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

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 ALRAHABI[1], 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 quanti-quali 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 esploratoria 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 ZARDETTO[1]
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 Internacional242
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 SOCAL, 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 *joggeuse* 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 VANNI[1], 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 Century514
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 LAVRENTIEV
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 TENDERO[1], 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 *The Origin of Species* : 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 *Rhapsodie*.... 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

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

Barbara Cordella, Francesca Greco, Paolo Meoli,
Vittorio Palermo, Massimo Grasso
Sapienza University of Rome – barbara.cordella@uniroma1.it; francesca.greco@uniroma1.it;
paolomeoli3@libero.it; vittorio.palermo2511@gmail.com; massimo.grasso@uniroma1.it

Abstract 1
The paper explores the professors and students' representation of professional training in Clinical Psychology in the faculty of Medicine and Psychology of the Sapienza University of Rome in order to understand whether the educational context supports students in developing their ability to enter the job market. To this aim, an Emotional Text Mining of the interviews of 30 students and 17 teachers of the Clinical Psychology Master of Science was performed. Both corpora underwent the analysis procedure performed with T-Lab, i.e. a cluster analysis with a bisecting *k*-means algorithm followed by a correspondence analysis on the keyword per cluster matrix, and the results were compared. The results show 4 clusters and 3 factors for each corpus, highlighting a relationship between student and professor representations. Both of them split the training process, distinguishing the educational process from the professional one. The emotional text mining of the interviews turned out to be an enlightening tool letting their latent dimensions emerge, setting the process and outcome of the academic training, and it proved to be very useful for educational purposes.

Abstract 2
La ricerca ha esplorato la rappresentazione della formazione in Psicologia Clinica dei professori e degli studenti della facoltà di Medicina e Psicologia della Sapienza Università di Roma al fine di comprendere se il contesto formativo supporti gli studenti nello sviluppo di competenze utili all'inserimento nel mercato del lavoro. A questo scopo è stata effettuata un'Emotinal Text Mining delle interviste di 30 studenti e di 17 professori del Corso di Laurea Magistrale in Psicologia Clinica con T-Lab (analisi dei cluster con algoritmo bisecting k-means seguita da un'analisi delle corrispondenze sulla matrice cluster per parole-chiave). I risultati mostrano 4 cluster e 3 fattori in entrambi i corpora, evidenziando una relazione tra le rappresentazioni degli studenti con quelle dei professori per quanto concerne il processo di apprendimento, distinguendo e mantenendo separati gli aspetti formativi da quelli professionali. L'Emotional Text Mining risulta essere uno

strumento utile ad evidenziare le dimensioni latenti che organizzano il processo e i risultati dell'apprendimento accademico.
Keywords: Education, Clinical Psychology, Job Market, Youth Unemployment, Emotional Text Mining.

1. Introduction

The problem of youth unemployment is relevant nowadays. In Italy, 25% of young people under 30 years of age are unemployed and this percentage grows to 40% for under 25s (Mckinsey & Company, 2014). But why is this percentage so high? According to Mckinsey's study (*ibidem*), it shows that the figure of 40% for youth unemployment does not rely on the economic cycle but on "structural causes". Among other causes, education is one of the relevant factors of youth unemployment, and is a protection factor for poverty and quality of life, as stated by ISTAT (2017). Graduates are less likely to become poor although the employability and the wages depend on the type of degree. 80% of young graduates in psychology are employed after four years (Anpal Servizi, 2017). Psychologists are more likely to become entrepreneurs than employees. Most probably, the length of time needed to get into the job market is connected to the mismatch between the educational system and enterprise (McKinsey & Company, 2014). Young people's skills are considered appropriate by 70% of Schools and Universities, but only by 42% of employers. The effectiveness of education depends in part on the representation of the professional training characterizing the University. Several studies were performed in order to investigate students' representation in the Psychology Faculty in order to improve the training process (e.g., Carli et al., 2004; Paniccia et al., 2009). Due to the change in the educational plan that took place over the past decade, this study aims to understand whether the present educational context supports students in developing their ability to enter the job market, performing an emotional text mining (Cordella et al., 2014; Greco, 2016) of the interviews of students and teachers of the Master Degree in Clinical Psychology at the Sapienza University of Rome.

