



Università
di Catania

ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

Consiglio Nazionale
delle Ricerche

ME.TE. DIGITALI

MEDITERRANEO IN RETE TRA TESTI E CONTESTI

ATTI DEL XIII CONVEGNO ANNUALE
AIUCD 2024



28 - 30 MAGGIO
MONASTERO DEI BENEDETTINI
P.ZZA DANTE, 32 CATANIA

ISBN 978-88-942535-8-0



Copyright ©2024 AIUCD
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). Ogni altro diritto rimane in capo ai singoli autori.
This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). All other rights retained by the legal owners.

A cura di: Di Silvestro Antonio; Spampinato Daria (2024). Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania.

Editing: Denise Bruno; Christian D'Agata; Laura Mazzagufò; Francesca Prado; Eliana Vitale; Alessandro Zammataro.

Ultimo accesso agli URL in data 15 maggio 2024.

Si prega di notificare all'editore ogni omissione o errore si riscontri: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria[at]aiucd.org)
Please notify the publisher of any omissions or errors found: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria[at]aiucd.org)

Il programma della conferenza AIUCD 2024 è disponibile online <https://aiucd2024.unict.it/programma/>
The AIUCD 2024 Conference Program is available online <https://aiucd2024.unict.it/programma/>

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato di Programma di AIUCD 2024.

All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2024 Program Committee.

Chair

Antonio Di Silvestro (Università di Catania)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Comitato di programma / Program committee

Emmanuela Carbé (Università di Siena)

Massimo Cultraro (CNR Istituto di Scienze del Patrimonio Culturale)

Christian D'Agata (Università di Catania)

Antonio Di Silvestro (Università di Catania)

Greta Franzini (Eurac Research)

Maurizio Lana (Università del Piemonte Orientale)

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Marco Mazzone (Università di Catania)

Ouafae Nahli (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Marianna Nicolosi-Asmundo (Università di Catania)

Marina Paino (Università di Catania)

Giuseppe Palazzolo (Università di Catania)

Jonathan Prag (University of Oxford Merton College)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Rachele Sprugnoli (Università di Parma)

Francesco Stella (Università di Siena)

Segreteria scientifica / Scientific Secretariat

Liborio Barbarino (Università di Catania)

Denise Bruno (Università di Catania)

Giulia Cacciatore (Università di Catania)

Giuseppe Canzoneri (Università di Catania)

Elisa Conti (Università di Catania)

Milena Giuffrida (Università di Catania)

Miryam Grasso (Università di Catania)

Francesca Prado (Università di Catania)

Emilio M. Sanfilippo (CNR Istituto di Scienze e Tecnologie della Cognizione)

Eliana Vitale (Università di Catania)

Alessandro Zammataro (Università di Catania)

Comunicazione istituzionale: Claudia Cantale (Università di Catania) e Area Per la Comunicazione dell'Università di Catania (ACOM).

Institutional communication: Claudia Cantale (University of Catania) and the Area for Communication of the University of Catania (ACOM)

Supporto tecnico: Rosario Agrò, Area della Terza Missione dell'Università di Catania, per la consulenza e la progettazione grafica dei materiali informativi del convegno.

Technical support: Rosario Agrò, Third Mission Area of the University of Catania, for advice and graphic design of the conference information materials.

Enti organizzatori / Organisers

AIUCD; Università di Catania: Dipartimento di Scienze Umanistiche; CNR Istituto di Scienze e Tecnologie della Cognizione; CINUM: Centro di Informatica Umanistica dell'Università di Catania.

Supporter

CLARIN-IT; Neperia Group; Storage; programma Piaceri 2020-2022, Linea 1; Parmalat-Sole.

Chair di area/ Track chair

Le culture digitali nel Mediterraneo

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Paola Moscati (CNR Istituto di Scienze del Patrimonio Culturale)

Archivi ed edizioni digitali

Christian D'Agata (Università di Catania)

Greta Franzini (Eurac Research)

Analisi computazionale dei testi

Angelo Mario Del Grosso (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Simone Reborra (Università di Verona)

Ontologie e Semantic Web

Marianna Nicolosi Asmundo (Università di Catania)

Francesca Tomasi (Università di Bologna)

Preservazione della memoria e del patrimonio digitale

Fabio Ciraci (Università del Salento)

Anna Maria Marras (Università di Torino)

