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, registra-tori 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
Cédrick 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 Tolouse, 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 Mingo: 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

Abstracts

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

Daniela Laricchiuta[1], Francesca Greco[2], Fabrizio Piras[3], Barbara Cordella[4], Debora Cutuli[5], Eleonora Picerni[6], Francesca Assogna[7], Carlo Lai[8], Gianfranco Spalletta[9], Laura Petrosini[10]

[1]Sapienza University, IRCCS Fondazione Santa Lucia – daniela.laricchiuta@uniroma1.it
[2]Sapienza University of Rome, Prisma S.r.l. – francesca.greco@uniroma1.it
[3]IRCCS Fondazione Santa Lucia - f.piras@hsantalucia.it
[4] Sapienza University of Rome – barbara.cordella@uniroma1.it
[5]Sapienza University of Rome, IRCCS Fondazione Santa Lucia – debora_cutuli@yahoo.it
[6]Sapienza University, IRCCS Fondazione Santa Lucia – eleonora.picerni@uniroma1.it
[7]IRCCS Fondazione Santa Lucia - f.assogna@hsantalucia.it
[8]Sapienza University of Rome – carlo.lai@uniroma1.it
[9]IRCCS Fondazione Santa Lucia – g.spalletta@hsantalucia.it
[10]Sapienza 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 cerebro-vascular 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	Factor 2	Factor 3
			Experience	Thought	Aim
1	Think	34,24%		Reasoning	Speculate
			0,03	0,46	-0,64
2	Feel	18,09%	Personal	Feeling	Speculate
			-0,13	-0,65	-0,18
3	Relationship	23,15%	Social		Choice
			0,52	0,08	0,24
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 Statistque 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.