

JADT' 18

PROCEEDINGS OF THE
14TH INTERNATIONAL CONFERENCE
ON STATISTICAL ANALYSIS OF TEXTUAL DATA

JADT' 18

PROCEEDINGS OF THE
14TH INTERNATIONAL CONFERENCE
ON STATISTICAL ANALYSIS OF TEXTUAL DATA

(Rome, 12-15 June 2018)

Vol. I

UniversItalia
2018

PROPRIETÀ LETTERARIA RISERVATA

Copyright 2018 - UniversItalia - Roma

ISBN 978-88-3293-137-2

A norma della legge sul diritto d'autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilm, registratori o altro. Le fotocopie per uso personale del lettore possono tuttavia essere effettuate, ma solo nei limiti del 15% del volume e dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5 della legge 22 aprile 1941 n. 633. Ogni riproduzione per finalità diverse da quelle per uso personale deve essere autorizzata specificatamente dagli autori o dall'editore.

Program Committee

Ramón Álvarez Esteban: Univ. of León, E
Valérie Beaudouin: Telecom ParisTech, F
Mónica Bécue: Poly. Univ. of Catalunya, E
Sergio Bolasco: Sapienza Univ. of Rome, I
Isabella Chiari: Sapienza Univ. of Rome, I
François Daoust, UQÀM, Montreal, CDN
Anne Dister, FUSL, Bruxelles / UCL, Louvain, B
Jules Duchastel: UQÀM, Montreal, CDN
Serge Fleury: Univ. Paris 3, F
Cédric Fairon: UCL, Louvain, B
Luca Giuliano: Sapienza Univ. of Rome, I
Serge Heiden, ENS, Lyon, F
Domenica Fioredistella Iezzi, Univ. of Tor Vergata, I
Margareta Kastberg, Univ. of Franche Comté, F
Ludovic Lebart: CNRS / ENST, Paris, F
Jean-Marc Leblanc: Univ. of Créteil, F

Alain Lelu: Univ. of Franche Comté, F
Dominique Longrée, Univ. of Liège, B
Véronique Magri: Univ. of Nice Sophia-Antipolis, F
Pascal Marchand: Univ. of Toulouse, F
William Martinez: Univ. of Lisboa, P
Damon Mayaffre: CNRS, Nice, F
Sylvie Mellet: CNRS, Nice, F
Michelangelo Misuraca: Univ. of Calabria, I
Denis Monière: Univ. of Montréal, CDN
Bénédicte Pincemin: CNRS, Lyon, F
Céline Poudat: Univ. of Nice Sophia-Antipolis, F
Pierre Retinaud: Univ. of Toulouse, F
André Salem: Univ. Paris 3, F
Monique Slodzian: Inalco, F
Arjuna Tuzzi: Univ. of Padua, I
Mathieu Valette: Inalco, F

Organising Committee

Domenica Fioredistella Iezzi: Univ. of Tor Vergata, I
Sergio Bolasco: Sapienza Univ. of Rome, I
Livia Celardo: Sapienza Univ. of Rome, I
Isabella Chiari: Sapienza Univ. of Rome, I
Francesca della Ratta: ISTAT, I
Fiorenza Deriu: Sapienza Univ. of Rome, I
Francesca Dolcetti: Sapienza Univ. of Rome, I

Andrea Fronzetti Colladon: Univ. of Tor Vergata, I
Francesca Greco: Sapienza Univ. of Rome, I
Isabella Mingò: Sapienza Univ. of Rome, I
Michelangelo Misuraca: Univ. of Calabria, I
Arjuna Tuzzi: Univ. of Padua, I
Maurizio Vichi: Sapienza Univ. of Rome, I
Francesco Zarelli: ISTAT, I

Local Organisation

Francesco Alò, Giulia Giacco,
Paolo Meoli, Vittorio Palermo, Viola Talucci

Table of contents

Introduction XVII

Acknowledgements XIX

Invited Speakers

GERMAN KRUSZEWSKI

Memorize or generalize? Searching for a compositional RNN in a haystack

Adam Liška XXIII

BING LIU

Scaling-up Sentiment Analysis through Continuous Learning XXIV

PASCAL MARCHAND

La textométrie comme outil d'expertise :

application à la négociation de crise. XXV

GEORGE K. MIKROS

Author Identification Combining Various Author Profiles. Towards a Blended

Authorship Attribution Methodology XXVI

ROBERTO NAVIGLI

From text to concepts and back: going multilingual

with BabelNet in a step or two XXVII

Contributors

MOTASEM ALRAHABI¹, CHIARA MAINARDI¹

Identification automatique de l'ironie et des formes apparentées dans un

corpus de controverses théâtrales 1

MOHAMMAD ALSADHAN, SASCHA DIWERSY,

AGATA JACKIEWICZ, GIANCARLO LUXARDO

Migrants et réfugiés : dynamique de la nomination de l'étranger 10

R. ALVAREZ-ESTEBAN, M. BÉCUE-BERTAUT, B. KOSTOV, F. HUSSON, J-A

SÁNCHEZ-ESPIGARES

Xplortext, a R package. Multidimensional statistics for textual data science. 19

ELENA, AMBROSETTI, ELEONORA MUSSINO, VALENTINA TALUCCI

L'evoluzione delle norme: analisi testuale delle politiche sull'immigrazione in

Italia 26

MASSIMO ARIA, CORRADO CUCCURULLO

A bibliometric meta-review of performance measurement, appraisal, management research 35

