

# JADT' 18

PROCEEDINGS OF THE  
14<sup>TH</sup> INTERNATIONAL CONFERENCE  
ON STATISTICAL ANALYSIS OF TEXTUAL DATA



# JADT' 18

PROCEEDINGS OF THE  
14<sup>TH</sup> 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 ALRAHABI<sup>1</sup>, CHIARA MAINARDI<sup>1</sup>**

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 ZARDETTO<sup>1</sup>**

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  
*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<sup>1</sup>, 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 POU DAT**

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

<b>SANDRO STANCAMPIANO</b>	
Misurare, Monitorare e Governare le città con i Big Data .....	748
<b>FADILA TALEB, MARYVONNE HOLZEM</b>	
Exploration textométrique d'un corpus de motifs juridiques dans le droit international des transports .....	755
<b>JAMES M. TEASDALE</b>	
The Framing of the Migrant: Re-imagining a Fractured Methodology in the Context of the British Media. ....	763
<b>MARJORIE TENDERO<sup>1</sup>, CECILE BAZART</b>	
Results from two complementary textual analysis software (Iramuteq and Tropes) to analyze social representation of contaminated brownfields .....	771
<b>MATTEO TESTI, ANDREA MERCURI, FRANCESCO PUGLIESE</b>	
Multilingual Sentiment Analysis.....	780
<b>JUAN MARTÍNEZ TORVISCO</b>	
A linguistic analysis of the image of immigrants' gender in Spanish newspapers.....	788
<b>FRANCESCO URZÌ</b>	
Lo strano caso delle frequenze zero nei testi legislativi euroistituzionali.....	796
<b>SYLVIE VANDAELE</b>	
Les traductions françaises de <i>The Origin of Species</i> : pistes lexicométriques .	805
<b>PIERRE WAVRESKY, MATTHIEU DUBOYS DE LABARRE, JEAN-LOUP LECOEUR</b>	
Circuits courts en agriculture : utilisation de la textométrie dans le traitement d'une enquête sur 2 marchés .....	814
<b>MARIA ZIMINA, NICOLAS BALLIER</b>	
On the phraseology of spoken French: initial salience, prominence and lexicogrammatical recurrence in a prosodic-syntactic treebank <i>Rhapsodie</i> ....	822

*Abstracts*

<b>FILIPPO CHIARELLO, GUALTIERO FANTONI, ANDREA BONACCORSI, SILVIA FARERI</b>	
What kind of contributions does research provides? Mapping issue based statements in research abstracts .....	833
<b>FILIPPO CHIARELLO, GIACOMO OSSOLA, GUALTIERO FANTONI, ANDREA BONACCORSI, ANDREA CIMINO, FELICE DELL'ORLETTA</b>	
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





# Looking Through the Lens of Social Sciences: The European Union in the EU-Funded Research Projects Reporting

Matteo Gerli

University for Foreigners of Perugia – matteogerli81@gmail.com

## Abstract

In the last decades, European integration and scientific production have come to be deeply intertwined as a result of the Europeanization of many research activities. On one side, European institutions promote the realization of research projects aiming at developing a type of knowledge “close” to the end users’ interests; on the other side, the resulting knowledge contributes to conditioning the practices that take place in the European and national institutions, according to a circular process that brings the innovations to feed back into the system that expresses them. The purpose of this paper is to explore this relationship by examining two peculiar scientific products realized by researchers operating within the broad domain of the Socio-economic Sciences and Humanities (SSH), as a part of the research projects financed by the Seventh Framework Programme (2007-2013) of the European Union: *final reports* and *policy briefs*. In other words, it aims to analyse all reports as a whole using some automatic text analysis tools, while incorporating some supplementary variables which help to define the broader context of scientific production.

**Keywords:** European Union, International Research Projects, Socio-economic Sciences and Humanities, Textual Data Exploration, Quantitative Discourse Analysis, IRaMuTeQ.

