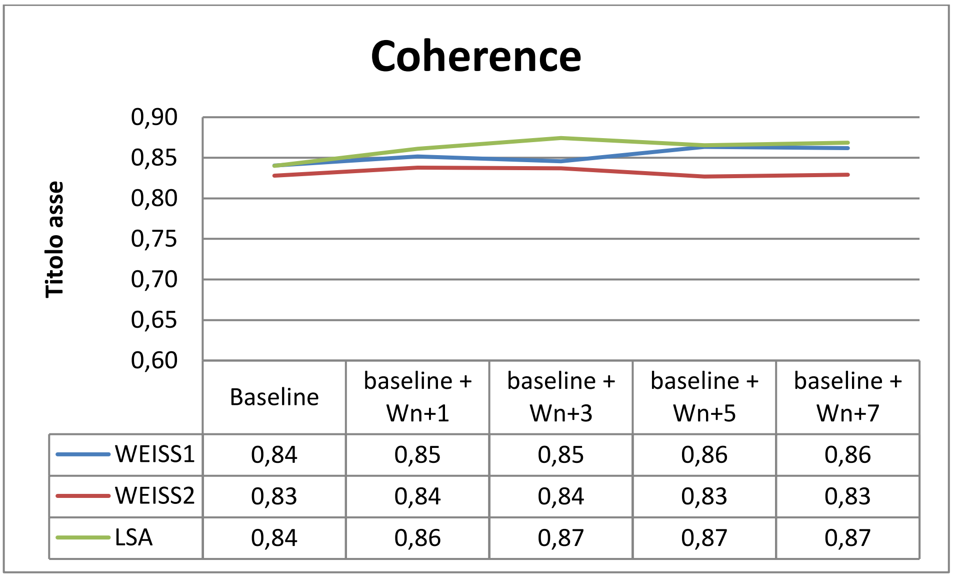


Figure 2.8 AUC values of ROC curves computed on values of coherence in the semantic VF task.

From visual inspection, WEISS1 (green line) seems to compute the highest sensitivity, and LSA (blue line) the highest specificity, while WEISS2 measures seem to be outperformed by the formers with respect to both specificity and sensitivity.



	Baseline	baseline + Wn+1	baseline + Wn+3	baseline + Wn+5	baseline + Wn+7
WEISS1	0,84	0,85	0,85	0,86	0,86
WEISS2	0,83	0,84	0,84	0,83	0,83
LSA	0,84	0,86	0,87	0,87	0,87

Figure 2.9 AUC values of classifiers based on the best set of parameters from the Troyer- and Koren-derived algorithms ("Baseline"), plus the values from the coherence function (at different in-list distances).

Figure 2.9 summarizes the AUC values of Figure 2.8. The highest value (AUC = .87) was scored by the procedure using LSA and considering as "baseline" the best classifier based on number of switches and mean size of cluster for that space (i.e., Troyer-derived measures with threshold at 90th quantile, AUC = .84), and the mean cosine similarity between words with 3 words distance.

The second-best classifier was the one based on the WEISS1 (AUC = .86) based on number of switches and mean size of cluster for that space (i.e., Koren-derived measures with threshold at the 70th quantile, AUC = .83), and the mean cosine similarity between words with 5 words distance.

The third best classifier was the one based on the WEISS2 (AUC = .84) based on number of switches and mean size of cluster for that space (i.e., Troyer-derived measures with threshold at the 10th quantile, AUC = .83), and the mean cosine similarity between subsequent words.

Comparison of ROC performances

How do the identified indexes of verbal fluency compare to each other? Swets and colleagues (2000) suggested that AUCs values >0.9 are ‘‘excellent,’’ >0.80 ‘‘good,’’ >0.70 ‘‘fair,’’ and <0.70 ‘‘poor’’. However, as pointed out by Youngstrom (2014), AUCs greater than .90 are likely to indicate design flaws rather than exceptional discriminative validity. Table 2.9 reports the best AUCs values for all the SVF indexes adopted in this study. As we can see, all measures of verbal fluency, including those derived by the manual annotation, have fair to good classifying performances. Measures based on predict-models (WEISS1 and WEISS2) have robustly good performances (always >.80), while measures derived from LSA seem less stable and to be affected by the scoring methodology, having performances ranging from fair (AUC = .74 for Koren-derived measures) to good (AUC =.87 for Coherence).

Method	Rater/Semantic Space	Best AUC value
Troyer	Manual	.76
	WEISS1	.82
	WEISS2	.83
	LSA	.84
Koren	Manual	.72
	WEISS1	.84
	WEISS2	.81
	LSA	.74
Baseline + Coherence	WEISS1	.86
	WEISS2	.84
	LSA	**.87**

Table 2.9 Best AUC values for all SVF indexes adopted in the study.

2.4 Discussion

The primary aim of this study was to investigate the predictive ability of measures derived from Distributional Semantic Models to discriminate the verbal production of people with and without SSD to an SVF test. Our hypothesis was that an unsupervised algorithm able to extrapolate the statistical distribution of words/concepts in a large text corpus would create a model of human knowledge (the semantic space) more fit to detect implicit nuanced associations made by human subjects engaged in a SVF task than human annotations based on strict, well-defined, taxonomical categories, and hence able to discriminate people with and without SSD. To do so, we administered an SVF task to a group of people with and without SSD, whereby participants were asked to produce as many words as possible in the category "animal", in a time limit of 60 seconds. Verbal outputs were analysed by: i) manually computing the number of switches and the mean size of semantic clusters, following the methodology proposed both by Troyer and colleagues (1997), and the revised methodology by Koren and colleagues (2005); ii) automatically identifying and counting semantic clusters and switches by means of measures of semantic similarities derived from Distributional Semantic Models; iii) automatically computing a measure of coherence by measuring the semantic similarity between words at different distances in the list of produced items. The resulting measures, considering indexes of integrity of the semantic store (mean size of clusters) and cognitive flexibility (mean number of switches between clusters), were compared between groups. Moreover, for each measures of verbal fluency (both manual and automated), a classifier able to categorize subjects into groups was developed, and we used ROC graphs (and their resulting AUCs) to rank the performance of the classifiers. In particular, for the first two algorithms (those derived by the methods proposed by Troyer and Koren) we draw ROC curves with the estimated fitted values of a logistic regression having the interaction of the variables *Number of switches* and *Size of clusters* as dependent variables, and the pre-identified *Group* memberships as categorical dependent variable. Having identified the best set of predictors on the base of the resulting AUCs, we used such predictors as "baseline" for a new set ROC curves, obtained by adding the resulting cosine values of the "coherence" algorithm as interacting independent variables.

Standard quantitative measures

Quality of data collected was confirmed by standard quantitative measures: consistently with previous studies administering verbal fluency tasks to schizophrenic patients, people with SSD in our cohort produced, on average, a significant lower number of words than Control participants (Bokat & Goldberg, 2003). Moreover, words produced by the patients' group had, on average, a higher frequency than those produced by Control participants. This finding is in line with a theory assuming a semantic store deficit in schizophrenia: results of previous research employing psycholinguistic tasks manipulating frequency of items (Rossell & David, 2006) concluded that patients with schizophrenia have fewer low-frequency items stored in semantic memory compared to HPs. According to a network model of semantic memory (such as that proposed by Collins and Loftus (1975), concepts with higher frequencies are endowed with stronger weights and higher connection strengths: any damage to the semantic memory system, such as in patients with temporal lobe atrophy, would affect items with lower connection strengths first, preserving those with the higher ones. An early work employing multidimensional scaling and pathfinder analyses (Paulsen et al., 1996) suggested that the semantic network of early-onset schizophrenic patients is less organized and contains atypical links. Previous studies on spreading activation in schizophrenic patients using a semantic priming paradigm (for example Moritz et al., 2002) found that excessive automatic spreading of activation in the semantic network may be a major contributor to the manifestation of formal Thought Disorder (ThD). Taking all these pieces of evidence together, it seems reasonable to suppose that an anomalous spreading activation of a disorganized and damaged semantic store could explain the poor performance in terms of frequency of the items produced of people with SSD in an SVF task.

Number of switches and size of clusters

In terms of clustering and switching, it is worth noting that SVF scores manually computed and scores obtained through the automatic systems were different from each other. In fact, our algorithms were not developed to reproduce the human intuition in annotating SVF outputs, i.e., they were not designed as a mere computational implementation of the classical scoring approach based on *ad-hoc* taxonomical categories. Rather, we adopted a different approach to the notion of "semantic similarity", leveraging on the potentialities of the Distributional Hypothesis of language. Notwithstanding these differences, the results of the study showed also some interesting similarities between the results of the automatic models and human annotations, speaking for the construct validity of the proposed approach (convergent validity). In line with the literature analysing SVF in the schizophrenic population using a manual approach to calculate clusters and switches (Robert et al., 1998; Bozikas et al., 2005), our Control participants produced more switches than SSD Participants. The same pattern of results was found applying the automated procedure, irrespective of the algorithm used and of the semantic space of reference adopted. Moreover, differently from values computed with the manual procedure, which did not identify any significant differences between groups, significant differences in the mean size of clusters were visible in NLP-derived measures computed following Koren's procedure using WEISS1 and setting thresholds at 10^{th}, 30^{th} and 50^{th} quantiles, whereby HPs always produced clusters bigger in size than SSD participants. These results speak for a greater sensitivity of NLP-derived measures of verbal fluency compared to the manual annotation and are in line with a theoretical framework assuming a degraded semantic store of people with SSD. This means that, once entered a semantic cluster, people with SSD seem unable to successfully explore it, presumably due to the degraded nature of the store and the impaired activation of its network: thus, in order to fulfil the task at hand, they are forced to switch to another semantic field. This strategy results to be less effective in the overall economy of the task, presumably because it is costly in terms of cognitive effort and, given the reduced prefrontal functioning of people with SSD (Everett et al., 2001; Stirling et al., 2006; Zalla et al., 2001), plays against their executive performance.

Coherence

An interesting finding is related to the semantic coherence of the items produced during SVF as represented by the cosine values between word vectors. In a previous study measuring semantic coherence between adjacent words in a "animal" SVF task with LSA-derived measures of semantic relatedness (Elvevåg et al., 2007), coherence scores of patients with high levels of clinically rated ThD were significantly lower than patients with low levels of clinically rated ThD, as well as to that of healthy control participants. No comparison was made between the schizophrenic group as a whole and HPs. However, a small effect was indeed found between a group of healthy control and a group of schizophrenic patients in the mean similarity of words with the immediately preceding word calculated using CoVec, a vector-based method (Pauselli et al., 2018). Contrary to these results, we did not find any difference in the cosine similarity of adjacent words between the two Groups, indicating a nearly identical level of coherence between subsequent words as produced by HPs and people with SSD. On the contrary, significant differences between the two groups were found in the cosine similarities calculated according to the WEISS1 and LSA spaces of words at 5 and 3 in-list distance, respectively, whereby HPs showed lower coherence than SSD participants. These results show that, proceeding through the exploration of the semantic store to perform the SVF task, HPs produced words that were weakly related to each other, an indication of a structured and functioning semantic network in this group: the spreading of semantic activation of HPs proceed as they delve into the semantic store, making items available and ready for retrieval. Differently, this process seems to be less performing in SSD patients, whose responses tend in fact to be bound to a single area of the semantic space, leading to significant difference between groups in the values of cosine similarity between any word and the third and the fifth subsequent word. We interpret this finding as an indication of an impaired semantic store, which is further grounded considering the (previously discussed) impact of word frequency.

Performance of classifiers

In line with our initial hypothesis, comparison of AUCs showed that scoring based on measures of semantic similarity derived from Distributed Semantic Models systematically outperforms manual scoring in the classification task. Among the formers, the best classifier (AUC = .87) was the one based on the number of switches and mean size of clusters computed following Troyer's methods and considering a threshold set at the 90th quantile of the cosine values of our population as derived by the LSA space, to which the mean cosine similarity of words with 3 words of distance was added. By adding a measure of coherence to the computed measure of number of switches and mean size of clusters, we maximized the discriminative performance of the proposed algorithm. We interpret this result as reflecting a limit of the taxonomical approach at the basis of human annotation. In particular, the categories (and subcategories) proposed in Troyer's original work (such as geographical living environment, type of human use, and zoological categories) excludes all other kinds of associations and strategies that participants might want to use to group words, such as, for example, multi-words expressions and idioms. The limitation of a super-imposed model of knowledge to the investigation of semantic properties of words have been already proven (Jones et al, 2006): in this sense, models of knowledge based on the distributional properties of words do not impose pre-defined constraints and thus appears flexible enough to "learn" those semantic information that are salient for word representation. Moreover, from a practical point of view, the advantage of adopting automatic SVF test scoring lies also in the paucity of resources that are needed: an automated algorithm can compute the desired measures faster, and results are internally consistent. One may claim that the sole number of words could be discriminative of the two groups. However, this seems not to be the case: we trained a classifier using only number of unique words produced, and the result was AUC = .81, which is remarkably lower than the classification ability of our proposed method (.87).

The setting of the semantic space parameters deserves a consideration apart. It is known that several factors impact on the performance of semantic spaces (Lenci, 2018): among these, we find the characteristics of the training corpus, the dimension of the co-occurrence window used to train the model, as well as the chosen dimensionality reduction (vector dimensionality).

In our case, the choice of using ItWac as training corpus may have had an impact on the resulting performance of the algorithms (Berardi et al., 2015): being its content mainly encyclopaedic, it may have facilitated the production of the most educated group (Control participants, in our case).

In our study, we used two semantic spaces created with a predict algorithm using a 5- (WEISS2) and 9-word (WEISS1) window, respectively, and a semantic space created with LSA based on whole documents as contexts. As previously pointed out (Baroni & Lenci, 2010), minimal windows have been proven appropriate for syntactic tasks, while larger window sizes are best for semantic tasks. Given the semantic task at hand, it is thus no surprise that only the algorithm based on WEISS1 was able to identify significant differences in the mean size of clusters of the two groups. However, when considering all the measures of verbal fluency adopted in this experiment altogether, such as coherence, cognitive flexibility (number of switches) and semantic store integrity (mean size of clusters), contrary to our predictions, the LSA space was overall the fittest to distinguish the two groups of participants (AUC = .87) with a classifier combining all these measures. In this sense, the overall best performance of the LSA space prompts further consideration on the kind of semantic associations computed by the two different types of semantic space. More specifically, what kind of "relatedness" are we talking about when we talk about "semantic relatedness" in this context?

In his work, Sahlgren (2008) reports the foundational distinction proposed in Saussure's *Cours de linguistique générale* (1916/1987) between paradigmatic and syntagmatic relationships of linguistic elements as applied to the distributional hypothesis of language implemented in word vectors. In this theoretical framework, paradigmatic relations would relate entities that occur in the same context, but not at the same time (a relation "*in absentia*"), while syntagmatic relations would relate entities that co-occur in a linear, sequential combination in the text (i.e., "*in praesentia*"). Similarly, the importance of what is consider "context" by the algorithm creating the model has been described and it has been proven that semantic space models created by incorporating both contextual information and word-order are best equipped to explain mediated priming effects (Jones et al., 2006)

In our specific case, on one side, we have a "categorical" verbal fluency task, that would predict the production of words sharing the same context, but not at the same time (like "*cane*" and "*ghepardo*" in the sentence "*il cane|ghepardo è un animale*"). On the other, we have an LSA space created on a document-based matrix of co-occurrences, i.e., word vectors derived from this space are created on a matrix having entire documents as co-occurrence contexts. We thus hypothesize that the nature of the fluency task under investigation prompt participants to produce words paradigmatically related: in this sense, the architecture of our LSA space, based on a document-based co-occurrence matrix, would best fit to capture this kind of relationship, compared to predict-models based on narrow co-occurrence windows. In this latter case, being the semantic spaces derived from co-occurrences computing according to a narrow moving-window sliding across the text, their vector representations would be more likely to capture syntagmatic relationships (like "*cane*" and "*collare*" in the sentence "il |*cane*| porta il |*collare*|"). These results support the speculation by Sahlgren that LSA would captures associative relations, while moving-window models like word2vec would capture semantic relations.

Moreover, differently from LSA models, which traditionally adopt a PMI-based method to smooth the negative sampling distribution, we adopted a PPMI-based method: this latter is known to outperform the formers on similarity tasks (Bullinaria & Levy, 2007). In this sense, we align with the considerations of Levy and colleagues (2015) stating that a significant part of the performance of word embedding are due to design choice and hyper-parameters settings, rather than to the algorithms themselves. Nonetheless, the performance of both predict models with this combined set of predictors was similarly very good, with AUCs range from .84 to .86.

If we consider the broad aim of creating tools to support clinicians in the diagnosis of people with schizophrenia disorders, our algorithms outperform previous work employing supervised binary classification task based on linguistic features of written dataset from schizophrenic patients such (Sarioglu Kayi et al., 2017), which, by using Latent Dirichlet Allocation (LDA) and incorporating Part-Of-Speech (POS) tagging, reached an accuracy of AUC = 81.7. However, differently from their work, our semantic spaces are non-tagged, implying a practical advantage in terms of pre-processing required to perform the analyses.

Conclusions

To conclude, this study showed that a fine-grained analysis of SVF tests can provide important measures of cognitive performance in people with SSD by analysing verbal fluency outputs by taking into consideration additional scoring that mirror the different cognitive processes at stake, we can better characterise the performance of participants, highlighting the mutual interaction of semantic store integrity and executive functions to perform the task.

Furthermore, this study represents a proof-of-concept that a fast, sensitive, and reliable tool based on distributional semantics can outperform human annotations for the purpose of classifying subjects on the basis of their performance. Our results challenge the shared notion of the superiority of predict-models over count-models, given the superior performance of LSA compared to word2vec.

As pointed out by Iter and colleagues (2018), unlike other medical fields that can rely on hard metrics such as blood pressure or blood glucose levels, no objective metrics of speech irregularities for schizophrenia are currently used in psychiatry, despite the central role of speech assessment as a diagnostic marker. Such metrics would be in fact desirable to support the diagnosis, as well as monitoring the progress and the response to treatments of people with psychiatric disorders such as SSD.

3 Distributional semantic models applied to a generative associative naming task in people with Schizophrenia Spectrum Disorders

3.1 Introduction

Idiosyncratic or "loose" associations are considered a core feature of schizophrenia (see Chapter 1) and past studies found that schizophrenic patients (Kent & Rosanoff, 1910; Mefferd, 1979; Shakow & Jellinek, 1965), as well as subjects with high levels of psychoticism (Merten, 1993) and high levels of schizotypal thinking (Duchêne et al., 1998), respond with more idiosyncratic word associations than HPs to word association tests (WATs). Firstly introduced to assess personality traits as a projective technique, the WAT has later developed into a diagnostic tool (Singh, 2017). Among the different versions commonly used to assess associative processes, the generative associative naming task ideated by Spinnler and Tognoni (Spinnler & Tognoni, 1987) combines a word association task with a fluency task. In fact, differently from traditional WATs which require a single-word response to a given cue, here subjects are asked to produce as many words as possible related to the index stimulus, in a given timeframe. Such a process is deemed to rely on both the access to the semantic store and executive control (Ralph et al. 2016). Hence, a task combining word association and verbal fluency is an ideal choice to assess the differential contribution of the two cognitive systems in determining the poor verbal fluency observed in SSD patients (see Chapter 1 for details of cognitive impairments in this population).

To the best of our knowledge this tool, however, has never been administered to the schizophrenic population. Perhaps the most important limitation of word association tasks is that they rely on pre-compiled lists, against which the production of each subject must be compared. For native English speakers, the list of Kent-Rosanoff (1910) is still used, along with the one developed by Jenkins and Palermo (1967) and, for the Italian population, the work by Parisi & Pizzamiglio (1967). Only words that are within the list are considered acceptable and computed in the total score. A reduced performance to a generative associative naming task is considered indicative of the presence of neurocognitive deterioration. However, the "total word" measure does not provide specific information as to the degree to which the different cognitive systems implied in the task are compromised (Mayr, 2002). Systems considering clustering and switching (Troyer et al., 1997; Koren et al., 2005), proposed for fluency tasks, may be better suited to a discriminative analysis of WAT outputs.

We have already explored the limitations of traditional scoring systems (rigidity of a prescriptive scoring approach and time-constraints of the manual rating – see Chapter 2), and how NLP can help overcome such limitations. In this sense, measures of semantic relatedness derived by calculating the cosine distance between word vectors of LSA semantic spaces (see Chapter 1) have already been proven their usefulness in studying clinical populations (Holshausen et al., 2014; Pakhomov & Hemmy, 2014; Farzanfar et al., 2018; Iter et al., 2018). However, recent advances of NLP techniques, such as semantic spaces generated by neural networks, might be better suited for the task, given the recognized superiority of predict-based models over count-based models in psycholinguistics tasks (Baroni et al., 2014a; Mandera et al., 2017).

The objective of the present study is twofold. Firstly, we will analyse the results of a generative associative naming task administered to a sample of people with SSD and a matched group of people without psychiatric disorders by considering number and mean size of the produced semantic clusters. These two metrics will be computed by a manual annotator and three NLP-based semantic systems following the procedures of two possible scoring systems (Koren et al., 2005; Troyer et al., 1997). An additional measure of semantic coherence will also be computed. In absence of previous literature considering this specific task on this population, or the application of such measures of clustering and switching to a generative associative naming task, we will base our predictions on the general literature on semantic VF tasks in this population (Bokat & Goldberg, 2003), as well as works on WAT in SSD (Baskak et al., 2008; Elvevåg et al., 2007; Johnson & Shean, 1993; Merten, 1993), and the results presented in Chapter 2. We thus hypothesize to find: (i) a reduced number of total and unique items produced by SSD participants compared to HPs; (ii) SSD participants to produce fewer and smaller-in-size clusters than HPs; and (iii) given that "loose associations" contribute to the very definition of "conceptual disorganization" in SSD, we expect word associations in people with SSD to be less "coherent" than in HPs.

Secondly, we aim at comparing the sensibility and specificity of automatic procedures *vis-à-vis* human-derived measures in a classification task. According to previous literature, the best performance of NLP-measures to different semantic relatedness tasks was obtained by predict-models (Baroni et al., 2014). Given that our classifiers will leverage on measures of semantic relatedness to categorize subjects, we hypothesize that those derived from predict-based models (WEISS1 and WEISS2) will perform better compared to those derived from a count-based model (LSA). Given the natural limits of a reference list, we predict that automatic scores will outperform the traditional manual approach in characterizing the production of SSD participants and HPs and hence in classifying subjects in the two groups.

3.2 Materials and methods

3.2.1 Subjects

Thirty-four persons with a diagnosis of SSD according to DSM-5 (APA, 2013) were recruited. Participants were recruited from the outpatients' service and the residential facilities of the IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, between February 2018 and April 2019. Diagnoses were made by the treating clinicians (staff psychiatrists). Participants were 18-65 years old, able to give informed consent, right-handed, and had normal or correct-to-normal visual acuity. The diagnosis on Axis I had to be unique, but co-morbidities on Axis II were admitted. Exclusion criteria were: additional neurological disorders, head trauma with cognitive sequalae, mental retardation, substance abuse in the 3 months preceding the enrolment. The final sample of people with SSD (26 males and 8 females) had a mean age of 49.32 years (SD = 10.19) and a mean years of illness of 23.42 (SD = 12.14, N = 33 – for one SSD participant, it was not possible to date the age of onset); at the time of recruitment, the group of patients had been on treatment with at least one anti-psychotic medication (M = 2.5, SD = 1.62) for at least the previous 6 months.

Thirty-four healthy volunteers, matched by age-class (18-30, 31-40, 41-50, 51-65) and gender, were recruited among the hospital staff and through public announcements also in Brescia. Level of education was not matched at recruitment between the two groups and was found to be significantly different ($W = 816.5$, $p < .01$). The exclusion criteria for the control group were: any documented psychiatric disorders or being first-degree relative of a patient with diagnosis of SSD. All subjects were native speakers of Italian.

After a complete description of the study, informed consent to participation was obtained from all subjects. In case of patients with support administration, the participation to the study was first discussed with the patient; then, written consent was obtained from both the patient and the appointed administrator. The study was approved by the IRCCS Ethical Committee (Opinion 61/2017) and followed the principles of the Helsinki Declaration.

3.2.2 Task

As part of a wider assessment battery, the generative associative naming task by Spinnler and Tognoni (1987) was administered to study participants. This task combines aspects of fluency and of word association tests, requiring the subject to name as many words as possible in a time constraint of 120 seconds, whose meaning are in relationship with an index item. The index words used to cue responses are four, selected so as to explore different fields of the subject' semantic knowledge, and namely: an animal ("gatto"/*cat*), a concrete object ("scarpa"/*shoe*), a meteorological event ("pioggia"/*rain*), and an abstract word ("sciopero"/*strike*). All sessions were audio recorded and later transcribed by a training psychologist. The transcripts were converted to plain ASCII text to make them machine-readable, and hand-edited to enforce standard spelling.

3.2.3 Description of the experimental (independent) variables

Manual scoring

Participants' productions were manually scored by a training psychologist, not blind to the experimental design. Performance was rated by counting the number of correct answers, whereby a word is considered acceptable based on the reference lists of the standardization work by Spinnler and Tognoni (1987) as reported in Bandera and colleagues (1991). Words were collected from a sample of N = 322 persons (130 males and 192 females) with age 40-90 and different levels of education (N = 58 with 4 or less, N = 189 with 5 to 11, and N = 75 with 12 or more years of formal education). The range of production was quite variable across subjects, ranging from 7 to 108 words per index word. The reference list contains a total number of 1,012 words, divided as follows: "*gatto*" = 210; "*scarpa*" = 224; "*pioggia*" = 220; "*sciopero*" = 358. Number of switches and mean size of clusters were scored following the approach of Troyer's (Troyer et al., 1997) as well as the revised scoring approach proposed by Koren (Koren et al., 2005), described in details in Chapter 2.

Automatic scoring

A detailed description on how the numerical representations of word meaning (semantic spaces: WEISS1, WEISS2, and LSA) were obtained, together with a description of the algorithms used to compute measure of verbal fluency (Troyer-derived, Koren-derived, and coherence) are reported in Chapter 2. To instruct the algorithm in the identification of a semantic switch, we considered the distribution of semantic relatedness values in our sample for this specific task. We computed cosine distance values between all adjacent words (N = 4,228) produced by the study cohort (N = 68) and selected the 10^{th}, 30^{th}, 50^{th}, 70^{th} and 90^{th} quantiles of the values distribution as thresholds marking the switch to a different semantic cluster. For each semantic space, five thresholds were computed. Switches were identified applying only thresholds specific for that semantic space (numerical values are reported in Table 3.1).

	10^{th}	**30^{th}**	**50^{th}**	**70^{th}**	**90^{th}**
WEISS1	.11	.13	.13	.15	.17
WEISS2	.18	.20	.22	.23	.25
LSA	.24	.26	.28	.30	.33

Table 3.1. Thresholds for the three semantic spaces to be applied to the generative associative naming task

Standard quantitative measures (number of words, repetitions, and unique words) were collected. Mean frequency and length of the produced words were also considered. Frequency estimates were obtained from Subtlex-IT, an annotated corpus of around 130 million words of the Italian language based on movie subtitles, with a frequency range from .01 – 17,854 per million and mean length of the word produced and available at http://crr.ugent.be/subtlex-it.

3.2.4 Data analyses

Results of both the manual and automated procedures (based on WEISS1, WEISS2, and LSA) used to compute mean number of switches and mean cluster size, as well as coherence measures, were entered as continuous dependent variables in a series of general linear regression models, having group membership (pre-identified by the clinical staff) as categorical predictor. Moreover, albeit known to have minimal effect on the mean size of the produced clusters in animal fluency tests (Troyer, 2003), but given the proven effect of education on semantic VF tasks (Spinnler and Tognoni, 1985; Novelli et al., 1986), *Education* was included as covariate in the regression models in order to evaluate its possible confounding effect.

To test the predictive ability of the different models (manual scoring, WEISS1, WEISS2, and LSA) to categorize subjects, we compared Area Under the Curve (AUC) values derived from

Receiver Operating Characteristics (ROC) graphs (see Chapter 2 for a detailed description of ROC graphs and AUC values). ROC curves were created based on the fitted values of logistic regressions having measures of verbal fluency as independent variables, and group as categorical dependent variable. In particular, we considered the interaction of mean size of clusters and mean number of switches as computed for all the considered thresholds and semantic spaces.

Moreover, we created an additional set of classifiers by applying a logistic regression model having group membership (pre-identified by the clinical staff) as categorical dependent variable and fluency indexes (number of switches*means size of clusters, and the best predictor of the first step with cosine similarity between words at different distances) calculated manually and based on the three semantic spaces (WEISS1, WEISS2, and LSA) as continuous independent variables.

We then compared our models' performances, with AUC = 1 indicating perfect accuracy. All data were analysed using the R software V3.6 (R CoreTeam, 2019).

3.3 Results

3.3.1 Verbal fluency data

The final dataset used for the analysis consists of 4,228 words (1,671 unique words), as produced by the cohort of 68 subjects described above.

3.3.2 Standard quantitative measures

Table 3.2 reports mean values, standard deviations, as well as the partial estimated coefficients for *Group*, of a set of standard quantitative values.

index	HPs M	HPs SD	SSD Participants M	SSD Participants SD	estimate	Std.Error	t value	p-value	sign
Number of items	77.56	14.64	48.59	19.93	-25.35	4.41	-5.75	<.001	***
Unique words	75.5	13.94	46.65	19.02	-25.19	4.19	-6.02	<.001	***
Repetitions	2.06	1.79	1.94	2.12	-.16	.51	-.32	.749	
Mean word-length	7.12	.27	6.84	.44	-.21	.09	-2.22	.030	*
Mean frequency	4,914.26	1,889.22	6,595.72	4,871.33	1,400.66	963.6	1.45	.151	

Table 3.2 Mean, SD, and estimated coefficients for group of the SVF standard quantitative measures

Once having partialled out the effect of Education, which was significant on measures of total number of items produced (p = .027), unique words (p = .018), and mean word length (p = .043) the effect of *Group* was still present. In particular, results show that, on average, people with SSD produced significantly (t = -5.75, p < .001) fewer words (M = 48.59, SD = 19.93) than HPs (M = 77.56, SD = 14.64). Once having subtracted the number of repetitions, non-words, and words outside the SS, people with SSD still produced significantly less (t = -6.02, p < .001) unique words (M = 46.65, SD = 19.02) than HPs (M = 75.5, SD = 13.94). With respect to the mean length of words produced, people with SSD used words that were significantly (t = -2.22, p <.05) shorter (M = 6.84; SD = .44) than HPs (M = 7.12; SD = .27). No significant differences were observed in the mean frequency of words (p = .151), and number of repetitions (p = .749).

Results of the manual scoring procedure

Table 3.3 reports mean values, standard deviations, as well as the estimated the partial coefficients for *Group* on estimates based on the manual scoring assessment.

Index	HPs Mean	HPs SD	SSD Participants Mean	SSD Participants SD	Estimate	Std.error	T value	p.value	Sign.
Troyer									
Number of switches	40.92	14.31	25.74	15.12	-14.09	3.84	-3.67	< .001	***
Size of clusters	.83	.26	.97	.6	.09	.12	.73	.471	
Koren									
Number of clusters	18.24	6.62	9.9	5.64	-7.97	1.61	-4.96	< .001	***
Size of clusters	1.47	.23	1.74	.87	.23	.17	1.39	.168	

Table 3.3 Mean, SD, and estimated coefficients on the manual-scoring values for *Group*.

Once having partialled out the effect of *Education*, which was not significant for any of the variables under investigations, results the showed that SSD participants produced significantly less switches than HPs, both when considering values calculated according to Troyer's method and when considering values calculated on the basis of the Koren's method (t = -3.67, p < .001 and t = -4.96, p

< .001, respectively). No significant difference was found in the mean size of cluster following either approach (p =.471 and p = .168, respectively).

3.3.4 Results of the automated scoring procedure

Tables 3.4 and 3.5 report the results concerning the two indexes of verbal fluency (number of switches and mean size of clusters) calculated with the Troyer-derived automatic procedure.