LAURA ASCONE

Textual Analysis of Extremist Propaganda and Counter-Narrative: a quantitative investigation 44

LAURA ASCONE, LUCIE GIANOLA

Analyse de données textuelles appliquée à des problématiques de sécurité et d'enquête judiciaire 52

SIMONA BALBI, MICHELANGELO MISURACA, MARIA SPANO

A two-step strategy for improving categorisation of short texts 60

CHRISTINE BARATS, ANNE DISTER, PHILIPPE GAMBETTE, JEAN-MARC LEBLANC, MARIE PERES

Appeler à signer une pétition en ligne : caractéristiques linguistiques des appels 68

MANUEL BARBERA, CARLA MARELLO

Newsgroup e lessicografia: dai NUNC al VoDIM 76

IGNAZIA BARTHOLINI

Techniques for detecting the normalized violence in the perception of refugee / asylum seekers between lexical analysis and factorial analysis 83

PATRIZIA BERTINI MALGARINI, MARCO BIFFI, UGO VIGNUZZI

Dal corpus al dizionario: prime riflessioni lessicografiche sul Vocabolario storico della cucina italiana postunitaria (VoSCIP) 90

MARCO BIFFI

Strumenti informatico-linguistici per la realizzazione di un dizionario dell'italiano postunitario 99

ANNICK FARINA, RICCARDO BILLERO

Comparaison de corpus de langue « naturelle » et de langue « de traduction » : les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues 108

FELICE BISOGNI, STEFANO PIRROTTA

Il rapporto tra famiglie di anziani non autosufficienti e servizi territoriali: un'analisi dei dati esplorativa con l'Analisi Emozionale del Testo (AET).... 117

ANTONELLA BITETTO, LUIGI BOLLANI

Esperienza di analisi testuale di documentazione clinica e di flussi informativi sanitari, di utilità nella ricerca epidemiologica e per indagare la qualità dell'assistenza 126

GUIDO BONINO, DAVIDE PULIZZOTTO, PAOLO TRIPODI

Exploring the history of American philosophy in a computer-assisted framework 134

MARC-ANDRE BOUCHARD, SYLVIA KASPARIAN

La classification hiérarchique descendante pour l'analyse des représentations sociales dans une pétition antibilinguisme au Nouveau-Brunswick, Canada 142

LIVIA CELARDO, RITA VALLEROTONDA, DANIELE DE SANTIS, CLAUDIO SCARICI, ANTONIO LEVA

Analysing occupational safety culture through mass media monitoring..... 150

BARBARA CORDELLA, FRANCESCA GRECO, PAOLO MEOLI, VITTORIO PALERMO, MASSIMO GRASSO

Is the educational culture in Italian Universities effective? A case study..... 157

MICHELE A. CORTELAZZO, GEORGE K. MIKROS, ARJUNA TUZZI

Profiling Elena Ferrante: a Look Beyond Novels 165

FABRIZIO DE FAUSTI, MASSIMO DE CUBELLIS, DIEGO ZARDETTO¹

Word Embeddings: a Powerful Tool for Innovative Statistics at Istat 174

Gibbons A. (1985). *Algorithmic Graph Theory*. Cambridge University Press. . 182

VIVIANA DE GIORGI, CHIARA GNESI

Analisi di dati d'impresa disponibili online: un esempio di data science tratto dalla realtà economica dei siti di e-commerce 183

ALESSANDRO CAPEZZUOLI, FRANCESCA DELLA RATTA, STEFANIA MACCHIA, MANUELA MURGIA, MONICA SCANNAPIECO, DIEGO ZARDETTO

The use of textual sources in Istat: an overview..... 192

FRANCESCA DELLA RATTA, GABRIELLA FAZZI, MARIA ELENA PONTECORVO, CARLO VACCARI, ANTONINO VIRGILLITO

Twitter e la statistica ufficiale: il dibattito sul mercato del lavoro 200

SAMI DIAF

Gauging An Author's Mood Using Hidden Markov Chains 209

MARC DOUGUET

Les hémistiches répétés 215

FRANCESCA DRAGOTTO, SONIA MELCHIORRE

«Mangiata dall'orco e tradita dalle donne». Vecchi e nuovi media raccontano la vicenda di Asia Argento, tra storytelling e Speech Hate 223

CRISTIANO FELACO, ANNA PAROLA

Il *cosa* e il *come* del processo narrativo. L'uso combinato della Text Analysis e Network Text Analysis al servizio della precarietà lavorativa 233

ANA NORA FELDMAN

Hablando de crisis: las comunicaciones del Fondo Monetario Internacional 242

VALERIA FIASCO

Brexit in the Italian and the British press: a bilingual corpus-driven analysis 250