## 1. Introduction

The European Research Policy plays a strategic role for thousand of researchers and research institutions which operate within the EU borders. Thanks to the concomitant decrease in national public funds for scientific activities (see for instance, Vincent-Lacrin, 2006; 2009), the European research agenda has dramatically increased its appeal among scholars and consequently its ability to have an impact on the directions and processes of scientific knowledge production. Indeed, starting from the 90s, the European Commission has equipped itself with new means to combine and manage, on the basis of medium to long-term planning cycles, the whole set of scientific and technological initiatives financed by the European budget: the *framework*

*programme* (Ippolito, 1989; Ruberti and André, 1995; Guzzetti, 1995; Menéndez and Borrás, 2000; Borrás, 2000; Banchoff, 2002; Cerroni and Giuffredi, 2015). In short, the underlying logic is that of the programmatic intersection between research activities and other European policies, so that the promotion of scientific excellence complements the need to foster the creation of cross-border and interdisciplinary collaborations intended for producing a type of knowledge “close” to the end users’ interests.

As it was observed in previous studies (Adler-Nissen and Kropp, 2015), European integration and scientific production have come to be deeply intertwined: on one side, the progress of integration process influenced (and still influences) research activities through the promotion of particular forms of knowledge and research questions (as far as we are concerned, mainly through the realization of cross-national and cross-disciplinary research projects); on the other side, the resulting knowledge contributes to conditioning the practices that take place in the European and national institutions, according to a circular process that brings the innovations to feed back into the system that expresses them. Social Sciences and Humanities, which are less directly involved in the production of knowledge with a clear practical usability, are by no means unconcerned with this kind of phenomenon. At this regard, the *Journal of European Integration* has recently published a special issue on the relationship between social sciences and European integration, hosting some important articles that have highlighted the existence of several “crossroads” between the European Union and the scientific community’s “itineraries”<sup>1</sup>: Rosamond (2015), for instance, observed how certain theories on the political and economic integration (in particular that of the Hungarian Béla Balassa, from the economics side, and the neofunctionalism, from the political science side) had been informing the “strategic narrative” adopted by the European Commission during the 60s and 70s to legitimize its newly-formed institutional role and its economic policy position, according to a quite peculiar two-ways traffic of influences process, being the economic integration theorized while it was happening; Deem (2015) pointed out the existence of a relationship between the birth of a new field on higher education studies, the simultaneous evolution of national university systems and the launch of the so-called *Bologna process* at European level; Vauchez analysed, through a sociogenetic approach, the historical process through which the *acquis communautaire* «has been formulated, stretched, criticized, revised and finally naturalized as the most rigorous and objective measure of Europe against other possible methods» (2015: 196) thanks to the work of those who have been defined

---

<sup>1</sup> *Journal of European Integration*, 37 (2015).

“methodological entrepreneurs”, that is European officials who have politically invested and succeeded in establishing Europe’s cognitive and technical equipment.

Looking beyond such individual cases, what is really relevant to our purpose is the underlying idea about the possibility of studying science production from a sociological point of view, basically by rejecting what was traditionally regarded as an *internal/external division* (Adler-Nissen and Kropp, 2015: 161-163), and thus admitting that even scientific and academic concepts can be formulated in conjunction with political-economic ambitions and practical problems (see Bohme *et al.*, 1983; Funtowicz and Ravetz 1993; Slaughter and Leslie 1997; Gibbons *et al.*, 1994; Ziman, 2000; Albert and McGuire, 2014), such as those above mentioned. This does not mean that science is equal to politics or economics (Breslau, 1998); what it does mean is that, in order to understand science production, one needs to recognize that “non-academic” resources (such as, for instance, financial or material resources, ideas and beliefs, symbolic resources, political or normative resources, people, etc.) may overstep scientific boundaries and be used for the production of new knowledge. Bourdieu (1975, 1984, 1990, 1992, 1994, 1995, 2001) described this phenomenon through the concept of “fields interrelations”. In few words, the social world is composed of multiple semi-autonomous *fields*, basically microcosms characterized by different *stakes, rules of the game* and particular resources which one needs to possess to get access to the game itself and its specific advantages. He conceptualized these sphere as partially independent, by which he means that, even though each field develops its own institutions, hierarchies, problems, tacit or explicit rules, they necessarily interact and affect each other. This is particularly true for cultural fields (art, cinema, religion, science, journalism, etc.), since they are structurally dependent and subordinated to political and economic fields. Going straight to the point, this is to say that, if one is dealing with a sociological analysis of a cultural product (e.g. a text), thus one neither can just consider its formal characteristics, nor be limited to its context of production. Instead, one should use a “relational approach”, taking into account both the *internal* features of the product and its *external* determinants. In engaging with this broad issue, this paper will try to further contribute to the understanding of the topic by examining two peculiar scientific products realized by researchers operating within the broad domain of the Socio-economic Sciences and Humanities (SSH), as a part of the research projects financed by the Seventh Framework Programme (2007-2013) of the European Union: *final reports* and *policy briefs*. By using some automatic text analysis tools, it will thus statistically explore the contents of such documents not *per se*, but in connection with some variables, which help to define the broader