Troyer – Number of switches

	HPs		**SSD Participants**						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	28.56	10.45	16.68	9.92	-12.5	2.66	-4.7	p <.001	***
30th	34.82	12.24	19.74	11	-15.53	3.05	-5.1	p <.001	***
50th	34.82	12.24	19.74	11	-15.53	3.05	-5.1	p <.001	***
70th	41.82	11.47	23.74	12.04	-17.68	3.08	-5.74	p <.001	***
90th	47.24	12.16	27.62	13.08	-18.91	3.3	-5.73	p <.001	***
WEISS2									
10th	27.91	10.71	16.88	9.55	-11.7	2.65	-4.41	p <.001	***
30th	32.5	11.68	19.44	11.34	-13.69	3.01	-4.55	p <.001	***
50th	37.68	12.11	21.44	11.32	-16.68	3.07	-5.43	p <.001	***
70th	40.29	11.73	22.38	11.57	-17.95	3.05	-5.88	p <.001	***
90th	44.5	11.83	25.12	11.53	-19.43	3.06	-6.35	p <.001	***
LSA									
10th	28.09	9.93	15.56	8.31	-12.15	2.4	-5.07	p <.001	***
30th	33.12	10.66	17.97	8.93	-14.8	2.57	-5.75	p <.001	***
50th	37.26	10.66	20.71	9.8	-16.1	2.68	-6.01	p <.001	***
70th	41.56	10.49	22.74	10.1	-17.67	2.67	-6.62	p <.001	***
90th	46.94	10.36	26.68	11.15	-19.29	2.8	-6.89	p <.001	***

Table 3.4. Mean, SD, and estimated coefficients for Group of number of switches computed by the Troyer-derived computational algorithm.

Once having partialled out the effect of *Education*, which was not significant for any of the variables under investigation, significant differences in the mean number of switches produced were found for all the thresholds (10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th} quantiles), irrespective of the semantic space of reference used, with people with SSD producing always fewer switches than healthy control participants (p <.001).

Troyer – Size of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	2.87	.95	3.1	1.19	.54	.26	2.06	.044	*
30th	2.35	.77	2.55	.79	.43	.19	2.26	.027	*
50th	2.35	.77	2.55	.79	.43	.19	2.26	.027	*
70th	1.88	.39	2.07	.52	.29	.12	2.54	.014	*
90th	1.65	.28	1.76	.37	.17	.08	2.07	.042	*
WEISS2									
10th	3	1.12	3.01	1.06	.3	.27	1.12	.267	
30th	2.57	.97	2.65	.94	.35	.23	1.51	.136	
50th	2.15	.67	2.3	.62	.33	.16	2.11	.039	*
70th	1.97	.5	2.2	.59	.37	.14	2.7	.009	**
90th	1.77	.4	1.9	.43	.25	.1	2.46	.016	*
LSA									
10th	2.89	.93	3.27	1.32	.45	.3	1.52	.133	
30th	2.42	.65	2.83	1.26	.46	.26	1.77	.081	
50th	2.11	.48	2.4	.87	.34	.18	1.84	.070	
70th	1.86	.29	2.13	.53	.28	.11	2.49	.015	*
90th	1.64	.24	1.77	.29	.17	.07	2.48	.016	*

Table 3.5 Mean, SD, and estimated coefficients for Group of mean cluster size computed by the Troyer-derived computational algorithm.

Education was found to have a significant effect on all the measures of cluster size derived by WEISS1 and WEISS2. Specifically, it was the only significant predictor for the mean size of clusters as calculated according to the 10^{th} and 30^{th} quantile threshold with WEISS2 (t = 2.96, p = .004; and t = 3.14, p = .002, respectively), with higher level of education correlated with bigger cluster size. In all other cases, once have partialled out the effect of *Education*, the effect of *Group* was still significant. Namely, the results of the regression model on mean size of clusters calculated using WEISS1 indicated significant differences at all thresholds, whereby people with SSD always produced on average clusters bigger in size than healthy volunteers (p <.05).

We found the same pattern of results using WEISS2 and having set a threshold at the 50^{th}, 70^{th}, as well as 90^{th} quantile, with significant differences (ranging from p = .009 to p = .039) between people with and without SSD, whereby healthy control participants produced on average cluster smaller in size than SSD Participants.

When considering estimates from the LSA-derived semantic space, the effect of *Education* was not significant for any of the variables under investigation, and we found the same pattern of differences, with HPs producing smaller clusters than SSD participants, (p <.05) for the thresholds set at the 70^{th} and 90^{th} quantiles.

Tables 3.6 and 3.7 report the estimated fitted values for *Group* for the two indexes of verbal fluency (number of switches and mean size of cluster) calculated by means of the Koren-derived procedure.

Koren – Number of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	12.21	3.65	7.62	4.3	-4.36	1.04	-4.19	p <.001	***
30th	13	3.62	8.03	4.34	-4.66	1.04	-4.48	p <.001	***
50th	13	3.62	8.03	4.34	-4.66	1.04	-4.48	p <.001	***
70th	13.47	3.47	8.35	4.57	-4.48	1.04	-4.31	p <.001	***
90th	12.88	3.51	8.5	4.47	-3.58	1.02	-3.52	p <.001	***
WEISS2									
10th	11.94	4.27	7.88	4.26	-3.8	1.11	-3.41	.001	**
30th	12.47	3.9	8.32	4.6	-3.96	1.12	-3.55	p <.001	***
50th	13.18	3.88	8.38	4.26	-4.5	1.06	-4.23	p <.001	***
70th	13.12	3.39	8.5	4.26	-4.29	1	-4.28	p <.001	***
90th	13.06	3.58	8.56	4.13	-4.11	1	-4.1	p <.001	***
LSA									
10th	11.06	3.9	6.24	3.43	-4.58	0.96	-4.77	p <.001	***
30th	12.06	4.26	6.79	3.67	-4.87	1.03	-4.71	p <.001	***
50th	12.97	4.27	7.53	3.78	-4.81	1.03	-4.65	p <.001	***
70th	12.76	4.03	7.68	3.79	-4.55	1.01	-4.51	p <.001	***
90th	12.76	3.32	8.09	4	-4.05	0.94	-4.31	p <.001	***

Table 3.6 Mean, SD, and estimated coefficients for Group of number of clusters computed by the Koren-derived computational algorithm.

Once having partialled out the effect of *Education*, which was significant only for number of clusters calculated with WEISS1 and having set the threshold at the 90th quantile (t = 2.17, *p* = .003), the results of the regression model on the number of switches indicated significant differences between the two Groups when this metric was computed using each and every semantic spaces, as well as considering thresholds set at any quantile: overall, healthy volunteers produced more switches than SSD Participants, with significance from *p* = .001 below.

Koren – Size of clusters

	HPs		SSD Participants						
Threshold	**mean**	**SD**	**mean**	**SD**	**estimate**	**Std.Err**	**t**	**p**	**sign**
WEISS1									
10th	5.13	1.72	5.3	1.95	.55	.47	1.18	.243	
30th	4.3	1.3	4.73	1.62	.74	.37	2	.049	*
50th	4.3	1.3	4.73	1.62	.74	.37	2	.049	*
70th	3.65	.9	4.09	1.28	.58	.29	2.04	.046	*
90th	3.32	.73	3.48	.98	.29	.22	1.31	.196	
WEISS2									
10th	5.43	2.09	5.45	3.17	.48	.69	.7	.488	
30th	4.85	1.78	5.11	3.26	.74	.67	1.11	.271	
50th	4.14	1.4	4.37	1.77	.5	.41	1.24	.220	
70th	3.86	1.09	4.24	1.79	.62	.38	1.62	.110	
90th	3.52	.96	3.68	1.16	.43	.26	1.63	.107	
LSA									
10th	5.47	1.7	6.62	2.96	1.4	.63	2.23	.029	*
30th	4.75	1.32	5.79	2.62	1.17	.54	2.15	.035	*
50th	4.14	1.04	4.74	1.6	.7	.35	1.98	.052	
70th	3.8	.8	4.38	1.51	.66	.32	2.1	.039	*
90th	3.38	.67	3.74	1.17	.45	.25	1.82	.074	

Table 3.7 Mean, SD, and estimated coefficients for Group of size of clusters computed by the Koren-derived computational algorithm.

Education was found to have a significant effect on some measures of cluster size derived by WEISS1 and WEISS2. Specifically, it was the only significant predictor for the mean size of clusters as calculated according to the 10th with WEISS1and according to the 90th quantile with WEISS2 (t = 2.20, p = .031; and t = 2.79, p = .007, respectively). For the mean size of clusters computed with WEISS1 according to the 30th and 50th quantiles, *Education* was a significant predictor (t = 2.30, p = .025 in both cases) along with *Group* (t = 2, p = .049 in both cases).

In all other cases, the effect of *Group* was the only significant predictor, and namely for size of clusters calculated using WEISS1: having set the threshold at the 70th quantiles, we found that on average HPs produced clusters smaller in size than SSD Participants (p <.05).

The same pattern, with HPs producing on average smaller clusters than SSD participants, was found using the LSA-derived semantic space and having set the threshold at the 10th, 30th, and 70th quantile (p <.05).

3.3.5 Comparison between automated and manual scoring at a classification task

We used the fitted values of logistic regressions with measures of verbal fluency as independent variables, and group as categorical dependent variable to create ROC curves. In particular, we considered the interaction of mean size of clusters and mean number of switches as computed for all the considered thresholds and semantic spaces.

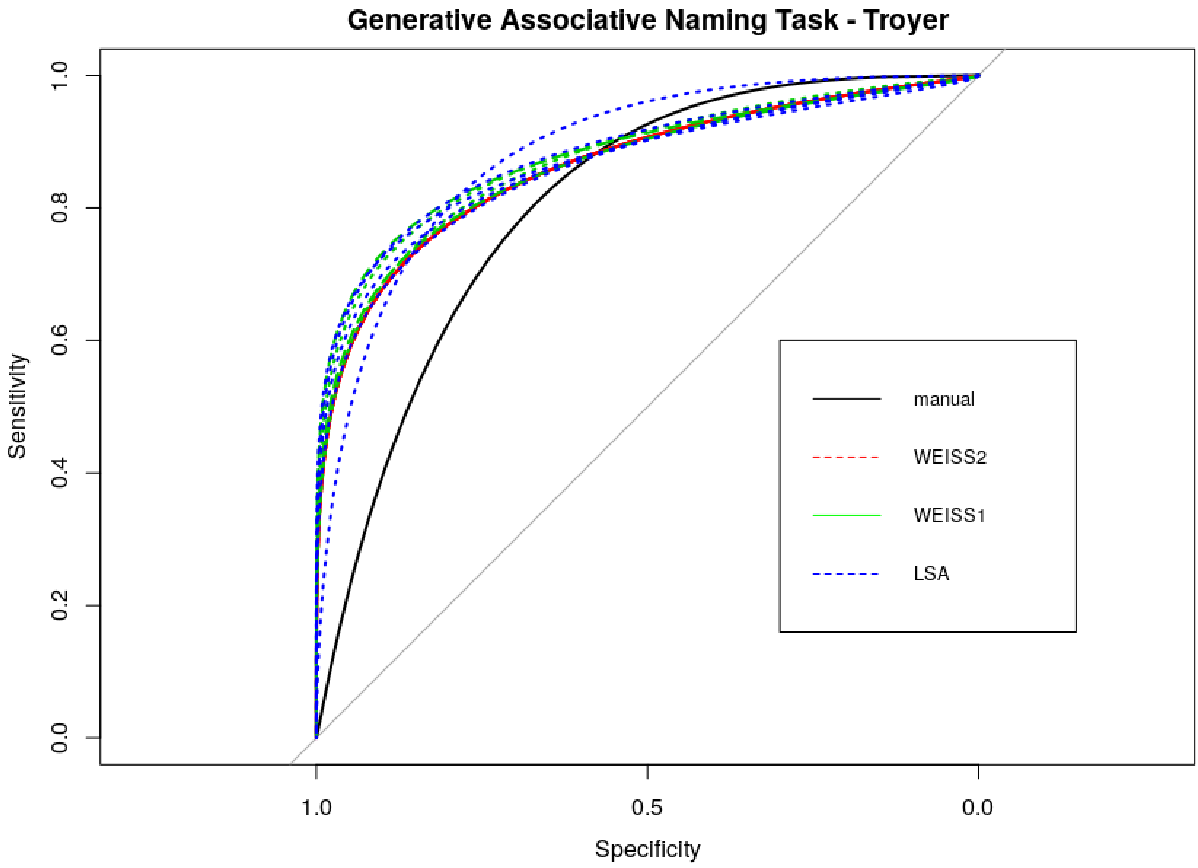


Figure 3.1 ROC curves for Troyer-derived measures of verbal fluency as estimated fitted values with *Group* as dependent variables

From visual inspection, Troyer-derived measures based on NLP methods (red, green, and blue lines) show higher specificity but lower sensitivity than those derived by manual ratings (black line) in the classification task. The overall performance of NLP-derived measures appears the fittest to the task, though. WEISS2 (red line) seems very robust, as the estimated values are homogeneous for all thresholds, and with good specificity, while the models based on LSA (blue lines) appear to have a higher AUC, with higher sensitivity. WEISS1, albeit being more performant than the manual rater, appears to be associated with a smaller AUC than the other two semantic spaces.

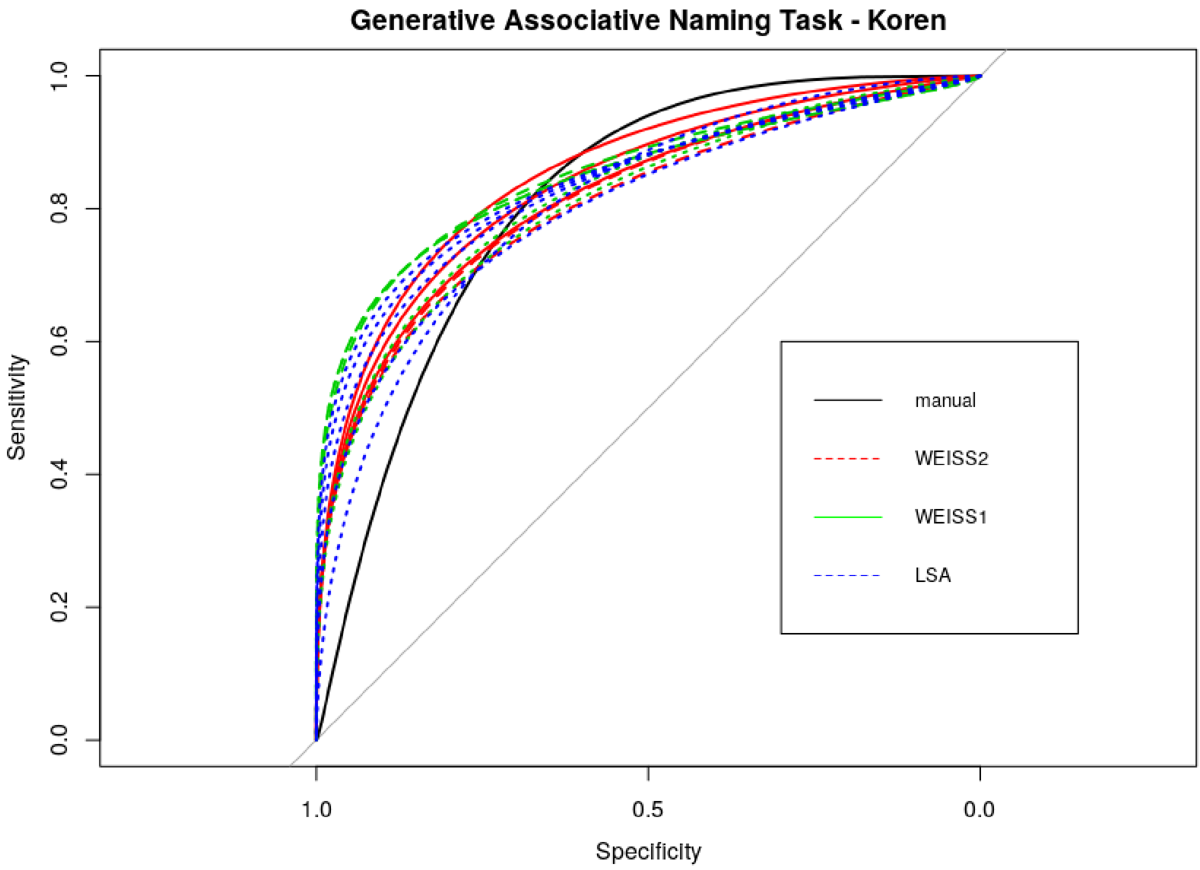


Figure 3.2. ROC curves for Troyer-derived measures of verbal fluency as estimated fitted values with Group as dependent variables

From visual inspection, Koren-derived measures based on NLP methods (red, green, and blue lines) appear, on average, to outperform the measures obtained from the manual rater (black line) in the classification task, albeit this latter shows the highest sensitivity. WEISS1 (green line) computes the highest specificity.

Figures 3.3 and 3.4 summarizes AUC values (Y-axis) of ROC curves computed based on the fitted values of the logistic regression model, with the Number of switches and Mean size of clusters calculated by the Troyer-derived and the Koren-derived algorithms (Figure 3.3 and 3.4, respectively) as continuous independent variables, and Group as dichotomous dependent variable. On the X-axis, we report the different thresholds used to define shifts.

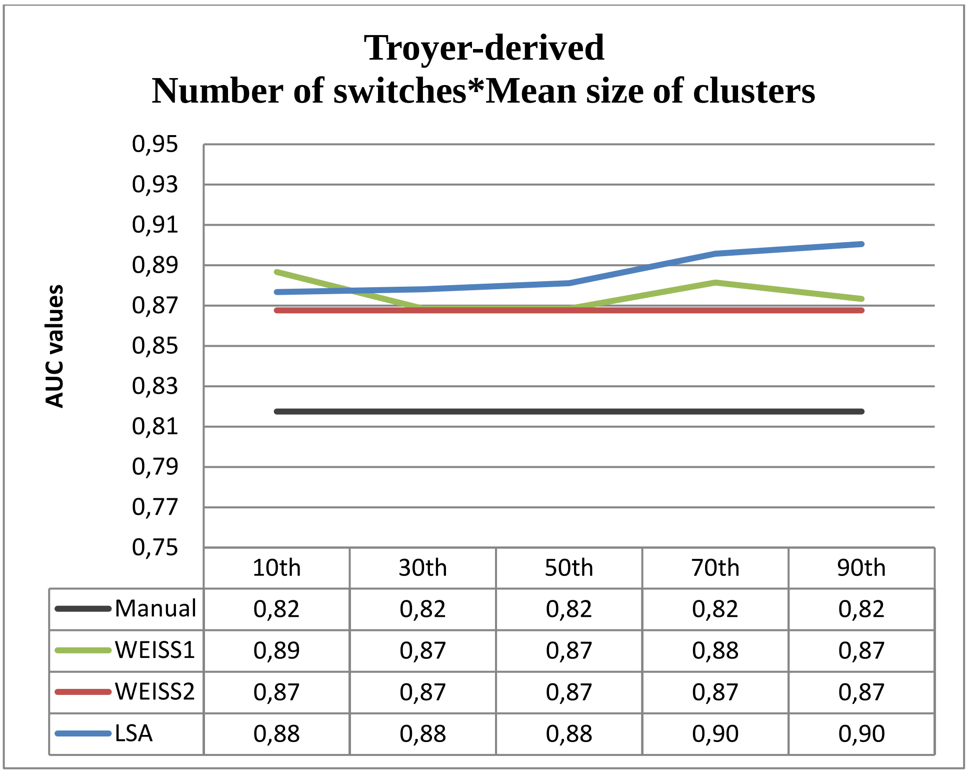


Figure 3.3 AUC values of Troyer-derived for number of switches.

The highest value (AUC = .90) was scored by the procedure using LSA and having set a threshold at .30 and .43 (70th and 90th percentile). This system outperformed the manual method (AUC = .82), as well as the best performances of both the algorithms based on WEISS1 (AUC = .89) and WEISS2 (AUC = .87).

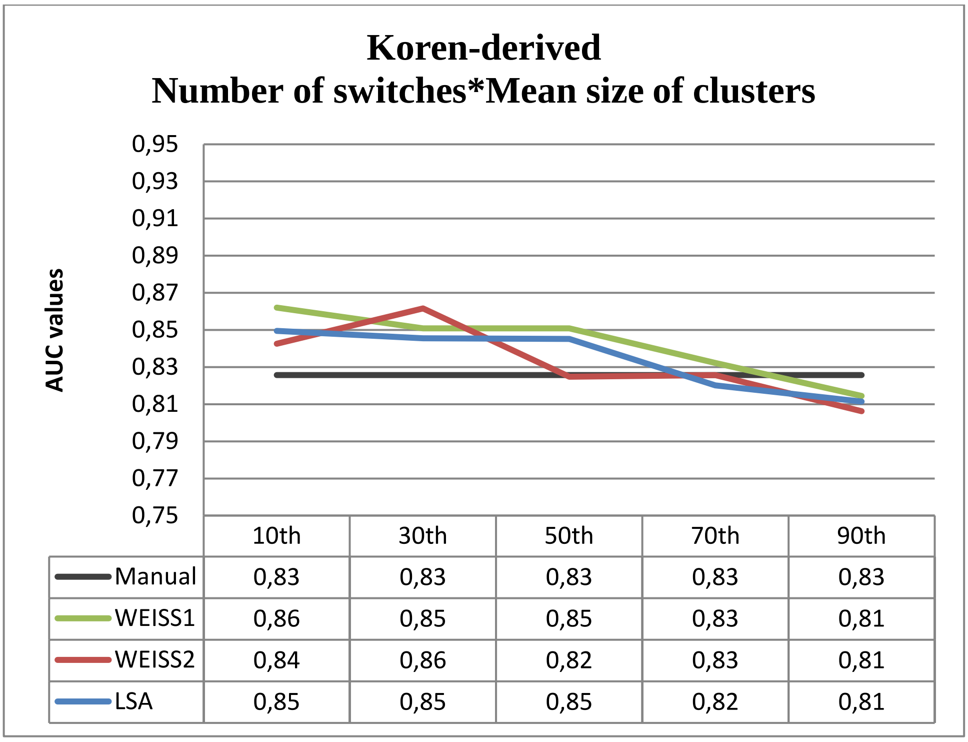


Figure 3.4 AUC values of Koren-derived for number of switches.

The highest values (AUC = .86) were scored by the procedure using WEISS1 and WEISS2 and having set a threshold at the 10th and 30th percentiles (.11 and .20, respectively). This system outperformed both the manual method (AUC = .83), as well as the best performances the algorithm based on and LSA (AUC = .85). A steep decrease in the AUC values of NLP-derived measures is visible for values computed according to the highest thresholds (particularly, from the 50th quantile onwards), yielding to NLP performances lower than the manual rating at the 90th quantile.

3.3.6 Analysis of coherence

Table 3.8 reports the estimated fitted values for *Group* in explaining cosine values between words at 1-, 3-, 5-, and 7-word in-list distances, as computed by the "Coherence" function on the basis of the three semantic spaces.

	HPs		SSD Participants						
	M	**SD**	**M**	**SD**	**estimate**	**Std.Error**	**t value**	**p-value**	**sign**
					WEISS1				
word N+1	.13	.02	.14	.02	.01	.01	2.79	.009	**
word N+3	.09	.02	.1	.03	.01	.01	1.61	.105	
word N+5	.08	.02	.08	.02	0	.01	.86	.227	
word N+7	.08	.02	.08	.02	0	0	.03	.833	
					WEISS2				
word N+1	.21	.03	.22	.03	.02	.01	2.69	.007	**
word N+3	.15	.02	.16	.04	.01	.01	1.65	.112	
word N+5	.14	.03	.14	.03	.01	.01	1.22	.395	
word N+7	.13	.02	.13	.03	0	.01	-.21	.977	
					LSA				
word N+1	.27	.03	.29	.04	.03	.01	2.88	.005	**
word N+3	.22	.03	.24	.04	.02	.01	2.42	.018	*
word N+5	.21	.03	.22	.04	.01	.01	1.36	.179	
word N+7	.21	.02	.2	.04	0	.01	-.46	.644	

Table 3.8 Fitted values for *Group* on cosine similarities between words at different distances, calculated according to the three different semantic spaces

Once having partialled out the effect of *Education*, which was significant for coherence values derived from WEISS1 and WEISS2 for adjacent words (t = 2.18, $p = .033$; and t = 2.06, $p = .049$, respectively), statistical differences were found in the mean cosine values between SSD participants and HPs for adjacent words: in this case, HPs always produced words pairs that are less related than those produced by SSD Participants ($p < .01$). The same pattern was found with respect to the mean cosine values of semantic relatedness between w_n and the 3^{th} consecutive words calculated with LSA, whereby HPs produced words less correlated than SSD participants ($p < .05$). No statistical difference was found between the mean cosine values for the two Groups for words with 5 and 7 interpolated items, as calculated on measures derived by all the three semantic spaces.

Figure 3.5 depicts the categorization ability of the estimated fitted values of the regression models having *Group* as dependent variable and measures of *Coherence* as independent variable. To maximize the performance of the classifier, we selected the best predictors based on number of switches and mean size of clusters for each semantic space (baseline) from the previous analyses and added it as a predictor to compute ROC curves.

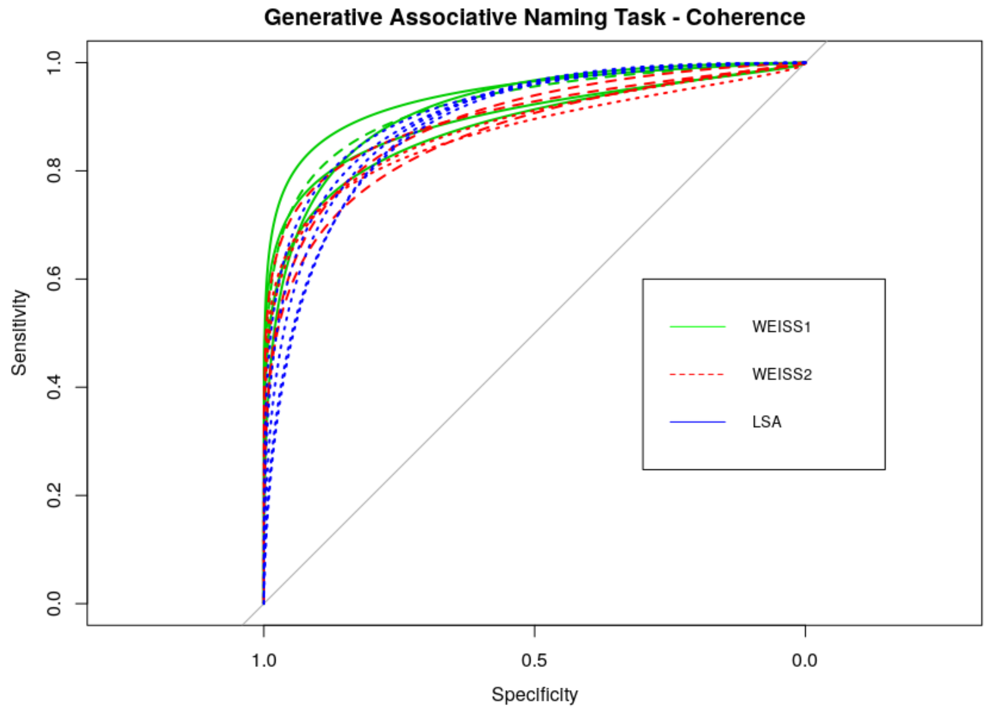
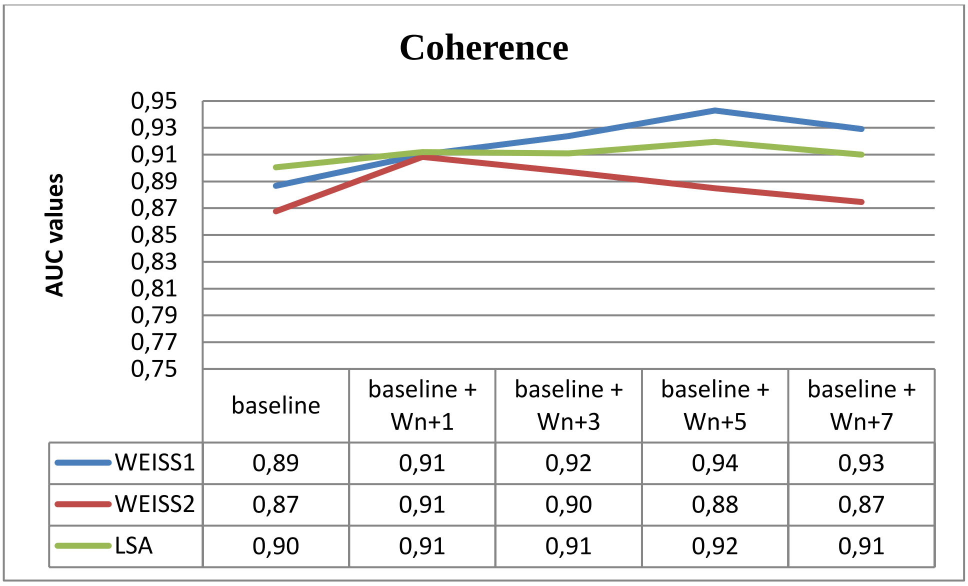


Figure 3.5 AUC values of ROC curves computed on values of classifiers based on the best set of parameters from the Troyer- and Koren-derived algorithms, plus the values from the coherence function (at different in-list distances) in a generative associative naming task.

From visual inspection, WEISS1 (green lines) seems to compute the highest AUC, with a good ratio between sensitivity and specificity, while the performance of the classifiers based on WEISS2 (red lines) and LSA (blue lines) seem to lag behind.



	baseline	baseline + Wn+1	baseline + Wn+3	baseline + Wn+5	baseline + Wn+7
WEISS1	0,89	0,91	0,92	0,94	0,93
WEISS2	0,87	0,91	0,90	0,88	0,87
LSA	0,90	0,91	0,91	0,92	0,91

Figure 3.6 AUC values of classifiers based on the best set of parameters from the Troyer- and Koren-derived algorithms ("baseline"), plus the values from the coherence function (at different in-list distances).

Figure 3.7 summarizes AUC values derived from ROC curves having as predictors coherence values, plus the best combination of number of switches and mean size of clusters for each semantic space derived from previous analyses ("baseline"). The highest value (AUC = .94) was scored by the procedure using WEISS1 and considering the mean cosine similarity between words with 5-word in-list distance. Its performance outperformed both the best AUC of WEISS2 (AUC = .91), as well as the best AUC of LSA (AUC = .92).

Comparison of the performances of verbal indexes in the classification task

In order to identify the set of verbal fluency scores that best discriminate the verbal performance of the two Groups, we compared the best AUC values, keeping in mind the rating suggested by Swets and colleagues (2000) as well as the limitations pointed out by Youngstrom (2014) (see Chapter 2). Table 3.9 report the best AUCs values for all the verbal fluency score indexes adopted in this study. As we can see, all measures of verbal fluency, including those derived by the manual annotation, show good classification performances.

Method	Rater/Semantic Space	Best AUC value
Troyer	Manual	.82
	WEISS1	.89
	WEISS2	.87
	LSA	.90
Koren	Manual	.83
	WEISS1	.86
	WEISS2	.86
	LSA	.85
Baseline + Coherence	**WEISS1**	**.94**
	WEISS2	.91
	LSA	.92

Table 3.9 Best performances of classifiers according to their AUC values

3.4 Discussion

In this study, we administered an associative generative naming task to a group of people with SSD and a matched HPs group. Participants were asked to produce as many words as possible after four clue words ("gatto"/*cat*, "scarpa"/*shoe*, "pioggia"/*rain*, "sciopero"/*strike*) in a time limit of 120 seconds per word. The aim of this study was firstly to characterize the verbal outputs of participants by identifying measures of integrity of the semantic store (mean size of clusters) and cognitive flexibility (mean number of switches between clusters). Secondly, we aimed at comparing the predictive ability of such measures in a categorization task.