2. Methodology

We know that a person's behaviour depends not only on their rationale thinking but also, and sometimes most of all, on their emotional and social way of mental functioning (Carli, 1990; Moscovici, 2005). Namely, people consciously categorize reality and, at the same time, unconsciously symbolize it emotionally (Fornari, 1976). These two thinking processes are the product of the double-logic way of the functioning of the mind (Matte Blanco, 1981) which allows people to adapt to their social environment. According to this

socio-constructivist approach, based on a psychodynamic model, the unconscious processes are social, as people generate interactively and share the same emotional meanings. The socially shared emotional symbolization sets the interactions, behaviours, attitudes, expectations and communication processes, and for this reason, the analysis of the narrations allows for the acquisition of the latent emotional meaning of the text (Salvatore & Freda, 2011). If the conscious process sets the manifest content of the narration, namely *what* is narrated, the unconscious process can be inferred through *how* it is narrated, that is to say, the words chosen to narrate and their association within the text. We consider that people emotionally symbolize an event, or an object, and socially share this symbolisation. The words they choose to talk about this event, or object, is the product of the socially-shared unconscious symbolization (Greco, 2016). According to this, it is possible to detect the associative links between the words to infer the symbolic matrix determining the coexistence of these terms in the text. To this aim, we performed a multivariate analysis based on a bisecting *k*-means algorithm (Savaresi et Boley, 2004) to classify the text, and a correspondence analysis (Lebart et Salem, 1994) to detect the latent dimensions setting the cluster per keywords matrix. The interpretation of the cluster analysis results allows for the identification of the elements characterizing the emotional representation of education, while the results of correspondence analysis reflect its emotional symbolization (Cordella et al., 2014; Greco, 2016). The advantage connected with this approach is to interpret the factorial space according to words polarization, thus identifying the emotional categories that generate professional training representations, and to facilitate the interpretation of clusters, exploring their relationship within the symbolic space.

3. Data collection and analysis

In order to explore the emotional representation of the education in the Master of Science in Clinical Psychology, we interviewed 30 students (13% of students) and 17 teachers (71% of teachers) of the Sapienza University of Rome accordingly to their voluntary participation. We used an open-questions interview for students and teachers. Students' interviews resulted in a medium size corpus of 57.387 tokens, and teachers' interviews resulted in a small size corpus of 28.746 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 ($TTR_{students}$ = 0,09; $Hapax_{students}$ = 50,3%; $TTR_{teachers}$ = 0,147; $Hapax_{teachers}$ = 53,8%). According to the size of the corpus, both lexical indicators highlight its richness and indicate the possibility to proceed with the analysis. First, data were cleaned and pre-processed by the software T-Lab (Lancia, 2017) and keywords were selected.

Due to the size of the corpus and the hapax percentage, in order to choose the keywords, we used the selection criteria proposed by Greco (Cordella et al., 2014; Greco, 2016). In particular, we used stem as keywords instead of type, filtering out the lemma of the open-questions of the interviews. 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 did not have at least two keywords co-occurrence. The eta squared value was used to evaluate and choose the optimal solution. To finalize the analysis, a correspondence analysis on the keywords per clusters matrix was made (Lebart et Salem, 1994) in order to explore the relationship between clusters, and to identify the emotional categories setting professional training representations both for students and teachers.

4. Main results and discussion

The results of the cluster analysis show that the keywords selected allow the classification on an average of 96% for both corpuses. The eta squared values was calculated on partitions from 3 to 9, and they show that the optimal solution is four clusters for both corpora. The correspondence analysis detected three latent dimensions. In table 1 and 2, we can appreciate the emotional map of the professional training emerging from the interviews of the teachers and the students and cluster location in the factorial space.

Table 1 – Cluster coordinates on factors of the teachers' corpus (the percentage of explained inertia is reported between brackets above each factor)

Cluster (CU in Cl %)		Factor 1 1 (26,53%) Motivation	Factor 2 (19,03%) Outcome	Factor 3 (14,56%) Role
1	Training Group (22,3%)	Group -0,21	Competence 0,51	Teacher -0,50
2	Clinical Training (33,7%)	Institution 0,33	Competence 0,23	Professional 0,39
3	Institutional Obligations (20,2%)	Institution 0,65	Degree -0,66	Teacher -0,38
4	Student Orientation (23,8%)	Group -0,79	Degree -0,39	Professional 0,16

CU in Cl = context units classified in the cluster.