VIVIANA FINI, GIUSEPPE LUCIO GAETA, SERGIO SALVATORE

Textual analysis to promote innovation within public policy evaluation 259

ALESSIA FORCINITI, SIMONA BALBI	
A proposal for Cross-Language Analysis: violence against women and the Web	268
BEATRICE FRACCHIOLLA, OLINKA SOLENE DE ROGER	
La verbalisation des émotions	276
LUISA FRANCHINA, FRANCESCA GRECO, ANDREA LUCARIELLO, ANGELO SOCIAL, LAURA TEODONNO	
Improving Collection Process for Social Media Intelligence: A Case Study .	285
ANDREA FRONZETTI COLLADON, JOHANNE SAINT-CHARLES, PIERRE MONGEAU	
The impact of language homophily and similarity of social position on employees' digital communication	293
MATTEO GERLI	
Looking Through the Lens of Social Sciences: The European Union in the EU- Funded Research Projects Reporting	300
LUCIE GIANOLA, MATHIEU VALETTE	
Spécialisation générique et discursive d'une unité lexical L'exemple de <i>joggeuse</i> dans la presse quotidienne régionale	312
PETER A. GLOOR, JOAO MARCOS DE OLIVEIRA, DETLEF SCHODER	
The Transparency Engine – A Better Way to Deal with Fake News	319
FRANCESCA GRECO, LEONARDO ALAIMO, LIVIA CELARDO	
Brexit and Twitter: The voice of people.....	327
FRANCESCA GRECO, GIULIO DE FELICE, OMAR GELO	
A text mining on clinical transcripts of good and poor outcome psychotherapies	335
FRANCESCA GRECO, DARIO MASCHIETTI, ALESSANDRO POLLI	
DOMINIO: A Modular and Scalable Tool for the Open Source Intelligence	343
LEONIE GRÖN, ANN BERTELS, KRIS HEYLEN	
Is training worth the trouble? A PoS tagging experiment with Dutch clinical records.....	351
FRANCE GUERIN-PACE, ELODIE BARIL	
Les outils de la statistique textuelle pour analyser les corpus de données d'enquêtes de la statistique publique.....	359
SERGE HEIDEN	
Annotation-based Digital Text Corpora Analysis within the TXM Platform	367
DANIEL HENKEL	
Quantifying Translation : an analysis of the conditional perfect in English- French comparable-parallel corpus.....	375
DANIEL DEVATMAN HROMADA	
Extraction of lexical repetitive expressions from complete works of William Shakespeare.....	384

OLIVIER KRAIF, JULIE SORBA

Spécificités des expressions spatiales et temporelles dans quatre sous-genres romanesques (policier, science-fiction, historique et littérature générale) 392

CYRIL LABBE, DOMINIQUE LABBE

Les phrases de Marcel Proust 400

LUDOVICA LANINI, MARÍA CARLOTA NICOLÁS MARTÍNEZ

Verso un dizionario *corpus-based* del lessico dei beni culturali: procedure di estrazione del lemmario 411

DANIELA LARICCHIUTA, FRANCESCA GRECO, FABRIZIO PIRAS, BARBARA CORDELLA, DEBORA CUTULI, ELEONORA PICERNI, FRANCESCA ASSOGNA, CARLO LAI, GIANFRANCO SPALLETTA, LAURA PETROSINI

"The grief that doesn't speak": Text Mining and Brain Structure 419

GEVISA LA ROCCA, CIRUS RINALDI

Icone gay: tra processi di normalizzazione e di resistenza. Ricostruire la semantica degli hashtag..... 428

LUDOVIC LEBART

Looking for *topics*: a brief review..... 436

GAËL LEJEUNE, LICHAO ZHU

Analyse Diachronique de Corpus : le cas du poker..... 444

JULIEN LONGHI, ANDRE SALEM

Approche textométrique des variations du sens..... 452

LAURENT VANNI¹, DAMON MAYAFFRE, DOMINIQUE LONGREE

ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables 459

LUCIE LOUBERE

Déconstruction et reconstruction de corpus... À la recherche de la pertinence et du contexte 467

HEBA METWALLY

L'apport du *corpus-maquette* à la mise en évidence des niveaux descriptifs de la chronologie du sens. Essai sur une Série Textuelle Chronologique du *Monde diplomatique* (1990-2008). 474

JUN MIAO, ANDRE SALEM

Séries textuelles homogènes..... 491

SILVIO MIGLIORI, ANDREA QUINTILIANI, DANIELA ALDERUCCIO, FIORENZO AMBROSINO, ANTONIO COLAVINCENZO, MARIALUISA MONGELLI, SAMUELE PIERATTINI, GIOVANNI PONTI SERGIO BOLASCO, FRANCESCO BAIOCCHI, GIOVANNI DE GASPERIS

TaLTaC in ENEAGRID Infrastructure..... 501

ISABELLA MINGO, MARIELLA NOCENZI

The dimensions of Gender in the International Review of Sociology. A lexicometric approach to the analysis of the publications in the last twenty years 509

ADIEL MITTMANN, ALCKMAR LUIZ DOS SANTOS

The Rhythm of Epic Verse in Portuguese From the 16th to the 21st Century 514

DENIS MONIERE, DOMINIQUE LABBE

Le vocabulaire des campagnes électorales 522

CYRIELLE MONTRICHARD

Faire émerger les traces d'une pratique imitative dans la presse de tranchées à l'aide des outils textométriques 532