The verbal productions of participants were analysed by computing the mean size of semantic clusters and the mean number of switches between them following two methodologies (Troyer et al., 1997; Koren et al, 2005). These measures were derived from four different sources: two vector space models created by a predict-model (word2vec) with different hyper-parameters (9-word window size and 400 dimensions for WEISS1, 5-word window size and 200 dimensions for WEISS2), one vector space model derived from a count-based algorithm (LSA), and measures computed by a manual annotator. Moreover, a measure of coherence obtained by computing the semantic similarity between words at different distances was derived from our three semantic spaces. For each measures of verbal fluency (both manual and automated) we developed a classifier (logistic regression models having the *Group* as dependent variable) and assessed their predictive ability by comparing their AUCs of ROC curves.

3.4.1 Standard quantitative measures

The quality of the collected data was confirmed by standard quantitative measures: people with SSD produced significantly less words than HPs, both in terms of overall number of items and of unique words (both with $p < .001$), in line with previous literatures (Bokat & Goldberg, 2003; and see Elvevag et al., 2001 for a concise overview). No significant differences were found in the number of repetitions and mean frequency of words produced by the two groups. However, the mean word length was statistically different between people with and without SSD ($p < .05$), whereby healthy volunteers produced, on average, longer words than SSD participants. Length of words is known to be inversely correlated with word frequency but not linearly (Sigurd et al., 2004), such as that the shortest words are also the most frequent. It is known that longer words take more time to be retrieved from semantic memory (Le Dorze & Durocher, 1992) and that high-frequency words are produced at the beginning of a fluency task (Crowe, 1998). Given the presence of working memory dysfunction in people with SSD (Gold et al., 2002) paired with the known psychomotor slowing in the this population (Morrens et al., 2007), this could explain the fact the, by the end of the time available, people with SSD would have produced only those set of short words readily available in their semantic store.

3.4.2 Number of switches and size of clusters

In terms of number of switches between clusters, fluency scores computed manually and by means of the proposed algorithms were aligned in patterns, whereby people with SSD always showed a significantly lower number of switches than HPs. When considered as a measure of cognitive flexibility, this result, taken alone, is in line with the literature that speaks for a frontal impairment in schizophrenia (Everett et al., 2001; Stirling et al., 2006; Zalla et al., 2001). The consistency of manual- and NLP-derived measures confirms the construct validity of the automated approach.

Significant differences in the mean size of clusters were visible only when considering NLP-derived measures, computed following both Troyer's and Koren's procedures and using all three semantic spaces at different thresholds. On the contrary, values computed with the manual procedure did not identify any significant differences between groups, despite a similar pattern was observable. These results speak for a greater sensitivity of NLP-derived measures of verbal fluency compared to those based on the manual annotation: the formers are able to catch subtle or "latent" similarity in meaning, which the manual scoring system, based on pre-compiled lists, may overlook. Most notably, contrary to our predictions, HPs always produced clusters *smaller* in size than SSD participants. We interpret this result as due to the nature of the task under investigation, rather than an exceedingly fitness of the semantic store of SSD participants. In a generative associative naming task, such as the one employed in the present study, participants are free to explore and retrieve words from the semantic store related to the given cue word, and not necessarily limited to a single semantic category. For example, in response to the cue word "cat", participants (both HPs and SSD participants) produced items such as "collare" (*collar*, an inanimate noun object), or "adorabile" (*adorable*, a modifier), or "dormire" (*to sleep*, a verb). If evaluated in terms of number of items produced, such flexibility appears the best strategy to complete the task at hand. People with SSD, however, seems to be unable to consistently implement such strategy, and tend to remain bound to the same semantic (sub)field: this approach will generate words strongly correlated in terms of semantic relatedness, which are considered clusters by our algorithms, but appears to be a suboptimal strategy leading to a lower total number of items produced in the given timeframe.

3.4.3 Coherence

With respect to the semantic coherence of the items produced, results indicate significant differences between groups for adjacent words and, when assessed by means of the LSA-derived measures, also for words with 3 interpolated items. Contrary to our prediction, in these cases healthy participants show a *lower* index of semantic coherence than SSD participants. Also, in this case, this finding seems to call into question the interpretation of the verbal production of the task under investigation. In fact, loose associations between words are permitted or, perhaps, even required by tasks such as generative associative naming task: the best strategy to ensure a steady flow of items for 120sec seems to be producing words loosely associated with each other, rather than pre-organize the semantic store around a prototypical semantic cluster to be explored. If we look at the semantic store as a network of interconnected nodes, each node defining a concept (Collins & Loftus, 1975), we can speculate that here subjects are adopting the cognitive equivalent of "breadth-first search" heuristic (Cormen et al., 2002) to produce the responses. This means that, starting from the concept node representing the cue word provided at the beginning of the task, subjects begin exploring the network by getting to those nodes immediately reachable from there. These will be the nodes in the immediate surroundings connected by links having equal weights. The opposite strategy, a "depth-first" search, is, as the name implies, to search "deeper" in the network whenever possible. In this way, subjects would try and reach nodes within a specific branch of the network (what we would call a semantic clusters) until exhausted and would then "backtrack" to the original node to restart the process. It follows that careful attention to the specific task at hand must be paid for the interpretation of a measure of cognitive performance such as the mean size of clusters: the oversimplification "the bigger the clusters, the better" can be misleading if the task at hand, as in this case, prompt subjects to adopt a breadth-first strategy, recruiting frontal executive functions. In this sense, the contribution of prefrontal functions (such as cognitive flexibility) appears fully appreciable in this task, differently from the integrity of the semantic store.

Taken this difference into consideration, our results are not directly comparable to previous work assessing semantic coherence in a single-word association task using LSA measure of semantic relatedness (Elvevåg et al., 2007), which found schizophrenic participants to produce words less semantically related to the index item than HPs.

3.4.4 Performance of classifiers

In line with our hypothesis, NLP-derived measures always outperformed scoring derived by manual annotations in the classification task.

The best performance (AUC = .94) was obtained by on a classifier based on values of number of switches and mean size of cluster computed according to Troyer's method at the 10th quantile as derived from WEISS1 (a predict-model), adding to it a measure of "coherence" in order to maximize its discriminative performance. The next best classifier (AUC = .90) was the one based on the number of switches and mean size of clusters computed following Troyer's method and considering a threshold set at the 90th quantile of the LSA space. Manual-derived measures, on the contrary, scored the lowest (AUC = .83). We interpret these results as an evidence of NLP-derived measures of semantic relatedness being able to catch subtle nuances of meaning, which a rigid list-based approach, such as the original scoring system proposed by Spinnler and Tognoni (1987), cannot but overlook.

In particular, when a measure of "coherence" was added to indexes of cluster size and number of switches as predictors, the performance of the classifier improved consistently, and reached a remarkable high performance. By comparing the different values of cosine proximity at different intervals (w_{n+1}, w_{n+3}, … etc.), we were able to appreciate the different temporal "trajectories" that the two groups were following when exploring the semantic store during the task, and how the content of the semantic store unfolded following the spread of the activation process, that would otherwise remain unseen considering simply the count of words. Proceeding through the test, the intact and functional architecture of the semantic store of healthy subjects enables them to moves further, bridging "*wide gaps in making associative responses*" (Wilson et al., 1953). This appears to be the best strategy, that is, the one that eventually lead to a higher number of words produced, and highlights the contribution of diverged thinking (Forthmann et al., 2018) in a generative associative naming task. Contrary from our expectations, our cohort of SSD participants appears in fact to be highly "coherent" in their production, to the point to be rather inflexible. In this sense, with this proposed measure of "coherence", we are operationalizing and quantifying the otherwise opaque notion of "rigidity" that has been used to describe the speech of people with SSD, observed in production but also, for example, in the poor performance in the interpretation of metaphors and idioms in this population (Tavano et al., 2008).

Nonetheless, regardless of the better performance, results derived from the semantic spaces created with word2vec must be taken with a note of caution, given the observed confounding effect of education. Education is known to exert a significant effect on verbal fluency tasks (Novelli et al., 1991), presumably for the advantage that formal education gives in abstract reasoning and in the manipulation of metacognitive materials. Previous studies assessing neuropsychological functioning in monozygotic twins discordant for schizophrenia (Goldberg et al., 1990) found that ill twins completed on average almost 2 years less education than their unaffected cotwins. This result has been interpreted as indicative of the difference between educational potential versus actual attainment in individual with schizophrenia, presumably a consequence of the disease progress. Schizophrenia onset is typically in late adolescence or early adulthood and might reasonably expect to truncate educational attainment in affected individuals. In this sense, as noted by Resnik (Resnik, 1992) matching for education may lead to systematic mismatching, i.e., the selection of atypical groups of high-achieving patients or low-achieving controls. From this point of view, cosine values computed with the LSA-derived space, despite the relatively lower performance, showed to be relatively independent from this effect.

This result needs to be interpreted by taking into consideration the *kind* of distributional information mediated by the different algorithms we used. As already pointed out (see Section 2.4), a fundamental distinction must be drawn between paradigmatic and syntagmatic relationships (Sahlgren, 2008), whereby the former would relate entities that co-occur "*in absentia*" (such as words that share the same context, but not at the same time), while the latter would characterize the relationship between words "*in presentia*" (as between words within the same sentence).

In the case of semantic spaces derived from co-occurrences considering the whole document as context (as for our LSA space), word vectors are more likely to capture their paradigmatic relationships. On the contrary, semantic spaces derived from co-occurrences computed according to a sliding-window with size comparable to that of a single sentence (word2vec), word vector representations would be more likely to capture their syntagmatic relationships.

This consideration can help explain the better performance of word2vec compared to LSA in an associative naming task: as observed, the items produced in response to the cue word "cat", ("*collare*", "*adorabile*", or "*dormire*") are in a paradigmatic relationship. In this sense, the best performance of the WEISS1 model is likely to be found in its incorporation of distributional properties of words significantly affected by the level of education, most likely in their syntagmatic relationship.

3.4.5 Conclusions

To conclude, this study showed that applying measures of verbal fluency, such as number of switches and mean size of clusters, to an associative generative naming task can provide meaningful insight into the cognitive processes underlying the performance of people with SSD. This work highlighted the predominant contribution of frontal executive functions *vis-a-vis* the integrity of the semantic store to the SSD responses in this task. Moreover, in the last years, the application of NLP-derived measures to the study of mental disorders yielded remarkable results (Corcoran et al., 2018; de Boer et al., 2018; Pauselli et al., 2018; Sarioglu Kayi et al., 2017) and our study further support the timeliness of a re-conceptualization of language in the context of mental health, considering that relying on existing tests for language assessment in the psychiatric field have been deemed to provide "grossly inadequate view of language abilities" (Elvevåg et al., 2016).

Moreover, our results proved the advantage of NLP-derived measures compared to the traditional manual scoring system in a participant-classification task. Finally, our findings support the idea that, despite the hype around predict models, LSA-derived measures have still something to say, as they did not lag behind the measures derived from predict-models in our classification task and were consistently independent from the effect of the level of education.

In order to understand the meaning of these results, it is important to keep in mind our operational definition of the outcome, i.e., our diagnoses. In our case, one could argue that the application of a rather broad inclusion criterion such as the decision to recruit participants with any diagnosis within the schizophrenia spectrum might have led to a quite heterogeneous sample. In fact, the psychopathological construct of SSD is remarkably complex, and the clinical presentation of the disorder can be strikingly different (the effect of the sampling design on the diagnostic efficiency of a test is known – Youngstrom, 2014). In a real clinical setting, it is unlikely that a clinician is faced with decision whether a subject has schizophrenia versus no mental issue at all. What our best AUC measure is telling us is simply that it is likely that "there may something wrong with the speech of the subject", which could deserve further investigation. In this sense, we believe that the proof-of-concept here proposed could add to the current clinical diagnostic test commonly used in the clinical practice, which are deemed to reach AUC = .70/.80 in real diagnostic conditions (Youngstrom, 2014).

Further research should aim at optimizing the classifier, both in terms of sensitivity, by subgrouping people with SSD according to psychotic symptoms, as well as of specificity.

4 Processing of Argument Structure and Syntactic Complexity in People with SSD

4.1 Introduction

Early works on language in SSD carried out using tools developed to test language comprehension and production of aphasic patients found distinct language profiles in these two groups (DiSimoni et al., 1977), to the point that performance of schizophrenic patients was found to be indistinguishable from that of HPs (Rausch et al., 1980). These early findings led some researchers to conclude that poor performance on aphasia screening batteries by people with SSD is merely a function of intellectual impairment (Oh et al., 2002). In the past few decades, however, a number of studies have consistently confirmed the presence of an impaired processing of complex syntax in schizophrenic participants, both in terms of production and comprehension, to the point that the most recent diagnostic guidelines (APA, 2013) compared the severely disrupted language production in this population, also called schizophrenic “word salad”, to receptive aphasia. Consequently, lexical and syntactic aspects of language production in SSD have been object of renewed attention, also thanks to the progress of linguistic theory and psycholinguistic tools.

With respect to verb production, early studies employing verb generation tasks found a specific difficulty for action verbs versus mental state verbs in this population (Marvel et al., 2004). Similarly, on action (verb) fluency tasks (Woods et al., 2007), performance of people with SSD was approximately one standard deviation below HPs, suggestive of a possible impaired mental representation of (lexical) actions. Further studies supported these findings (Kambanaros et al., 2010) and suggested a possible specific verb impairment at the lemma level, postulating a deficit in mapping semantics to (verb) word forms.

No studies so far, however, have taken into consideration the Argument Structure Complexity Hypothesis (ASCH - Thompson, 2003) as a theoretical framework to explain the production and comprehension of verbs in this population. Previous studies showed that verb argument structure recruit bilateral hemispheric areas following the scalability of its argument structure. In particular, neuroimaging studies on HPs showed that verbs differing in the number of their obligatory arguments are processed differently in the brain, with a bilateral activation in case of verb having more than one argument (Thompson et al., 2007; Thompson et al., 2010) thus requiring a functional inter-hemispheric connectivity. Moreover, previous studies on agrammatic subjects (Kim & Thompson, 2000) showed that the selection and retrieval of a verb involves the automatic access to its lexical entry (which includes information about their argument structure), and as the number of syntactic arguments increases, so does the verb selection difficulty. As for people with SSD, neuroimaging studies have consistently found a reduced lateralization of language functions (Sommer et al., 2003; Weiss et al., 2006; Spaniel et al., 2007) as well as a reduce inter-hemispheric connectivity (Chang et al., 2019), which has been proposed as specific phenotype of the disorder. For these reasons, in people with SSD we should expect to see an increasing difficulty in accessing the lemma representations of verb argument structure, recruiting a bilateral hemispheric activation, along the direction proposed by the ASCH.

With respect to syntactic complexity as source of processing demands, previous studies indicated the presence of a deficit in the comprehension and production of complex syntactic structures in this population. However, the linguistic material used in the different studies appears to be rather heterogeneous, with rather different level of complexity, and not always generalizable. For example, Morice and McNicol (1985) used a modified version of the Token test (De Renzi & Vignolo, 1962) including complex sentences with one and two levels of clausal embedding, left and right branching, embedded verbless clauses and non-embedded coordinated clauses, such as (1):

(1) *Before touching the yellow circle, pick up the circle above the square that is next to the yellow square*

(Example of a complex sentence with two level of clausal embedding, three dependent clauses – adverbial, verbless, and full relative – and left and right branching, with resulting structure: A [sb V O] V O [A [rel V A]]). The Authors found a deficit in the comprehension of syntactically complex sentences in schizophrenia. Likewise, a poor comprehension accuracy concerning the meaning of syntactically complex sentences (Condray et al., 2002) was found in SSD participants by testing the comprehension of simple sentences *vs* subject- *vs* object-relative sentences such as in (2):

(2) *The reporter that the senator attacked admitted the error*

Similarly, Bagner and colleagues (2003) found that increasing sentence length (simple short and long sentences) and complexity (center-embedded subject-relative *vs* center-embedded object-relative sentences) such as in (3):

(3) *The candidate that the governor endorsed lost the campaign*

had a greater impact on language comprehension performance in SSD patients than in control participants. Tavano and colleagues (2008) used a picture matching test utilising as linguistic prompt either locative, active and passive negative, relative, and dative sentences, such as in (4):

(4) *The girl gives the ball to the boy*

The Authors hypothesized that an impaired access to syntactic structure could explain the poor performance of schizophrenic participants to the task. Previous studies adopting the theoretical framework of Chomsky's Universal Grammar (Moro et al., 2015) have demonstrated the presence of an impaired syntactic (but not semantic) knowledge in SSD patients. Participants were asked to identify grammatical (syntactic) anomalies embedded in either short or long sentences, and violating: i) the locality principle with question formation, as in (5); ii) the locality principle with clitic constructions, as in (6); and iii) wrong contrastive focus involving inversion) as in (7):

(5) **Who does John want to contact the nurse before meeting*

(6) **Of these pictures, Maria of-them$_{clitic}$ thinks that Gianni wants to see two*

(7) **Not arrives Gianni but leaves*

or a form of semantic errors resulting from a contradiction in the computation of the whole sentence meaning, as in (8):

(8) *I will dry the laundry with water*

Given the wide heterogeneity of the approaches adopted in the previous literature, it is difficult, if not impossible, generalizing a distinctive effect of sentence complexity in this population.

According to a generative approach to syntax (Chomsky, 1965; Chomsky, 1981), active SVO sentences are considered as having canonical word order, while, in sentences with non-canonical word order, the object is moved from its original position, and surfaces in the clause-initial position via syntactic operation. It follows that we can devise three pairs of sentence types differing only by the type of syntactic movement involved in their construct (Fig.4.1):

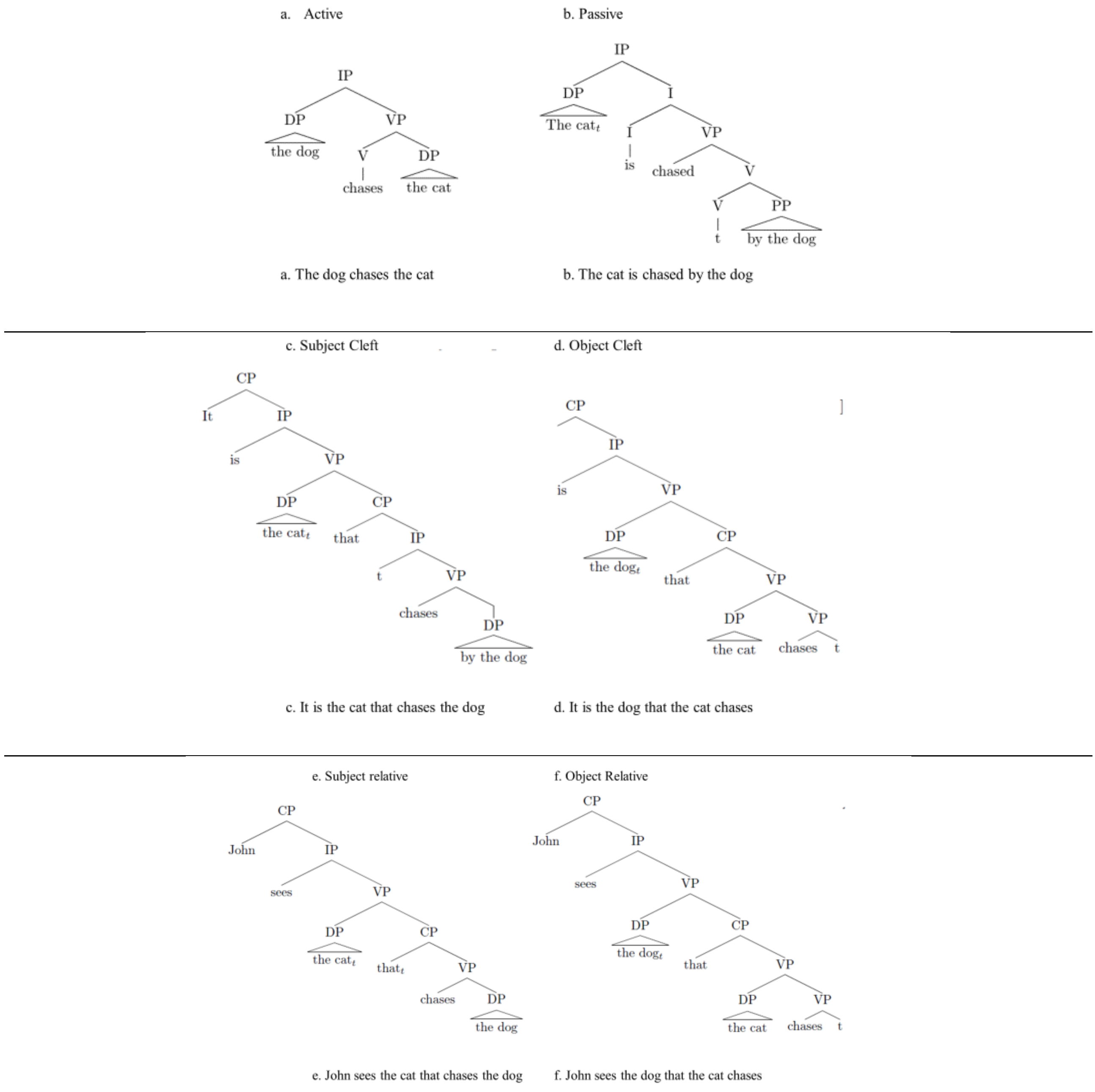


Figure 4.1 Differnt types of syntactic movements.

1. Active sentences (Fig. 4.1a) have canonical word order, i.e., the d(eep)-structure is identical to the s(urface)-structure. On the contrary, the s-structure of passive sentences (Fig. 4.1b) results from a syntactic operation that moves the object across the verb, eventually surfacing in pre-verbal position;
2. Cleft sentences are complex sentences with a main clause and a dependent clause, in which one constituent is moved up along the syntactic structure, in order to be put into focus. In the case of subject-cleft sentences (Fig. 4.1c), the subject of the main clause is moved via syntactic movement across the clause boundary. The s-structure, however, does not change. On the contrary, when the focused element is the object, as in object-cleft sentences (Fig. 4.1d), such operation implies a movement of the constituent across the subject of the main clause, to a pre-verbal position;
3. Relative clauses are subordinate clauses in which an element (typically, a relative pronoun) is interpreted by an antecedent in the main clause. In subject relative clauses (Fig. 4.1e) the subject of the dependent clause is moved via syntactic operation across the sentence boundary in the object position of the main clause. However, this operation does not change the surface structure of the sentence. On the contrary, in the case of object relative clauses (Fig. 4.1f) the direct moves across the subject in a pre-verbal position.

The aim of the present study is twofold. Firstly, it aims at assessing the presence of any impairment in verb and sentence processing with respect to verb argument structure (in terms of the number of arguments selected by verbs, i.e., one-, two-, or three-argument verbs) compared to HPs that were matched by age and gender. Secondly, it aims at assessing the presence of any impairment in the processing of syntactic complexity (with respect to order of constituents in the sentence, i.e., canonical *vs* non canonical, and type of sentences), compared to matched HPs. As the Northwestern Assessment of Verb Structure (NAVS) (Cho-Reyes & Thompson, 2012) addresses both fields, it appears to be the best tool for an in-depth characterization of language performance of people with SSD, exploring the processing of lexical properties of verbs, as well as the performance of these patients with respect to syntactic complexity. The value of the proposed framework relies in the assumption that linguistic complexity is independent from the clinical state of the subject. As such, both ASCH and syntactic complexity as framed here should be extendable to different clinical population, from aphasia to SSD. Based on the available evidences, we expect patients with SSD to: i) show a greater difficulties in producing three- and two-argument verbs *vs* one-argument verbs, than HPs; and ii) be less accurate than HPs on production and comprehension tasks involving a manipulation of syntactic complexity: in particular, we expect greater difficulties for non-canonical *vs* canonical sentences, and specifically for Passive *vs* Active forms, Object *vs* Subject Cleft sentences, and Object *vs* Subject

Relative sentences; and iii) with respect to sentence type, based on previous literature (Lelekov et al., 2000; Bagner et al., 2003), we expected an effect of sentence type in this group, with a decreased accuracy for sentences with a complex syntactic structure.

4.2 Materials and methods

4.2.1 Participants

Thirty-six participants with a diagnosis of SSD according to DSM-5 (APA, 2013), aged 18-65, were recruited from the outpatient unit and the residential facilities of the IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy between February 2018 and April 2019. Three participants were excluded, leaving a final sample of thirty-three participants. Reasons for exclusion were: unexpected technical (audio-recorder) issues preventing the subsequent test scoring (N = 1), refusal or lack of cooperation while undergoing the NAVS assessment (N = 2).

Clinical diagnoses of SSD was made by the treating clinicians (staff psychiatrists). All SSD participants were able to give informed consent, had normal or corrected-to-normal visual acuity, and were right-handed. The diagnosis on the Axis I had to be unique, while co-morbidities on the Axis II were admitted. Exclusion criteria included neurological disorders, head trauma with cognitive sequelae, mental retardation, and substance abuse in the 3 months preceding the enrolment. At time of recruitment, SSD participants were on treatment with at least one anti-psychotic medication (M = 2.45, SD = 1.62) for at least the previous 6 months. The mean years of illness of the SSD participants was 23.41 (SD = 12.23, N = 32 – for one SSD participant, it was not possible to date the age of onset).

A paired sample of thirty-three healthy control participants was recruited among the hospital staff and the general population, matched by age and gender. Education was found significantly different in the two groups (SSD < Control participants: $t = -3.40$, $p < .01$) and, for this reason, it was added as fixed independent variable to all statistical models.

Exclusion criteria for the control group were the same as for the SSD sample, as well as any documented psychiatric disorders or being first-degree relative of a patient with diagnosis of SSD. All Participants were native speakers of Italian.

After a complete description of the study, informed consent to participation was obtained from all participants. In case of patients with support administration, the participation to the study was first discussed with the patient, then, written consent was obtained both from the patient and the appointed administrator. The study was approved by the IRCCS Ethical Committee (Opinion 61/2017) and followed the principles of the Helsinki Declaration.

4.2.2 Task

As part of a larger assessment battery, study participants were administered the Northwestern Assessment of Verbs and Sentences (NAVS) (Cho-Reyes & Thompson, 2012) in its Italian adaptation (Barbieri et al., 2019). The first three tasks (Verb Naming Task – VNT; Verb Comprehension Task – VCT; and Argument Structure Production Task – ASPT) were developed to assess production (VNT and ASPT) and comprehension (VCT) performance of verbs either in isolation (VNT and VCT) or within a sentence context (ASPT), with respect to the number of verb arguments (one-, two- and three-argument verbs).

a) b) c)

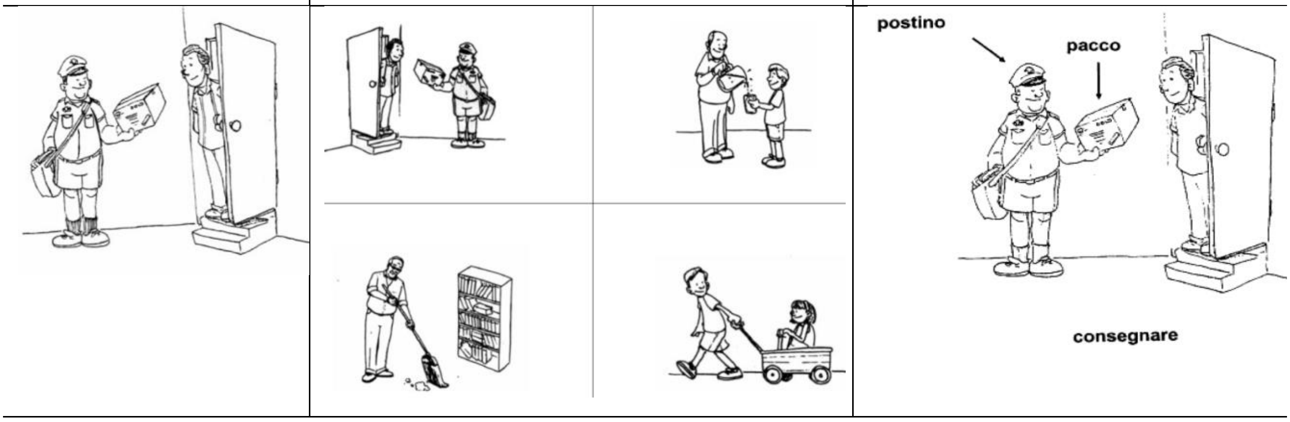


Fig. 4.2 Example items from VNT (a), VCT (b), and ASPT (c). In all pictures, the target verb is "*consegnare*" (to deliver). In (a), the participant was instructed to name the action using one word; in (b) the participant had to indicate which one out of the four pictures illustrated "*consegnare*" (top left), with distractors bearing same argument structure (as for "*versare*," to pour, top right) or different argument structure (as for "*spazzare*", to mop, in the bottom left corner, or "*tirare*", to pull, in the bottom right corner); in (c), participants were instructed to describe the picture using a sentence that employed all the provided words (i.e. "*Il postino consegna un pacco all'uomo*", The postman is delivering a parcel to the man).

The last two tasks (Sentence Production Priming Task – SPPT; and Sentence Comprehension Task – SCT) were designed to study production and comprehension of sentences according to different syntactic complexity. Complexity was designed either as a function of the order of constituents (canonical *vs* non-canonical) or of the type of syntactic movement involved: in case of passive, the moved element is a noun phrase, while in the case of object-cleft and object-relative sentences the moved element is a wh-phrase.

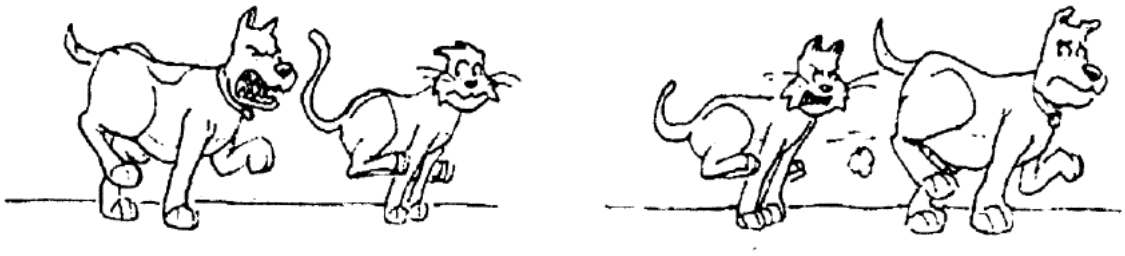


Fig. 4.3 Example item from the SPPT and SCT. The figures depict the action "*rincorrere*" (to chase). In the SPPT, participants heard a prime sentence, "*Il gatto è rincorso dal cane*" (the cat is chased by the dog), matching the picture on the left, and were instructed to produce a similar sentence for the picture on the right (i.e. "*Il cane è rincorso dal gatto*", the dog is chased by the cat). In the SCT, the sentence "*Il gatto è rincorso dal cane*" (the cat is chased by the dog), was provided and participants had to indicate which picture matched the sentence (i.e. the picture on the left).

A paper-and-pencil version of the test was employed. In the VNT task (Fig.4.2a), participants were presented with pictures depicting target verbs and were instructed to name the action using a verb. In the VCT task (Fig.4.2b), the researcher named a target verb and participants had to indicate which one out of four pictures illustrated the target item. In the ASPT task (Fig.4.2c), participants were required to generate a sentence that described a given picture by employing words that were provided in written form. In the SPPT (Fig.4.3), participants heard a prime sentence, matching a picture on the left side of the stimulus book, and were instructed to produce a similar sentence for the picture pair on the right side. In the SCT task (Fig.4.3), a sentence was provided, and participants had to indicate which of two pictures matched the stimulus sentence (same pictures used for the SPPT task).

4.2.3 Data analyses

All the analyses were performed in R, v.3.6 (R CoreTeam, 2015). Linear mixed-effects regression analyses (Baayen et al., 2008) were run on accuracy data (as continuous dependent variable, i.e., correct responses/total number of items) of all tasks. Random intercepts for participants were also included. Post-hoc comparisons for number of arguments (one- *vs* two-argument verbs, one- *vs* three-argument verbs, and one- *vs* three-argument verbs) in the case of tasks on verb argument structure (VNT, VCT, and ASPT), and for sentence type and word order in the case of task of syntactic complexity (SPPT and SCT) were carried out, estimating marginal means (least-squares means), adjusted with Tukey's method. All contrasts were estimated with respect to the base level which, in our case, was: i) Group: HPs; ii) Type: active sentences; iii) Order: canonical. Stated differently, differences are estimated compared to the data produced by HPs when processing active sentences having canonical word order.