Lista dei revisori /List of reviewers

Maristella Agosti (Università di Padova), **Stefano Allegrezza** (Università di Bologna), **Chiara Alzetta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Liborio Barbarino** (Università di Catania), **Nicola Barbuti** (Università di Bari Aldo Moro), **Stefano Bazzaco** (Università di Verona), **Benedetta Bessi** (Università Ca' Foscari di Venezia), **Andrea Bolioli** (ricercatore indipendente), **Paolo Bonora** (Università di Bologna), **Federico Boschetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Dominique Brunato** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Flavia Bruni** (Università Gabriele D'Annunzio di Chieti-Pescara), **Marina Buzzoni** (Università Ca' Foscari di Venezia), **Alberto Campagnolo** (Université Catholique de Louvain/KULeuven), **Anna Cappellotto** (Università di Verona), **Emmanuela Carbé** (Università di Siena), **Vittore Casarosa** (CNR Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Università di Pisa), **Fabio Ciotti** (Università di Roma "Tor Vergata"), **Fabio Ciraci** (Università del Salento), **Elisa Conti** (Università di Catania), **Salvatore Cristofaro** (CNR Istituto per il Lessico Intellettuale Europeo e Storia delle Idee), **Christian D'Agata** (Università di Catania), **Elisa D'Argenio** (HUN-REN Hungarian Research Centre for Linguistics), **Mauro De Bari** (Università di Bari Aldo Moro), **Riccardo Del Gratta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Angelo Mario Del Grosso** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Saulo Delle Donne** (Università del Salento), **Giorgio Maria Di Nunzio** (Università di Padova), **Antonio Di Silvestro** (Università di Catania), **Filippo Diara** (Università di Torino), **Giulia Fabbris** (Università Ca' Foscari di Venezia), **Riccardo Fedriga** (Università di Bologna), **Franz Fischer** (Università Ca' Foscari di Venezia), **Greta Franzini** (Eurac Research), **Francesca Frontini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Daniele Fusi** (Stuttgart University & VeDPH – Università Ca' Foscari di Venezia), **Carola Gatto** (Università del Salento), **Lucia Giagnolini** (Università di Bologna), **Emiliano Giovannetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Milena Giuffrida** (Università di Catania), **Edmondo Grassi** (Università San Raffaele di Roma), **Miryam Grasso** (Università di Catania), **Alessandro Iannella** (Università di Cagliari - Università di Pisa – Università di Torino), **Paola Italia** (Università di Bologna), **Maurizio Lana** (Università del Piemonte Orientale), **Pietro Maria Liuzzo** (Bibliotheca Hertziana), **Dominique Longrée** (Université de Liège), **Francesco Mambrini** (Università Cattolica del Sacro Cuore di Milano), **Tiziana Mancinelli** (Istituto Italiano di Studi Germanici), **Anna Maria Marras** (Università di Torino), **Cristina Marras** (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee), **Federico Meschini** (Università della Toscana), **Alessio Miaschi** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Paolo Monella** (Università Sapienza di Roma), **Ouafae Nahli** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Marianna Nicolosi-Asmundo** (Università di Catania), **Giuseppe Palazzolo** (Università di Catania), **Valentina Pasqual** (Università di Bologna), **Gianluca Pavan** (Università di Roma "Tor Vergata"), **Giulia Pedonese** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Jonathan Prag** (University of Oxford Merton College), **Simone Reborra** (Università di Verona), **Giulia Renda** (Università di Bologna), **Roberto Rosselli Del Turco** (Università di Torino), **Enrica Salvatori** (Università di Pisa), **Emilio M. Sanfilippo** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Eva Sassolini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Pietro Sichera** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Daniele Silvi** (Università di Roma "Tor Vergata"), **Elena Spadini** (University of Basel), **Daria Spampinato** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Linda Spinazzè** (Università Ca' Foscari di Venezia), **Rachele Sprugnoli** (Università di Parma), **Francesco Stella** (Università di Siena), **Cecilia Tamagnini** (Università di Bologna), **Timothy Tambassi** (Università Ca' Foscari di Venezia), **Francesca Tomasi** (Università di Bologna), **Marco Venuti** (Università di Catania), **Fabio Vitali** (Università di Bologna).