ALBERT MORALES MORENO

Evolución diacrónica de la terminología y la fraseología jurídico-administrativa en los Estatutos de autonomía de Catalunya de 1932, 1979 y 2006 541

CEDRIC MOREAU

Comment penser la recherche d'un signe pour une plateforme multilingue et multimodale français écrit / langue des signes française ? 556

JEAN MOSCAROLA, BORIS MOSCAROLA

Conclusion ADT et visualisation, pour une nouvelle lecture des corpus Les débats de 2ème tour des Présidentielles (1974-2017) 563

MAURIZIO NALDI

A conversation analysis of interactions in personal finance forums 571

STEFANO NOBILE

Analisi testuale, rumore semantico e peculiarità morfosintattiche: problemi e strategie di pretrattamento di corpora speciali 578

DANIEL PELISSIER

L'individu dans le(s) groupe(s) : focus group et partitionnement du corpus 586

BENEDICTE PINCEMIN, CELINE GUILLOT-BARBANCE, ALEXEI**LAURENTIEV**

Using the First Axis of a Correspondence Analysis as an Analytical Tool. Application to Establish and Define an Orality Gradient for Genres of Medieval French Texts 594

CELINE POUDAT

Explorer les désaccords dans les fils de discussion du Wikipédia francophone 602

MATTHIEU QUIGNARD, SERGE HEIDEN, FREDERIC LANDRAGIN,**MATTHIEU DECORDE**

Textometric Exploitation of Coreference-annotated Corpora with TXM: Methodological Choices and First Outcomes 610

PIERRE RATINAUD

Amélioration de la précision et de la vitesse de l'algorithme de classification de la méthode Reinert dans IRaMuTeQ 616

LUISA REVELLI

- Il parametro della *frequenza* tra paradossi e antinomie:
il caso dell'*italiano scolastico* 626

PIERGIORGIO RICCI

- How Twitter emotional sentiments mirror on the Bitcoin
transaction network 635

CHANTAL RICHARD, SYLVIA KASPARIAN

- Analyse de contenu versus méthode Reinert : l'analyse comparée d'un corpus
bilingue de discours acadiens et loyalistes du N.-B., Canada 643

VALENTINA RIZZOLI, ARJUNA TUZZI

- Bridge over the ocean: Histories of social psychology in Europe and North
America. An analysis of chronological corpora 651

LOUIS ROMPRE, ISMAÏL BISKRI

- Les « itemsets fréquents » comme descripteurs de documents textuels 659

CORINNE ROSSARI, LJILJANA DOLAMIC, ANNALENA HÜTSCH, CLAUDIA RICCI, DENNIS WANDEL

- Discursive Functions of French Epistemic Adverbs: What can Correspondence
Analysis tell us about Genre and Diachronic Variation? 668

VANESSA RUSSO, MARA MARETTI, LARA FONTANELLA, ALICE TONTODIMAMMA

- Misleading information in online propaganda networks 676

ELIANA SANANDRES, CAMILO MADARIAGA, RAIMUNDO ABELLO

- Topic modeling of Twitter conversations 684

FRANCESCO SANTELLI, GIANCARLO RAGOZINI, MARCO MUSELLA

- What volunteers do? A textual analysis of voluntary activities in the Italian
context 692

S. SANTILLI, S. SBALCHIERO, L. NOTA, S. SORESI

- A longitudinal textual analysis of abstract presented at Italian Association for
Vocational guidance and Career Counseling'
Conferences from 2002 to 2017 700

JACQUES SAVOY

- A la poursuite d'Elena Ferrante 707

JACQUES SAVOY

- Regroupement d'auteurs dans la littérature du XIXe siècle 716

STEFANO SBALCHIERO, ARJUNA TUZZI

- What's Old and New? Discovering Topics in the American Journal of
Sociology 724

NILS SCHAETTI, JACQUES SAVOY

- Comparison of Neural Models for Gender Profiling 733

LIONEL SHEN

- Segments répétés appliqués à l'extraction de connaissances trilingues 740

SANDRO STANCAMPIANO	
Misurare, Monitorare e Governare le città con i Big Data	748
FADILA TALEB, MARYVONNE HOLZEM	
Exploration textométrique d'un corpus de motifs juridiques dans le droit international des transports	755
JAMES M. TEASDALE	
The Framing of the Migrant: Re-imagining a Fractured Methodology in the Context of the British Media.	763
MARJORIE TENDERO¹, CECILE BAZART	
Results from two complementary textual analysis software (Iramuteq and Tropes) to analyze social representation of contaminated brownfields	771
MATTEO TESTI, ANDREA MERCURI, FRANCESCO PUGLIESE	
Multilingual Sentiment Analysis.....	780
JUAN MARTÍNEZ TORVISCO	
A linguistic analysis of the image of immigrants' gender in Spanish newspapers.....	788
FRANCESCO URZÌ	
Lo strano caso delle frequenze zero nei testi legislativi euroistituzionali.....	796
SYLVIE VANDAELE	
Les traductions françaises de <i>The Origin of Species</i> : pistes lexicométriques .	805
PIERRE WAVRESKY, MATTHIEU DUBOYS DE LABARRE, JEAN-LOUP LECOEUR	
Circuits courts en agriculture : utilisation de la textométrie dans le traitement d'une enquête sur 2 marchés	814
MARIA ZIMINA, NICOLAS BALLIER	
On the phraseology of spoken French: initial salience, prominence and lexicogrammatical recurrence in a prosodic-syntactic treebank <i>Rhapsodie</i>	822