4.3 Results

4.3.1 Accuracy – all tasks

Table 4.1 and 4.2 report the accuracy rate of healthy controls and SSD participants for all the NAVS tasks.

	HPs		SSD participants	
	M	**SD**	**M**	**SD**
VNT				
1 argument	99.39	3.48	90.91	15.88
2 arguments	97.27	8.01	78.48	22.24
3 arguments	93.07	11.36	63.34	25.78
VCT				
1 argument	99.39	3.48	99.39	3.48
2 arguments	100.00	-	98.48	4.42
3 arguments	100.00	-	98.3	4.64
ASPT				
1 argument	100.00	-	98.79	6.96
2 arguments	98.58	4.33	92.33	13.67
3 arguments	99.76	1.39	94.18	11.15

Table 4.1 Accuracy rate for the NAVS lexical tasks (VNT, VCT, and ASPT)

	HPs		SSD participants	
	M	**SD**	**M**	**SD**
SPPT				
Active	99.39	3.48	96.97	10.15
Passive	98.18	7.69	78.79	35.33
Subject-cleft	99.39	3.48	82.42	32.69
Object-cleft	92.73	19.9	52.73	45.23
Subject-relative	99.39	3.48	75.76	34.19
Object-relative	91.52	18.05	46.67	46.28
SCT				
Active	100.00	-	96.97	17.41
Passive	100.00	-	87.88	23.42
Subject-cleft	98.79	4.85	95.76	17.86
Object-cleft	87.27	19.25	64.85	31.24
Subject-relative	100.00	-	95.15	18.05
Object-relative	93.33	15.55	61.12	33.52

Table 4.2 Accuracy rate for the NAVS syntactic complexity tasks (SPPT and SCT)

Control participants

Healthy control participants performed nearly at ceiling on all tasks.

For the VNT mean percentage correct production of one-, two-, and three-argument verbs were 99.39%, (SD = 3.48), 97.27% (SD = 8.01), and 93.07% (SD =11.36), respectively. For the VCT mean percentage correct production of one-, two-, and three-argument verbs were 99.39%, (SD = 3.48), 100%, and 100%, respectively. For the ASPT mean percentage correct production of verbs and verb arguments were 100%, 98.58% (SD = 4.33), and 99.76% (SD = 1.39) for one-, two-, and three-argument verbs, respectively.

On the SPPT mean percentage correct production of Active, Passive, Subject Cleft, Object Cleft, Subject Relative, and Object Relative sentences were 99.39% (SD = 3.48), 98.18% (SD = 7.69), and 99.39% (SD = 3.48), 92.73% (SD = 19.9), 99.39% (SD = 3.48), and 91.52% (SD = 18.05), respectively. For the SCT mean percentage correct comprehension of Active, Passive, and Subject Relative sentences was100%, whereas that of Subject Cleft, Object Cleft, and Object Relative sentences were 98.79% (SD = 4.85), 87.27 (SD = 19.25) and 93.33% (SD = 15.55), respectively.

SSD Participants

For the VNT, mean percentage correct production of one-, two-, and three-argument verbs were 90.91%, (SD = 15.88), 78.48% (SD = 22.24), and 63.34% (SD =25.78), respectively. For the VCT mean percentage correct production of one-, two-, and three-argument verbs were 99.39%, (SD = 3.48), 98.48% (SD = 4.42) and 98.3% (SD = 4.64), respectively. For the ASPT, mean percentage correct production of verbs and verb arguments were 98.79% (SD = 6.96), 92.33% (SD = 13.67), and 94.18% (SD = 11.15) for one-, two-, and three-argument verbs, respectively.

On the SPPT, mean percentage correct production of active, passive, Subject Cleft, Object Cleft, Subject Relative, and Object Relative sentences were 96.97% (SD = 10.15), 78.79% (SD = 35.33), 82.42% (SD = 32.69), 52.73% (SD = 45.23), 75.76% (SD = 34.19), and 46.67% (SD = 46.28), respectively. For the SCT, mean percentage correct comprehension of active, passive, Subject Cleft, Object Cleft, Subject Relative, and Object Relative sentences were 96.97% (SD = 17.41), 87.88% (SD = 23.42), 95.76% (SD = 17.86), 64.85% (SD = 31.24), 95.15% (SD = 18.05), and 61.12% (SD = 33.52), respectively.

4.3.2 Verb Argument Structure

4.3.2.1 Production of single verbs - VNT

Figure 4.4 depicts the accuracy rate for the VNT task for one-, two-, and three-argument verbs in the two Groups.

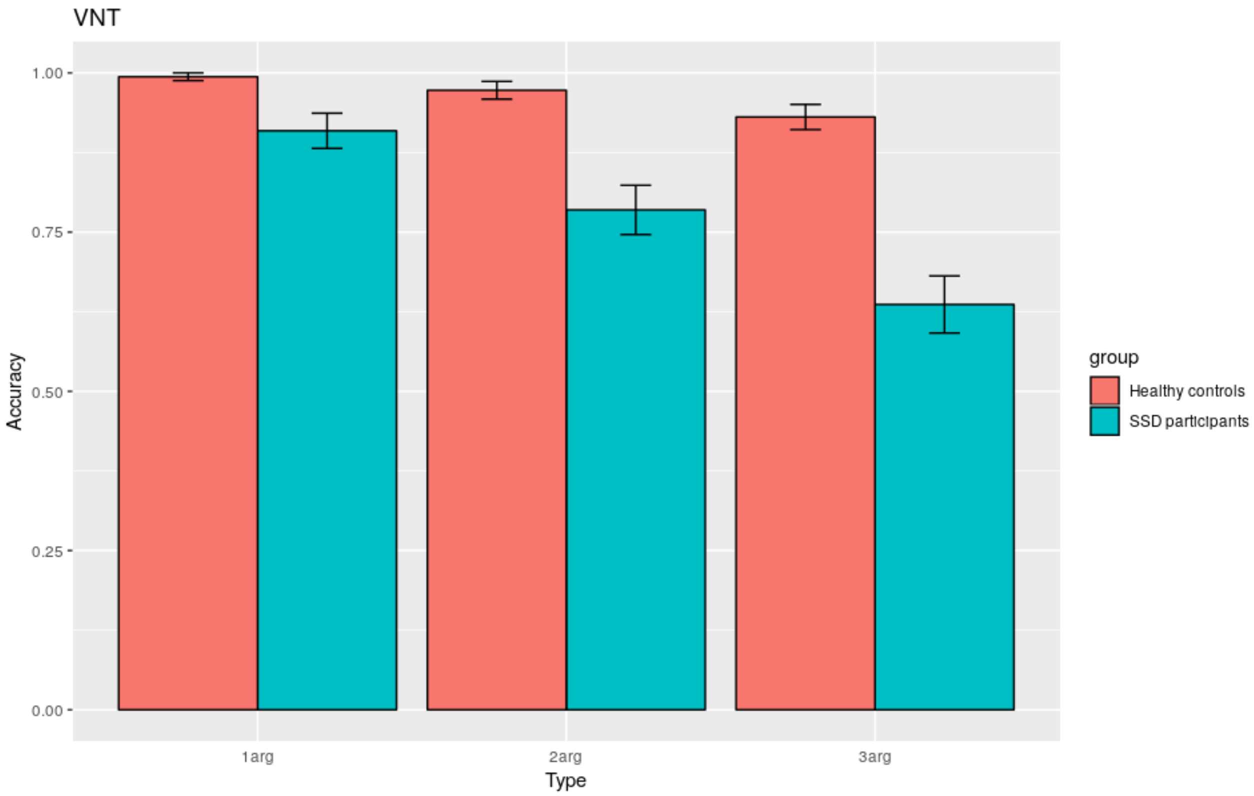


Figure 4.4. Accuracy rate for the Verb Naming Task (VNT). Bars represents standard error of the mean.

Table 4.3 reports the fixed effects for all the variables on the VNT, together with the significance test.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	.73	.05	63.25	13.38	<.001	***
Verbs with 2 args	-.07	.02	128	-3.35	.001	**
Verbs with 3 args	-.17	.02	128	-7.24	<.001	***
Group (SSD)	-.15	.03	65.43	-4.27	<.001	***
Education	.01	.00	62.99	2.78	.007	**
2args : Group (SSD)	-.10	.04	128	-2.37	.019	*
3args : Group (SSD)	-.21	.05	128	-4.52	<.001	***

Table 4.3 Estimated fixed parameters for the Verb Naming Task (VNT)

Once having partialled out the effect of *Education* (t = -2,37, $p = .007$), the interaction between the variable *Number* of arguments and the variable *Group* was still significant for both two- (t = -2.37, $p < .05$) and three-argument verbs (t = -4.52, $p < .001$). This means that the effect of argument complexity acts differently in the two Groups: in our case, the higher the number of arguments, the less accurate are SSD participants compared to HPs in the VNT.

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by number of arguments. On the contrary, SSD participants showed a significantly increased accuracy rate for one-than two-argument verbs (t = 4.040, $p <.001$), for one- than three-argument verbs (t = 8.29, $p <.001$), as well as for two- than three-argument verbs (t = 5.37, $p <.001$).

4.3.2.2 Comprehension of single verbs - VCT

As performance of both SSD participants and HPs at the VCT was nearly at ceiling (Fig.4.5), with accuracy rate over 98% for both Groups in all conditions (one-, two, and three-argument verbs), no further analyses were conducted on this task.

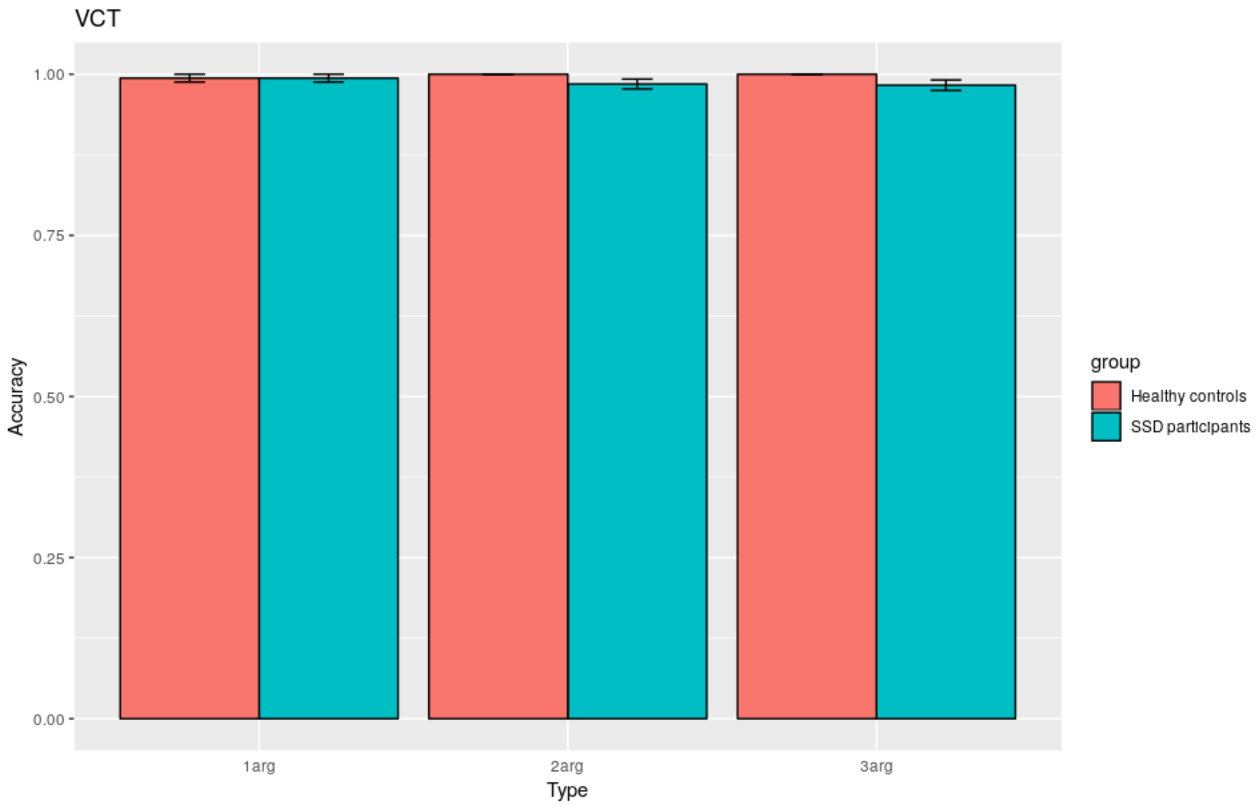


Figure 4.5 Accuracy rate for the Verb Comprehension Task (VCT). Bars represents standard error of the mean.

4.3.2.3 Production in context - ASPT

Figure 4.6 depicts the accuracy rate for the ASPT for one-, two-, and three-argument verbs in the two groups.

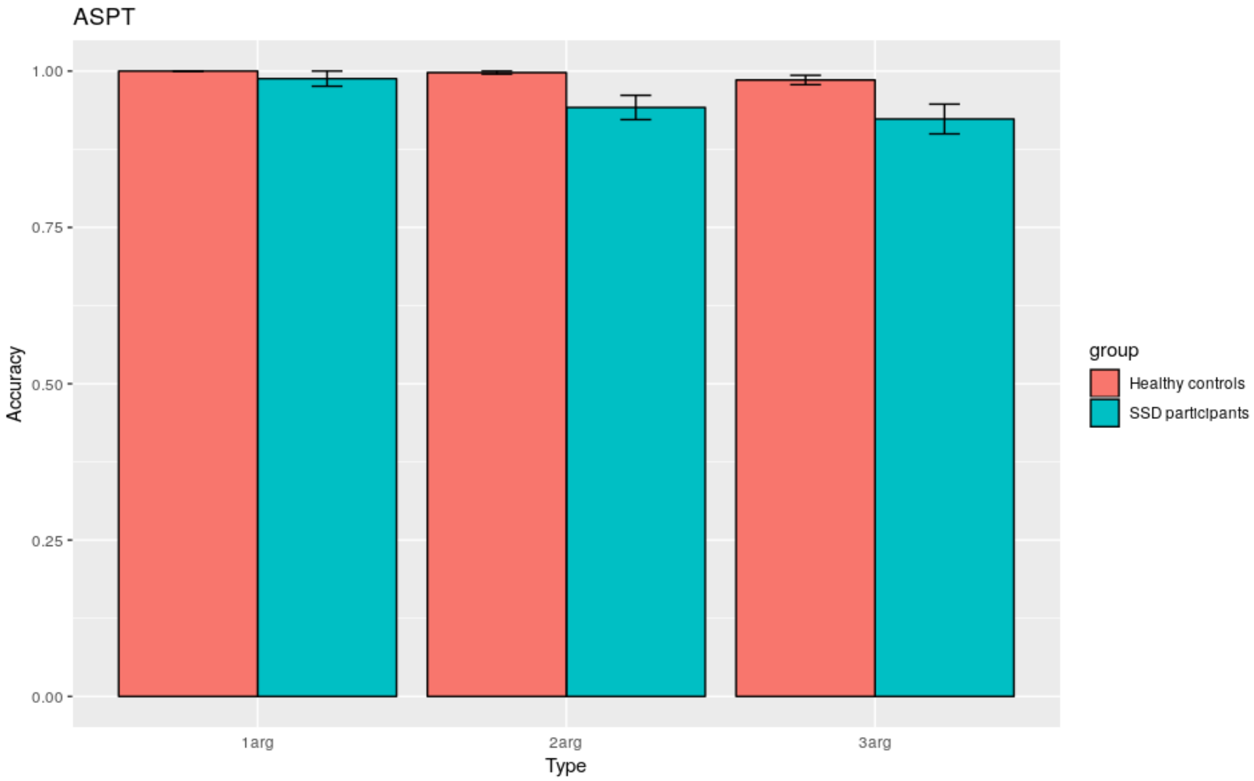


Figure 4.6. Accuracy rate for the Argument Structure Production Task (ASPT). Bars represents standard error of the mean

Table 4.4 reports the fixed effects for all the variables on the ASPT, together with the significance test.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	.91	.03	63.5	29.91	<.001	***
Verbs with 2 args	-.04	.01	128	-3.51	.001	***
Verbs with 3 args	-.02	.01	128	-2.09	.038	*
Group (SSD)	-.03	.02	67.81	-1.34	.185	
Education	.01	.00	63	2.19	.032	*
2args : Group (SSD)	-.05	.02	128	-2.25	.026	*
3args : Group (SSD)	-.04	.02	128	-1.87	0.063	

Table 4.4 Estimated fixed parameters for the Argument Structure Production Task (ASPT).

Once having partialled out the effect of *Education* ($t = 2.19$, $p < .05$), the interaction between the variable *Number of arguments* and the variable *Group* was still significant, but only for two-argument verbs ($t = 2.247$, $p < .05$). This means that the effect of argument complexity acts differently in the two groups, but only for two-argument verbs: the complexity of the verb argument structure, in case of verbs with three arguments, does not act differently in the two Groups on the dependent variable *Accuracy*.

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by number of arguments. On the contrary, SSD participants showed a significantly higher accuracy rate for one- than two-argument verbs (t = 4.06, p <.001), for one- than three-argument verbs (t = 2.8, p <.05), but not for two- than three-argument verbs (p = .26).

4.3.3 Syntactic complexity

4.3.3.1 Production – SPPT

Figure 4.7 depicts the accuracy rate to the SPPT by sentence type in the two groups.

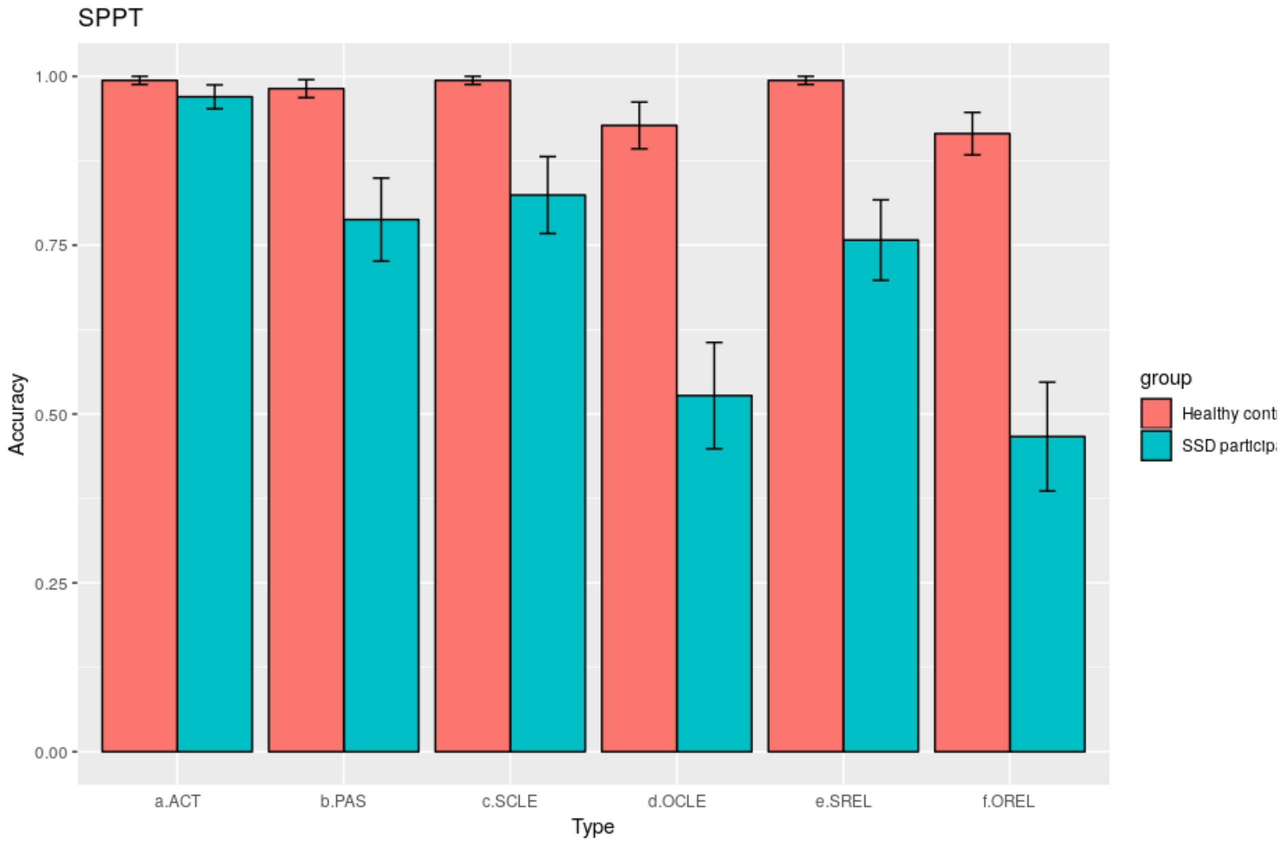


Figure 4.7. Accuracy rate for the Sentence Production Comprehension Task (SPPT). a.ACT = Active sentences; b. PAS = Passive sentences; c.SCLE = Subject Cleft sentences; d.OCLE = Object Cleft sentences; e.SREL = Subject Relative sentences; fOREL = Object Relative sentences.

4.3.3.1.1. Sentence type

Table 4.5 reports the fixed effects for all the variables on the SSPT by sentence type, together with the significance test.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	.71	.08	63	9.05	<.001	***
Object Cleft	-.25	.04	320	-7.14	<.001	***
Object Relative	-.29	.04	320	-8.16	<.001	***
Passive	-.1	.04	320	-2.72	.007	**
Subject Cleft	-.07	.04	320	-2.04	.042	*
Subject Relative	-.11	.04	320	-2.98	.003	**
Group(PTS)	-.21	.05	63	-4.15	<.001	***
Education	.012	.01	63	1.86	.067	
Object Cleft : Group (SSD)	-.38	.07	320	-5.27	<.001	***
Object Relative : Group (SSD)	-.42	.07	320	-5.95	<.001	***
Passive : Group (SSD)	-.17	.07	320	-2.38	.018	*
Subject Cleft : Group (SSD)	-.14	.07	320	-2.04	.042	*
Subject Relative: Group (SSD)	-.21	.07	320	-2.98	.003	**

Table 4.5 Estimated fixed parameters for the Sentence Structure Production Task (SPPT)

With respect to sentence type, the fixed effect of *Education* was found to be not significant (p = .07). A significant two-level interaction was found between *Group* and all types of sentences. Results showed that the interaction between *Group* and sentences type significantly affects the accuracy rate in the two groups, with SSD participants performing worse than HPs in the production of Object Cleft (t = -5.27, p <.001), Object Relative (t = -5.95, p <.001), Subject Relative (t = -2.98, p <.001), Passive (t = -2.38, p <.05) and Subject Cleft (t = -2.04, p <.05) sentences.

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by sentence type. On the contrary, SSD participants showed a significantly higher accuracy rate for Active than Passive sentences (t = 3.60, p <.001), and a reduced accuracy for Object Cleft *vs* Subject Cleft sentences (t = -5.89, p <.001), and of Object Relative compared to Subject relative sentences (t = -5.76, p < .001).

4.3.3.1.2 Word order

Table 4.6 reports the fixed effects for all the variables on the SSPT by word order, together with the significance test.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	.78	.08	65.60	9.94	<.001	***
Non-canonical	-.15	.02	328	-6.92	<.001	***
Group (SSD)	-.11	.5	89.64	-1.94	.056	
Education	.01	.01	63	1.86	.067	
Non-canonical : Group (SSD)	-.20	.04	328	-4.57	<.001	***

Table 4.6 Estimated fixed parameters for word order in the Sentence Production Priming Task (SPPT).

With respect to word order, the fixed effect of *Education* was found to be not significant (*p* = .06). A significant two-level interaction was found between *Group* and canonicity of word order. Compared to HPs, SSD participants were significantly less accurate in the production of non-canonical sentences (t= -4.57, *p* <.001) than of canonical sentences.

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by sentence type. On the contrary, SSD participants showed a significantly better performance in the production of sentences with canonical word order than sentences with non-canonical word order (t = 8.11, *p* <.001).

4.3.4 Syntactic complexity - Comprehension (SCT)

Figure 4.8 depicts the accuracy rate to the SCT by sentence type in the two groups.

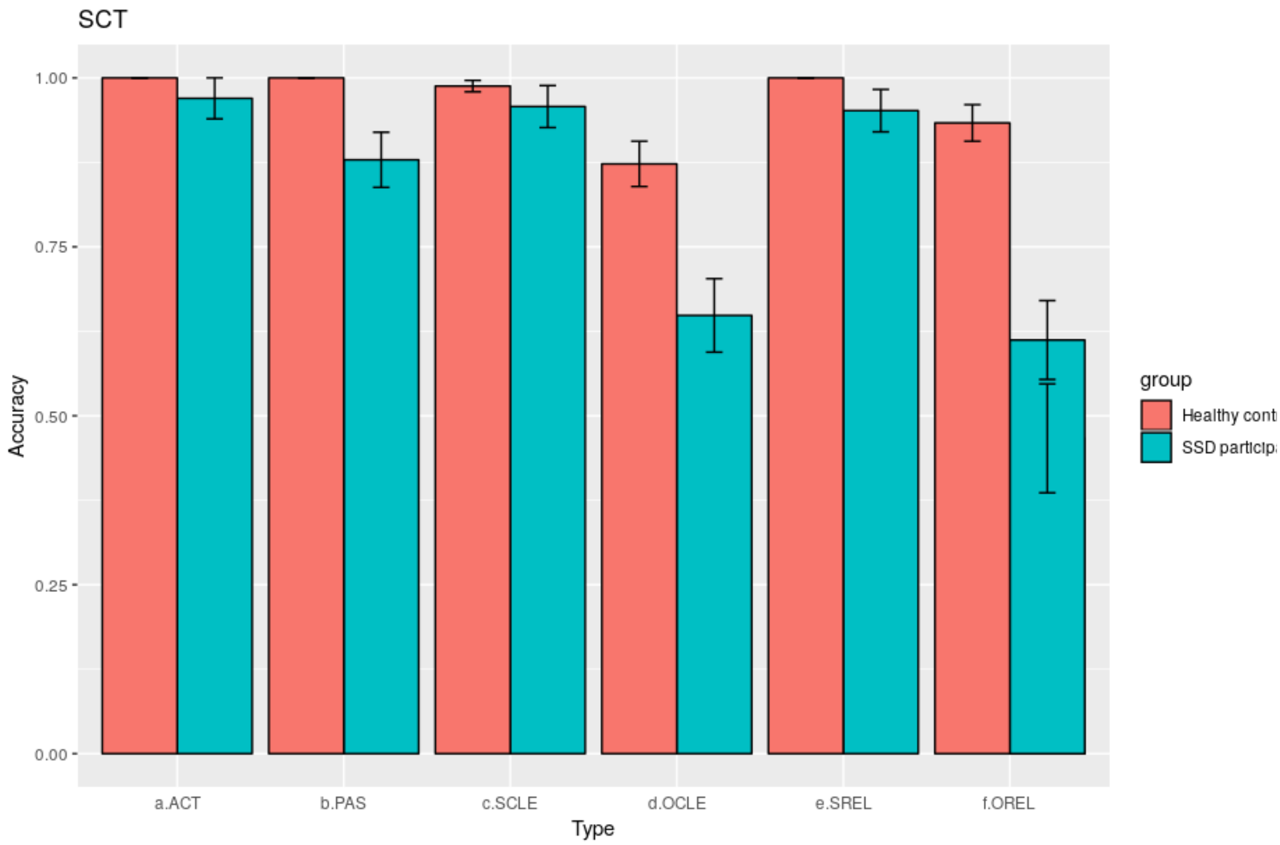


Figure 4.8. Accuracy rate for the Sentence Comprehension Task (SCT).

4.3.4.1 Sentence type

Table 4.7 reports the fixed effects for all the variables on the SCT by sentence type, together with the significance test.

	Estimate	Std. Error	df	t value	Pr(>|t|)	Sign.
Intercept	.83	.07	105.90	12.10	< .001	***
Object Cleft	-.07	.05	325.70	-1.39	.165	
Object Relative	-.05	.05	325.70	-.92	.361	
Passive	.00	.05	324.80	.00	1.00	
Subject Cleft	.00	.05	325.70	.04	.969	
Subject Relative	.00	.05	324.80	.00	1.00	
Group (SSD)	.01	.05	244.60	.10	.918	
Education	.01	.00	64.18	3.13	.003	**
Object Cleft : Group (SSD)	-.20	.06	325.70	-3.56	< .001	***
Object Relative : Group (SSD)	-.22	.06	325.70	-3.83	< .001	***
Passive : Group (SSD)	-.06	.06	324.80	-1.08	.281	
Subject Cleft : Group (SSD)	-.02	.06	325.70	-.33	.745	
Subject Relative: Group (SSD)	-.01	.06	324.80	-.22	.829	

Table 4.7 Estimated fixed parameters for sentence type in the Sentence Comprehension Task (SCT).

Once having partialled out the effect of *Education* ($t = 3.13$, $p < .01$), a significant two-level interaction was found between the variable *Group* and two types of sentences, namely Object Clefts and Object Relatives. Compared to HPs, SSD Participants were significantly less accurate in the comprehension of Object Cleft ($t = -3.56$, $p < .001$) and Object Relative sentences ($t = -3.83$, $p < .001$).

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by sentence type. Similarly, SSD Participants showed no differences in the accuracy rate of Active versus Passive sentences ($p = .2971$), but a significant difference was found for Object versus Subject Cleft sentences ($t = -8.79$, $p < .001$) and for Object versus Subject Relative sentences ($t = -8.62$, $p < .001$).

4.3.4.2 Word order

Table 4.8 reports the fixed effects for all the variables on the SCT by word order, together with the significance test.

Fixed effects	Estimate	Std. Error	df	t value	Pr(>|t|)	
Intercept	.83	.06	75.83	13.24	< .001	***
Non-canonical	-.04	.03	333.27	-1.23	.219	
Group (SSD)	-.01	.04	113.75	-.26	.795	
Education	.01	.00	64.28	3.13	.003	**
Non-canonical : Group (SSD)	-.15	.04	333.30	-4.24	< .001	***

Table 4.8 Estimated fixed parameters for word order in the Sentence Comprehension Task (SCT).

Once having partialled out the effect of *Education* (t = 3.13, p <.01), the interaction between the fixed variables (*Group* and *Word order* canonicity) was found to be still significant (t = -4.24, p < .001), meaning that the effect of the variable canonicity acts differently in the two Groups. In particular, SSD Participants were significantly less accurate in the comprehension of non-canonical sentences.

Post-hoc group comparisons showed no differences in the accuracy rate of HPs by sentence type. On the contrary, SSD participants showed a significantly better performance in the comprehension of sentences with canonical word order than sentences with non-canonical word order (t = 10.36, p <.001).

4.4 Discussion

Aim of this experiment was to investigate the syntactic abilities of people with SSD and to compare it to those of a group of HPs matched by gender and age, and, in particular, to address a possible impairment of the processing of verbs and sentence structures. To do so, we administered the Northwestern Assessment of Verb Structure (NAVS) (Cho-Reyes & Thompson, 2012) in its Italian adaptation (Barbieri et al., 2019) to a cohort of people with SSD matched by age and gender to a group of HPs. The first part of the NAVS battery is structured so as to enable the assessment of both production and comprehension of verbs in isolation or embedded in sentences, according to their different number of arguments. The second part assesses the production and comprehension of different sentences types, taking into consideration two levels of syntactic complexity: sentence types (Active *vs* Passive, Subject Cleft *vs* Object Cleft, Subject Relative *vs* Object Relative sentences) and word order canonicity (canonical *vs* non canonical). To our knowledge, the present study is the first attempt to address the Argument Structure Complexity Hypothesis (ASCH) (Thompson, 2003), coupled to a fine-grained assessment of productive and receptive syntax in SSD participants. The analyses of results highlighted some effects that are worth to discuss. It must be noted that our experimental sample of SSD participants included a wide range of neurocognitive profiles, from nearly intact individuals to severely compromised participants. Such heterogeneity is indeed observable in the high standard deviations of accuracy scores.