SOMMARIO

PREFAZIONE	I
<i>Antonio Di Silvestro, Daria Spampinato</i>	
RELAZIONI DEI KEYNOTE SPEAKERS	VI
INFORMATICA UMANISTICA, INFOCRAZIA, E INTELLIGENZE ARTIFICIALI	VII
<i>Giuseppe Savoca</i>	
THE MEDIEVAL MEDITERRANEAN IN...DATA? INTERPRETATION, CONJECTURE AND DIGITAL METHODS	XIII
<i>Tara L. Andrews</i>	
RELAZIONI DEGLI INVITED SPEAKERS	XXI
RICOSTRUZIONE DEL TESTO E BANCHE DATI. LA FILOGIA DIGITALE ALLA PROVA DELL'ESEGESI ANTICA DELLA <i>COMMEDIA</i>	XXII
<i>Vittorio Celotto, Andrea Mazzucchi</i>	
IL MEDITERRANEO: UN MARE DI OPPORTUNITÀ E SFIDE	XXX
<i>Salvatore Capasso</i>	
MEDITERRANEO TRA TESTI E CONTESTI	1
COMBINING GENERATIVE AI AND ARCHAEOLOGY TO BUILD DATA-DRIVEN STORIES	2
<i>Francesca Buscemi, Angelica Lo Duca</i>	
DIGITAL PRESERVATION E SOSTENIBILITÀ AMBIENTALE	9
<i>Adele Gorini</i>	
IL PROGETTO DIGITALE SU ELISA CHIMENTI (LAI-ALEEF): LA PROBLEMATICITÀ DI UN PROFILO MEDITERRANEO TRA RETI E FRONTIERE	15
<i>Ada Desideri, Bianca Vallarano</i>	
MEDITERRANEO VERDE FANTASTICO: UN REPERTORIO DIGITALE ATTRAVERSO LA COLLEZIONE BOTANICA DI ULISSE ALDROVANDI (1522-1605)	21
<i>Sara Obbiso</i>	
PER LA DIGITALIZZAZIONE DEL PATRIMONIO LINGUISTICO E CULTURALE ITALIANO IN EGITTO: I PERIODICI ITALIANI (1892- 1940)	26
<i>Wafaa El Beih</i>	
PIATTAFORME WIKI PER L'INSEGNAMENTO UMANISTICO: SPERIMENTAZIONI IN CORSO NEL LICEO DE COSMI DI PALERMO	32
<i>Antonino Fiorino, Paolo Monella, Francesca Saieva, Antonella Sorci</i>	
THE ORGANIZATION AND MANAGEMENT OF THE MAGIC PROJECT FOR ANCIENT MANUSCRIPTS DIGITIZATION: CONNECTIONS BETWEEN MEDITERRANEAN CULTURES	38
<i>Stefania Conte, Andrea Mazzucchi, Guido Russo, Augusto Tortora, Giorgia Tortora</i>	
TOWARDS A RESEMANTISATION OF THE CONCEPT OF MODELLING IN DIGITAL HUMANITIES	43
<i>Cristina Marras, Arianna Ciula, Øyvind Eide, Patrick Sahle</i>	
VALORIZZARE UN ARCHIVIO 'MEDITERRANEO': STUDI PER UN'EDIZIONE CRITICA DIGITALE DELLE OPERE DI GIOVANNI COMISSO	48
<i>Marco Borrelli</i>	
ARCHIVI E MUSEI DIGITALI PER IL PATRIMONIO CULTURALE	55
A WORKFLOW FOR GLAM METADATA CROSSWALK	56
<i>Arianna Moretti, Ivan Heibi, Silvio Peroni</i>	
DIGITALIZZAZIONE E MODELLAZIONE DELLA <i>DRAMMATURGIA</i> DI LEONE ALLACCI	63
<i>Luca Giovannini, Giorgia Gallucci</i>	