context of production. In its exploratory character, this study does not have strong hypothesis to be tested. Nevertheless, following Bourdieu's approach, it aims to give an original perspective through which observing the relationship between the field of social sciences and the public policy field of the European Union (Gerli, 2017).

## 2. The corpus and methodology

Unlike the studies discussed earlier, which are mainly based on micro-sociological observation, our investigation covers a macro-sociological analysis of a quit large *corpus* made of 46.513 graphic forms, equal to 3.025.960 occurrences. It is an *ad-hoc* constructed *corpus*: it contains 360 texts, of which 205 belonging to *final reports* and 155 to *policy briefs*, which were collected from the digital database CORDIS<sup>2</sup>, the main institutional source of information related to the research projects financed by the European Union. The choice to focus on these documents is not accidental, but depends on their strict relevance to our research objectives. In fact, both include a summary of the project results and conclusions, with a description of their potential socio-economic impact (EC 2010), even though *policy brief* is strictly designed for policy makers (both European and national ones), while *final report* is addressed to a wider audience, which may include (at least potentially) lay people as well. In this perspective, they represent an effective "shortcut" through which empirically observe the way in which the research groups awarded a grant "actualized" the inputs they received from the Commission. This is, to resume the previous discussion, to analyse how European institutions and social scientists contribute together to the definition and resolution of some EU-related issues.

With regard to the methodology, both simple and multivariate analyses were performed with the IRaMuTeQ software (Lebart *et al.*, 1998; Bolasco, 2013). In particular, the lexicographical analysis was used for a first exploration of the *corpus*, that is to identify and format texts units, turn texts into text segments (TS) and classify words by their frequency. The multivariate analysis, instead, was performed to detect the associations between textual data and the following supplementary variables related to what in the 7FP was defined as *macro-activity* (MA) and *financing scheme* (FS)<sup>3</sup>. Going into more details, the 7FP included eight macro-activities: *Growth, employment and competitiveness in a knowledge society* (MA1); *Combining economic, social and environmental goals in Europe: towards sustainable development* (MA2); *Major*

---

<sup>2</sup>[http://cordis.europa.eu/projects/home\\_it.html](http://cordis.europa.eu/projects/home_it.html).

<sup>3</sup> For more details: Decision No 1982/2006/EC of the European Parliament and of the Council of 18 December 2006.

*trends in society and their implications* (MA3); *Europe in the world* (MA4); *The citizen in the European Union* (MA5); *Socio-economic and scientific indicators* (MA6); *Foresight studies* (MA7); *Strategic activities* (MA8). As for the financing schemes, the 7FP included five main different types, which differed from each other by the research team size and the type of purposes to be achieved (the first three mainly focused on the development of new knowledge, while the last two were mainly thought for the coordination and support of research activities and policies): *Small or medium-scale focused project* (FS1); *Small or medium-scale focused research project aimed at international cooperation* (FS2); *Large-scale integrating project* (FS3); *Coordination action* (FS4); *Support action* (FS5). Additionally, we also took into account the *starting year* of the project and the *geographic area* in which the coordinating institution was located.