Effect of number of arguments

In line with our hypothesis, which predicted an increased difficulty for SSD participants in producing three- and two-argument verbs *vs* one-argument verbs, compared to HPs, we found a significant effect of the number of arguments on verb production performance in people with SSD.

This effect was particularly visible in the verb-naming task (VNT): verbs with complex lexical entries were more difficult to retrieve for people with SSD and, in particular, this group showed greater impairment for three-argument verbs compared to one-argument verbs (p <.001), three-argument verbs compared to two-argument verbs (p <.001), as well as for two-argument verbs compared to one-argument verbs (p <.001).

As for production of the verb argument structure in a sentence context, results from the ASPT showed a reduced accuracy rate for this population, but only for one-argument verbs compared to two-argument verbs (p < .001) and for one-argument verbs compared to three-argument verbs (p <.05). No significant effect was found for one- *vs* two-argument verbs (p = .06)

Our results are indeed compatible with the Argument Structure Complexity Hypothesis postulated by Thompson (2003) and point to a selective difficulty, for SSD participants, in the retrieval of verbs with a complex argument structure. In this sense, our work complements previous studies on verb naming in SSD, by taking into account a detailed characterization of the verb argument structure. It is interesting to note that the results from the assessment of verb production, both alone and within sentences, are consistent with the literature on agrammatic speakers (Kim & Thompson, 2000), who perform more poorly on verb naming and reading as the number of thematic roles of the verb increases (Barbieri et al., 2013; Cho-Reyes & Thompson, 2012; Kemmerer & Tranel, 2000; Kim & Thompson, 2000, 2004; Luzzatti et al., 2002; Thompson et al., 2007).

The interpretation of the present results must take into consideration previous neuroimaging findings on the healthy population, and in particular studies showing different patterns of activation in the processing of verbs with more than one obligatory argument (Thompson et al., 2007; Thompson et al., 2010). Activation of the supramarginal and angular gyri, limited to the left hemisphere, was observed only when verbs with two obligatory arguments were compared to verbs with a single argument, while bilateral activation was detected when both two- and three-argument verbs were compared to one-argument verbs. These findings suggest that posterior peri-sylvian regions are engaged for processing argument structure information associated with verbs, with increasing neural tissue in the inferior parietal region associated with increasing argument structure complexity. These findings are consistent with processing accounts, which suggest that these regions are crucial for semantic integration. In this sense, the recognised incomplete lateralization of language in people with SSD (Artiges et al., 2000; Sommer et al., 2003; Spaniel et al., 2007; Weiss et al., 2006) and the reduce inter-hemispheric connectivity (Chang et al., 2019), appears to compromise the retrieval of verbs with complex argument structures.

Effect of sentence type

Secondly, the results of tasks requiring the production (SPPT) and the comprehension (SCT) of sentences highlighted the presence of a significant effect for sentence type on the performance of SSD participants. In particular, SSD participants showed a significantly lower accuracy in the production of Passive *vs* Active sentences, for Object Cleft *vs* Subject Cleft sentences, as well as for Object Relative *vs* Subject relative sentences (all comparison with $p < .001$). We interpret these results as an evidence of an effect of syntactic movement in the d-structure of sentences on the performance of participants. The increased difficulty, or the computational load associated with this increased difficulty, can be interpreted as an increased demand on the resources available to the executive functions (Fernandez-Duque, 2009). It follows that in a clinical population where a prefrontal functional damage is observed, such is the case of people with SSD, the performance to tasks posing a significant load on frontal function would be reduced, as observed in this work.

Effect of canonicity

Thirdly, the results from SPPT and SCT highlighted that, overall, non-canonical sentences were significantly more difficult ($p <.001$) for the SSD group than canonical sentences, both in terms of production and comprehension. Differently from sentences displaying a canonical word order (such as active, SVO sentences), in sentences with non-canonical word order a syntactic operation moves the object from its original, post-verbal position, to the clause-initial position. This operation is involved in Passive, Object-cleft, as well as Object relative sentences. Indeed, according to our results, these types of sentences are those displaying a selective deficit in production and comprehension in people with SSD.

Our results are in line with the literature on syntactic processing deficits in agrammatic aphasia and support the finding of Lelekov and colleagues (2000), who found that both schizophrenic and aphasic patients suffer from selective impairment for non-canonical sentences with complex syntax in a comprehension task. However, with respect to this literature, our data provide a more precise depiction of the language profile of people with SSD based on a detailed characterization of syntactic complexity, grounded in the linguistic theory.

We interpret these results as an evidence of an effect of syntactic movement in the d-structure of sentences on the performance of participants. The increased difficulty, or the computational load associated with this increased difficulty, can be interpreted as an increased demand on the resources required for executive functions (Fernandez-Duque, 2009). It follows that in a clinical population where a prefrontal functional damage is hypothesized, such as for people with SSD, the performance to tasks posing a significant load on frontal function would be reduced, as it emerged from the present study.

Effect of Education

Finally, although it was not aim of the present experiment to test the effect of years of formal education on the performance to the NAVS test, we found a significant effect of this variable on verb production tasks (VNT and APST with $p < .01$ and $p < .05$, respectively) and sentence comprehension (SCT) tasks. The effect of Education, however, did not account for the full variance of our data, implying a specific effect of the number of arguments and syntactic complexity variables. No effect emerged on sentence production in terms of sentence type or word order. We interpret this finding as supporting of a general effect of formal education on the subjects' ability to deal with meta-linguistic tasks, such as the ones addressed by the NAVS.

Conclusions

In summary, the present study showed that verbs with complex lexical entries are more difficult to process in verb naming, sentence production, as well as sentence comprehension tasks for people with SSD than for HPs. Most notably, deficits in the production of verb lexical entries appear in line with an extension of the ASCH hypothesis (Thompson, 2003) also to SSD patients, while the reduced performance of non-canonical sentence production and comprehension are in line with a generative approach to syntactic complexity (Chomsky, 1981). Moreover, the observed patters appear to bear some functional relation to those observed in agrammatic aphasia, pointing to an organic psychopathological impairment associated to a prefrontal functional damage. Finally, having confirmed the ability of the NAVS to capture linguistic deficit patterns in people with SSD, this study supports its use for the assessment of language performance also in people with SSD.

5 Effects of disrupted lexical representations of verbs on anomalous-sentence processing in Schizophrenia Spectrum Disorders[9]

5.1 Introduction

The presence of verbal "incoherencies" in the spontaneous speech of people with SSD in the form of abnormal noun-verb associations (such as the famous example reported by Oh et al., 2002: "*The pond fell in the front doorway*") have attracted the attention of researchers interested in the relationship between language and cognition. By means of neurophysiological measures, it has been argued (Kuperberg et al., 2006) that such language dysfunction in SSD could be attributable to a deficit in combining syntactic and lexical-semantic information to build the overall meaning of the sentence, as the amplitude of N400 waveform (typically associated with semantic processing) in response to animacy-violated sentences[10] was normal in schizophrenia, while that of P600 (influenced by the processing cost occurring when plausible semantic relationships conflict with syntax) was reduced in patients relative to HPs. In other words, patients failed to detect sentences containing a violation, whose identification would require both access to the lexical properties of the verb and the integration of such information in the context of the sentence.

However, to our knowledge, no works along this hypothesis have taken in consideration the effect of the presence of the "disorder of the self" to explain such integration difficulties in this population. In fact, a specific deficit that would affect the attribution of the "agency" of actions has been hypothesized in SSD: what appears to be unstable is the basic experience of being a self, which

[9] The preliminary results of this study were presented during the workshop "Medici e linguisti III" at the Università degli Studi di Napoli Federico II, 13-14 December 2018. The proceedings of the conference are under review for publication by Aracne Editore.

[10] For example: "*For breakfast the eggs would only eat toast and jam*". In this case, the thematic grid of the verb "to eat" poses a semantic restriction on the grammatical subject (it mandatorily requires an animated Agent).

facilitates a sense of "ipseity", resulting in a markedly diminished sense of "mine-ness" of one's own thoughts, actions, and body (Henriksen & Noordgard, 2014). Self-Disorder (SD) appears to be stable in time (Nordgaard et al., 2018) and have been suggested as fundamental trait features of schizophrenia (Postmes et al., 2014). Studies investigating the association between SD and neurocognitive dysfunction in schizophrenia (Haug et al., 2012) found that the level of self-disorder was associated only with verbal memory (as tested assessing immediate memory short stories orally presented) but not with other functions (working memory, executive functions, psychomotor speed, or visual memory). Authors interpreted this finding as suggestive of a link between verbal memory and the self-disorder (albeit the directionality of the effect was not clear). The test employed in the study assessed the ability to process incoming verbal information and organize it efficiently in relation to preexisting self-knowledge for recall. In this sense, it is possible that a deficit in verbal memory may hinder the ability to comprehend, direct, remember and reason about one's own thoughts and self-knowledge, functions that can be seen as related to several aspects of SDs, or the sense of self.

In language, the notion of "agency" is usually conveyed by the construct of "Agent", one of the Thematic Roles, which are stored in the (lemma) argument structure of the verb (Bock & Levelt, 1994). It is the verb that it said to "assign" each of its arguments to a specific Thematic Role, thus identifying the function of each participant in the action. According to Dowty (1991), two fundamental Thematic Roles can be identified: "Proto-Agent" and "Proto-Patient", whereby Agent is the participant in an event that causes things to happen, while the Patient is the entity passively involved in the event expressed in the predicate (in this latter category, the thematic role of "Theme" is usually integrated). In this sense, a better characterization of the cognitive processes at the basis of Thematic Role comprehension might shed some light on the link between SD and language in people with SSD.

Eye-tracking techniques provide researchers with the opportunity to collect quantitative measures that are thought to reflect cognitive processes underlying reading comprehension. However, reading studies investigating eye movement in SSD are limited. In order to understand the psychophysiological processes as the base of the comprehension deficits of written materials by schizophrenic patients (Hayes & Maree O'Grady, 1998), research addressing eye movements during reading tasks found significant differences in the overall reading rate and in saccadic movements of schizophrenic participants compared to HPs, although the lack of a significant difference in the mean fixation duration and in the number of regressions between the two groups led researchers to conclude that schizophrenic patients did not have significant difficulties in visual and cognitive processing of words (Roberts et al., 2013). The results of a gaze-contingent task (Whitford et al., 2013), which identified reduced forward saccade amplitude and a reduced effect of window size reduction in schizophrenia patients compared to controls, suggested that deficits in language, oculomotor control, and cognitive control contribute to reading deficits in schizophrenia. However, to the best of our knowledge, reading performance of individuals with SSD when processing anomalous text (i.e., semantic violations) has not been investigated so far.

The aim of the present study is to assess the ability of people with SSD to identify the presence of semantic violations on the (Proto)-Agent of verbs, compared to semantic violations on a different thematic role (i.e., the Theme). In particular, we are interested in testing the ability of people with SSD to identify a semantic violation on the "animacy" feature of the Agent. Given the presence of SD (Henriksen & Noordgard, 2014) and based on previous results on electrophysiological measures of tolerance to semantic violations of people in SSD (Kuperberg et al., 2006), we hypothesize that SSD participants will be less sensitive to a violation of the Agent animacy feature. In particular, we expect that the interaction between semantic violations and different Thematic Roles (Agent vs Theme) will act differently in the two Groups, and that such difference will be visible on eye-tracking indexes of access to the lexical properties of the verb.

5.2 Materials and methods

5.2.1 Participants

Thirty-seven participants with a diagnosis of SSD according to DSM-5 (APA, were recruited. SSD participants were recruited from the outpatients' service and the residential facilities of the IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, between February 2018 and April 2019. Diagnoses were made by the treating clinicians (staff psychiatrists). Patients were required not to be overtly psychotic, so that they were able to participate in a lengthy interview and in the neurocognitive assessment. Patients were aged 18-65, able to give informed consent, right-handed, and had normal or correct-to-normal visual acuity. The diagnosis on Axis I had to be unique, but co-morbidities on Axis II were admitted. Exclusion criteria were: neurological disorders, head trauma with cognitive sequalae, mental retardation, substance abuse in the 3 months preceding the enrolment. Seven participants were excluded, leaving a final sample of thirty participants. Reasons for exclusion were: unexpected technical (software) issues preventing the acquisition of eye-movement data (N = 2), refusal to undergo the eye-tracking experiment (N = 2), nystagmus (N = 1), poor quality of reading glasses (N = 2). The group of participants with SSD had a mean age of 47.48 years (SD = 9.80) and a mean education of 10.76 years (SD = 3.64). At the time of the recruitment, patients were on treatment with at least one anti-psychotic medication for at least the previous 6 months and the mean years of illness of the participants was 23.58 (SD = 11.85, N = 29 – for one SSD participant, it was not possible to date the age of onset).

Thirty-two control participants, matched by age-class (18-30, 31-40, 41-50, and 51-65), gender, and education, were recruited among the hospital staff and through public announcements also in Brescia. Two participants were excluded from the analysis (refusal to complete the data acquisition), leaving a final paired sample of thirty HPs. The exclusion criteria for the control group were: any documented psychiatric disorders or being first-degree relative of a patient with diagnosis of SSD. The healthy control group was matched by age (M = 47.7, SD = 10.58), gender (20 males and 10 females), and education (M = 12.43, SD = 3.77).

All participants were native speakers of Italian.

After having presented the study, informed consent to participation was obtained from all participants. In case of patients with support administration, the participation to the study was first discussed with the patient; then, written consent was obtained both from the patient and the appointed administrator. The study was approved by the IRCCS Ethical Committee (Opinion 61/2017) and followed the principles of the Helsinki Declaration.

5.2.2 Stimuli

A set of 112 sentences was presented to participants. Items were validated through a rating study: twenty healthy subjects had been asked to rate each sentence on a 5-point scale, whereas 1 was "Totally unacceptable" and 5 was "Totally acceptable". Sentences that were rated differently from the intended categorization were either removed or replaced. In half (N = 56) of the stimuli the subject was an Agent argument and in the other half (N = 56) a Theme (i.e., the grammatical subject of an unaccusative verbs). Half of the stimuli (N = 56) were generated to be semantically acceptable, and half (N = 56) not acceptable. The semantic anomaly was placed on the animacy feature of the Subject, hence making the Subject-Verb relationship congruent or incongruent (see Table 5.1). Semantic violations always affected the grammatical subject in pre-verbal position to avoid a possible confounding effect related to the word position within the sentence. This means that sentences in the "Theme" condition were the result of underlying syntactic movements (Burzio, 1986; Perlmutter, 1978), contrary to sentences in the "Agent" condition. Given that frequent words are processed faster than less frequent words (Brysbaert et al., 2011), and that long words slow down reading times (Rayner et al., 2011), frequency and length of target words (verbs) were matched across conditions. Target verbs were placed roughly at the center of the sentence. The same experimental list was administered to all participants, with sentences appearing in random order. Table 5.1 reports examples of the stimuli.

Condition	**Example**	**N. of stimuli**
Agent – correct	Ogni sera Bianca telefona al nipotino *(Every evening Bianca calls her nephew)*	28
Agent – with violation	* Ogni sera il sonno telefona al nipotino *(Every evening the sleep calls her nephew)*	28
Theme – correct	A volte il bambino cade dalle scale *(Sometimes the child falls from the stairs)*	28
Theme – with violation	* A volte il sole cade dalle scale *(Sometimes the sun falls from the stairs)*	28

Table 5.1. Conditions, examples, and numbers of experimental stimuli. Target verbs are underlined. The grammatical subject can be either coherent or incoherent. A literal translation is reported in brackets.

Length of target verbs in the correct condition (M = 7.50, SD = 1.93) was similar to that of target verbs in the condition with violation (M = 7.48, SD = 1.96). The length difference was tested through a t-test (p = .96).

Frequencies of target words were taken from Subtlex-IT (available at http://crr.ugent.be/subtlex-it; Crepaldi et al. 2013), an annotated corpus of around 130 million words of the Italian language based on movie subtitles, with a frequency range from .01 – 17,854 tokens per million. Frequency of target verbs in the correct condition (M = 3,027.96, SD = 12,694.94) was comparable to that of target words in the condition with violation (M = 3,024.54, SD = 12,694.96). Given that the distribution of word frequency in a corpus is not normal, but it rather Zipfian (Zipf, 1949), the difference between the frequency distribution was tested applying a two-sided Kolmogorov-Smirnov test. The result was not significant (p = 0.76), indicating that frequency of verbs was matched between conditions.

Twenty-four sentences containing morphological violations were included as fillers and randomly presented to participants along with the target sentences. The subset of morphologically violated sentences was structured as follows: eight sentences reported a violation on the subject-verb number agreement (i.e., the subject was singular while the verb was plural, and vice-versa; e.g. "*Nel bosco le volpi scappa dal cacciatore*"); eight sentences presented a violation on the "determiner-NP" relationship (e.g. "*Nel mare le pesci nuotano liberi*"), and eight sentences presented a violation on the "verb- antecedent clitic" (e.g. "*Il cane insegue il gatto e la rincorre fin sotto il tavolo*").

The complete set of experimental stimuli is reported in Annex 2.

5.2.3 Eye-tracking procedure and experimental setting

An eye-tracking device (Eye-link 1000 Plus® manufactured by SR Research Ltd., Canada) was employed to collect data on eye-movement. Eye-link is a corneal reflection system, which assesses changes in gaze position by measuring both the reflection of an infrared illumination on the cornea and the pupil size, by means of a video camera sensitive to the light in the infrared spectrum. The desktop camera sampled the pupil at a frequency of 1,000Hz with a single 35mm lens. All participants used a chin-rest support to stabilize the head position. The left eye was recorded. The experimental stimuli were presented on a Display PC via Experiment Builder®, an object-oriented programming suite specifically designed to administer stimuli while recording eye movements.

The eye-tracker was calibrated by means of a three-point grid before each acquisition, or any time the subject moved the head from the chinrest. Sentence presentation was preceded by a fixation point and triggered manually by the researcher at the moment the participant was fixating it. Sentences were presented in white Courier New font, against a black background. Participants were asked to silently read each sentence and then evaluate their acceptability by pressing either a green (acceptable) or red (not acceptable) key on the keyboard. Each sentence remained on the screen until the key was pressed.

A training session was carried out the beginning of each session including eight “ambiguous” sentences as resulted from the pilot phase. During the training session, participants were instructed to stop and ask any questions if needed. Fixation and accuracy data from the training sessions were excluded from the subsequent analysis. The duration of the whole experimental session was quite variable, ranging from 15 minutes for HPs, up to two hours (split in two consecutive sessions) for patients. A dedicated room in the Neurophysiology and Neuropsychology Laboratory of the IRCCS San Giovanni di Dio Fatebenefratelli, Brescia was set up.

5.2.4 Data analysis

We adopted four indexes of reading process as dependent variables: first fixation duration, gaze duration, total fixation duration, and regression movements.

1. *First Fixation Duration* (FFD) is the duration in milliseconds of the first fixation on the target item and it is considered as a proxy of early word processing (Liversedge et al., 2007).
2. *Gaze Duration* (GD) is the sum of all the fixations made on the target word before the eyes leave the region. This metrics is considered a measure of lexical access (Rayner et al., 2011) which, in case of verbs, includes processing of the argument structure (Bock & Levelt, 1994).
3. *Total Fixation Duration* (TFD) is the cumulative duration in milliseconds of all the fixations on the target word, including any regressions. Total fixation duration is generally considered to gain insight into any possible processing difficulties associated with the integration of the word meaning within the sentence (Rayner et al., 2006).
4. *Probability of Regressions,* in our study, is the probability to register a go-back movement starting from the verb. Regression movements are associated with integration difficulties (Reichle et al., 1998).

A score for *Accuracy* in the detection of violations was also computed as the percentage of correct responses for each condition (correct *vs* with violation) for either thematic roles, i.e. the concordance between the expected categorization of the sentence in either "acceptable" or "non-acceptable" and the participant's response. In other words, there were four possible combinations of answers subjects can produce: (i) Subjects could categorize as acceptable sentences in the correct condition (i.e., with no violation of the Thematic Role, either Theme or Agent); (ii) Subjects could categorize as non-acceptable sentences having a violation of the Thematic Role (i.e., the Agent or the Theme are not coherent with the verb); (iii) Subjects could categorize as non-acceptable sentences in the correct condition and; (iv) Subjects could categorize as acceptable sentences in presence of a violation of the Thematic Role. In this work, we considered accurate only those responses fulfilling conditions (i) and (ii).

Statistical models were run so as to have the indexes of reading process outlined above, along with accuracy, as dependent variables (data on fixations – i.e., FFD, GD, and TFD – were logarithmically transformed to reduce the skewness of their distribution, thus obtaining a Gaussian-like distribution), and *Thematic Role*, *Condition*, and *Group* as categorical independent variables. Considering that each of the three independent variables have two levels (*Agent* vs *Theme*, *Correct* vs *With Violation*, and *SSD Participants* vs *HPs*, respectively), our experiment has a mixed 2x2x2 factorial design.

Linear mixed-effects regression analyses (Baayen et al., 2008) and generalized linear mixed-effects model (Baayen, 2008a) were conducted, with *Group*, *Thematic Role* of the subject, and *Condition* as two-level fixed effects. Moreover, two and three-level interactions were included in the model. Random intercepts for items and participants were also included. All models were refitted after having identified and excluded atypical outliers using a 2.5 SD criterion over the model standardized residuals (model criticism: Baayen et al., 2008). Statistics of the refitted models are reported. Logistic mixed-effects regression analyses (Jaeger, 2008) were run on accuracy and the regression data. Mixed effect regression analyses were carried out taking participants and items as crossed random effects (Baayen et al., 2008). Alpha level was set at .05. All the analyses were performed in R, v.3.6 (R CoreTeam, 2019).

All contrasts (having three categorical variables as independent variables) were estimated with respect to the base level, which, in our case, was: i) Group: HPs; ii) Condition: correct; iii) Role: Agent. Stated differently, differences were estimated compared to the data produced by HPs when reading correct sentences having an Agent as grammatical subject. This scheme applies for all the models of the present Chapter.

5.3 Results

5.3.1 First Fixation Duration

Table 5.2 and Figure 5.1 summarize the mean durations (in milliseconds) of FFD on target verbs, per conditions, in the two groups.

	HPs				SSD participants			
	Agent		Theme		Agent		Theme	
	M	SEM	M	SEM	M	SEM	M	SEM
Correct	261.93	3.36	266.88	3.69	310.79	5.29	313.18	4.92
With violation	274.72	3.75	268.63	3.74	319.50	5.72	314.63	5.18

Table 5.2. Mean durations and standard error of the mean (in ms) of FFD on target verbs, per conditions, in the two groups.

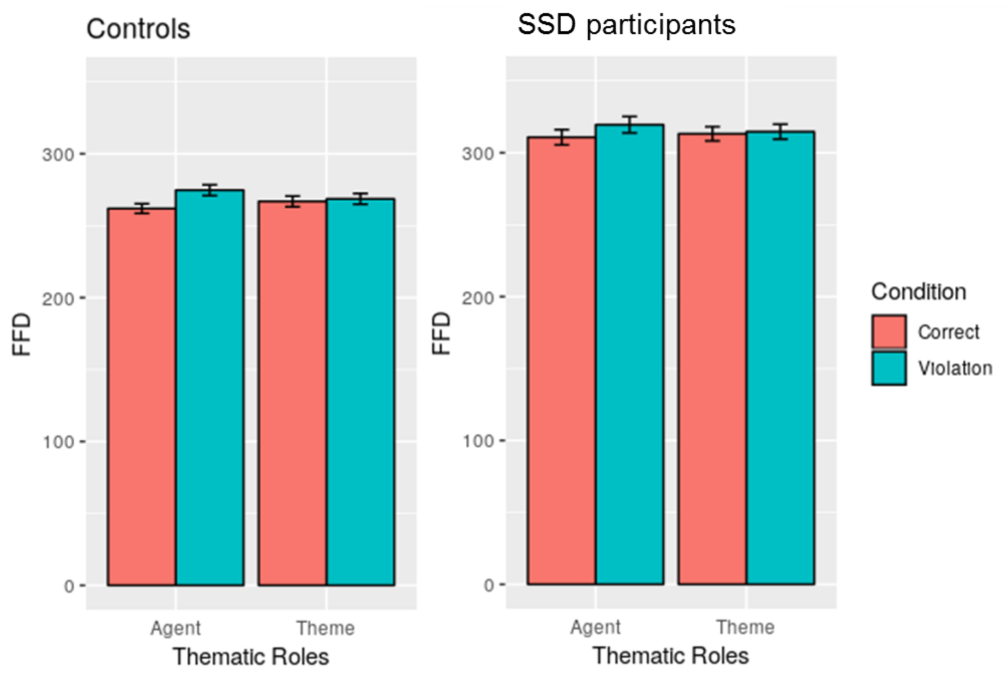


Fig. 5.1 Mean FFD on target verbs per Groups, Thematic Role, and Conditions.

Table 5.3 shows the estimated fixed parameters of the model, together with significance tests.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	5.53	.03	77.72	193.53	< .001	***
Condition (violated)	.03	.02	5893.00	1.95	.051	
Role (Theme)	.01	.02	510.40	.49	.622	
Group (SSD)	.15	.04	76.29	3.67	< .001	***
Condition (violated) * Role (Theme)	-.04	.02	5871.00	-1.80	.073	
Condition (violated) * Group (SSD)	-.03	.02	5849.00	-1.08	.279	
Role (Theme) * Group (SSD)	-.01	.02	5848.00	-.38	.707	
Condition * Role * Group	.02	.03	5847.00	.76	.450	

Table 5.3. Estimated fixed parameters for the analysis on FFD.

Of the three main effects considered, only that of *Group* ($p < .001$) was found to significantly affect FFD. Stated differently, this means that irrespective of *Condition* and *Role*, SSD participants show longer FFD than HPs. Neither a two- nor a three-level interaction between the fixed effects was found, indicating that no interaction of the three main effects emerges on early-stage measure of word processing in this cohort.

5.3.2 Gaze Duration

Table 5.4 and Figure 5.2 summarize the mean duration (in milliseconds) of GD on target verbs, per conditions, in the two groups.

	HPs				SSD Participants			
	Agent		Theme		Agent		Theme	
	M	SEM	M	SEM	M	SEM	M	SEM
Correct	362.81	7.12	382.22	7.88	519.02	16.11	484.17	12.40
With violation	420.78	9.35	421.18	9.96	545.40	16.08	527.72	13.05

Table 5.4 Mean durations and standard error of the mean (in ms) of GD on target verbs, per conditions, in the two groups.

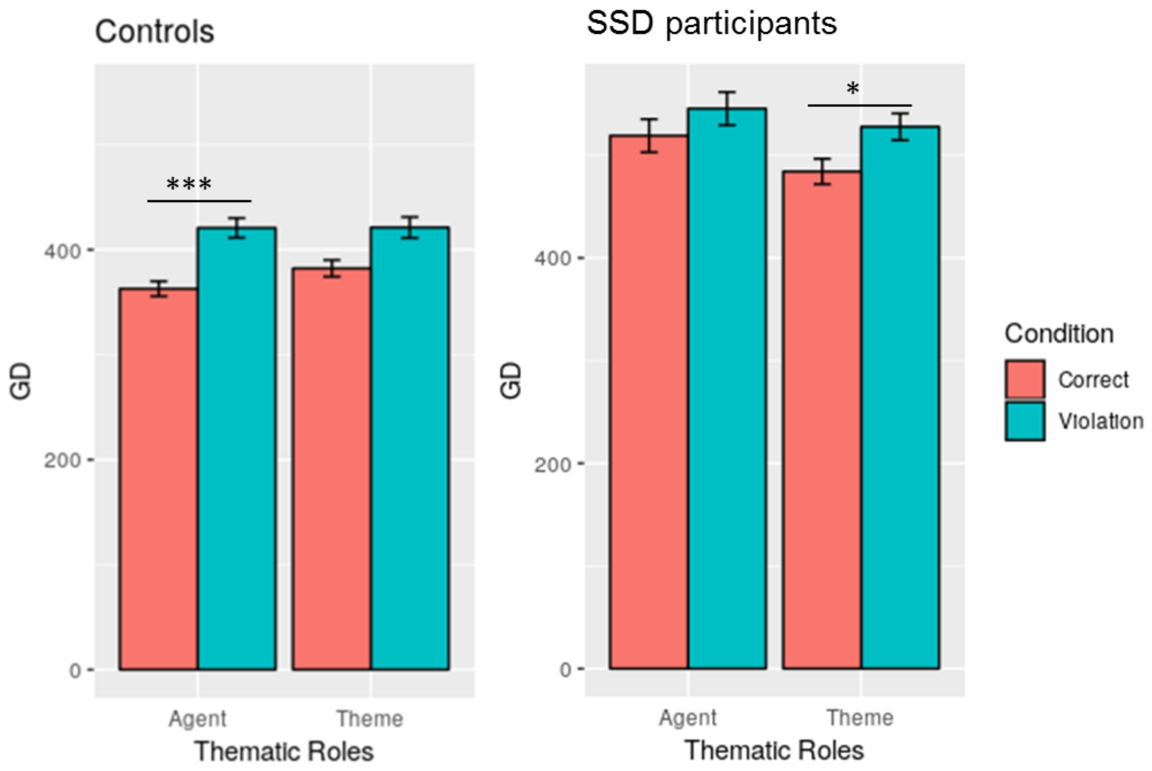


Figure 5.2 Mean GD by Group, Thematic Role, and Condition (bars indicate the standard error of the mean).

Table 5.5 shows the estimated fixed parameters of the model, together with significance tests.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	5.79	.06	93.41	97.86	< .001	***
Condition (violated)	.14	.02	5829.22	7.16	< .001	***
Role (Theme)	.06	.04	72.20	1.31	.194	
Group (SSD)	.25	.07	64.47	3.36	.001	**
Condition (violated) * Role (Theme)	-.09	.03	5839.91	-3.23	.001	**
Condition (violated) * Group (SSD)	-.08	.03	5807.75	-3.06	.002	**
Role (Theme) * Group (SSD)	-.07	.03	5807.72	-2.70	.007	**
Condition * Role * Group (SSD)	.1	.04	5807.64	2.58	.010	**

Table 5.5 Estimated fixed parameters for FFD.

We found a significant three-level interaction ($t = 2.58$, $p < .01$) on GD. In other words, the effects of *Condition* and *Role* interacted differently in the two Groups: sentences carrying a semantic violation of the Agent (e.g., "*Ogni sera il sonno telefona al nipotino*") did not cause a similar reduced GD across the two groups.

Post-hoc comparisons showed that HPs present significantly longer GDs ($t = -7.16$, $p < .001$) when reading non-correct sentences with semantic violation on the Agent *vis-à-vis* when reading the corresponding correct sentences, but not when reading sentences with violation on the Theme ($z = -2.54$, $p = .11$) *vis-à-vis* when reading correct sentences with a grammatical subject of the same role. In other words, HPs were more sensitive to violation on the Agent than on the Theme.

On the contrary, SSD participants showed the opposite pattern: GDs in this group were significantly longer when reading sentences with a violation of the Theme subject ($t = -3.306$, $p < .05$) *vis-à-vis* correct sentences with a grammatical subject of the same role; conversely, GDs were not significantly longer when reading sentences with a violation of the Agent subject ($p = .07$) *vis-à-vis* when reading correct sentences with a grammatical subject of the same role. This finding suggests that sentences carrying an animacy violation of the Agent subject appeared to be more tolerated in the SSD population.

5.3.3 Total Fixation Duration

Table 5.6 and Figure 5.3 summarize the mean durations (in milliseconds) of TFD on target verbs, per conditions, in the two groups.