FROM DATA COMPLEXITY TO USER SIMPLICITY: A FRAMEWORK FOR LINKED OPEN DATA RECONCILIATION AND SERENDIPITOUS DISCOVERY	67
<i>Marco Grasso, Giulia Renda, Marilena Daquino</i>	
GIUSEPPE CHIARINI: UN'OPERA INEDITA	73
<i>Elena Almangano, Mirko Castaldi, Eleonora De Longis, Daniele Pasqualetti</i>	
HERITRACE: TRACING EVOLUTION AND BRIDGING DATA FOR STREAMLINED CURATORIAL WORK IN THE GLAM DOMAIN	78
<i>Arcangelo Massari, Silvio Peroni</i>	
LIBRI E BIBLIOTECHE TRA MUSEABILITÀ E MUSEALIZZAZIONE DIGITALE: SOGNO O REALTÀ?	84
<i>Nicola Barbuti, Mauro De Bari</i>	
LISTENING2PAINTING: AN AUDIO AUGMENTED REALITY APPROACH FOR ARTS	89
<i>Nicola Orio, Daniel Zilio, Andrea Micheletti</i>	
LUOGHI COMUNI: METODI E STRATEGIE DI SVILUPPO SOFTWARE IN AMBITO GLAM, DALLE VOCI DI AUTORITÀ ALL'ESPLORAZIONE CARTOGRAFICA	92
<i>Herbert Natta, Gianluca Rossi, Roberta Maggi</i>	
MESSAGGISTICA ISTANTANEA E ARCHIVI DIGITALI. QUALI SOLUZIONI? BEST PRACTICES E CONSIDERAZIONI DAL CONTESTO INTERNAZIONALE	98
<i>Alessia Del Bianco</i>	
NUOVE INTERAZIONI CON COLLEZIONI DIGITALI: L'“ARCHIVIO DIGITALE DEL CAPITOLO DI LATERZA”	104
<i>Stefania Riso, Nicola Barbuti</i>	
PRESERVARE E VALORIZZARE LA MEMORIA DI ARCHIVI STORICI DI EX-OSPEDALI PSICHIATRICI	110
<i>Grazia Serratore</i>	
PRESERVAZIONE DEL PATRIMONIO CULTURALE E CLOUD COMPUTING: CARATTERISTICHE E CRITICITÀ	117
<i>Manuela Grillo</i>	
PRESERVING CULINARY TRADITIONS. A CROWDSOURCED DIGITAL COLLECTION OF COOKBOOKS	122
<i>Giulia Renda, Giulia Manganelli, Mila Fumini, Marilena Daquino</i>	
SERIOUS GAMES E GAMIFICATION: A CHE PUNTO SONO LE ISTITUZIONI CULTURALI ITALIANE?	128
<i>Vincenzo Colaprice</i>	
THE TREE OF PHILOSOPHERS: DESIGN AND IMPLEMENTATION OF A DIGITAL RESOURCE FOR THE HISTORY OF ACADEMIC PHILOSOPHY	133
<i>Guido Bonino, Nicola Ruschena</i>	
THINKING OUTSIDE THE BLACK BOX: INSIGHTS FROM A DIGITAL EXHIBITION IN THE HUMANITIES	138
<i>Sebastian Barzaghi, Alice Bordignon, Bianca Gualandi, Silvio Peroni</i>	
EDIZIONI SCIENTIFICHE DIGITALI	143
IL PROGETTO CORR<SI>CA: EDIZIONE DIGITALE DELLA CORRISPONDENZA CANIONI	144
<i>Anna Giaufret, Beatrice Dal Bo, Elena Margherita Vercelli, Laura Bonanno</i>	
IL PROGETTO “MAXIMHUM”: ITALIA UMANISTICA E MOSCOVIA CINQUECENTESCA DIALOGANO IN DIGITALE	152
<i>Francesca Romoli, Letizia Ricci, Angelo Mario Del Grosso</i>	
L'ARCHIVIO DI GIUSEPPE FAVA: CONSERVAZIONE E VALORIZZAZIONE ATTRAVERSO IL DIGITALE	159
<i>Giuseppe Davide Di Mauro, Marzia D'Amico</i>	
L'ARCHIVIO DIGITALE DI UNA CASA EDITRICE: L'ESEMPIO DEL SAGGIATORE E DELLA SUA PRIMA PUBBLICAZIONE	165
<i>Giada Di Pino</i>	
MARCARE LA POESIA DEL NOVECENTO: UNO STUDIO PER <i>OSSI DI SEPPIA</i>	171
<i>Chiara Cauzzi, Martina Corti, Anna Guadagnoli, Maria Grazia Schiaroli</i>	
METASCRIP: A FRAMEWORK PROPOSAL FOR SCREENPLAY ENCODING	175
<i>Erica Andreose, Giorgia Crosilla, Leonardo Zilli</i>	
OPENDATA: OPENGADDA	181
<i>Eleonora Pasquale, Martina Pensalfini</i>	
PAUL KLEE, <i>TUNISREISE</i> E <i>BILDNERISCHE FORMLEHRE</i> : UN CASO STUDIO DI DISCEPT (DIGITAL SCHOLARLY EDITIONS PLATFORM AND ALIGNED TRANSLATIONS)	186
<i>Hansmichael Hohenegger, Tiziana Mancinelli, Fabio Ciotti, Federico Boschetti, Angelo Mario Del Grosso, Eleonora De Longis</i>	