As a whole, our sample (of non-probabilistic type) involves 223 research projects out of 251 realized in 2007-2013 (equal to 88.8%) and broadly covers all macro-activities and financing schemes above mentioned. In Tab. 1, a description of the *corpus* and its main subsets is provided.

Tab. 1: Description of the corpus

Type	Number of texts	Graphic forms	Occurrences
Final report	205	42.047	2.441.168
Policy Brief	155	19.795	584.792
Corpus	360	46.513	3.025.960

### 3. The main findings

At first glance, the most frequent “full” words used in the SSH research reports do not provide particularly relevant insights. The first ten (*social, policy, research, European, project, EU, countries, public, national, Europe*) concerns the “general context of meaning” where discourses on Europe and related issues took shape. Ten words that, without having a clear disciplinary connotation, define some “semantic coordinates” common to all research projects carried out. Interesting enough, it is the wide use of the words *country/es* (freq.=10.531) and *national* (freq.=5.527) which, compared with the words *European* (freq.=9.190), *EU* (freq.= 8.563) and *Europe* (freq.=5.408), prove the great importance of the “national” level of analysis, mainly in a comparative way. Scrolling down the list, we can also recognise some typical words of the socio-economic lexicon (*economic, market, growth, employment, financial*), the socio-political lexicon (*people, education, State, young, groups, cultural, society, governance*), and the methodological one, namely related to the operative context of the research activities (*date, case, results, impact, analysis, study*). Yet these are terms that, at this early stage of the analysis, do not provide any clear “message”.

At a closer look, however, we can identify some specific words which are, in a broad sense, linked to the political macro-orientations defined by the Lisbon Strategy (European Council 2000), demonstrating the “osmosis” existing between European institutions and social sciences. Here some examples: *innovation* (freq.=5.793), cornerstone of industrial competitiveness and economic growth (EC 2003, 2006); *development* (freq.=5.176), to be understood, among the various meaning, mainly as sustainable development (EC 2005, 2009); *education* (freq.= 3.490) and *knowledge* (freq.=3.221), which, together with the already mentioned “innovation”, represent the “three sides” of the so-called “knowledge triangle”, from the European Commission’s perspective, the ground for a greater economic and social dynamism.

For the aim of this study, what is of particular interest is also the geographical scope of the research activities. Indeed, the most frequent toponyms refer to EU based countries. Among these, the five main sponsors and recipients of the framework programs (*Germany, UK, France, Italy and Spain*) are placed at the top of the ranking. As for the extra-European countries, several of them are placed in Asia (e.g. *China, Japan, India, Vietnam and Thailand*), North Africa (*Morocco, Tunisia, Egypt and Libya*) and South America (*Brazil, Argentina, Colombia, Peru and Chile*). This is indicative of a globalization process, which is affecting both European institutions and researchers by expanding their interests (“political”, with regard to the first ones, and “scientific”, for the second ones) beyond the European borders. What matters is that they are moving together insofar we can suppose the existence of a clear synergy between the emergence of a new multipolar area of political, commercial and cultural influence, in which the European Union is now required to act, and the production of knowledge on topics with a potential “global” added value.

### **3.1 The main semantic groups and their connections with the “context”**

To go deeper in the analysis, and to explore the relationship between the selected texts and some variables related to their context of production, we performed a Descending Hierarchical Analysis (DHA). Indeed, this method allowed us, first, to identify clusters with similar vocabulary within text segments and, then, to visualize them in conjunction with the supplementary variables (Camargo and Justo 2013; Curbelo, 2017). In Fig. 1, the output of the DHA is summarised.