The teachers' corpus first factor (table 1) represents the motivation in teaching, focusing on the group of students and their specific needs or on the Institutional generic scopes; the second factor focuses on the training outcome, the degree or the professional skills; and the third factor reflects the role of the academic professor that could represent oneself as a teacher or a

professional. As regards the students corpus (table 2), the first factor represents the approach to university experience, which can be perceived as an individual experience or a social one (relational); the second factor explains how students experience vocational training, perceiving it as the fulfilment of obligations or the construction of professional skills that requires personal involvement; and the third factor reflects the outcome of the educational training that can focus on professional skills development or on the achievement of qualifications.

Table 2 – Cluster coordinates on factors of the students' corpus (the percentage of explained inertia is reported between brackets above each factor)

Cluster (CU in Cl %)		Factor 1 (23,2%) Approach	Factor 2 (15,3%) Training	Factor 3 (14,0%) Outcome
1	Idealized Product (27,6%)	Individual -0,56	Fulfilment 0,45	Skills -0,43
2	Professional Education (20,8%)	-0,04	Construction -0,63	Skills -0,24
3	Group Identity (26,3)	Relational 0,69	Fulfilment 0,22	-0,01
4	Empty Degree (25,3%)	Individual -0,32	0,01	Qualifications 0,59

CU in Cl = context units classified in the cluster.

Table 3 – Teachers' Cluster (the percentage of context units classified in the cluster is reported between brackets)

Cluster 1 (22,3%) Training Group		Cluster 2 (33,7%) Clinical Training		Cluster 3 (20,2%) Institutional Obligations		Cluster 4 (23,8%) Student Orientation	
keyword	CU	keyword	CU	keyword	CU	keyword	CU
studente	59	psicologia	94	scuola	29	domanda	42
cercare	43	lavoro	81	persona	28	idea	40
corso	43	clinico	54	laurea	19	organizzazione	33
teoria	32	insegnare	36	università	18	aggiungere	32
lezione	21	contesto	29	trovare	17	processo	30
modalità	21	problema	27	specializzazione	16	rispetto	29
organizzazione	20	intervento	27	importante	16	orientare	21
intervento	19	diverso	25	entrare	15	parlare	21
relazione	17	conoscenza	22	scegliere	14	Corso di laurea	20
				percorso	14	Attività	18
modello	16	interno	22			didattiche	

CU = context units classified in the cluster.

The four clusters of both corpuses are of different sizes (tables 1 and 2) and reflect the representations of the professional training (table 3 and 4). Regarding the teachers' corpus (table 3), the first cluster represents the group of students as a tool to teach professional skills, focusing on the group process where relational dynamics are experienced; the second cluster focuses on clinical training, teaching skills marketable in the job market; the third cluster focuses on the teachers' institutional obligations regardless of the students' training needs; and the fourth cluster represents students' orientation as a way to support students in managing their academic training regardless of professional skills. As regards the students' corpus (table 4), in the first cluster the good training involves students' adherence to lesson tasks regardless of critical thinking on the theoretical model proposed; in the second cluster, learning professional skills is strictly connected to the ability to get and respond to market demand; the third cluster reflects the relevance of belonging to a group of colleagues supporting the construction of a professional identity that, unfortunately, seems unconnected to professional skills development; and the fourth cluster represents professional training as a process in which the degree achievement is the main goal, regardless of the job market demand.

Table 4 – Students' Cluster (the percentage of context units classified in the cluster is reported between brackets)

Cluster 1 (27,6%) Idealized Product		Cluster 2 (20,8%) Professional Education		Cluster 3 (26,3) Group Identity		Cluster 4 (25,3%) Empty Degree	
keyword	CU	keyword	CU	keyword	CU	keyword	CU
esperienza	116	pensare	89	scelta	154	vivere	26
triennale	44	esame	71	studiare	153	trovare	85
percorso	43	psicologia	65	frequentare	104	tesi	20
professione	41	seguire	55	rapporto	102	sentire	91
università	37	realtà	55	piacere	98	riuscire	30
possibilità	35	vedere	55	colleghi	97	prendere	33
capire	33	iniziare	53	parlare	74	persone	105
diverso	31	triennale	53	organizzare	68	maniera	23
senso	30	lavoro	44	domanda	55	livello	35
vivere	25	interessante	44	aggiungere	36	laboratorio	18

CU = context units classified in the cluster.