PAVES-E: PER UNA HYPEREDIZIONE DELL'OPERA DI CESARE PAVESE	191
<i>Christian D'Agata, Angelo Mario Del Grosso, Laura Nay, Giuseppe Palazzolo, Antonio Sichera, Daria Spampinato</i>	
PER UN'EDIZIONE DIGITALE DI <i>SE QUESTO È UN UOMO</i>	197
<i>David Tagliacozzo</i>	
PER UN'EDIZIONE SCIENTIFICA DIGITALE DELLO <i>SPECULUM GUY OF WARWICK</i>	204
<i>Omar Khalaf, Sibilla Siano</i>	
PER UNA LETTURA ANTROPOLOGICA DI VERGA: TRA CODIFICA E GEOREFERENZIAZIONE	210
<i>Giovanna Zisa</i>	
PROGETTO DI EDIZIONE GENETICA DIGITALE DEL <i>CANZONIERE</i> MANOSCRITTO DI U. SABA (1919-20)	215
<i>Marina Buzzoni, Davide Cucurnia, Cristina Fenu, Roberto Rosselli Del Turco, Giulia Tancredi</i>	
RAPPRESENTARE LA STORIA SACRA: UN'IMPRESA IERI, UNA SFIDA OGGI. PROPOSTA DI EDIZIONE SCIENTIFICA DIGITALE DEL "COMPENDIUM" DI PIETRO DI POITIERS	221
<i>Franz Fischer, Agnese Macchiarelli</i>	
TOWARDS AN INTEGRATED DIGITAL EDITION OF THE <i>LEGES LANGOBARDORUM</i>	226
<i>Marina Buzzoni, Roberto Rosselli Del Turco</i>	
UN <i>CORPUS</i> ONLINE DELLA LETTERATURA SECONDARIA (1872- 1890) DEL VERISMO ITALIANO	232
<i>Denise Bruno, Giuseppe Canzoneri, Antonio Di Silvestro, Daria Spampinato, Alessandro Zammataro</i>	
UN PROGETTO DI EDIZIONE DIGITALE <i>IMAGE-BASED</i> DELLE <i>MERAVIGLIE D'ORIENTE</i> NEL MS COTTON VITELLIUS A.XV	240
<i>Andreea Mihaela Toma</i>	
UNA PROPOSTA DI CODIFICA IN XML/MEI PER TESTI MUSICALI AUTOGRAFI DI VINCENZO BELLINI	246
<i>Laura Mazzagufò</i>	
VERISMO DIGITALE. PER UN'EDIZIONE DIGITALE COMMENTATA DELLE OPERE DI VERGA, CAPUANA, DE ROBERTO	252
<i>Liborio Pietro Barbarino, Elisa Conti, Christian D'Agata, Miryam Grasso, Ninna Maria Lucia Martines, Eliana Vitale</i>	
VERSO L'HYPEREDIZIONE. LO SVILUPPO DI PIRANDELLO NAZIONALE TRA DIDATTICA E RICERCA	260
<i>Milena Giuffrida, Christian D'Agata, Giulia Cacciatore, Fabrizio Lo Presti</i>	
VOCI DALL'INFERNO: DANTE PER DIRE IL LAGER - DIGITALIZZARE E STUDIARE LE TESTIMONIANZE	267
<i>Angelo Mario Del Grosso, Marina Riccucci, Elvira Mercatanti</i>	
DIZIONARI E DIGITALIZZAZIONE DI BANCHE DATI	274
IL VIVER (VOCABOLARIO DELL'ITALIANO VERISTA)	275
<i>Gabriella Alfieri, Marco Biffi, Stephanie Cerruto, Giovanni Salucci</i>	
L'INFORMATIZZAZIONE DEL GDLI: RISULTATI, PROSPETTIVE, SFIDE FUTURE	281
<i>Eva Sassolini, Sebastiana Cucurullo, Marco Biffi</i>	
LA DIGITALIZZAZIONE DEL DIZIONARIO LATINO LANA 1978	287
<i>Francesca Michelone</i>	
LEXICAD: PIATTAFORMA LESSICOGRAFICA DIGITALE PER L'ITALIANO DELLE ORIGINI	293
<i>Salvatore Arcidiacono, Antonella Zammataro</i>	
STRATEGIE PER LA CREAZIONE E LA CONDIVISIONE DI UNA COLLEZIONE DIGITALE DI TESTI GRECO-LATINI	298
<i>Vincenzo Ortoleva, Maria Rosaria Petringa, Salvatore Cammisuli, Mariarosaria Villareale</i>	
THE CORR<SI>CA PROJECT: ENHANCING AND "QUERYING" THE CANIONI FAMILY CORRESPONDENCE	303
<i>Tiziana Pasciuto, Selenia Anastasi, Daniele Zolezzi, Simonetta Acacia, Giada D'Ippolito, Chiara Storace, Maria Tolaini</i>	
XML/TEI E DIZIONARI <i>BORN-DIGITAL</i> : UNA PROPOSTA PER LE RISORSE LESSICOGRAFICHE DELLA RETE LEXICAD/PLUTO	310
<i>Giuseppe Leonardo Zappalà</i>	
ANALISI COMPUTAZIONALE, INTELLIGENZA ARTIFICIALE E LINGUISTICA	318
ANALISI COMPUTAZIONALE DEI REPORT DI SOSTENIBILITÀ: LA VAGHEZZA COME STRATEGIA DI GREENWASHING	319
<i>Erica Cutuli</i>	
ANALISI STILOMETRICA APPLICATA ALLE CAPACITÀ EMULATIVE DI GPT-4	325
<i>Marco De Cristofaro, Mariangela Giglio</i>	
C'È UN TESTO IN QUESTA CHAT? INTELLIGENZA ARTIFICIALE E COOPERAZIONE INTERPRETATIVA	332
<i>Daniele Silvi</i>	