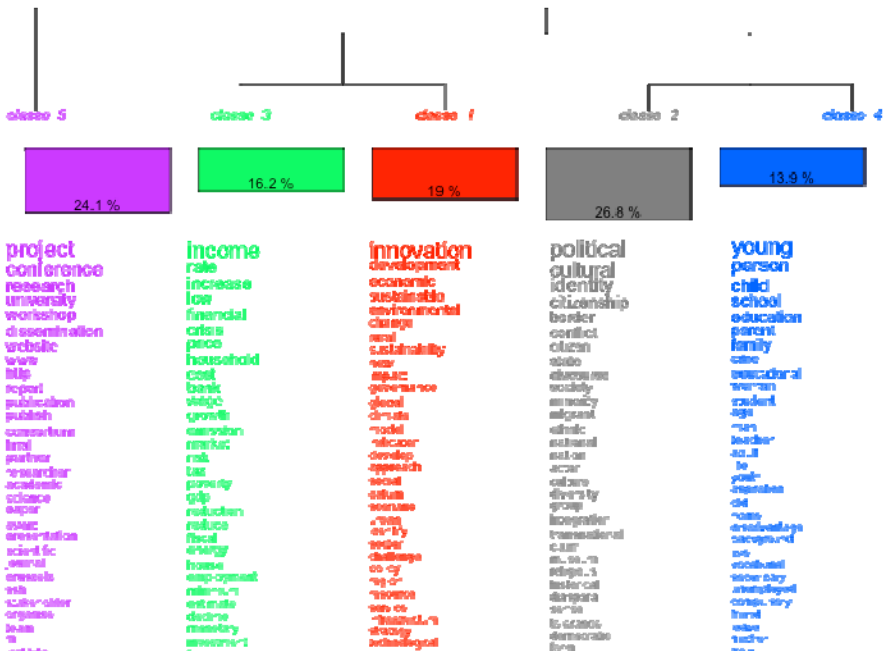


Fig. 1: Dendrogram of top-down hierarchical classification (Reinert method) of the corpus

As it can be easily seen in Fig. 1, the DHA algorithm allowed the identification of five clusters, each with its own specific semantic content. Following Reinert (1987), they can be interpreted as “lexical words”, namely specific semantic structures which, in our case, refer to different and even competing scientific representations of the European Union and related issues. The second cluster has the greater representation (26,8% of the SSH discourses) and identifies a semantic sphere characterized by a language mainly oriented towards political and social issues. Indeed, the most central word in this cluster is *political*, followed by *cultural*, *identity*, *citizenship*, *border*, *conflict*, *citizen*, *State* and so on. Immigration (*migrant*) and related issues appear to be particularly relevant as well. The fifth cluster (24,1%) delineates a quite peculiar semantic sphere based on a set of words (such as *project*, *conference*, *research*, *university*, *workshop*, *dissemination*, *website*, etc.) strictly linked with the management and realization of European research projects and, more in general, with scientific research and related activities. The first cluster, third in terms of representativeness (19%), refers to the relationship between economic development and environmental protection, being the most central word *innovation*, followed by *development*, *economic*, *sustainable*, *environmental*, *change*, *rural* and so on. This interpretation seems to be supported by the presence of several words that refer to the need for a change with respect to a situation that is perceived as not desirable (*change*,

*impact, strategy, challenge, need, solution, improve, step, etc.*). The third cluster (16,2%), instead, covers a semantic area mainly related to the economy and the market. It is a language that involves two main branches, the one of the real economy (*income, price, household, wage, firm, energy, poverty, etc.*), and the one of the finance (*financial, bank, risk, monetary, credit*), but above all it is characterized by the large presence of technical terms and acronyms (*gdp, estimate, asset, inflation, emu, Eurozone, insurance, macroeconomic, etc.*). Finally, the fourth linguistic cluster (13,9%) includes words essentially associated to the relationship between education, training and employment, as shown by the presence of terms such as *young, person, child, school, education, aspiration, background, vocational* and *compulsory*. It is a cluster that differs from the others due to the greater concreteness of the language, as proved by the recurring use of words referring to “concrete” social actors (*child, parent, student, teacher, mother, friend, volunteer, etc.*).

Fig. 2, resulting from a Lexical Correspondences Analysis (LCA), shows the relationship between clusters (left side) and between clusters and the supplementary variables (right side). The main aim here was to verify whether or not SSH discourse exhibits clear evidence of “adaptability” with regard to the macro-activities and the financing schemes, as defined by the European Commission.