Students and teachers seem to have similar representations of the training process: the academic need of building a network, highlighted by the students' cluster on *group identity*, and the teachers' cluster on *training group* and *student orientation*; the relevance of achieving a qualification, highlighted by the students' cluster on *empty degree* and the teachers' cluster on *institutional obligation*; and the development of professional skills marketable in the job market reflected by the teachers' cluster on *clinical training* and the

students' cluster on *professional education* in line with what it was found by Carli and colleagues (2004) and Paniccia and colleagues (2009) by means of a similar methodology, the emotional textual analysis (Carli et al., 2016). The awareness of the psychological demand of the labour market is an indicator of the professional training process's effectiveness. Nevertheless, students and teachers split the academic achievement from the development of professional skills. This could be a critical aspect, possibly explaining young graduates' difficulty in entering the job market, focusing more on academic context rather than on market demand. As a consequence, during the training process, students do not develop the connection between professional training (what they are learning) and professional skills (what they are going to do in the future).

5. Conclusion

Although the study results could not be generalized, due to the participants' selection criteria and the methodology we used, they highlight professional training representation characteristics, which are the elements influencing the rate of unemployment among young psychologists. Even though it is not possible to quantify the relevance of the characteristics of the representation, the emotional text mining, allowing for the identification of the words association explanatory of the education representation, allows for hypotheses definition and the identification of the resources and the issues pertaining the professional training in a specific context.

The interpretation of the text mining results lets the social unconscious process emerge, setting the education useful to defining the type of psychological intervention able to support the representation transformation toward a more effective training process. In this particular case study, the intervention would aim to develop the connection between professional qualification achievement and the professional skills development, which are currently split.

References

Anpal Servizi (2017), L*'inserimento occupazionale dei laureati in psicologia, dell'università La Sapienza di Roma, Direzione e studi analisi statistica - SAS.*

Carli R. (1990). Il processo di collusione nelle rappresentazioni sociali. *Rivista di Psicologia Clinica*, 4: 282-296.

Carli R., Dolcetti F. and Dolcetti (2004). L'Analisi Emozionale del Testo (AET): un caso di verifica nella formazione professionale. In Purnelle G., Fairon C. and Dister A., editors, *Actes JADT 2004: 7es Journées internationales d'Analyse statistique des Données Textuelle*s, pp. 250-261.

Carli R., Paniccia R.M., Giovagnoli F., Carbone A. and Bucci F. (2016).

Emotional Textual Analysis. In L. A. Jason and D. S. Glenwick, editors, Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods. Oxford University Press.

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),* pp. 173-184.

Fornari F. (1976). *Simbolo e codice: Dal processo psicoanalitico all'analisi istituzionale.* Feltrinelli.

Greco F. (2016). *Integrare la disabilità. Una metodologia interdisciplinare per leggere il cambiamento culturale*. Franco Angeli.

ISTAT (2017). *Rapporto annuale 2017*. ISTAT

Lancia F. (2017). *User's Manual : Tools for text analysis. T-Lab version Plus 2017.*

Lebart L. and Salem A. (1994). *Statistique Textuelle*. Dunod

Matte Blanco I. (1981). *L'inconscio come insiemi infiniti: Saggio sulla bi-logica.* Einaudi

McKinsey & Company (2014). Studio ergo Lavoro, come facilitare la transizione scuola lavoro per ridurre in modo strutturale la disoccupazione giovanile in italia. Report di Ricerca "Studio ergo Lavoro", McKinsey & Company, https://www.mckinsey.it/file/2785/download?token=a3VfesjU.

Moscovici S. (2005). *Le rappresentazioni sociali.* Il Mulino.

Paniccia R.M., Giovagnoli F., Giuliano S., Terenzi V., Bonavita V., Bucci F., Dolcetti F., Scalabrella F. and Carli R. (2009). Cultura Locale e soddisfazione degli studenti di psicologia. Una indagine sul corso di laurea "intervento clinico" alla Facoltà di Psicologia 1 dell'Università di Roma "Sapienza". *Rivista di Psicologia Clinica, Supplemento n.* 1: 1-49.

Salvatore S. and Freda M. F. (2011). Affect, unconscious and sensemaking: A psychodynamic, semiotic and dialogic model. New Ideas, *Psychology*, Vol. 29, pp. 119–135.

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.