GENERE E GEOPOLITICA NELLE DISCIPLINE UMANISTICHE DIGITALI IN ITALIA. LE CONFERENZE AIUCD (2012-2023)	
<i>Selenia Anastasi</i>	336
GLI LLM COME LETTORI MODELLO ARTIFICIALI	342
<i>Fabio Ciotti</i>	
I SIMILI SI ATTRAGGONO. LA VALUTAZIONE LETTERARIA SULLE PIATTAFORME DI DIGITAL SOCIAL READING	348
<i>Gabriele Vezzani, Simone Rebora, Massimo Salgaro</i>	
IL <i>DISTANT READING</i> È L'ORNITORINCO	354
<i>Pietro Mazzarisi</i>	
L'IMPIEGO DELL'INTELLIGENZA ARTIFICIALE PER LA RICOSTITUZIONE DELLE AGGREGAZIONI ARCHIVISTICHE E L'ARRICCHIMENTO DEI METADATI NEGLI ARCHIVI DIGITALI	361
<i>Stefano Allegrezza</i>	
MACCHINE PER LEGGERE: LA TEXT ANALYSIS COME STRUMENTO PER IMPARARE A LEGGERE I CLASSICI DELLA NARRATIVA... E AD AMARLI	367
<i>Fabio Ciotti</i>	
PRESERVARE LA DIVERSITÀ NELL'ERA DELL'INTELLIGENZA ARTIFICIALE: IL DILEMMA ETICO DI BIAS E DISCRIMINAZIONI NEGLI ALGORITMI	371
<i>Gianluca Pavani</i>	
<i>QUI PRO QUO?</i> DATI TESTUALI E STRUMENTI PER LA RISOLUZIONE DI COREFERENZE IN LATINO	377
<i>Roberta Grazia Leotta, Eleonora Delfino</i>	
STRUMENTI DIGITALI PER LA TRASCRIZIONE E LA LEMMATIZZAZIONE DI TESTI IN ITALIANO ANTICO	382
<i>Emiliano Degl'Innocenti, Alessia Spadi, Federica Spinelli, Lucia Francalanci, Michela Perino, Irene Falini, Francesco Coradeschi, Francesco Pinna</i>	
TESTI ALLOGRAFICI: CONTATTI TRA LINGUE E SCRITTURE DEL MEDITERRANEO	388
<i>Antonio Pagliara, Federico Boschetti, Daniele Baglioni</i>	
THE DARK MIRROR OF ARTIFICIAL INTELLIGENCE: HOW AI AFFECTS CLIMATE CHANGE	393
<i>Mauro De Bari</i>	
UN SISTEMA DI CLASSIFICAZIONE AUTOMATICA DI IMMAGINI RELATIVE A MATERIALI LIBRARI ANTICHI E MODERNI	398
<i>Nicola Barbuti, Tommaso Caldarola</i>	
UNCOVERING THE SPREAD OF LEXICAL INNOVATION IN ITALIAN TWEETS	405
<i>Greta Franzini, Paolo Brasolin, Stefania Spina</i>	
ORGANIZZAZIONE DELLA CONOSCENZA CON SEMANTIC WEB	410
AFFINARE IL CONTESTO: ESTRAZIONE DI INFORMAZIONI STRUTTURATE PER L'ARRICCHIMENTO DEI CONTESTI ARCHIVISTICI	411
<i>Lucia Giagnolini, Paolo Bonora, Francesca Tomasi</i>	
CLEF 2.0. SOLUZIONI PER LA CATALOGAZIONE NATIVA LINKED DATA DEL PATRIMONIO DIGITALE CULTURALE ITALIANO	417
<i>Marilena Daquino, Laurent Fintoni, Sebastiano Giacomini, Francesca Tomasi</i>	
L'ONTOLOGIA BIGRAFO: VERSO UN MODELLO SEMANTICO PER L'OPERA DI FRANCO FORTINI	423
<i>Laura Antonietti, Emilio M. Sanfilippo, Emmanuela Carbé</i>	
LOST IN DATIFICATION? THE JOURNEY OF DATA FROM THE PRIMARY SOURCE TO THE FINAL INTERPRETATION	429
<i>Enrica Bruno, Sofia Baroncini, Francesca Tomasi</i>	
ORBIS DIOECESIVM. CREATING AUTHORITY DATA ON THE LEGAL-HISTORICAL CHANGES OF CATHOLIC DIOCESES	435
<i>Benedetta Albani, Rowan Dorin, Yohan Park</i>	
PER L'INTEROPERABILITÀ E LA SOSTENIBILITÀ DELLE RISORSE DIGITALI DANTESCHE: IL PROGETTO LIDA	441
<i>Cesare Concordia, Gaia Tomazzoli, Nicola Aloia, Carlo Meghini, Luca Trupiano</i>	
PER UN'ANALISI DEI PERSONAGGI TRA LETTERATURA, FILOSOFIA E ONTOLOGIA APPLICATA	448
<i>Emilio M. Sanfilippo, Gaia Tomazzoli, Michele Paolini Paoletti, Jansan Favazzo, Roberta Ferrario</i>	
REPRESENTING TEXTS AS LOD: A SYSTEMATIC LITERATURE REVIEW	455
<i>Michela Bandini, Valeria Quochi</i>	
STRUCTURING AUTHENTICITY ASSESSMENTS ON HISTORICAL DOCUMENTS USING LLMs	463
<i>Andrea Schimmenti, Valentina Pasqual, Francesca Tomasi, Fabio Vitali, Marieke van Erp</i>	
VALORIZZAZIONE DI REGISTRAZIONI STORICHE DI CANTO LIRICO NEL WEB SEMANTICO	469
<i>Marcello Ranieri, Angelo Pompilio</i>	