Abstracts

FILIPPO CHIARELLO, GUALTIERO FANTONI, ANDREA BONACCORSI, SILVIA FARERI	
What kind of contributions does research provides? Mapping issue based statements in research abstracts	833
FILIPPO CHIARELLO, GIACOMO OSSOLA, GUALTIERO FANTONI, ANDREA BONACCORSI, ANDREA CIMINO, FELICE DELL'ORLETTA	
Technical sentiment analysis: predicting the success of new products using social media.....	835

FIorenza DERIU, DOMENICA FIOREDISTELLA IEZZI
 Citizens and neighbourhood life: mapping population sentiment in Italian cities..... 837

FRANCESCA DI CARLO, ROSY INNARELLA, BRIZIO LEONARDO TOMMASI
 Vax network: profiling influential nodes with social network analysis on twitter..... 838

DAVIDE DONNA
 Alteryx 840

VALERIO FICCADENTI, ROY CERQUETI, MARCEL AUSLOOS
 Complexity of US President Speeches 841

PETER A. GLOOR
 Measuring the Dynamics of Social Networks with Condor 842

IOLANDA MAGGIO, DOMENICA FIOREDISTELLA IEZZI, MATTEO FATIGHENTI
 "BIG DATA" Words Trend Analysis using the multidimensional analysis of texts 844

MARIO MASTRANGELO
 Itinerari turistici, network analysis e text mining 845

MARIA FRANCESCA ROMANO, GUIDO REY, ANTONELLA BALDASSARINI PASQUALE PAVONE
 Text Mining per l'analisi qualitativa e quantitativa dei dati amministrativi utilizzati dalla Pubblica Amministrazione..... 847

ALESSANDRO CESARE ROSA
 Taglio cesareo e Vbac in Italia al tempo dei Big Data: una proposta di ulteriore contributo informativo..... 849

“The grief that doesn’t speak”: Text Mining and Brain Structure

Daniela Laricchiuta¹, Francesca Greco², Fabrizio Piras³, Barbara Cordella⁴,
 Debora Cutuli⁵, Eleonora Picerni⁶, Francesca Assogna⁷,
 Carlo Lai⁸, Gianfranco Spalletta⁹, Laura Petrosini¹⁰

¹Sapienza University, IRCCS Fondazione Santa Lucia – daniela.laricchiuta@uniroma1.it

²Sapienza University of Rome, Prisma S.r.l. – francesca.greco@uniroma1.it

³IRCCS Fondazione Santa Lucia - f.piras@hsantalucia.it

⁴Sapienza University of Rome – barbara.cordella@uniroma1.it

⁵Sapienza University of Rome, IRCCS Fondazione Santa Lucia – debora_cutuli@yahoo.it

⁶Sapienza University, IRCCS Fondazione Santa Lucia – eleonora.picerni@uniroma1.it

⁷IRCCS Fondazione Santa Lucia - f.assogna@hsantalucia.it

⁸Sapienza University of Rome – carlo.lai@uniroma1.it

⁹IRCCS Fondazione Santa Lucia – g.spalletta@hsantalucia.it

¹⁰Sapienza University of Rome, IRCCS Fondazione S. Lucia – laura.petrosini@uniroma1.it

Abstract

Contemporary neurosciences have shown that emotions, thought and language involve the functioning of connected brain areas, which allow the recognition and expression of one's own feelings. The scope of this pilot study is to investigate the link among the verbal expression of emotional experiences (assessed with the Toronto Structured Interview for Alexithymia - TSIA -), the linguistic structure and the brain structure. To this aim, 9 healthy adult subjects of both sexes were interviewed by means of the TSIA and the cortical and subcortical structural measures were detected. The TSIA transcripts were analysed by using a cluster analysis and, subsequently, a correspondence analysis, and the values of factors were correlated with cortical and subcortical structural measures as well as TSIA scores, evidencing significant associations. The study highlighted that in healthy subjects it is possible to identify a link between the manner in which people express their experiences, recognize and use their emotions and the brain structural correlates.

Abstract

Le neuroscienze contemporanee hanno evidenziato come le emozioni, il pensiero e il linguaggio coinvolgono il funzionamento di aree cerebrali differenti connesse tra loro, le quali consentono il riconoscimento e l'espressione dei propri sentimenti. Scopo di questo studio pilota è di indagare il nesso tra l'espressione verbale delle proprie esperienze emotive (valutata con la Toronto Structured Interview for Alexithymia – TSIA -), la

struttura del linguaggio utilizzato e la struttura cerebrale. A questo scopo 9 soggetti sani di entrambi i sessi sono stati intervistati con la TSIA e sono state rilevate le misure strutturali corticali e sottocorticali. Le interviste sono state sottoposte ad analisi dei cluster e successivamente ad analisi delle corrispondenze e i valori dei fattori sono stati correlati con le misure strutturali corticali e sottocorticali e con i punteggi della TSIA. I risultati evidenziano associazioni significative, che mettono in luce come in soggetti sani sia possibile individuare un nesso tra il modo in cui le persone raccontano le proprie esperienze, la loro capacità di riconoscere e utilizzare le loro emozioni e la loro struttura cerebrale.