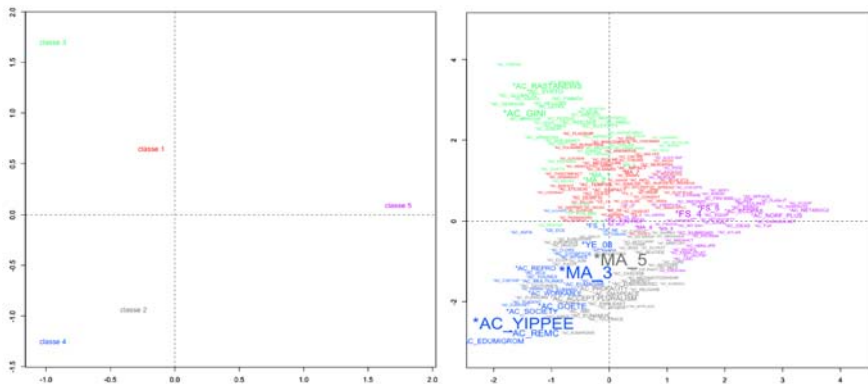


Fig. 2: Association between clusters and supplementary variables

The first two factors summarize together 67,5% of the total inertia: the first one (39,97%) marks a clear opposition between cluster 5 (positive half-plane) and the other four clusters (negative half-plane); the second factor (27,47%), instead, highlights a significant opposition between clusters 1 and 3 (positive half-plane) and clusters 2 and 4 (negative half-plane). As a whole, we can distinguish three different (partially autonomous) semantic contexts, arising from the association between the “cultural” and “socio-political” discourses



(third quarter), the “economic” discourse and that on “innovation” and “sustainable development” (forth quarter), and finally the discourse on “research activities” (in-between the first and the second quarters).

As far as the relationship between discourses (clusters) and supplementary variables, Fig. 3 and 4 show the most significant categories (those with a larger chi-square and a lower p-value), referring to the “macro-activity” and “financing scheme” variables. As shown in the first figure, MA1 and MA2 categories are only significant in the definition of clusters 1 (innovation) and 3 (economics); MA5 is the most relevant for cluster 2 (politics); similarly, MA3 category is the only significant for cluster 4 (culture); and finally, MA4 and MA8 categories predominate on cluster 5 (research activities). In short, these results strongly support the thesis of adaptability, insofar the different scientific representations of the European Union emerged from the analysis resulted strongly associated with the macro-activities defined by the European Commission.

Cluster	Category	Chi2	%	p-value
1	MA2	1226.7	25,7	<0.0001
	MA7	762.9	36,5	<0.0001
2	MA5	5220.0	54,8	<0.0001
3	MA1	1282.4	28,9	<0.0001
	MA2	1414.2	27,0	<0.0001
4	MA3	5238.5	33,0	<0.0001
5	MA4	839.9	33,6	<0.0001
	MA8	534.9	43,7	<0.0001

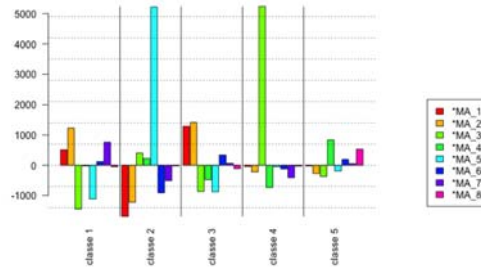


Fig. 3: Chi2 significance of variable “macro-activity” by cluster

On the other hand, the role of the “financing scheme” variable resulted much less significant in discriminating the five clusters, except for categories FS4 and FS5, which are the most significant for cluster 5, and category FS1, which instead clearly prevail on cluster 4. Nothing relevant emerged in relation to the variables “geographic area” and “starting year”.