PUBLIC HISTORY E ARCHEOLOGIA DIGITALE	473
ARCHEOLOGIA E RILIEVO 3D: UNA RIFLESSIONE SULLE METODOLOGIE. DUE CASI STUDIO DI AREA MEDITERRANEA	474
<i>Graziana D'Agostino, Pietro Maria Militello</i>	
DAL CATASTO BORBONICO ALLA GENOMICA. PIATTAFORME DIGITALI E INTERDISCIPLINARITÀ: IL PROGETTO "WE ARE WHAT THEY WERE" DI RIPOSTO	481
<i>Salvatore Spina</i>	
DRAWING HISTORY FROM THE CODICE PELAVICINO DOCUMENTS: GRAPH VISUALIZATION FOR HUMAN RESEARCHERS	489
<i>Natthida Wiwatwicha</i>	
GESTIONE INFORMATICA DELLA DOCUMENTAZIONE ARCHEOLOGICA "MINORE". METODOLOGIE E APPLICAZIONI NELL'AMBITO DEL PROGETTO STORAGE	494
<i>Marianna Figuera, Erica Platania</i>	
GESTIRE L'IMMATERIALE. CONTESTI SENSORIALI A SERVIZIO DEL PATRIMONIO ARCHEOLOGICO	501
<i>Serena D'Amico</i>	
GODSCAPES: MODELING SECOND MILLENNIUM BCE POLYTHEISMS IN THE EASTERN MEDITERRANEAN THANKS TO THE STORAGE PROJECT	507
<i>Nicola Laneri, Marianna Nicolosi-Asmundo, Daniele Francesco Santamaria, Chiara Pappalardo</i>	
IL PROGETTO STORAGE: DAI DATI AL WEB	512
<i>Simone Faro, Pietro Maria Militello, Marianna Nicolosi-Asmundo</i>	
LA MODELLAZIONE 3D AL SERVIZIO DELL'ARCHEOLOGIA: NUOVE PROSPETTIVE PER L'APPLICAZIONE AD EDIFICI MULTIPIANO DI ETÀ PROTOSTORICA	517
<i>Dario Puglisi, Marco Chiricallo</i>	
MLS CON SENSORE LIDAR APPLE PER LO SCAVO ARCHEOLOGICO: APPLICAZIONI PRATICHE	523
<i>Luigi M. Calì, Antonello Fino, Gian Michele Gerogiannis</i>	
ODONIMI D'ITALIA E DIGITAL PUBLIC HISTORY: LE PROBLEMATICHE DI UNA SCHEDATURA PARTECIPATA	528
<i>Enrica Salvatori, Vittore Casarosa, Riccardo Chiari</i>	
OPENSTREETMAP: UNO STRUMENTO E UNO SPAZIO PER LA DIGITAL PUBLIC HISTORY?	535
<i>Camilla Zucchi</i>	
PRESERVARE LA MEMORIA: IL PROGETTO STORAGE E L'ARCHIVIO DELL'EX ISTITUTO DI ARCHEOLOGIA	539
<i>Giovanni Fragalà, Pietro Maria Militello</i>	
UN ATLANTE DIGITALE PER LA STORIA MARITTIMA DEL REGNO DI SARDEGNA	545
<i>Giampaolo Salice</i>	
UN FUTURO PER LA MEMORIA. STRUMENTI, MODELLI E SINERGIE PER L'INTEGRAZIONE DEI DATI NEL PORTALE DELLE FONTI PER LA STORIA DELLA REPUBBLICA ITALIANA	549
<i>Michela Tardella, Roberta Maggi, Giorgia Lodi, Riccardo Albertoni, Herbert Natta, Gianluca Rossi, Tiziana Pasciuto, Paola Ciandrini, Luca Sinopoli, Maria Teresa Artese, Isabella Gagliardi, Eleonora Lattanzi, Sara Ventroni, Elisa Tizzoni, Alessandro Russo, Margherita Porena</i>	
INFRASTRUTTURE PER LA RICERCA UMANISTICA	555
COPHI EDITOR: FROM GREEKSCHOOLS TO THE PROJECTS MULTIVERSE	556
<i>Simone Zenzaro</i>	
DIGITAL HUMANITIES AND HERITAGE SCIENCE: MOVING FROM LANDSCAPING TO A DYNAMIC RESEARCH OBSERVATORY IN AN OPEN SCIENCE CLOUD	559
<i>Roberta Bianca Luzietti, Alessia Spadi, Nicola Giampietro, Giacomo Mancuso, Alessandra Caravale, Antonio D'Eredità, Marta Caradonna, Paola Moscati, Valeria Quochi, Monica Monachini, Emiliano Degl'Innocenti</i>	
FUNZIONI E SOSTENIBILITÀ DI UNA PIATTAFORMA DIGITALE PER LE LINGUE ARCAICHE	566
<i>Michele Mallia, Riccardo Del Gratta, Valeria Quochi</i>	
INFRASTRUTTURE DI RICERCA COME STRUMENTI DI "INTERCULTURALITÀ DIGITALE"	572
<i>Salvatore Cristofaro, Vittoria Fabiani, Cristina Marras, Enrico Pasini, Pietro Sichera, Mingyang Yu</i>	
MATERIALI DIDATTICI COME OGGETTI DIGITALI FAIR: UNA METODOLOGIA CONDIVISA PER LA FORMAZIONE IN H2IOSC	577
<i>Giulia Pedonese, Francesca Frontini, Roberta Ottaviani, Federico Boschetti, Alessia Spadi, Lucia Francalanci, Alessia Scognamiglio, Pietro Restaneo, Antonina Chaban, Jana Striova, Laura Benassi</i>	
RETHINKING SCHOLARLY DIGITAL OBJECTS AS CULTURAL HERITAGE: THE KNOT PROJECT	582
<i>Laurent Fintoni</i>	