	HPs				SSD Participants			
	Agent		Theme		Agent		Theme	
	M	SEM	M	SEM	M	SEM	M	SEM
Correct	602.62	14.84	597.64	14.38	1197.90	48.01	1118.41	42.85
With violation	620.27	14.49	635.11	16.35	1379.39	52.42	1470.16	54.04

Table 5.6 Mean durations and standard error of the mean (in ms) of TFD on target verbs, per conditions, in the two group

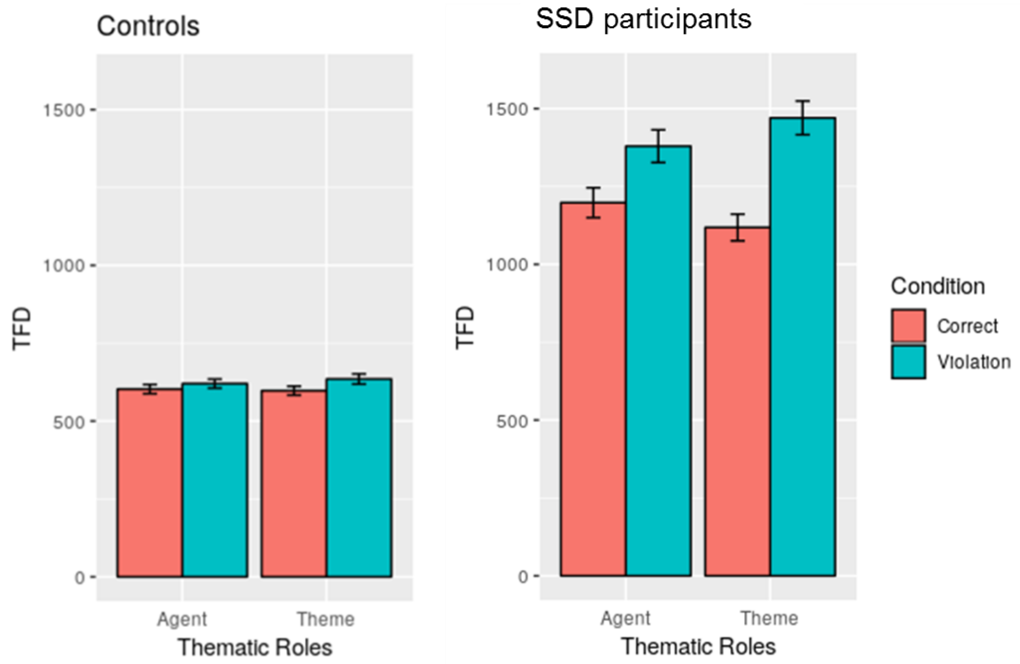


Fig. 5.3 Mean TFD per Groups, Thematic Role, and Condition.

Table 5.7 shows the estimated fixed parameters of the model, together with significance tests.

Fixed effects	Estimate	SD	df	t	p value	Sign
Intercept	6.21	.09	82.43	70.23	< .001	***
Condition	.07	.03	5898.00	2.85	.004	**
Role	-.01	.05	80.55	-.11	.911	
Group	.48	.12	62.68	4.13	<.001	***
Condition * Role	-.02	.04	5910.00	-.59	.558	
Condition * Group	.12	.04	5876.00	3.22	.001	**
Role * Group	-.04	.04	5876.00	-.99	.323	
Condition * Role * Group	.06	.05	5876.00	1.23	.219	

Table 5.7 Estimated fixed parameters for TFD.

The interaction between the variables of *Condition* and *Group* was found to significantly affect TFD (t = -.3.22, *p* < .01). This means that the effect of a semantic violation on TFDs, irrespective of the Thematic Roles it affects, acts differently in the two group. In particular, SSD participants showed significantly longer TFDs in the case of semantic violations vis-à-vis correct sentences, as compared to HPs. However, no three-level interaction between our three fixed effects was found, indicating that the effects of *Condition* and *Theme* did not interact differently in the two Groups on later-stage processes.

5.3.4 Probability of regressions

Table 5.8 and Figures 5.4 report the probability of a go-back movement (regression) from the verb by Condition and Role for the two groups.

	HPs		**SSD Participants**	
	Agent	**Theme**	**Agent**	**Theme**
Correct	71%	73%	80%	76%
With violation	75%	73%	84%	84%

Table 5.8 Probability of regression from target verbs backwards, per conditions, in the two groups.

Overall, HPs showed a lower probability of regression from the verb in all conditions compared to SSD participants.

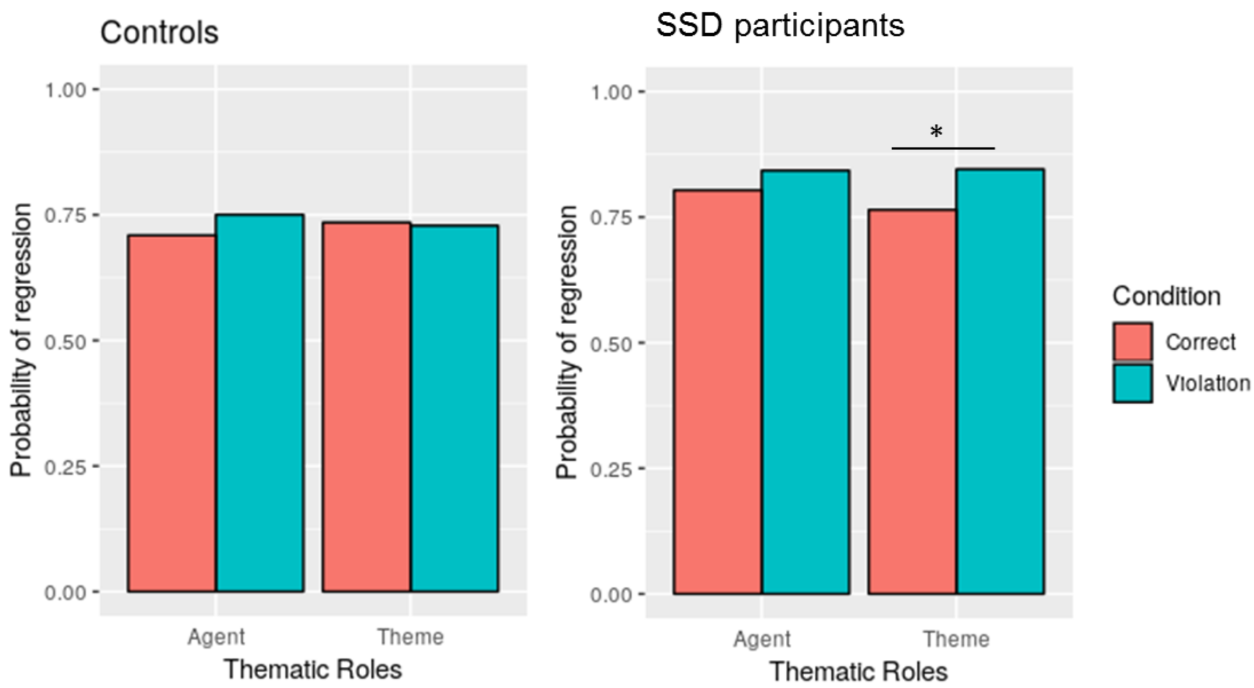


Fig. 5.4 Probability of regression from the target verb by Groups, Thematic Role, and Condition.

Table 5.9 shows the estimated fixed parameters of the model, together with significance tests.

Fixed effects	Estimate	SD	Z value	p value	Sign
Intercept	1.12	.26	4.39	< .001	***
Condition	.30	.13	2.38	.017	*
Role	.16	.20	.80	.426	
Group	.68	.33	2.05	.041	*
Condition * Role	-.32	.18	-1.79	.074	
Condition * Group	.11	.19	.59	.558	
Role * Group	-.43	.19	-2.30	.022	*
Condition * Role * Group	.58	.27	2.13	.033	*

Table 5.9 Estimated effect for probability of regression after having read the verb.

We found a significant three-level interaction ($z = 2.13$, $p < .05$) between the fixed effects on the probability of regression.

Post-hoc comparisons showed that the interaction between *Condition* and *Role* did not affect the probability of regression in HPs: in particular, HPs did not show a higher probability of re-reading verbs in sentences with a violation on the Agent compared to sentences without a violation on the same Thematic Role ($z = -2.38$, $p = .19$). The same pattern of result emerged with respect to sentences with a violation on the Theme compared to sentences without a violation on the same Thematic Role ($z = 0.16$, $p = 1$). In other words, the presence of a violation did not affect the probability of regression in HPs, irrespective of the Thematic Role of the grammatical subject.

The interaction between *Condition* and *Role* did not affect the probability of regression in SSD participants either, when reading sentences with violation on the Agent ($z = -2.79$, $p = 0.07$). However, a significant interaction was found for sentences with a violation on the Theme subject ($z = -4.57$, $p < .001$). This means that a semantic violation on the Theme subject significantly increased the probability of a regression movement in people with SSD, but that, at the same time, a semantic violation on the Agent subject did not equally increased the probability of a go-back movement in this group.

5.3.5 Accuracy

Table 5.10 and Figure 5.5 report the accuracy by Condition and Role for the two groups.

	HPs		SSD Participants	
	Agent	**Theme**	**Agent**	**Theme**
Correct	92%	88%	87%	87%
With violation	97%	98%	87%	86%

Table 5.10 Accuracy, per conditions, in the two groups

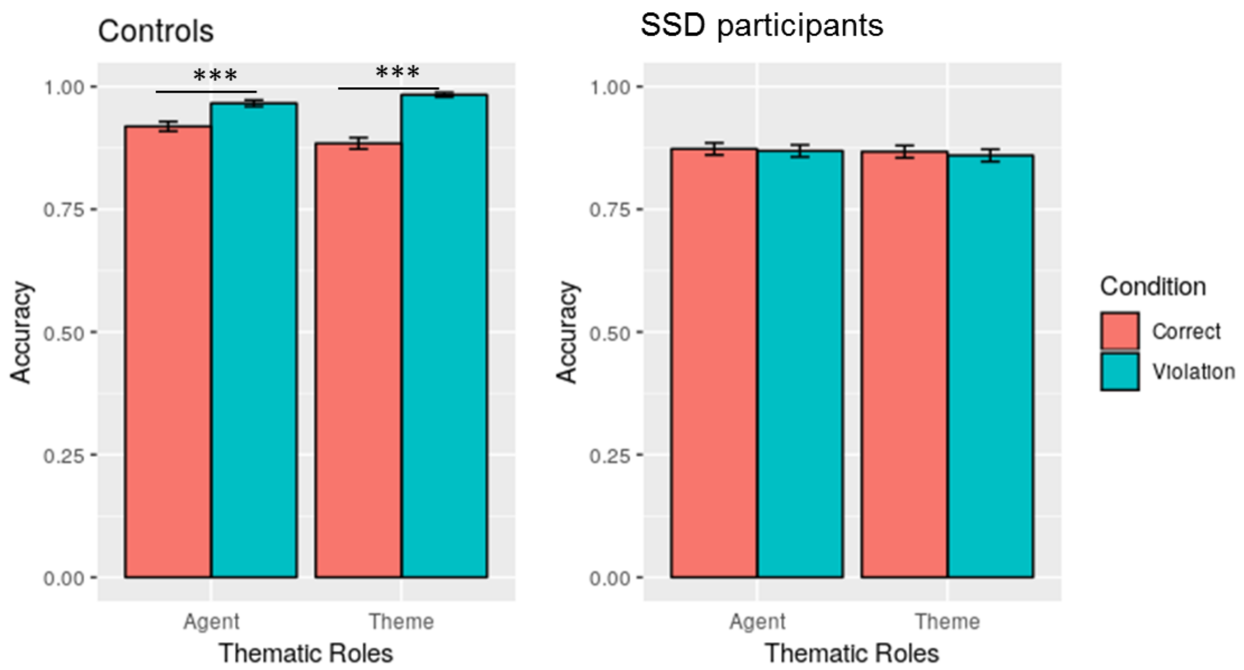


Fig. 5.5 Accuracy by Groups, Thematic Role, and Condition (bars indicate the standard error of the mean).

Figure 5.5 indicates that HPs were, in general, more accurate than SSD participants in categorizing sentences.

Table 5.11 shows the estimated fixed parameters of the model, together with significance tests.

Fixed effects	Estimate	SD	Z value	p value	Sign
Intercept	2.79	.24	11.50	< .001	***
Condition	1.12	.25	4.57	< .001	***
Role	-.40	.25	-1.62	.106	
Group	-.52	.29	-1.80	.072	
Condition * Role	1.01	.39	2.62	.009	**
Condition * Group	-1.04	.29	-3.56	< .001	***
Role * Group	.41	.24	1.70	.089	
Condition * Role * Group	-1.22	.45	-2.72	.006	**

Table 5.11 Estimated effect for probability of regression after having read the verb

A three-level interaction between fixed predictors was found to be significant ($z = -2.72$, $p < .01$) with respect to accuracy scores. This result shows that *Condition* (correct or with violation) of *Role* (Agent or Theme) acted differently in the two Groups with respect to explicit judgments.

Post-hoc comparisons showed that HPs were significantly more accurate in categorizing sentences containing a violation than sentences with no violation, both concerning the Agent subject ($z = -4.57$, $p < .001$) and the Theme subject ($z = -7.09$, $p < .001$).

On the contrary, in SSD participants accuracy did not differ significantly when categorizing sentences carrying a semantic violation on the subject *vis-à-vis* sentences without violation, neither on the Agent subject ($z = -0.48$, $p = 1$), nor on the Theme subject ($z = 0.77$, $p = .992$). However, a significant three-way interaction suggests that, in HPs, the advantage for Theme-violation sentences was larger than that for Agent-violation sentences.

5.4 Discussion

The aim of the present study was to assess the sensibility of people with SSD to semantic violations on the pre-verbal (Proto)-Agent of verbs, as compared to semantic violations on the Theme.

To do so, we collected both implicit (eye-movements) and explicit (accuracy) measures in individuals with SSD when engaged in a silent reading task. Stimuli were presented as full sentences whose grammatical subject was either an Agent or a Theme. In half of the stimuli, grammatical subjects carried a violation of the animacy aspect, so as to make the relationship with the verb not acceptable. To our knowledge, the present study is the first attempt to address eye-movement signatures in an anomaly detection task of sentences with a semantic violation of different Thematic Roles in a group of individuals with SSD.

First fixation duration

Our results indicate that only the main variable *Group* has a significant ($p <.001$) effect on early processing measures of reading, namely FFDs. In this case, SSD participants showed significantly longer first fixation durations on the target verb compared to HPs, irrespective both of the Thematic Role of the grammatical subject and of the presence (or absence) of a semantic violation. In fact, the effect on this measures, which could reflect difficulties in local processing (Boland et al., 2004), was experimentally controlled given that length and frequency of the target word was matched across conditions. As previous studies on eye-movement during reading in schizophrenia (Roberts et al., 2013; Whitford et al., 2013) did not provide information on possible correlation between slow reading rates and psychopathological symptoms, we can speculate that this main effect of group could derive from the presence of the overall motor retardation caused by negative symptoms (Chapter 1).

Gaze Duration

In line with our hypothesis, we found an effect of the experimental manipulation on measures of lexical access such as gaze duration. In particular, the results of our statistical model indicate that the effects of *Condition* and *Role* interact differently in the two groups on this measure. What we observed is that sentences carrying a violation of the animacy feature of the Agent seemed to be more tolerated by people with SSD relative to HPs, insomuch this combination of variables does not appear to interfere with the gaze duration in this latter group in the same way.

By adopting a "lexicalist" approach, we assume that the verb thematic grid is part of its lemma representation (Bock & Levelt, 1994). It follows that, should the lexical representation of the verb be somehow impaired, the identification of a semantic violation on one of its arguments would be deteriorated. Moreover, as schizophrenia has been consistently linked with a disturbed sense of Agency (Haug et al., 2012; Henriksen & Noordgard, 2014; Postmes et al., 2014; Nordgaard et al., 2018), the linguistic structure that conveys the concept of agency might exhibited some form of anomalous processing. This is exactly what we observe in our sample of SSD participants: sentences containing a semantic violation of the Agent are processed differently compared to what observed in control participants. In other words, when reading sentences such as "*Ogni sera il sonno telefona al nipotino*", individuals with SSD do not seem to detect a semantic violation, as they do not increase the processing time of the target verb, visually processing the anomalous stimulus as if it was correct.

With respect to the previous literature, these results further support and extend previous studies on word-monitoring tasks (Kuperberg et al., 1998) highlighting an insensitivity to linguistic violations (presumably due to a deficit in the use of linguistic context to process speech), as well as studies using electrophysiological responses (Kuperberg et al., 2006) on critical verbs that are suggestive of abnormalities in combining semantic and syntactic information online to build up propositional meaning. Differently from the previous literature, we further focused our investigation on the integration of syntactic and semantic information of language information by taking into consideration a single class of semantic anomalies (the violation of the animacy feature of the preverbal subject, in two different Thematic Roles). Given the observed effect of different Thematic Roles in interaction with semantic violation, we suggest that in people with SSD the top-down linguistic elaboration needed in reading could be affected by a disrupted sense of agency, coherent with a phenomenological approach to schizophrenia (Henriksen & Noordgard, 2014).

One may argue that these results may be linked, at least in part, to the fact that constructs implying a syntactic movement (unaccusative verbs with pre-verbal subject theme) pose an additional cognitive load onto the reader. In fact, the task employed in the present study requires the subject to store and integrate syntactic (pre-verbal movement of the subject), morphological (argument structure), as well as semantic information (animacy of the thematic roles) in the working memory.

In this sense, the impaired working memory and executive functions in people with SSD (Gold et al., 2002), might have negatively affected the performance of individuals with SSD. However, if this was the case, we should have always found, irrespective of violations, longer GD for sentences with a Theme subject compared to sentences with an Agent subject. This was not the case, ruling out this alternative, syntax-centered, explanation.

An alternative interpretation is, however, possible. In fact, one may claim that, when reading sentences with unaccusative verbs depicting changes of states, it is more difficult to exclude the implausibility of inanimate grammatical subjects. In other words, it could be that "the sun can fall from the sky" is more acceptable than "the sleep can give a call". However, if this were the case, we would have found the same effect for both groups, i.e., the semantic anomaly on the Theme subject would have been difficult to recognize both by HPs and SSD participants, and this would have led to significantly longer GD for sentences with an anomalous grammatical subject in both groups. On the contrary, we observe a double dissociation for group and thematic role: while HPs take significantly longer to judge an anomalous Agent as such, they can easily spot an anomalous Theme. The reverse pattern is found in the experimental group, where for SSD participants it is easier to identify an anomalous Theme than an anomalous Agent.

Total Fixation Duration

The results of our statistical model indicated the presence of an interaction between the effects of our main variables *Group* and *Condition* on TFD, a measure of cognitive processing previously linked to integration difficulties of the word meaning within the sentences (Rayner et al., 2006). In other words, we found that the effect of the presence of a semantic violation (either on the Agent or the Theme pre-verbal subject) in the sentences acted differently in the two groups. Specifically, we found significant ($p < .01$) longer TFD in people with SSD compared to HPs. These results are indeed in line with the model proposed by Kuperberg and colleagues (2006) of a specific integration difficulty between syntactic and semantic information in sentence comprehension in SSD population.

Probability of regressions

Go-back movements during reading have been also linked to integration difficulties (Reichle et al., 1998). Overall, our group of SSD participants showed higher probability of regression in all conditions compared to healthy control participants.

Coherently with previous literature indicating a significant effect of integration load on neurophysiological measures of reading in people with SSD (Kuperberg et al., 2006), we found that the presence of a semantic violation did affect the probability of regression movement of the eye from the verb backwards in our experimental sample. More specifically, the main variables *Role* and *Condition* interacted differently in the two groups (p <.05), whereby SSD participants showed a significant higher probability to re-read a sentence containing a semantic violation on the Theme grammatical subject but, conversely, the presence of the same type of semantic violation did not affect the probability of regression when the element carrying the semantic violation was an Agent subject. These results are indeed in line with our findings on GD measures, and demonstrate the psychophysiological reality of Thematic Roles.

Accuracy

As for accuracy, an explicit measure of cognitive processes, our results show that a semantic violation of the grammatical subject is processed differently in the two groups. In details, HPs are significantly more accurate in categorizing sentences with a semantic violation (both on the Agent and the Theme), while individuals with SSD do not significantly differ in categorizing sentences with a semantic violation from sentences without violation.

Conclusions

To conclude, in this study we showed the specific contribution of a disrupted top-down linguistic processing to the reading performance in individuals with SSD. In particular, our results support the hypothesis of an impaired lexical representation of verbs implying an animated Agent as grammatical subject in people with SSD. This disrupted representation, which we link to the presence of a "disordered self", affects the implicit (gaze duration and probability of regression) elaboration of sentences. The effect observed in these two specific measures of sentence processing might be an instance of a more general integration difficulty, as clearly emerging in the measure of total fixation duration and might reflect an overall impairment in integrating semantic and syntactic information online, as suggested by Kuperberg and colleagues (2006).

Our study hence supports previous works pointing to a specific contribution of language impairments in the clinical presentation of this population. A limitation of the present study is that the experimental set of stimuli investigating the semantic plausibility of the Agent was not balanced between animate and inanimate nouns. In this sense, it is not possible to distinguish whether the specific tolerance to anomalous Agent showed by SSD participants is limited to a tolerance to inanimate subjects performing actions (for example, the "sleep" that "gives a call") or to animate subjects changing to implausible states (for example, the "owl" that "shines"). This study identifies an overall difficulty in processing the concept of "agentivity", and further research investigating language disturbances in people with SSD could indeed help to shed light on the specific (dis)ability showed by this clinical population in the recognition of the non-agentivity of inanimate objects, thus contributing to the explanation of, for example, delusion of persecution (e.g., being intentionally followed by the wind).

Moreover, our work could partly contribute to the understanding of the connection between language and schizophrenia: the so-called schizophrenic “word salad” could be partially explained by this anomalous lexical representation of verbs, which eventually produces the characteristic “incoherent speech”, made of bizarre and sometimes nonsensical word associations.

As eye-tracking is a non-invasive methodology, it is a good candidate for screening purposes in population at high risk for psychosis onset, as both language and oculomotor dysfunctions are known to precede the illness onset (Obyedkov et al., 2019). Given that reading abilities could play a role on educational and occupational achievements (Duncan et al., 2007), these results may prove useful in the context of cognitive remediation, contributing to a better socio-economic function of people with SSD.

6 General Discussion

The present work investigated language production and comprehension in people with SSD throughout a set of experiments aimed at a better understanding of the anomalous linguistic features observed in this population.

The theoretical frameworks proposed so far to explain language dysfunctions in SSD can be grouped in three main categories. The first group of works give priority to a disordered semantic store (over other cognitive processes) in order to explain the observed language anomalies: this first line of research suggests the content of the semantic store to be deviant (Aloia et al., 1998) and disorganized (Bozikas et al., 2005) due to a lack of semantic or conceptual knowledge (Goldberg et al., 1998; Paulsen et al., 1996), as well an abnormal spreading activation across semantic nodes (Moritz et al., 2002). Second, deficits in working memory and/or executive functions have been proposed to account for the majority of the language deviances observable in SSD patients, affecting linguistic sequence processing (Lelekov et al., 2000), as well as the building up and using of linguistic contextual information (Kuperber et al., 2010). Third, some Authors have reversed the causal relation between language and cognition and have derived SSD symptoms from a fundamental breakdown of the language system at the neural level. Among these, we find Crow's theory of schizophrenia as "the price we pay for language" (1997), as well as the closely related "un-Cartesian" linguistic framework of schizophrenia by Hinzen and Rosselló (2015). According to this view, from an evolutionary perspective, language dysfunctions *and* schizophrenia would be strictly linked by a common genetic event, that led to the hemispheric specialization, relegating language to the left hemisphere. From this stance, it would follow that, at the ontogenetic level, a neuro-developmental failure in the definition of brain asymmetry would lead to incomplete lateralization (and hence an impairment) of those language processes relying on the bilateral activation of neural circuits, parallel with the core symptoms of schizophrenia (with which they could be ontologically linked, according to Hinzen & Rosselló, 2015).

The objective of this project was four-fold. First, we aimed at investigating the differential contribution of semantic store integrity and executive processes to verbal production of people with SSD. Based on the first theoretical framework speaking for a selective impairment of the semantic store in this population, we would expect people with SSD to produce words "loosely" associated, and not organized in consistent and coherent semantic clusters. Secondly, we aimed at assessing the predictive ability of NLP-derived methods to discriminate the verbal production of people with and without SSD. Thirdly, we aimed at examining the integrity of the lexical representation of verb argument structure and the production and comprehension performance to complex syntactic structures in this population. Here, as the elaboration of verb argument structures was found to recruit the right homologous of the Broca's area when processing verbs with more than one argument (Thompson et al., 2007), Crow's hypothesis (1997) would predict a specific deficit, in the SSD population, for verbs entailing two or more arguments. Following the first theoretical framework, we would expect to see both the effect of an impaired semantic store on verb naming task, as well as a selective impairment for items entailing complex lexical representation. Moreover, according to the second theoretical framework, in tasks entailing the manipulation of complex syntactic structures, we would expect to see the effect of reduced executive functioning (Lelekov et al., 2000). Lastly, we aimed at exploring the integrity of the semantic representation of the verb Thematic grid. Given the presence of a "disordered self" (Henriksen & Nordgaard, 2014) in SSD, affecting the attribution of the "agency" of actions, we should expect this to affect the ability of people with SSD to process the construct of "Agent", one of the Thematic role store in the (lemma representation) of the verb.

For its complex nature, the study of mental illnesses, and in particular schizophrenia, has always required a multi-disciplinary approach. At the same time, the multifaceted construct of language calls for an innovative and composite approach, given that current clinical tests tend to overlook many components of language. This change of perspective has been considered a mandatory step for a real advancement toward the understanding of the pathophysiological mechanisms of severe mental illnesses (Elvevåg et al., 2016). The present research adopted a combined approach, moving from different disciplines and making use of the tools offered by computational linguistics, cognitive science and neuropsychology. For our study, we recruited a sample of SSD participants at the IRCCS San Giovanni di Dio Fatebenefratelli, in Brescia, matched with a group of healthy volunteers. An experimental battery of neuropsychological and linguistic tests, comprising two verbal fluency tasks (Chapter 2 and 3), a composite battery for the assessment of verb argument structure and syntactic complexity (NAVS, Chapter 4), as well as an anomaly detection task paired with an eye-tracking study (Chapter 5) was administered to the whole cohort.

The results of Chapters 2 and 3 show that the application of a fine-grained analysis to verbal fluency tests, beyond the mere count of total words produced, can provide insightful measures concerning the participants' performance. By taking into consideration measures such as the mean size of semantic clusters and the number of switches between them which, as proposed by Troyer and colleagues (1997), which are considered mirroring the integrity of the semantic store and executive functioning, respectively, we have been able to show how different cognitive processes interact with each other in order to perform the task. In particular, the analysis of the semantic VF task (Chapter 2) highlighted the presence of a deficit in the semantic store of people with SSD, visible in the production of high-frequency words, smaller semantic clusters and overall higher coherence of adjacent words compared to HPs. These results appear to be indicative of an anomalous spreading activation in a deteriorated semantic store, which enable the retrieval only of those items with the strongest representation (Collin & Loftus, 1975). The significant effect of the coherence between near words further corroborates this stance, as it highlights the inability of SSD participants to activate semantic nodes that are weakly connected, remaining bound to a closed semantic field. This last result, notably, does not replicate previous studies (Elvevåg et al., 2007), which found significant differences in a comparable measure of coherence between subsequent words as produced by HPs and people with SSD. On the contrary, we found a nearly identical value of coherence for adjacent words, while significant differences emerging at a later stage, i.e., between words at 3 and 5 in-list distance, whereby HPs showed lower coherence than SSD participants. In other words, moving on with the task, HPs can successfully explore their semantic store and produce words weakly related to each other, an indication of a structured and functioning semantic network. On the contrary, this process seems to be less performing in SSD patients, whose responses tend in fact to be bound to a single area of the semantic space.

Despite some commonalities, the comparison of the results of Chapter 2 and Chapter 3 highlights some remarkable differences. In both studies we found that people with SSD produced significantly fewer words than HPs, in line with previous literature (Bokat & Goldber, 2003). Moreover, albeit different in the raw number, the pattern of results across manually- and NLP-derived measures was consistently aligned in both studies, speaking for the construct validity of the automated approach. However, contrarily from results of the SVF task, in the generative associative naming task, HPs always produced clusters *smaller* in size than SSD participants. This difference highlights the different nature of the tasks under investigation. In SVF, participants are asked to generate items pertaining to a general category, requiring them to be bound to a single semantic field (i.e., "animal" in our case) and a single grammatical class (nouns). Inversely, in a generative associative naming task, participants are free to explore and retrieve words from the semantic store related to the given cue word but not necessarily part of the same semantic category. In other words, while an SVF task prompts the production of words paradigmatically related to each other, a generative associative naming task limit the production of words syntagmatically related to the cue word. In a generative associative naming task the cognitive processes recruited are mainly executive, mostly needed when the task at stake requires a "breadth-first" search heuristic in the exploration of the semantic store; inversely, the integrity of the semantic store is critical in a task requiring a strategy that rewards in-depth exploration of clusters of meaning in order to retrieve words sharing semantic features, such as in a semantic VF. From the comparison of the two tasks, it thus emerges the different contribution of executive functions and the integrity of the semantic store, as well as the abilities of count- *vs* predict-models to capture different types of semantic relatedness. Finally, we assessed the predictive abilities of such measures by mean of a supervised binary classification task: we created ROC curves from the fitted values of a logistic regression model having group membership as categorical dependent variable and the different indexes of verbal fluency (both computed manually and derived from our different semantic spaces) as predictors, and we compared the resulting Area Under the Curve. In terms of performance, values derived from a document-based matrix of co-occurrences (such as the one adopted by our LSA space) appears to be the fittest to capture the paradigmatic relationships induced by a semantic VF task (as in Chapter 2), while a narrow window-based matrix (such as those generated by word2vec) appears the best to detect syntagmatic relationships produced in the context of a generative associative naming task (Chapter 3). In this sense, the combination of, on one hand, the computational architecture behind the semantic spaces and, on the other hand, the types of fluency task under investigation revealed to be crucial in determining the classification performance. Crucially, we have proved that the different implementations at the basis of either approach can have a significant impact on the kind of semantic relationships they are able to capture. It follows that the nature of the task (in terms of type of semantic relationships that can prompt) should be taken into careful consideration when choosing the NLP algorithm to apply. With both these works (Chapter 2 and 3), we showed that, by implementing a detailed analysis of verbal production through NLP techniques, it is possible to conceive a fast, sensitive, and reliable tool, able to discriminate people with and without SSD. Furthermore, given their highest discriminative value, we showed that NLP-derived measures of cognitive performances are more suitable to feed a computational classifier compared to manually derived measures. This is because semantic relatedness as computed employing vector space models appears to be more sensitive to the subtle relationships between words compared to the traditional taxonomic approach (as in Chapter 2) or to reference lists (as in Chapter 3). Overall, these results appear promising for future research considering that, despite the central role of the assessment of speech during the objective examination of patients, psychiatry as a medical discipline still lacks hard surrogate biomarkers (or "speech intermediate phenotype" following Elvevåg et al., 2016) that can support clinician in the (early) diagnosis of SSD. The models proposed in this work must be considered proof-of-concepts and, before being ready to be applied in a real clinical context, the assessment of their differential diagnostic value is warranted: testing the discriminatory power in a categorization task including participants with, for example, SSD *vs* people with personality disorders *vs* HPs, could be a proper test bench in this sense.