Cluster	Category	Chi2	%	p-value
1	FS2	186.3	25,7	<0.0001
	FS3	145.1	24,7	<0.0001
2	FS1	487.6	29,0	<0.0001
3	FS1	286.5	17,6	<0.0001
4	FS1	1245.0	16,7	<0.0001
5	FS4	2195.0	51,5	<0.0001
	FS5	1583.2	58,5	<0.0001

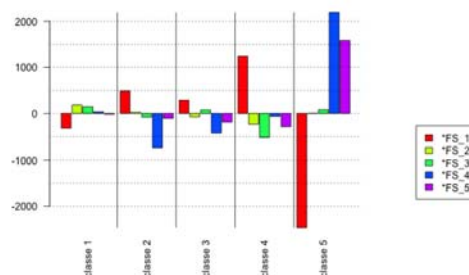


Fig. 4

#### 4. Conclusions

The findings presented herein indicate a close relationship between the programmatic framework, defined by the Commission, and the contents of the *final reports* and *policy briefs*, supporting the thesis of a co-construction of the European integration (Adler-Nissen, Kropp 2015). The scientific discourse has come to be structured around few semantic macro-aggregates arisen from DHA, which in turn resulted associated with the variables performed in LCA. Furthermore, the SSH linguistic space shows a clear cleavage between the economic discourse and the cultural discourse, which points out the existence of a lack of interaction between these two spheres. From a more “general” point of view, all this means that, in connecting the social sciences field with the policy field, the European research projects produced a scientific discourse that, on the whole, is structurally *homologous* with the “space of possibilities” inherent to the 7PQ.

#### References

- Adler-Nissen R., Kropp K. (2015). A Sociology of Knowledge Approach to European Integration: Four Analytical Principles. *Journal of European Integration*, 37(2): 155-173.
- Albert M., McGuire W. L. (2014). Understanding Changes in Academic Knowledge Production in a Neoliberal Era. *Political Power and Social Theory*, 27: 33-57.
- Banchoff T. (2002). The Politics of the European Research Area. *ACES Working Paper 3*, Paul H. Nitze School for Advanced International Studies.
- Böheme G., Van den Daele W., Hohlfeld R., Krohn W., Shafër W. (1983). *Finalization in Science. The Social Orientation of Scientific Progress*. Dordrecht: Riedel.
- Bolasco S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci.
- Borras S. (2000). *Science, Technology and Innovation in European Politics. Research Paper n. 5*, Roskilde University.
- Bourdieu P. (1975). The Specificity of Scientific Field and the Social Condition of the Progress of Reason. *Social Sciences Informations*, 6: 19-47.
- Bourdieu P. (1984). *Homo academicus*, trad. it. (2013) *Homo academicus*. Bari: Edizioni Dedalo.
- Bourdieu P. (1992). *Les règles de l'art*, trad. it. (2013) *Le regole dell'arte*. Milano: Il Saggiatore.
- Bourdieu P. (1994). *Raisons pratiques. Sur la théorie de l'action*, trad. it. (2009) *Ragioni pratiche*. Bologna: Il Mulino.
- Bourdieu P. (1995). *Champ politique, champ des sciences sociales, champ*