Keywords: Text mining, brain imaging, alexithymia, TSIA

1. Introduction

According to the multiple code theory, the emotional information is represented in verbal, non-verbal symbolic and non-verbal sub-symbolic multiple systems (Bucci, 1997). The verbal system is a communication and reflection code through which the emotional, private and subjective experience can be shared with others. It refers to the capacity of language to direct and regulate ourselves, activate imagination and emotions, stimulate actions and control them. The multiple channels of the non-verbal systems include representations and proceedings related to implicit elaboration associated with visceral, somatic, sensory and motor modalities. While in the non-verbal symbolic system the information is processed in images, in the sub-symbolic one, rapid and complex computations are carried out in an implicit continuous path. These computations contribute to recognize slight facial expressions modifications, identify body movement or vocal quality changes, and distinguish visceral states. The multiple code theory is in line with the contemporary neurosciences (LeDoux, 2012; Damasio et Carvalho, 2013), suggesting that in the presence of the affective experience it is possible to discriminate between emotions and feelings. Emotions occur at a physiological and motor-expressive level, involving bodily systems and subcortical and cortical somato-sensory brain areas. Feelings are based on complex symbolization and cognitive processes related to the functioning of prefrontal and associative cortices. The integration of emotions and feelings, as well as of the verbal and non-verbal systems, depends on the so-called referential processes, which transform the non-verbal symbolic and sub-symbolic materials into words, and *vice versa*.

The referential processes are the core factors contributing to the development, maintenance and promotion of health, since a deficit in these processes generates dysfunctional conditions and pathologies, characterized

by multifactorial and bio-psycho-social etiology as well as marked somatization. Among these dysfunctional conditions, alexithymia is a psychological construct represented by impairment in cognitive-emotional and affective processing (Bagby et al., 1994). It describes people with deficiencies in identifying or describing subjective emotions or feelings, difficulty in distinguishing between bodily sensations of emotional arousal and feelings, and limited affect-related fantasy and imagery. People with alexithymic traits have a tendency to focus on facts without affective involvement rather than inner experiences, exhibiting a “concrete and reality-based cognitive style”. They often avoid social situations, seem cold, show a lack of intimacy and warmth and are insecurely attached to others. Although alexithymia is not a psychological disorder *per se*, it is associated with a low quality of life and enhanced risk of psychological impairment and it is present in a broad spectrum of psychosomatic disorders (Taylor et Bagby, 2004). Neuroimaging studies have indicated that people with high alexithymic traits show less activation in brain areas associated with emotional awareness and volumetric variations in brain areas associated with emotional and somato-sensory and sensory-motor processing (Laricchiuta et al., 2015a, and see for a literature review Laricchiuta et al., 2015b). Therefore, the aim of this pilot study is to investigate a complex bio-psycho-social pattern of verbal expression of the emotional experiences, alexithymia levels and brain structure.

2. Data collection and analysis

A sample of 9 (males=5) healthy adult subjects of both sexes was recruited for the pilot study at the IRCCS Fondazione Santa Lucia, Rome. Participants were selected according to the following inclusion criteria: age between 18 and 70 years and suitability for structural Magnetic Resonance Imaging (MRI) scanning. Exclusion criteria included the suspicion of cognitive impairment or dementia; the subjective complaint of memory difficulties or of any other cognitive deficit, regardless of interference with daily activities; major medical illnesses; current or reported psychiatric or neurological disorders; known or suspected history of alcoholism or drug dependence and abuse; and MRI evidence of focal parenchymal abnormalities or cerebrovascular diseases.

To assess the cortical and subcortical structural measures, participants underwent an imaging protocol that included standard clinical sequences (FLAIR, DP-T2-weighted) and a volumetric whole-brain 3D high-resolution T1-weighted sequence, performed with a 3 T Allegra MR imager, with a standard quadrature head coil. Volumetric whole-brain T1-weighted images were obtained in the sagittal plane using a Modified Driven Equilibrium