The investigation of the syntactic abilities of people with SSD, in particular with respect to an impaired processing of verbs and their implement in a sentence structure (Chapter 4), showed that verbs with most complex lexical entries resulted in greater difficulty both verb naming, as well as for sentence production and comprehension in SDD participants *vis-à-vis* control participants. These findings appear to bear some functional relation to those observed in agrammatic aphasia (Barbieri et al., 2013; Cho-Reyes & Thompson, 2012; Kemmerer & Tranel, 2000; Kim & Thompson, 2000, 2004; Luzzatti et al., 2002; Thompson et al., 2007) where it has been observed that as the number of syntactic arguments increases, so does the difficulty in verb processing. Keeping in mind the different etiopathogenesis of the two conditions, these results point to similar dysfunctions affecting (at least partially) the same neural networks. As such, the outcomes of this work support the ASCH (Thompson, 2003) and extend its validity to the study of language in different clinical population beyond aphasia (i.e., SSD patients). Moreover, the specific deficit observed for complex argument structures suggest a disrupted mental representation of the grammatical class of verbs, bringing experimental support in favour of the literature that consistently reported a reduced performance on verb (action) processing in this population (Kambanaros et al., 2010; Marvel et al., 2004; Woods et al., 2007). Moreover, the results of the present study are in line with the concept of a distributed representation of verbs argument structure in the brain, as observed in the neuroimaging studies that proved the recruitment of bilateral cortical areas to process verbs with more than one argument (Thompson et al., 2007; Thompson et al., 2010). In fact, given the assumption of an incomplete lateralization of the language functions in people with SSD (Sommer et al., 2003; Weiss et al., 2006; Spaniel et al., 2007), the poor performance in our verb processing tasks can be interpreted as follow: when verbs subcategorize two or three arguments, the lack of a clear brain asymmetry specializing the brain areas on the two hemispheres, that presumably work in parallel, and the impaired inter-hemispheric connectivity (Chang et al., 2019), prevents SSD participants to deal with both verb arguments at the same time. Further neuroimaging studies would be needed to characterize the neural representation of verb processing in this population.

Finally, the collection of implicit (eye-movements) besides explicit (accuracy) measures of linguistic processing (Chapter 5) proved useful to identify a specific lexico-semantic deficit in SSD and enabled us to track down the temporal processing of sentences carrying semantic anomalies by SSD participants. In particular, by comparing the mean gaze duration (a measures of access to the lexical properties of a fixated item, Rayner et al., 2011) we wanted to test the impact of two different types of Thematic Roles ("Agent" vs "Theme"), occupying the pre-verbal position of grammatical subject, at the presence of a semantic violation of their animacy feature. The results showed that, when reading sentences such as "*Ogni sera il sonno telefona al nipotino*" (where the verb "*telefonare*" would require an animated Agent), individuals with SSD did not seem to detect a semantic violation, as they did not increase the processing time of the target verb, visually processing the anomalous stimulus as if it was correct. Along this same line, significant differences were found in the probability of regression movement from the verb backwards, again pointing to a misdetection of the semantic anomaly in the SSD population, contrary to HPs. Moreover, the significantly increased total fixation duration, a measure of processing related to the integration of the word meaning within the sentence (Rayner et al., 2006) supports a specific integration difficulty between syntactic and semantic information in sentence comprehension, in line with previous studies (Kuperberg et al., 2006). We have interpreted these results in the context of a lexicalist approach (Bock & Levelt, 1994), assuming that the verb thematic grid (i.e., the "*who does what to whom*" of a verb) is stored in its lexical representation, and in line with the assumption of the existence of a specific disorder of the Self (Henriksen & Noordgard, 2014) in schizophrenia. In this sense, what we see when we observed the inability of SSD participants to identify a semantic violation of the Agent would be the direct effect of the hypothesized disrupted sense of agency, which find its linguistic realization in the Thematic Role of "Agent".

Conclusions

Human language is an extraordinary symbolic system, whose complexity we can appreciate in full only when a brain damage, whether acquired or genetic, comes to disrupt its cognitive architecture. In the case of SSD, given the heterogeneity of symptoms and presentations across patients, it would be surprising to find a single explanatory factor. Rather, the interaction of different abnormal systems and processes appears more plausible. In this sense, it is reasonable to assume a fundamental breakdown in the organization of the language system in SSD, both at the neuroanatomical level and at the cognitive level (Crow, 1997; Hinzen & Rosselló, 2015). This study has tried to shed some light on a specific subset of language production and comprehension issues in this population. In particular, we were interested in understanding the cognitive procedures interacting during the processes of search and retrieval from the semantic store, needed in production and comprehension alike. Bock & Levelt's (1994) prediction concerning the consequences of a disruption of the lemma representation are quite revealing here, given that, taken all together, our results evidenced the central role of verb processing in the sentence frame. According to these Authors, given that the lemma representation (retrieved at the functional level) mediates between the conceptual information and the phonological syntactic form (realized at the positional level), a specific disruption of the lemma representation of verbs would affect its construction at the sentence level. This is exactly what we observed in the linguistic performance of our SSD participants in the studies reported in Chapter 4 and 5. Figure 6.1 attempts to summarize and illustrate this concept.

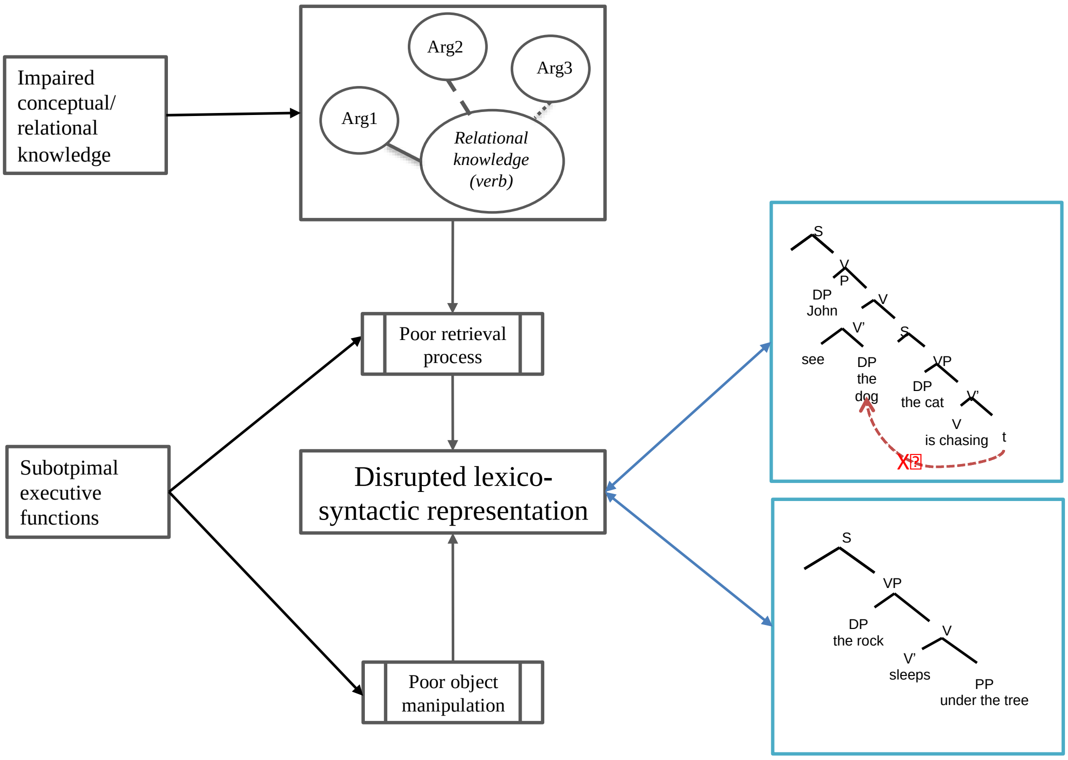


Fig.6.1 A tentative model to explain language breakdown in schizophrenia.

Assuming the existence of a pre-lexical/conceptual system, we can hypothesise this to be impaired both structurally (in terms of an altered connectivity between nodes) as well as functionally (in terms of an abnormal spreading activation) in people with SSD, as demonstrated by the poor performance of this clinical population in two different task of verbal fluency (Chapter 2 and 3). Furthermore, at this level, the "relational" knowledge between stored items, concisely identified with the grammatical class of verbs (Gentner, 1978), would be impaired in terms of both the number of mandatory arguments (nodes) and their referential characteristics. Suboptimal executive functions, as those observed in SSD, yield to poor retrieval processes (as seen in the outcome of a verbal fluency task), a decreased ability to manipulate mental objects (as seen, for example, in a sentence production priming task), and an impaired monitoring of perceptual stimuli against stored knowledge (as in the eye-tracking study). The result is a disrupted lexical-syntactic representation of language, which affects both the ability to produce a standard speech, and to monitor the receptive language to identify possible errors.

The results of the present work call for further investigations. In particular, in light of the literature underlying the presence of brain abnormalities in SSD, presumably affecting structures relevant for verb argument processing, further neuroimaging studies could help clarify the neurocognitive processes underlying verb argument structure processing in SSD. Moreover, all the studies included in the present dissertation were carried out limiting the recruitment to right-handers in order to reduce variance in the data. However, further studies in SSD should aim at including, or even targeting exclusively, left-handers, who in fact represent a considerable proportion of the SSD population. This could help to further characterize the contribution of asymmetrical brain development to language disturbances in SSD (Ribolsi et al., 2014).

Moreover, given the recognised effect of statistical distribution to language acquisition (Newport, 2016), further research on verb argument structure complexity should consider whether or one-argument verbs have a different statistical distribution in language compared to verbs with two or more obligatory arguments. Furthermore, given the promising results on verbal fluency scores and their ability to classify subjects (as reported in Chapters 2 and 3), NLP methodologies could be extended to enhance the representation of verb argument structure in terms, for example, of their selectional preference in the context of experimental paradigms investigating the ASCH (Thompson, 2003).

References

Abwender, D. A., Swan, J. G., & Bowerman, J. T. (2001). Qualitative analysis of verbal fluency output: review and comparison of several scoring methods. *Assessment*, *8*(3), 323–336.

Aloia, M. S., Gourovitch, M. L., Missar, D., Pickar, D., Weinberger, D. R., & Goldberg, T. E. (1998). Cognitive Substrates of Thought Disorder, II: Specifying a Candidate Cognitive Mechanism. *American Journal of Psychiatry*, *155*, 1677–1684.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th Edition). In *American Journal of Psychiatry*. https://doi.org/10.1176/appi.books.9780890425596.744053

Andreasen, N.C., & Olsen, S. (1982). Negative vs Positive Schizophrenia Definition and Validation. *Archives of General Psychiatry*, *39*(7), 789–794.

Andreasen, Nancy C. (1979). Thought, language and communication disorders. *Archives of General Psychiatry*, *36*, 1315–1321. https://doi.org/10.4324/9780203464199_chapter_4

Andreasen, Nancy C. (1986). Scale for the Assessment of Thought, Language, and Communication (TLC). *Schizophrenia Bulletin*, *12*(3).

Anselmetti, S., Poletti, S., Ermoli, E., Bechi, M., Cappa, S., Venneri, A., … Cavallaro, R. (2008). The brief assessment of cognition in schizophrenia. Normative data for the Italian population. *Neurological Sciences*, *29*(2), 85–92. https://doi.org/10.1007/s10072-008-0866-9

Artiges, E., Martinot, J. L., Verdys, M., Attar-Levy, D., Mazoyer, B., Tzourio, N., … Paillere-Martinot, M. L. (2000). Altered hemispheric functional dominance during word generation in negative schizophrenia. *Schizophrenia Bulletin*, *26*(3), 709–721. https://doi.org/10.1093/oxfordjournals.schbul.a033488

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

Badcock, J. C., Dragovic, M., Garrett, C., & Jablensky, A. (2011). Action (verb) fluency in schizophrenia: Getting a grip on odd speech. *Schizophrenia Research*, *126*(1–3), 138–143. https://doi.org/10.1016/j.schres.2010.11.004

Bagner, D. M., Melinder, M. R. D., & Barch, D. M. (2003). Language comprehension and working memory language comprehension and working memory deficits in patients with schizophrenia. *Schizophrenia Research*, *60*(2–3), 299–309. https://doi.org/10.1016/S0920-9964(02)00280-3

Bandera, L., Luzzatti, C., Santoro, E., & Spinnler, H. (1991). Prestazioni ad una prova di associazione libera durante l'invecchiamento. *Giornale Italiano di Psicologia*, *3*(18), 377–418.

Barbieri, E., Aggujaro, S., Molteni, F., & Luzzatti, C. (2013). Does argument structure complexity affect reading? A case study of an Italian agrammatic patient with deep dyslexia. *Applied Psycholinguistics*, *36*(3), 533–558. https://doi.org/10.1017/S0142716413000337

Barbieri, E., Brambilla, I., Thompson, C. K., & Luzzatti, C. (2019). Verb and sentence processing patterns in healthy Italian participants: Insight from the Northwestern Assessment of Verbs and Sentences (NAVS). *Journal of Communication Disorders*, *79*(March), 58–75. https://doi.org/10.1016/j.jcomdis.2019.03.001

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

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*, 238–247. Association for Computational Linguistics (ACL).

Baroni, M., Dinu, G., Kruszewski, G., & German, K. (2014). Don't count, predict! A systematic comparison ofcontext-counting vs. context-predicting semantic vectors. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*, 238–247. Association for Computational Linguistics (ACL).

Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*, *36*(4), 673–721. https://doi.org/10.1162/coli_a_00016

Baskak, B., Ozel, E. T., Atbasoglu, E. C., & Baskak, S. C. (2008). Peculiar word use as a possible trait marker in schizophrenia. *Schizophrenia Research*, *103*(1–3), 311–317. https://doi.org/10.1016/j.schres.2008.04.025

Bellani, M., Cerruti, S., & Brambilla, P. (2009). Language disorders in schizoprenia. *Epidemiologia e Psichiatria Sociale*, *18*(4), 314–317.

Berardi, G., Esuli, A., & Marcheggiani, D. (2015). Word embeddings go to Italy: A comparison of models and training datasets. *CEUR Workshop Proceedings*, *1404*. CEUR-WS.

Bleuler, E. (1958). *Dementia praecox, or the group of schizophrenias*. New York, NY: International Universities Press.

Bock, K., & Levelt, W. (1994). Language production: grammatical enconding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego: Academic Press.

Bokat, C. E., & Goldberg, T. E. (2003). Letter and category fluency in schizophrenic patients: A meta-analysis. *Schizophrenia Research*, *64*(1), 73–78. https://doi.org/10.1016/S0920-9964(02)00282-7

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond* (pp. 51–76). https://doi.org/10.4324/9780203509050

Bozikas, V. P., Kosmidis, M. H., & Karavatos, A. (2005). Disproportionate impairment in semantic verbal fluency in schizophrenia: Differential deficit in clustering. *Schizophrenia Research*, *74*(1), 51–59. https://doi.org/10.1016/j.schres.2004.05.001

Brambilla, P., Cerini, R., Gasparini, A., Versace, A., Andreone, N., Vittorini, E., … Tansella, M. (2005). Investigation of corpus callosum in schizophrenia with diffusion imaging. *Schizophrenia Research*, *79*(2–3), 201–210. https://doi.org/10.1016/j.schres.2005.04.012

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*(3), 890–907. https://doi.org/10.3758/s13428-011-0183-8

Burzio, L. (1986). *Italian Syntax: a Government-Binding Approach* (Dordrecht, Springer).

Callaway, E., & Naghdi, S. (1982). An Information Processing Model for Schizophrenia. *Archives of General Psychiatry*, *39*(3), 339–347. https://doi.org/10.1001/archpsyc.1982.04290030069012

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*(2), 117–175. https://doi.org/10.1016/0010-0277(85)90048-4

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, *119*(3), 933–949. https://doi.org/10.1093/brain/119.3.933

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, *26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachussetts: The MIT Press.

Chomsky, N. (1988). *Lectures on government and binding: The Pisa lectures*. Dordrecht, Foris Publication.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, Cortex, and Dopamine: A Connectionist Approach to Behavior and Biology in Schizophrenia. In *Psychological Review* (Vol. 99). https://doi.org/10.1037/0033-295X.99.1.45

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Condray, R., Steinhauer, S. R., Van Kammen, D. P., & Kasparek, A. (2002). The Language System in Schizophrenia: Effects of Capacity and Linguistic Structure. *Schizophrenia Bulletin*, *3*(28), 475–490. Retrieved from https://academic.oup.com/schizophreniabulletin/article-abstract/28/3/475/1839427

Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., … Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, *17*(1), 67–75. https://doi.org/10.1002/wps.20491

Cormen, T. H., Leiserson, C. E., Rivest, R. I., & Clifford, S. (2002). *Introduction to algorithms*. Cambridge, Massachussetts: The MIT Press.

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

Crow, T.J. (1997). Is schizophrenia the price Homo sapiens pays for language? *Schizophrenia Research, 28*, 127–141. https://doi.org/10.1192/bjp.192.4.289

Crow, T. J. (1980). Molecular pathology of schizophrenia: More than one disease process? *British Medical Journal*, *280*(6207), 66–68. https://doi.org/10.1136/bmj.280.6207.66

Crow, T.J. (1997). Schizophrenia as failure of hemispheric dominance for language. *Trends in Neurosciences, 20*, 339–343. https://doi.org/10.1016/S0166-2236(97)01193-4

Crow, T. J. (2008). The "big bang" theory of the origin of psychosis and the faculty of language. *Schizophrenia Research*, *102*(1–3), 31–52. https://doi.org/10.1016/j.schres.2008.03.010

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time:

Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401. https://doi.org/10.1076/jcen.20.3.391.810

de Boer, J. N., Voppel, A. E., Begemann, M. J. H., Schnack, H. G., Wijnen, F., & Sommer, I. E. C. (2018). Clinical use of semantic space models in psychiatry and neurology: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *93*, 85–92. https://doi.org/10.1016/j.neubiorev.2018.06.008

De Renzi, E., & Vignolo, L. A. (1962). The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, *85*(4), 665–678. https://doi.org/10.1093/brain/85.4.665

DeLisi, L. E. (2001). Speech disorder in schizophrenia: Review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophrenia Bulletin*, Vol. 27, pp. 481–496. https://doi.org/10.1093/oxfordjournals.schbul.a006889

Dempster, E., Viana, J., Pidsley, R., & Mill, J. (2013). Epigenetic studies of schizophrenia: Progress, predicaments, and promises for the future. *Schizophrenia Bulletin*, *39*(1), 11–16. https://doi.org/10.1093/schbul/sbs139

Dinu, G., Pham, N. T., & Baroni, M. (2013). *DISSECT-DIStributional SEmantics Composition Toolkit*. 31–36. Retrieved from http://clic.cimec.unitn.

DiSimoni, F. G., Darley, F. L., & Aronson, A. E. (1977). Patterns of Dysfunction in Schizophrenic Patients on an Aphasia Test Battery. *Journal of Speech and Hearing Disorders*, *42*(4), 498–513. https://doi.org/10.1044/jshd.4204.498

Doughty, O. J., & Done, D. J. (2009). Is semantic memory impaired in schizophrenia? A systematic review and meta-analysis of 91 studies. *Cognitive Neuropsychiatry*, *14*(6), 473–509. https://doi.org/10.1080/13546800903073291

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, *67*(3), 547. https://doi.org/10.2307/415037

Duchêne, A., Graves, R. E., & Brugger, P. (1998). Schizotypal thinking and associative processing: A response commonality analysis of verbal fluency. *Journal of Psychiatry and Neuroscience*, *23*(1), 56–60.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Zill, N. (2007). School Readiness and Later Achievement. *Psychological Association*, *43*(6), 1428–1446. https://doi.org/10.1037/[0012-1649.43.6.1428].supp

Elliott, R., & Sahakian, B. J. (1995). The Neuropsychology of Schizophrenia Relations with

Clinical and Neurobiological Dimensions. *Psychological Medicine*, *25*(3), 581–594. https://doi.org/10.1017/S0033291700033493

Elvevåg, B., Cohen, A. S., Wolters, M. K., Whalley, H. C., Gountouna, V. E., Kuznetsova, K. A., … Nicodemus, K. K. (2016, September 1). An examination of the language construct in NIMH's research domain criteria: Time for reconceptualization! *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, Vol. 171, pp. 904–919. https://doi.org/10.1002/ajmg.b.32438

Elvevåg, B., Fisher, J. E., Gurd, J. M., & Goldberg, T. E. (2002). Semantic clustering in verbal fluency: Schizophrenic patients versus control participants. *Psychological Medicine*, *32*(5), 909–917. https://doi.org/10.1017/S0033291702005597

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

Elvevåg, B., Weinstock, D. M., Akil, M., Kleinman, J. E., & Goldberg, T. E. (2001). A comparison of verbal fluency tasks in schizophrenic patients and normal controls. *Schizophrenia Research*, *51*, 119–126. Retrieved from www.elsevier.com/locate/schres

Everett, J., Lavoie, K., Gagnon, J. F., & Gosselin, N. (2001). Performance of patients with schizophrenia on the Wisconsin Card Sorting Test (WCST). *Journal of Psychiatry and Neuroscience*, *26*(2), 123–130.

Farzanfar, D., Statucka, M., & Cohn, M. (2018). Automated Indices of Clustering and Switching of Semantic Verbal Fluency in Parkinson's Disease. *Journal of the International Neuropsychological Society*, *24*(10), 1047–1056. https://doi.org/10.1017/S1355617718000759

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fernandez-Duque, D. (2009). *Cognitive and neural underpinnings of syntactic complexity*. 433–460. https://doi.org/10.1075/tsl.85.16cog

Fioravanti, M., Bianchi, V., & Cinti, M. E. (2012). *Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence*. Retrieved from http://www.biomedcentral.com/1471-244X/12/64

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005, June). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychology Review*, Vol. 15, pp. 73–95. https://doi.org/10.1007/s11065-005-6254-9

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2–3), 285–307. https://doi.org/10.1080/01638539809545029

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2018). Application of Latent Semantic Analysis to Divergent Thinking is Biased by Elaboration. *Journal of Creative Behavior*, *0*(0), 1–17. https://doi.org/10.1002/jocb.240

Fusar-Pioli, P., Deste, G., Smieskova, R., Barlati, S., Yung, A., Howes, O., … Borgwardt, S. (2012). Cognitive functioning in prodromal psychosis. *Archives of General Psychiatry, H. 69,* 562–571.

Galaverna, F., Bueno, A. M., Morra, C. A., Roca, M., & Torralva, T. (2016). Analysis of errors in verbal fluency tasks in patients with chronic schizophrenia. *European Journal of Psychiatry*, *30*(4), 305–320.

Gold, J., de Girolamo, G., Brambilla, L., Cappa, S., Mazzi, F., O'Donnell, K., … Pioli, R. (2002). Schizofrenia e funzioni cognitive: Una revisione critica della letteratura. *Rivista Di Psichiatria*, *37*(2), 53–60.

Gold, J. M., & Harvey, P. D. (1993). Cognitive deficits in schizophrenia. *Psychatric Clinics of North Americalinics of North America*, *16*(2).

Goldberg, T. E., Aloia, M. S., Gourovitch, M. L., Missar, D., Pickar, D., & Weinberger, D. R. (1998). Cognitive Substrates of Thought Disorder, I: The Semantic System. In *American Journal of Psychiatry* (Vol. 155).

Goldberg, T. E., Ragland, J. D., Torrey, E. F., Gold, J. M., Bigelow, L. B., & Weinberger, D. R. (1990). Neuropsychological Assessment of Monozygotic Twins Discordant for Schizophrenia. *Archives of General Psychiatry*, *47*(11), 1066–1072. https://doi.org/10.1001/archpsyc.1990.01810230082013

Goodkind, A., Lee, M., Martin, G. E., Losh, M., & Bicknell, K. (2018). Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics. *Proceedings Ofthe Society for Computation in Linguistics (SCiL)*, 12–22.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(3), 225–240. https://doi.org/10.1037/0278-7393.6.3.225

Günther, F., Dudschig, C., & Kaup, B. (2015). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental*

Psychology, *69*(4), 626–653. https://doi.org/10.1080/17470218.2015.1038280

Harris, Z. S. (1954). Distributional Structure. WORD, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Harvey, P. D., Bowie, C. R., Friedman, J. I., & J.I., F. (2001). Cognition in schizophrenia. *Current Psychiatry Reports*, *3*(5), 423–428. https://doi.org/10.1007/s11920-996-0038-7

Haug, E., Øie, M., Melle, I., Andreassen, O. A., Raballo, A., Bratlien, U., … Møller, P. (2012). The association between self-disorders and neurocognitive dysfunction in schizophrenia. *Schizophrenia Research*, *135*(1–3), 79–83. https://doi.org/10.1016/j.schres.2011.11.015

Hayes, R. L., & Maree O'Grady, B. (1998). Do People With Schizophrenia Comprehend What They Read? *Schizophrenia Bullettin*, *29*(3), 499–507. Retrieved from https://academic.oup.com/schizophreniabulletin/article-abstract/29/3/499/1928660

Henriksen, M. G., & Noordgard, J. (2014). Schizophrenia as a Disorder of the self. *Journal of Psychopathology*, *20*, 435–441. https://doi.org/10.1093/med/9780199660575.003.0009

Henry, J., Crawford, J., Henry, J. D., & Crawford, J. R. (2010). *A meta-analytic review of verbal fluency deficits in schizophrenia relative to other neurocognitive deficits schizophrenia relative to other neurocognitive deficits*. *6805*. https://doi.org/10.1080/13546800344000309

Hinzen, W. (2015). The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Frontiers in Psychology*, *6*(July), 1–17. https://doi.org/10.3389/fpsyg.2015.00971

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin and Review*, *23*(6), 1744–1756. https://doi.org/10.3758/s13423-016-1053-2

Holshausen, K., Harvey, P. D., Elvevåg, B., Foltz, P. W., & Bowie, C. R. (2014). Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex*, *55*(1), 88–96. https://doi.org/10.1016/j.cortex.2013.02.006

ICD-10. (1992). Tenth Revision of the International Classification of Diseases, Chapter V (F): Mental and Behavioural Disorders. Clinical descriptions and diagnostic guidelines. *WHO 1992; Geneva: 182/184.*, xii, 362 p.

Iter, D., Yoon, J., & Jurafsky, D. (2018, May 30). *Automatic Detection of Incoherent Speech for Diagnosing Schizophrenia*. 136–146. https://doi.org/10.18653/v1/w18-0615

Jablensky, A., Sartorius, N., Ernberg, G., Anker, M., Korten, A., Cooper, J. E., … Bertelsen, A.

(1992). Schizophrenia: manifestation, incidence and course in different cultures. A World Health Organization Ten Country Study. *Psychol., Med. Suppl.*, (20), 55–93.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. https://doi.org/10.1007/978-1-4614-7138-7

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive Impairment Following Frontal Lobe Damage and Its Relevance to Human Amnesia. *Behavioral Neuroscience*, *103*(3), 548–560. https://doi.org/10.1037/0735-7044.103.3.548

Jenkins, J. J., & Palermo, D. S. (1965). Further data on changes in word-association norms. *Journal of Personality and Social Psychology*, *1*(4), 303–309. https://doi.org/10.1037/h0021914

Johnson, D. E., & Shean, G. D. (1993). *Word association and schizophrenic symptoms* (Vol. 27).

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003

Kambanaros, M., Messinis, L., Georgiou, V., & Papathanassopoulos, P. (2010). Action and object naming in schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, *32*(10), 1083–1094. https://doi.org/10.1080/13803391003733578

Keefe, R. S. E., Harvey, P. D., Goldberg, T. E., Gold, J. M., Walker, T. M., Kennel, C., & Hawkins, K. (2008). Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). *Schizophrenia Research*, *102*(1–3), 108–115. https://doi.org/10.1016/j.schres.2008.03.024

Kemmerer, D., & Tranel, D. (2000). Verb retrieval in brain-damaged subjects: 1. Analysis of stimulus, lexical, and conceptual factors. *Brain and Language*, *73*, 347–392. https://doi.org/10.1006/brln.2000.2311

Kent, G. H., & Rosanoff, A. J. (1910). A Study of Association in Insanity. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(25), 695. https://doi.org/10.2307/2013279

Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language*, *74*(1), 1–25. https://doi.org/10.1006/brln.2000.2315

Kim, M., & Thompson, C. K. (2004). Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and Language*, *88*, 1–20. https://doi.org/10.1016/S0093-934X(03)00147-0

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

Koren, R., Kofman, O., & Berger, A. (2005). Analysis of word clustering in verbal fluency of school-aged children. *Archives of Clinical Neuropsychology*, *20*(8), 1087–1104. https://doi.org/10.1016/j.acn.2005.06.012

Kraepelin, E. (1886). *Compendio di psichiatria: per uso dei medici e degli studenti* (R. Burgia & C. Bonafigli, Eds.). Napoli: Vallardi.

Kuperberg, G. R., Kreher, D. A., & Ditman, T. (2010, February). What can Event-related Potentials tell us about language, and perhaps even thought, in schizophrenia? *International Journal of Psychophysiology*, Vol. 75, pp. 66–76. https://doi.org/10.1016/j.ijpsycho.2009.09.005

Kuperberg, G. R., McGuire, P. K., Tyler, L. K., & David, A. S. (1998). Reduced sensitivity to context in schizophrenic thought disorder: Evidence from online monitoring for words in linguistically-anomalous sentences. *Journal of Abnormal Psychology*, *107*(3), 423–434. https://doi.org/10.1016/s0920-9964(97)82377-8

Kuperberg, G. R., Sitnikova, T., Goff, D., & Holcomb, P. J. (2006). Making sense of sentences in schizophrenia: Electrophysiological evidence for abnormal interactions between semantic and syntactic processing. *Journal of Abnormal Psychology*, *115*(2), 251–265. https://doi.org/10.1037/0021-843X.115.2.251

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. In *Psychological Review* (Vol. 1).

Landrø, N. I., & Ueland, T. (2008). Verbal memory and verbal fluency in adolescents with schizophrenia spectrum disorders. *Psychiatry and Clinical Neurosciences*, *62*(6), 653–661. https://doi.org/10.1111/j.1440-1819.2008.01864.x

Le Dorze, G., & Durocher, J. (1992). The Effects of Age, Educational Level, and Stimulus Length on Naming in Normal Subjects. *Journal of Speech-Language Pathology and Audiology*, *16*(1), 21–29.

Lelekov, T., Franck, N., Dominey, P. F., Georgieff, N., & Williams, L. (2000). Cognitive sequence processing and syntactic comprehension in schizophrenia. In *NeuroReport* (Vol. 11).

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*, 151–171. https://doi.org/10.1146/annurev-linguistics

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb-noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language*, *81*(1–3), 432–444. https://doi.org/10.1006/brln.2001.2536

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marelli, M. (2017). Word-embeddings Italian Semantic spaces: A semantic model for psycholinguistic research. *Psihologija*, *50*(4), 503–520. https://doi.org/10.2298/PSI161208011M

Marvel, C. L., Schwartz, B. L., & Isaacs, K. L. (2004). Word production deficits in schizophrenia. *Brain and Language*, *89*(1), 182–191. https://doi.org/10.1016/S0093-934X(03)00366-3

Mayr, U. (2002). On the dissociation between clustering and switching in verbal fluency: Comment on Troyer, Moscovitch, Winocur, Alexander and Stuss. *Neuropsychologia*, *40*(5), 562–566. https://doi.org/10.1016/S0028-3932(01)00132-4

Mefferd, R. B. (1979). Word association: capacity of chronic schizophrenics to follow formal semantic, syntactic, and instructional rules. *Psychological Reports*, *45*(2), 431–442. https://doi.org/10.2466/pr0.1979.45.2.431

Merten, T. (1993). Word association responses and psychoticism. In *Personalities and individual Differences* (Vol. 14).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. Retrieved from http://arxiv.org/abs/1310.4546

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*(1), 90–100.

Mitchell, R. L. C., & Crow, T. J. (2005). Right hemisphere language functions and schizophrenia:

The forgotten hemisphere? *Brain*, *128*(5), 963–978. https://doi.org/10.1093/brain/awh466

Morice, R., & McNicol, D. (1985). The Comprehension and Production of Complex Syntax in Schizophrenia. *Cortex*, *21*(4), 567–580. https://doi.org/10.1016/S0010-9452(58)80005-2

Moritz, S., Woodward, T. S., Küppers, D., Lausen, A., & Schickel, M. (2002). Increased automatic spreading of activation in thought-disordered schizophrenic patients. *Schizophrenia Research*, *59*(2–3), 181–186. https://doi.org/10.1016/S0920-9964(01)00337-1

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

Morrens, M., Hulstijn, W., & Sabbe, B. (2007). Psychomotor slowing in schizophrenia. *Schizophrenia Bulletin*, *33*(4), 1038–1053. https://doi.org/10.1093/schbul/sbl051

Newport, E. L. (2016). Statistical language learning: computational, maturational and linguistic constraints. *Language and Cognition*, *8*(3), 447–461. https://doi.org/10.1017/langcog.2016.20.Statistical

Nicolson, R., Lenane, M., Sujatha Singaracharlu, M., Malaspina, D., Giedd, J. N., Hamburger, S. D., … Rapoport, J. L. (2000). Premorbid Speech and Language Impairments in Childhood-Onset Schizophrenia: Association With Risk Factors. *American Journal of Psychiatry*, *157*(5).