- journalistique*, trad. it. (2010) *Campo politico, campo delle scienze sociali, campo giornalistico*. In Cerulo M. (a cura di). *Sul concetto di campo in sociologia*. Roma: Armando.
- Bourdieu P. (2001). *Science de la science et réflexivité*, trad. it. (2003) *Il mestiere di scienziato*. Milano: Mondolibri.
- Breslau D. (1998). *In Search of the Unequivocal: The Political Economy of Measurement in U.S. Labor Market Policy*. London: Praeger.
- Camargo B. V., Justo A. M. (2013). R Interface for Multidimensional Analysis of Texts and Questionnaires, *IraMuTeQ tutorial*, available on: <http://www.iramuteq.org>.
- Cerroni A., Giuffredi R. (2015). L'orizzonte di Horizon 2020: il futuro europeo nelle politiche della ricerca. *Futuri*, 6: 29-39.
- Curbelo A. A. (2017). Analysing the (Ab)use of Language in Politics: the Case of Donald Trump. *Working Paper n. 2*. University of Bristol: SPAIS.
- Deem R. (2015). What is the Nature of the Relationship between Changes in European Higher Education and Social Science Research on Higher Education and (Why) Does It Matter?. *Journal of European Integration*. 37(2): 263-279.
- European Commission (2010). *Communicating research for evidence-based policymaking*. Bruxelles: Directorate-General for Research.
- European Commission (2003). *Politica dell'innovazione: aggiornare l'approccio dell'Unione Europea nel contesto della Strategia di Lisbona*. COM(2003) 112 definitivo, 11.03.2003.
- European Commission (2005). *Comunicazione della Commissione al Consiglio e al Parlamento europeo sul riesame della strategia per lo sviluppo sostenibile. Una piattaforma d'azione*. COM(2005) 658 definitivo, 13.12.2005.
- European Commission (2006). *Mettere in pratica la conoscenza: un'ampia strategia per l'innovazione per l'UE*. COM(2006) 502 definitivo, 10.05.2006.
- European Commission (2009). *Integrare lo sviluppo sostenibile nelle politiche dell'UE: riesame 2009 della strategia dell'Unione Europea per lo sviluppo sostenibile*. COM(2009) 400 definitivo, 24.07.2009.
- Funtowicz S., Ravetz J. (1993). Science for the Post-Normal Age. *Future*, 25: 735-755.
- Gerli M. (2017). Il campo sociale dei progetti di ricerca europei. Il caso delle SSH. *Studi Culturali*, 1: 127-150.
- Gibbons M., Limoges C., Nowotny H., Schwartzman S., Scott P. e Trow M. (1994). *The New Production of Knowledge*. London: Sage.
- Guzzetti L. (1995). *A Brief History of European Union Research Policy*. Luxembourg: Publications Office of the European Communities.
- Ippolito F. (1989). *Un progetto incompiuto. La ricerca comune europea: 1958-88*. Bari: Edizioni Dedalo.

- Lebart L., Salem A., Berry L. (1998). *Exploring Textual Data*. New York: Kluwer Academic.
- Menéndez L. S., Borrás S. (2000). *Explaining Changes and Continuity in EU Technology Policy: The Politics of Ideas*. In Dresner S. e Gilbert N. (eds), *Changing European Research System*. Aldershot: Ashgate.
- Reinert M. (1987). Classification descendante hiérarchique et analyse lexicale par contexte: application au corpus des poésies d'Arthur Rimbaud. *Bulletin de Méthodologie Sociologique*, 13: 53-90.
- Rosamond B. (2015). Performing Theory/Theorizing Performance in Emergent Supranational Governance: The Live Knowledge Archive of European Integration and the Early European Commission. *Journal of European Integration*, 37(2): 175-191.
- Ruberti A., André G. (1995). *Uno spazio europeo della scienza. Riflessioni sulla politica europea della ricerca*. Firenze: Giunti.
- Slaughter S., Leslie L.L. (1997). *Academic Capitalism: Politics, Policies and the Entrepreneurial University*. Baltimore: The John Hopkins University Press.
- Vaucher A. (2015). Methodological Europeanism at the Cradle: Eur-lex, the Acquis and the Making of Europe's Cognitive Equipement. *Journal of European Integration*, 37(2): 193-210.
- Vincent-Lacrin S. (2006). What is Changing in Academic Research? Trends and Futures Scenarios. *European Journal of Education*, 41(2): 169-202.
- Vincent-Lacrin S. (2009). Finance and Provision in Higher Education: A Shift from Public to Private?. *Higher Education to 2030 (vol. 2)*, Centre for Education Research and Innovation: OECD.
- Ziman J. (2000). *Real Science: What It Is, and What It Means*. Cambridge-New York: Cambridge University Press.



Finito di stampare in proprio  
nel mese di giugno 2018  
UniversItalia di Onorati s. r. l.  
Via di Passolombardo 421, 00133 Roma Tel: 06/2026342  
email: [editoria@universitaliasrl.it](mailto:editoria@universitaliasrl.it) – [www.universitaliaeditrice.it](http://www.universitaliaeditrice.it)