Fourier Transform (MDEFT) sequence (Echo Time/Repetition Time -TE/TR- = 2.4/7.92 ms, flip angle 15°, voxel size 1 x 1 x 1 mm³). All planar sequence acquisitions were obtained in the plane of the anterior-posterior commissure line. For the volumetric measures, T1-weighted images were processed and examined using the SPM8 software, specifically the VBM8 toolbox running in Matlab 2007b. For the cortical thickness, FreeSurfer imaging analysis suite (v5.1.0) was used for cortical reconstruction of the whole brain. The segmented, normalized, modulated and smoothed images were used for analyses. Then, participants were interviewed by using the Toronto Structured Interview for Alexithymia (TSIA, Bagby et al., 2006; Italian version Caretti et al., 2011), composed of 24 items referred to four factors of the alexithymia construct: the Difficulty in Identifying Feelings (DIF); the Difficulty in Describing Feelings (DDF); the Externally Oriented Thinking (EOT); and the Imaginal Processes (IP). Each item is assessed by a specific open-ended question and its response is 3-point scored (coded '0', '1', or '2'). The sum of scores results in a total score that ranges from 0 (low alexithymia levels) to 48 (high alexithymia levels). The transcripts of the TSIA responses were used to evaluate the linguistic structure by means of a multivariate analysis. Namely, the nine TSIA transcripts resulted in a medium size corpus of 62.792 tokens. In order to check whether it was possible to statistically process data, two lexical indicators were calculated: the type-token ratio and the hapax percentage (TTR= 0,10; Hapax= 51,0%). According to the large size of the corpus both lexical indicators highlighted its richness and indicated the possibility to proceed with the analysis. First, data were cleaned and pre-processed with the software T-Lab (Lancia, 2017) and keywords selected. In particular, we used lemmas as keywords instead of type, filtering out the lemma of the high rank of frequency and those of the low rank of frequency lower to 9 occurrences (for keyword election see Cordella et al., 2014; Greco, 2016). Then, on the context units per keywords matrix, we performed a cluster analysis with a bisecting k-means algorithm (Savaresi et Boley, 2004) limited to ten partitions, excluding all the context units that do not have at least two keywords co-occurrence. To finalize the text mining a correspondence analysis (Lebart et Salem, 1994) on the keywords per clusters matrix was made in order to explore the relationship between clusters and to identify the latent dimensions setting the interviews.

Then parametric associations between TSIA scores and regional volumes or cortical thickness, and between lexical scores (resulted by the correspondence analysis) and TSIA scores or brain structural measures were calculated by means of Pearson's correlations in order to identify the possible direction and extent of the linear relationship between the variables.

3. Main results

The results of the cluster analysis show that the 369 keywords selected allow for the classification of 96.8% of the corpus. According to the theoretical framework (Cordella et al., 2014) we choose the solution with four cluster. The correspondence analysis detected three latent dimensions. In table 1, we can appreciate how the clusters are placed in the factorial space produced by three factors. The first factor represents the experience that could be personal (negative pole) or social (positive pole); the second factor reflects the thought that could be made on feelings (negative pole) or on a rational reasoning (positive pole); and the third factor represents the aim of the thinking process that could lead to make a speculation (negative pole) or a choice (positive pole).

Table 1 – Cluster coordinates on factors (the percentage of explained inertia is reported between brackets under each factor).

Cluster	Label	CU classified	Factor 1 Experience	Factor 2 Thought	Factor 3 Aim
1	Think	34,24%	0,03	Reasoning	Speculate
2	Feel	18,09%	Personal	Feeling	Speculate
3	Relationship	23,15%	Social	-0,65	Choice
4	Remember	24,51%	Personal	Reasoning	Choice
			-0,73	0,23	0,30

CU = context units classified.

The four clusters are of different sizes and reflect the general approach to the emotional experience solicited by the TSIA. The first cluster reflects the reasoning on the life event and hypothesis resulting in a rationale thinking process; the second cluster highlights the capacity to reflect on the experience of personal feelings; the third cluster represents the relationships characterizing social life; and the fourth cluster gets back to memories, reasoning on personal choices that were made (table 2).

Table 2 – Cluster (the percentage of context units classified in the cluster is reported between brackets).

Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Think		Feel		Relationship		Remember	
keyword	CU	keyword	CU	keyword	CU	keyword	CU
pensare	104	sentire	100	persona	163	immaginare	84
vedere	71	proprio	77	parlare	163	vedere	71
scrivere	45	riuscire	77	sentimento	149	prendere	64
amico	38	momento	67	provare	116	mettere	61
chiedere	32	capire	66	persone	114	positivo	39
chiamare	26	situazione	51	capire	100	casa	31
trovare	26	vivere	50	cercare	80	bello	28
tempo	21	piacere	50	situazione	71	problemi	27
ragazzo	20	succedere	33	amico	70	tornare	27
diverso	20	rabbia	32	trovare	45	portare	24

CU = context units classified in the cluster.

The correlation coefficient between the lexical scores of the three factors (resulted from the correspondence analysis) and the TSIA scores, as well as the brain volumes and the cortical thickness are reported in table 3. Namely, DIF, DDF, EOT and total alexithymia scores were positively associated with the second factor. At neurobiological level, the first factor was negatively associated with volumes of right caudate and thickness of right medial orbitofrontal cortex, as well as positively associated with the thickness of right lateral occipital cortex. Finally, the third factor was negatively associated with volumes of middle anterior, central and middle posterior cerebral cortices, as well as with thickness of right postcentral cortex and left posterior cingulate cortex. Conversely, the third factor was positively associated with thickness of the right posterior cingulate cortex. Finally, the IP scores (TSIA) were positively correlated ($r = 0.72$; $p = 0.03$) with the left entorhinal cortical thickness values.

4. Discussion

Although this is a pilot study and it is not possible to generalise the findings, the present data suggest that the methodology proposed (in order to identify the connections among verbal expression, alexithymia levels and brain structure) seems to be promising for a deeper understanding of the bio-psycho-linguistic connections. In fact, results indicate that high alexithymia scores were associated with a thought modality characterized by a rational (and not emotional) reasoning. Furthermore, the tendency to be engaged in

personal (not social) experience was associated with large volumes of right caudate and thickness of right medial orbitofrontal cortex.