Nieuwenstein, M. R., Aleman, A., & De Haan, E. H. F. (2001). Relationship between symptom dimensions and neurocognitive functioning in schizophrenia: A meta-analysis of WCST and CPT studies. *Journal of Psychiatric Research*, *35*(2), 119–125. https://doi.org/10.1016/S0022-3956(01)00014-0

Niwa, Y., & Nitta, Y. (1995). *Co-occurrence vectors from corpora vs. distance vectors from dictionaries*. Retrieved from: https://doi.org/10.3115/991886.991938

Nordgaard, J., Nilsson, L. S., Sæbye, D., & Parnas, J. (2018). Self-disorders in schizophrenia-spectrum disorders: a 5-year follow-up study. *European Archives of Psychiatry and Clinical Neuroscience*, *268*(7), 713–718. https://doi.org/10.1007/s00406-017-0837-3

Novelli, G., Papagno, C., Capitani, E., & Laiacona, M. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su sogetti normali. *Archivio Di Psicologia, Neurologia e Psichiatria.*

Obyedkov, I., Skuhareuskaya, M., Skugarevsky, O., Obyedkov, V., Buslauski, P., Skuhareuskaya, T., & Waszkiewicz, N. (2019). Saccadic eye movements in different dimensions of schizophrenia and in clinical high-risk state for psychosis. *BMC Psychiatry*, *19*(1). https://doi.org/10.1186/s12888-019-2093-8

Oh, T. M., McCarthy, R. A., & McKenna, P. J. (2002). Is there a schizophasia? A study applying the single case approach to formal thought disorder in schizophrenia. *Neurocase*, *8*(3), 244–244. https://doi.org/10.1093/neucas/8.3.244

Okasha, A., & Okasha, T. (2000). Notes on mental disorders in Pharaonic Egypt. *History of Psychiatry*, *11*(44), 413–424. https://doi.org/10.1177/0957154X0001104406

Okruszek, Ł., Rutkowska, A., & Wilińska, J. (2013). Clustering and switching strategies during the semantic fluency task in men with frontal lobe lesions and in men with Schizophrenia. *Psychology of Language and Communication*, *17*(1), 93–100. https://doi.org/10.2478/plc-2013-0006

Overall, J. E., & Pfefferbaum, B. (1962). The brief psychiatric rating scale for children. *Psychological Reports*, *10*, 799–812.

Padó, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, *33*(2), 161–199. https://doi.org/10.1162/coli.2007.33.2.161

Pakhomov, S. V. S., & Hemmy, L. S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. *Cortex*, *55*(1), 97–106. https://doi.org/10.1016/j.cortex.2013.05.009

Pakhomov, S. V. S., Hemmy, L. S., & Lim, K. O. (2012). Automated semantic indices related to cognitive function and rate of cognitive decline. *Neuropsychologia*, *50*(9), 2165–2175. https://doi.org/10.1016/j.neuropsychologia.2012.05.016

Parisi, D., & Pizzamiglio, L. (1967). *Dati quantitativi sulle associazioni verbali di studenti italiani*. Roma: Bulzoni.

Parks, R. W., Lowenstein, D. A., Dodrill, K. L., Barker, W. W., Yoshii, F., Chang, J. Y., … Duara, R. (1988). Cerebral metabolic effects of a verbal fluency test:A PET scan study. *Journal of Clinical and Experimental Neuropsychology1*, *10*(5).

Paula, F., Wilkens, R., Idiart, M., & Villavicencio, A. (2018, May 30). *Similarity Measures for the Detection of Clinical Conditions with Verbal Fluency Tasks*. 231–235. https://doi.org/10.18653/v1/n18-2037

Paulsen, J. S., Romeroayb, R., Chanayb, A., Davisa>b, A. V, Heatona, R. K., Jesteaybvc, D. V, … Jeste, D. V. (1996). Impairment of the semantic network in schizophrenia. *Psychiatry*

Research, *63*, 109–121. https://doi.org/10.1016/0165-1781(96)02901-0

Pauselli, L., Halpern, B., Cleary, S. D., Ku, B., Covington, M. A., & Compton, M. T. (2018). Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. *Psychiatry Research*, *263*, 74–79. https://doi.org/10.1016/j.psychres.2018.02.037

Perlmutter, D. M. (1978). Impersonal Passives and the Unaccusative Hypothesis. *Annual Meeting of the Berkeley Linguistics Society*, *4*, 157. https://doi.org/10.3765/bls.v4i0.2198

Postmes, L., Sno, H. N., Goedhart, S., van der Stel, J., Heering, H. D., & de Haan, L. (2014) Schizophrenia as a self-disorder due to perceptual incoherence, S*chizophrenia Research*, 152(1), 41-50.

R CoreTeam. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Retrieved from https://www.r-project.org/

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. https://doi.org/10.1093/schbul/17.4.555

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2016, December 15). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, Vol. 18, pp. 42–55. https://doi.org/10.1038/nrn.2016.150

Rausch, M. A., Prescott, T. E., & DeWolfe, A. S. (1980). Schizophrenic and aphasic language: Discriminable or not? *Journal of Consulting and Clinical Psychology*, *48*(1), 63–70. https://doi.org/10.1037/0022-006X.48.1.63

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, Vol. 10, pp. 241–255. https://doi.org/10.1207/s1532799xssr1003_3

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye Movements and Word Skipping During Reading: Effects of Word Length and Predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528. https://doi.org/10.1037/a0020990

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). *Psychological Review Toward a Model of Eye Movement Control in Reading* (Vol. 105). Psychological Association, Inc.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinformcement. In *Classical conditioning II: current* *research and theory* (pp. 64–99). Retrieved from papers2://publication/uuid/51EED98C-39D3-4ECA-9CC8-F7E445CCB145

Resnik, S. M. (1992). Matching for Education in Studies of Schizophrenia. *Archives of General Psychiatry*, *49*, 246.

Ribolsi, M., Daskalakis, Z. J., Siracusano, A., & Koch, G. (2014). Abnormal asymmetry of brain connectivity in schizophrenia. *Frontiers in Human Neuroscience*, *8*(DEC), 1–11. https://doi.org/10.3389/fnhum.2014.01010

Robert, P. H., Lafont, V., Medecin, I., Berthet, L., Thauby, S., Baudu, C., … Darcourt, G. (1998). Clustering and switching strategies in verbal fluency tasks: Comparison between schizophrenics and healthy adults. *Journal of International Neuropsychological Society*, *4*(6), 539–546. https://doi.org/10.1017/S1355617798466025

Roberts, E. O., Proudlock, F. A., Martin, K., Reveley, M. A., Al-Uzri, M., & Gottlob, I. (2013). Reading in schizophrenic subjects and their nonsymptomatic first-degree relatives. *Schizophrenia Bulletin*, *39*(4), 896–907. https://doi.org/10.1093/schbul/sbr191

Rochester, S., & Martin, J. R. (1979). *Crazy talk — a study of the discourse of schizophrenic speakers.* New York, NY: Plenum Press.

Rollins, A. L., Bond, G. R., Lysaker, P. H., Grew, J. H., & Salyers, M. P. (2010). Coping with Negative and positive Symptoms of Schizophrenia. *American Journal of Psychiatric Rehabilitation*, *13*, 208–223.

Rong, X. (2016). *word2vec Parameter Learning Explained*. Retrieved from http://arxiv.org/abs/1411.2738

Rossell, S. L., & David, A. S. (2006). Are semantic deficits in schizophrenia due to problems with access or storage? *Schizophrenia Research*, *82*(2–3), 121–134. https://doi.org/10.1016/j.schres.2005.11.001

Ryan, J. O., Pakhomov, S., Marino, S., Bernick, C., & Banks, S. (2013). Computerized analysis of a verbal fluency test. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, *2*, 1–6.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, *20*(1), 33–53.

Sarioglu Kayi, E., Diab, M., Pauselli, L., Compton, M., & Coppersmith, G. (2017, July 27).

Predictive Linguistic Features of Schizophrenia. 241–250. https://doi.org/10.18653/v1/s17-1028

Saussure, F. de. (1986). *Corso di linguistica generale* (T. de Mauro, Ed.). Roma: Laterza.

Saykin, A. J., Shtasel, D. ., Gur, R., Kester, D. ., Mozley, L. ., Stafiniak, P., & Gur, R. C. (1994). First-Episode Schizophrenia. *Archives of General Psychiatry*, *51*.

Scott, W. A. (1962). Cognitive complexity and Cognitive flexibility. *Sociometry*, *25*(4), 405–414.

Shakow, D., & Jellinek, E. (1965). Composite index of the Kent-Rosanoff Free Association Test. *Journal of Abnormal Psychology*, *70*(6), 403–404. https://doi.org/10.1037/h0022749

Shapleske, J., Rossell, S. L., Woodruff, P. W. R., & David, A. S. (1999, January). The planum temporale: A systematic, quantitative review of its structural, functional and clinical significance. *Brain Research Reviews*, Vol. 29, pp. 26–49. https://doi.org/10.1016/S0165-0173(98)00047-2

Shenton, M. E., Dickey, C. C., Frumin, M., & McCarley, R. W. (2001, April 15). A review of MRI findings in schizophrenia. *Schizophrenia Research*, Vol. 49, pp. 1–52. https://doi.org/10.1016/S0920-9964(01)00163-3

Sigurd, B., Eeg-Olofsson, M., & Van De Weijer, J. (2004). Word length, sentence length and frequency - Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. https://doi.org/10.1111/j.0039-3193.2004.00109.x

Singh, S. (2017). The Diagnostic Utility of Kent-Rosanoff Word Association Test In relation to Response Entropy. *The International Journal of Indian Psychology*, *5*(121–140). https://doi.org/10.25215/0501.053

Smirnova, D., Clark, M., Jablensky, A., & Badcock, J. C. (2017). Action (verb) fluency deficits in schizophrenia spectrum disorders: linking language, cognition and interpersonal functioning. *Psychiatry Research*, *257*, 203–211. https://doi.org/10.1016/j.psychres.2017.07.044

Sommer, I. E. C., Ramsey, N. F., Mandl, R. C. W., & Kahn, R. S. (2003). Language lateralization in female patients with schizophrenia: An fMRI study. *Schizophrenia Research*, *60*(2–3), 183–190. https://doi.org/10.1016/S0920-9964(02)00300-6

Spaniel, F., Tintera, J., Hajek, T., Horacek, J., Dezortova, M., Hajek, M., … Höschl, C. (2007). Language lateralization in monozygotic twins discordant and concordant for schizophrenia. A functional MRI pilot study. *European Psychiatry*, *22*(5), 319–322. https://doi.org/10.1016/j.eurpsy.2006.11.002

Spinnler, H., & Tognoni, G. (1987). Standardizzazione e taratura italiana di test neuropsicologici. Ital J Neurol Sci, 6, 21-120.

Stirling, J., Hellewell, J., Blakey, A., & Deakin, W. (2006). Thought disorder in schizophrenia is associated with both executive dysfunction and circumscribed impairments in semantic function. *Psychological Medicine*, *36*(4), 475–484. https://doi.org/10.1017/S0033291705006884

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait. *Archives of General Psychiatry*, *60*(12), 1187. https://doi.org/10.1001/archpsyc.60.12.1187

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest*, *1*(1), 1–26. https://doi.org/10.1111/1529-1006.001

Tavano, A., Sponda, S., Fabbro, F., Perlini, C., Rambaldelli, G., Ferro, A., … Brambilla, P. (2008). Specific linguistic and pragmatic deficits in Italian patients with schizophrenia. *Schizophrenia Research*, *102*(1–3), 53–62. https://doi.org/10.1016/j.schres.2008.02.008

Thomas, P. (1996). Syntactic Complexity and Negative Symptoms in First Onset Schizophrenia. *Cognitive Neuropsychiatry*, *1*(3), 191–200.

Thomas, P., Kearney, G., Napier, E., Ellis, E., Leudar, I., & Johnston, M. (1996). Speech and language in first onset psychosis differences between people with schizophrenia, mania, and controls. *British Journal of Psychiatry*, *168*(MAR.), 337–343. https://doi.org/10.1192/bjp.168.3.337

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics*, *16*(2–3), 151–167. https://doi.org/10.1016/S0911-6044(02)00014-3

Thompson, C. K., Choy, J. W. J., Holland, A., & Cole, R. (2010). Sentactics®: Computer-automated treatment of underlying forms. *Aphasiology*, *24*(10), 1242–1266. https://doi.org/10.1080/02687030903474255

Thompson, C. K., Dickey, M. W., Cho, S., Lee, J., & Griffin, Z. (2007). Verb argument structure encoding during sentence production in agrammatic aphasic speakers: An eye-tracking study. *Brain and Language*, *103*(1–2), 24–26. https://doi.org/10.1016/j.bandl.2007.07.012

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. In *Journal of Artificial Intelligence Research* (Vol. 37). Retrieved from http://www.natcorp.ox.ac.uk/.

van Os, J., & Kapur, S. (2009). Schizophrenia. *The Lancet*, *374*(9690), 635–645. https://doi.org/10.1016/S0140-6736(09)60995-8

van Valin, R. D. (1997). Generalized semantic roles and the syntax-semantic interface. In *Empirical Issues in Formal Syntax and Semantics 2* (pp. 373–389).

Weiss, E. M., Hofer, A., Golaszewski, S., Siedentopf, C., Felber, S., & Fleischhacker, W. W. (2006). Language lateralization in unmedicated patients during an acute episode of schizophrenia: A functional MRI study. *Psychiatry Research - Neuroimaging*, *146*(2), 185–190. https://doi.org/10.1016/j.pscychresns.2005.11.003

Whitford, V., O'Driscoll, G. A., Pack, C. C., Joober, R., Malla, A., & Titone, D. (2013). Reading impairments in schizophrenia relate to individual differences in phonological processing and oculomotor control: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: General*, *142*(1), 57–75. https://doi.org/10.1037/a0028062

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The Measurement of Individual differences in originality. *The Psychological Bullettin*, *5*.

Wittgenstein, L. (1953). *Philosophical Investigations*. https://doi.org/10.1126/science.53.1377.478

Woods, S. P., Weinborn, M., Posada, C., & O'Grady, J. (2007). Preliminary evidence for impaired rapid verb generation in schizophrenia. *Brain and Language*, *102*(1), 46–51. https://doi.org/10.1016/j.bandl.2006.11.005

Wu, D. H., Waller, S., & Chatterjee, A. (2007). The functional neuroanatomy of thematic role and locative relational knowledge. *Journal of Cognitive Neuroscience*, *19*(9), 1542–1555. https://doi.org/10.1162/jocn.2007.19.9.1542

Wykes, T., & Leff, J. (1982). Disordered speech: Differences between manics and schizophrenics. *Brain and Language*, *15*(1), 117–124. https://doi.org/10.1016/0093-934X(82)90051-7

Youngstrom, E. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204–221. https://doi.org/10.1093/jpepsy/jst062

Zalla, T., Posada, A., Franck, N., Georgieff, N., & Sirigu, A. (2001). A component analysis of action planning processes in schizophrenia: A comparison with patients with frontal lobe damage.

Cognitive Neuropsychiatry, *6*(4), 271–296. https://doi.org/10.1080/13546800143000087

Annex 1 Additional materials for Chapter 2 and 3

A.1 Pre-processing

Table A.1 reports the code lines applied to the pre-processing of the itWac corpus.

1	convert the corpus in a standard coding system	`iconv -f ISO-8859-1 -t utf8 <itwac_preproc >itwac_preproc.utf8`
2	remove all special characters, with the exception of vowels with orthographically-marked stresses (à, è, é, ì, ò, ù) and tokenize it	`tr -cs '[A-Z][a-z][àèéìòù]' '[\012*]' <itwac_preproc.utf8 >itwac_preproc.utf8.tkn`
3	convert all characters to lower case	`tr '[A-Z]' '[a-z]' < itwac_preproc.utf8.tkn > itwac_preproc.utf8.tkn.low`
4	Run w2v on the pre-processed corpus	`./word2vec -train itwac_preproc.utf8.tkn.low -output itwac_preproc.utf8.tkn.low.vec -size 400 -window 5 -sample 1e-5 -negative 10 -hs 0 -binary 0 -cbow 1 -iter 3 min_count 100`

Table A.1. Itwac pre-processing coding lines

A.2 R functions “`Troyer-derived`” and “`Koren-derived`”

We developed two R-based functions to compute number of switches and mean size of clusters from verbal fluency tasks following Troyer’s proposed methodology. Table A.2 and A.3 report the annotated scripts of the “Troyer-derived” and “Koren-derived” functions, respectively.

```
Troyer-derived = function(resp, threshold,
ss){

  switches = 0
  bc = 1
  N_cluster = 0
  cluster_size = 0

  for (i in 2 : length(resp)){

    cluster = resp[bc : (i-1)]
    t = ss[resp[i], ]
    c_set = ss[cluster,]

    if (is.matrix(c_set) == FALSE){
      c = c_set
    }
    else{
      c = colMeans(c_set)
    }

    sim = t %*% c / sqrt(t%*%t * c%*%c)

    if (sim < threshold) {
      switches = switches +1
      bc = i
      print(resp[i])
      }
    }

  N_cluster = switches + 1
  mean_cs = length(resp)/N_cluster
  results <- c(N_cluster, mean_cs)
  return(results)
}
```

```
#resp is the list of words produced by the
subject (must be a vector), threshold is the
threshold, ss the semantic space

# set count switches to zero
# beginning of cluster to word N 1
# set count of cluster to zero
# set cluster size to zero

# set a for cycle to read all the words in
resp
# the cluster is made up of all those words
in resp from the beginning of the cluster
(bc) to i-1
# retrieve the vector of the target word
from the semantic space
# retrieve the vectors of the words in the
cluster

# if the cluster is made up of a single word
(i.e., first cycle) …
# ...no additional transformations are
needed (it remains a vector, not a matrix)
 # if the cluster is made up of more than
one words…
 # ... the mean of those vectors must be
computed

#compute cosine similarities between target
word and cluster


# if sim cosine is below threshold
# set a for cycle to read all the words in
resp

# update number of switches
# set the beginning of a new cluster
# print the word that set the switches

# compute mean cluster size
# list all results in a vector
# return results
```

Table. A.2 R code of the Troyer-derived algorithm

A detailed description of the first function is given below.

The function "`Troyer-derived`" stores three different counters for the parameters of interests that are initially set to zero and then updated iteratively at each cycle, and namely:

- number of switches ("`switches`");
- number of clusters ("`N_cluster`"); and
- mean cluster size ("`cluster_size`").

Moreover, it sets the beginning of the cluster ("`bc`") to 1, to avoid out of bound errors (see further). An iterative loop ("`for`" cycle) is initiated, whose length is equal to the length of the list of words produced by the subject. At each cycle, the function reads a word from the subject's response list. As the function compares each word with the one preceding it (or with the preceding cluster, as we will see later), it is set to start from the second item of the list. Within the "`for`" cycle, three variables are created:

- "cluster" is a character vector listing all words in the subject response from "bc" to the word preceding the one under investigation;
- "t" is the numerical representation of the target word (word vector) retrieved from the semantic space defined;
- "c_set" is a variable containing the numerical representations of all the words in the cluster.

An "if" function checks the class of "c_set": if this variable is a matrix (i.e., if the cluster is made up of more than one word), it calculates the mean vector and stores the result in a new variable, a vector called "c". Otherwise, it simply renames "c_set" to "c". This is done because the following instruction will take only vector as input.

Now, the function is ready to compute the cosine of the two vectors, "c" and "t", and to store the result in a new variable called "sim". Next, an "if" function compares "sim" to the pre-set threshold. If the value is lower than the threshold, this means the function has encountered a shift: the initial counter "switches" is updated (+1), a new cluster is set to begin ("bc = i") at the next iterative cycle, and the word causing a shift is printed. The function will not exit the "if" loop until the compared values are above threshold, meaning that the function is reading words that are part of a semantic cluster. Once out, it updates the number of cluster (+1), calculates the mean cluster size ("mean_cs"), and returns the desired values.

```
Koren-derived = function(resp,              #resp is the list of words produced by the
threshold, ss){                             subject (must be a vector), threshold is the
                                            threshold, ss the semantic space
  bc = 1                                    # beginning of cluster to word N 1
  N_cluster = 0                             # set count of cluster to zero
  cluster_size = 0                          # set cluster size to zero
                                            # set a for cycle to read all the words in
                                            resp
  for (i in 2 : length(resp)){              # the cluster is made up of all those words
    cluster = resp[bc : (i-1)]              in resp from the beginning of the cluster
                                            (bc) to i-1
    t = ss[resp[i], ]                       # retrieve the vector of the target word
                                            from the semantic space
    c_set = ss[cluster,]                    # retrieve the vectors of the words in the
                                            cluster
    if (is.matrix(c_set) == FALSE)    {     # if the cluster is made up of a single word
      c = c_set                             (i.e., first cycle) …
    }                                       # ...no additional transformations are
    else{                                   needed (it remains a vector, not a matrix)
      c = colMeans(c_set)                   # if the cluster is made up of more than one
    }                                       words…
                                            # ... the mean of those vectors must be
                                            computed
    sim = t %*% c / sqrt(t%*%t * c%*%c)     # compute cosine similarities between target
                                            word and cluster
                                            # if sim cosine is below threshold and
                                            cluster > 1 item
    if((sim <                               # update number of clusters
threshold)&(is.matrix(c_set) == TRUE)) {    # set the beginning of a new cluster
                                            # print the word that set the shift
      N_cluster = N_cluster+1               # record the size of the cluster
      bc = i                                #if sim cosine is below threshold but
      print(resp[i])                        cluster is < 1 item
cluster_size = cluster_size                 # set a new cluster
+(length(c_set)/ncol(ss))

    }
else if ((sim <                             # compute mean cluster size
soglia)&(is.matrix(c_set) ==   FALSE)) {    # list all results in a vector
      bc = i                                # return results
    }
  }

  mean_cs = cluster_size/N_cluster
  results <- c(N_cluster, mean_cs)
  return(results)
}
```

Table A.3 R code of Koren-derived algorithm

A.3 Measure of overall coherence

Table A.4 reports the annotated script of the “coherence” function.

```
coherence = function(resp, target, ss) {          # vector of words, integer,
  sim_tot = 0                                     semantic space
  count.loop = 0

  for (i in (1:length(resp))) {                   # set a for cycle to read all the
    c = ss[resp[i], ]                             words in resp
    count.loop = count.loop + 1                   # retrieve the vector of the cue
                                                  word from the semantic space
    if (isTRUE((i+target) > length(resp))) {
     break                                        # retrieve the target of the cue
     } else {t = ss[resp[(i+target)],]            word from the semantic space (1,
             sim = (t %*% c / sqrt(t%*%t * c%*%c))  3, 5, 7)
     }                                            # compute cosine
    if (is.nan(sim)) {sim <- 0}
    sim_tot = sim_tot + sim
    }

  mean_sim = sim_tot/count.loop
  return(mean_sim)                                # return results

  }
```

Table. A.4 R code for the `coherence` algorithm.

Annex 2 Additional materials for Chapter 5

Target stimuli

Table B.1 and B.2 report the complete list of target stimuli employed in the eye-tracking study in Chapter 5 (target stimuli and fillers).

AGENT - correct	AGENT – with violation	THEME - correct	THEME – with violation
In estate la luce brilla attraverso le ante	Nel cantiere l’architetto solleva la cisterna	Al mattino le stelle tramontano dietro alla collina	In paese le fontane cambiano a ogni generazione
Alla luce i diamanti luccicano nell’astuccio	In inverno la neve riposa nella grotta	In paese gli abitanti cambiano a ogni generazione	Davanti al camino la zia rimane calda
Ogni sera Bianca telefona al nipotino	Ogni sera il sonno telefona al nipotino	A volte il bambino cade dalle scale	A volte il sole cade dalle scale
Sulla roccia il leone ruggisce con fierezza	Sulla roccia l'arbusto ruggisce con fierezza	In estate il maestro parte per la montagna	In estate la tana parte per la montagna
Alla festa gli uomini ridono di gusto	Alla festa i bicchieri ridono di gusto	All’improvviso il ragazzo sviene per la paura	All’improvviso l’oro sviene per la paura
In cucina i gatti dormono al sole	In cucina le bucce dormono al sole	Tutti i giovedì il professore arriva in ritardo	Tutti i giovedì il cielo arriva in ritardo
In estate i calciatori giocano all'aperto	In estate le panchine giocano all'aperto	Ogni giorno molti soldati muoiono nel mondo	Ogni giorno molte scarpe muoiono nel mondo
Nella favola la regina abdica al trono	Nella favola la colonna abdica al trono	A mezzogiorno i pensionati vanno in banca	A mezzogiorno i negozi vanno in banca
A pranzo il papà sbuccia l’arancia	A pranzo il porto sbuccia l’arancia	Grazie alla pedana l’elefantino scende dal camion	Grazie alla pedana il cruscotto scende dal camion
Sul marciapiede il ciclista gonfia la ruota	Sul marciapiede il parchimetro gonfia la ruota	Alle sei gli operai tornano a casa	Alle sei i ristoranti tornano a casa
La mattina i postini consegnano le lettere	La mattina i cereali consegnano le lettere	Ogni giorno molti neonati nascono in ospedale	Ogni giorno molti treni nascono in ospedale
Nel fine settimana i cuochi cucinano le torte	Nel fine settimana i parchi cucinano le torte	A mezzanotte i fuochi scoppiano nel cielo	A mezzanotte gli innamorati scoppiano nel cielo
Tutti i giorni la mamma legge il giornale	Tutti i giorni il muro legge il giornale	All'aeroporto gli aerei decollano a ogni ora	In primavera la gallina sboccia nel campo
In autostrada il camionista guida con prudenza	In autostrada l'asfalto guida con prudenza	In primavera il fiore sboccia nel campo	Questa settimana l'avvocato tramonta al mattino
Ogni sera sua moglie regola la sveglia	Ogni sera la notizia regola la sveglia	Con la convalescenza i malati guariscono dalla malattia	Con la convalescenza gli ospedali guariscono dalla malattia
D’estate le onde bagnano la riva del mare	D’estate le tasse bagnano la riva del mare	Dopo le feste molte persone ingrassano di qualche chilo	Dopo le feste molti piatti ingrassano di qualche chilo
A pranzo gli atleti mangiano i maccheroni	A pranzo i gomiti mangiano i maccheroni	All’improvviso le rovine franano rumorosamente	All’improvviso le lavoratrici franano rumorosamente
Di giorno i cani abbaiano alle automobili	Di giorno le rose abbaiano alle automobili	Sul balcone le fragole spuntano nei vasi	Sul balcone le insegnanti spuntano nei vasi

La sera i pesci abboccano con facilità	La sera i dischi abboccano con facilità	Sul monte gli alberi fioriscono a primavera	Sul monte gli studenti fioriscono a primavera
Nel castello la spada scintilla accanto al camino	Nel castello lo stalliere scintilla accanto al camino	La sera i fulmini lampeggiano sulla montagna	La sera i cervi lampeggiano sulla montagna
A mezzogiorno le campane risuonano nella valle	A mezzogiorno le ballerine risuonano nella valle	Col caldo l'acqua evapora dalla pozzanghera	Col caldo il dottore evapora dalla pozzanghera
All'occorrenza i pompieri corrono in aiuto	Quando occorre i papaveri corrono in aiuto	Dopo la sconfitta il nemico fugge sulle montagne	Col maltempo il caffè rimane al porto
Nel cantiere la gru solleva la cisterna	Nel cielo il gufo brilla a primavera	Col maltempo la nave rimane al porto	Dopo un terremoto le macerie intervengono con tempestività
In ufficio i telefoni squillano senza sosta	In ufficio le mosche squillano senza sosta	Dopo un terremoto i pompieri intervengono con tempestività	Alle quattro i telefoni escono dall'ufficio
A Natale il nonno regala qualche caramella	Alla luce gli attori luccicano nell'astuccio	Alle quattro gli impiegati escono dall'ufficio	Il mercoledì il cane costa di meno
Sulla montagna il vento soffia indisturbato	Sulla montagna il falco soffia indisturbato	Il mercoledì il cinema costa di meno	In Niger le madri scarseggiano durante la carestia
Sul banco la cartella contiene i libri	Durante l'intervista il microfono regala una copia del film	In Niger le risorse scarseggiano durante la siccità	Grazie alle dighe i manichini sopravvivono all'inverno
In città i grattacieli luccicano sotto il sole	Sotto al tavolo la mattonella solleva una briciola	Grazie alle dighe i castori sopravvivono all'inverno	In estate il macellaio scade dopo pochi giorni
Nel cielo il sole brilla a primavera	D'estate il sasso riposa sotto l'albero	In America i mandati presidenziali scadono dopo quattro anni	Quando piove i lombrichi decollano con fatica

Table B.1 Experimental stimuli for the eye-tracking study Chapter 5)

Subject-verb concordance (number)	**Determiner-NP (gender and number)**	**Verb and clitic**
Proto-Theme Nel bosco le volpi scappa dal cacciatore In cucina la torta lievitano nel forno Nella lavatrice la centrifuga girano con grande rumore In inverno la neve cadono in montagna	Gender Nel mare le pesci nuotano liberi In cucina i pentole borbottano fumanti Sugli spalti le tifosi incitano la squadra In laboratorio gli operaie smontano la macchina	Gender Il cane insegue il gatto e la rincorre fin sotto il tavolo Finito il libro, Sandro la appoggia sulla libreria Mario accende la radio e lo sintonizza sul giornale locale Il nonno raccoglie le patate e li porta in cucina
Proto-Agent In casa il pittore dipingono le pareti In estate le zanzare punge le caviglie Oggi la mamma vanno al mercato Di sabato i turisti affolla il centro città	Number Al pomeriggio il professori correggono i compiti degli alunni Tutti i giorni il ballerini si allenano per quattro ore Oggi la tartarughe nuotano libere nell'oceano Nel pollaio la galline depongono le uova al mattino	Number L'automobile sbanda e il camion le urta L'impiegata ritarda e il capoufficio le chiama al telefono Il vaso vacilla e Giovanni li prende al volo L'ape si avvicina e il turista le scaccia

Table B.2 Fillers (morphological violations)

Acknowledgments

I would like to thank first and foremost my supervisors, Claudio Luzzatti and Marco Marelli, who have guided me through this PhD journey from the very start to the end: your wisdom, ingenuity, availability, respect, and sincere passion for knowledge and research have been a true source of inspiration. I own you a great part of what I have achieved during these years.

A heartwarming thanks to all the volunteering participants to this study. Especially to all those people who, despite their everyday battle with an unmerciful companion as SSD, have found the time and the strength to come and take part to our experiments.

I am also in debt with the clinical staff of the IRCCS San Giovanni di Dio Fatebenefratelli (Dr. Adorni, Dr. Andreose, Dr. Barovelli, Dr. Beneduce, Dr. Bignotti, Dr. Laffranchini, Dr. Nodari, Dr. Rillosi, Dr. Rossi, and particularly to Dr. Tura) for the clinical support and referral of participants. I cannot forget the priceless support of Dr. Bulgari, Dr. C. Ferrari, E. Ferrari, and A. Macis.

A special thanks to all my former colleagues at the Psychiatric Epidemiology and Evaluation Unit, who were at the beginning of this personal and professional journey, and especially to Dr. de Girolamo. Also, the Neurophysiology Lab group, who have generously welcomed and supported me in countless ways at the end of it, deserves a special mention.

Finally, I am so much obliged to my family, who has been there all the time: you are the toughest, funniest, and sweetest place I ever want to be.