Table 3 – Correlation coefficients between lexical factors and TSIA scores as well as cerebral structure values.

Variables	Factor 1	Factor 2	Factor 3
Difficulty Identifying Feelings (TSIA Factor 1)		.83	
Difficulty Describing Feelings (TSIA Factor 2)		.71	
Externally Oriented Style of Thinking, (TSIA Factor 3)		.68	
TSIA Total score		.77	
Right-Caudate	-.71		
Right Hemisphere Medialorbitofrontal Thickness	-.78		
Right Hemisphere Lateraloccipital Thickness	.70		
Left Hemisphere Posteriorcingulate Thickness			-.71
Mid Posterior Cortical Cortex			-.75
Mid Anterior Cortical Cortex			-.75
Central Cortical Cortex			-.78
Right Hemisphere Postcentral Thickness			-.78
Right Hemisphere Posteriorcingulate Thickness			.68

In the table are reported only the correlation coefficients with a $p < 0,05$.

Conversely, the tendency to be engaged in social (not personal) experiences was associated to great thickness of right lateral occipital cortex. The speculative thinking processes (negative pole of the third factor) was associated with large volumes of middle anterior, central and middle posterior cerebral cortex, as well as with great thickness of right postcentral cortex and left posterior cingulate cortex. Finally, thinking processes related to a choice was associated with great thickness of the right posterior cingulate cortex.

Overall the study indicates that the organizational factors of thought and language, conveying also the emotional meaning of the text, are related to the structure of cerebral areas involved in somato-sensory associative processes (postcentral and lateral occipital cortices), in emotional awareness (entorhinal and posterior cingulate cortices), and in emotional control and feelings (orbitofrontal cortex). Just such functions are compromised in the presence of high levels of alexithymia, because an altered referential process can lead to somato-sensorially perceive but not-verbally express the emotions.

Furthermore, in the present study most of associations were found between first and third factor (resulted from the correspondence analysis) and the macro-structural measures in the right brain hemisphere, totally fitting the

proposal of Bucci (1997) that suggests the right hemisphere as the neurophysiological substratum underlying the processing of emotional information and referential process. On this vein, alexithymia may be considered an embodiment process related to altered perception of physiological correlates (viscero- and somato-motor responses) of the emotional activation resulting in a deficit in the emotional awareness. In fact, a dysfunctional referential process can lead to a lack of words for the emotions, up to being without symbols for the somatic states.

References

- Bagby R.M., Parker J.D. and Taylor G.J. (1994). The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure. *J Psychosom Res.* 38(1):23-32.
- Bagby R.M., Taylor G.J., Parker J.D. and Dickens S.E. (2006). The development of the Toronto Structured Interview for Alexithymia: item selection, factor structure, reliability and concurrent validity. *Psychother Psychosom.* 75(1):25-39.
- Bucci W. (1997). *Psychoanalysis and cognitive science: A multiple code theory.* Guilford Press.
- Caretti V., Porcelli P., Solano L., Schimmenti A., Taylor G.J. and Bagby R.M. (2011). Reliability and validity of the Toronto Structured Interview for Alexithymia in a mixed clinical and nonclinical sample from Italy. *Psychiatry Research*, 187:432-436.
- Cordella B., Greco F. and Raso A. (2014). Lavorare con Corpus di Piccole Dimensioni in Psicologia Clinica: Una Proposta per la Preparazione e l'Analisi dei Dati. In Nee E., Daube M., Valette M. and Fleury S., editors, *Actes JADT 2014 (12es Journées internationales d'Analyse Statistique des Données Textuelles, Paris, France, Juin 3-6, 2014)*, pp. 173-184.
- Damasio A. and Carvalho G.B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat Rev Neurosci.*, 14(2):143-152.
- Greco F. (2016). *Integrare la disabilità. Una metodologia interdisciplinare per leggere il cambiamento culturale.* Franco Angeli.
- Lancia F. (2017). *User's Manual: Tools for text analysis. T-Lab version Plus 2017.*
- Laricchiuta D., Petrosini L., Picerni E., Cutuli D., Iorio M., Chiapponi C., Caltagirone C., Piras F. and Spalletta G. (2015a). The embodied emotion in cerebellum: a neuroimaging study of alexithymia. *Brain Struct Funct.*, 220(4):2275-2287.
- Laricchiuta D., Lai C. and Petrosini L. (2015b). Alexithymia: From Neurobiological Basis to Clinical Implications. In Bryant M.L., editor, *Handbook on Emotion Regulation: Processes, Cognitive Effects and Social Consequences*, Nova Science Publishers.

- Lebart L. and Salem A. (1994). *Statistique Textuelle*. Dunod
- LeDoux J. (2012). Rethinking the emotional brain. *Neuron*, 73(4): 653-676.
- Savaresi S. M. and Boley D. L. (2004). A comparative analysis on the bisecting K-means and the PDDP clustering algorithms. *Intelligent Data Analysis*, 8(4): 345-362.
- Taylor G.J. and Bagby R.M. (2004). New trends in alexithymia research. *Psychother Psychosom.*, 73(2):68-77.