THE ATLAS: A KNOWLEDGE GRAPH OF DIGITAL SCHOLARLY RESEARCH ON ITALIAN CULTURAL HERITAGE	588
<i>Marilena Daquino, Alessia Bardi, Marina Buzzoni, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, Francesca Tomasi, Roberto Rosselli Del Turco</i>	
INDICE DEGLI AUTORI	593

Uncovering the spread of lexical innovation in Italian tweets

Greta Franzini¹, Paolo Brasolin², Stefania Spina³

¹ Eurac Research, Institute for Applied Linguistics, Italy - greta.franzini@eurac.edu

² Eurac Research, Institute for Applied Linguistics, Italy - paolo.brasolin@gmail.com

³ Università per Stranieri di Perugia, Italy - stefania.spina@unistrapg.it

ABSTRACT

This paper outlines ongoing research into lexical innovation in contemporary Italian in the context of social media. To date, the study has used a dataset of 5.32M timestamped and geotagged tweets extracted from the 2022 Italian timeline, yielding 720 emerging word forms. Here, we describe the reproducible pipeline developed to extract candidate neologisms from our dataset and introduce a custom tool to visualise the emergence and spread of candidate neologisms in the time-period under investigation.

KEYWORDS

Twitter; Italian; social media; lexical innovation; language change.

1. INTRODUCTION

The evolution of language is often propelled by lexical innovation [7, 14]. This phenomenon involves the generation of new words¹ and their assimilation into established lexical frameworks [11]. The process typically unfolds through distinct stages, starting with limited usage in specific contexts, progressing to wider adoption, and, in certain instances, formal inclusion in dictionaries [8, 12].

The creation of novel words is influenced by diverse linguistic processes such as derivation, compounding, transcategorisation, blending, semantic shifts, and borrowing from other languages. Social media platforms provide an avenue to study the emergence of new lexical forms in everyday discourse, offering real-time insights from a varied demographic, complete with geotagged text to examine geographical variations [10].

Our research into the language change mechanisms of contemporary Italian has so far examined Twitter interactions from the 2022 Italian timeline [5]. First results have yielded a list of 720 emerging candidates classified into 14 categories of lexical creation. Here, we describe the reproducible pipeline we developed to identify new word forms in tweets and introduce a bespoke tool to dynamically visualise the emergence and spread of candidate neologisms in our dataset.

2. RELATED WORK

A key focus of previous research on lexical innovation in contemporary Italian [1, 3, 4, 16] has been the categorisation of the processes that lead to lexical creation. Traditionally, these include borrowing from other languages, forming new words from existing ones, changing grammatical categories and semantic shifting [22]. The *Osservatorio neologico della lingua italiana* (ONLI) [2] has meticulously monitored the appearance of new Italian words in newspapers, recording, to date, 2,986 new forms.

In recent years, social media has gained traction as a means of tracking new words in informal contexts [17, 18, 21]. By offering an unprecedented volume of conversational data from diverse speakers, social media facilitates robust evaluations of lexical creativity while enhancing the exploration of language variation and change [15, 18]. This approach has prompted studies to delve into the initial, less-documented phases of lexical innovation occurring after word creation [9, 10, 13] but before institutionalisation in dictionaries.

3. METHODOLOGY

Our research seeks to answer two research questions, namely *are Twitter conversations a reliable source to trace lexical innovation? And what are the linguistic processes leading to the creation of emerging words in Italian Twitter?* To this end, our work began with the sampling of 5.32M timestamped and geotagged tweets from the Italian timeline of 2022 using Twitter's advanced search query language [6]. These tweets were posted by 153,000 unique users for a total 71.5M tokens or 564M characters.

Tweets are complex structures. They may include geolocation information in the form of a latitude/longitude pair or a place. The latter is defined as an administrative region or a point of interest, identified by an ID, a country code, a

¹ The terms “word” and “form” are used interchangeably.

geographical bounding box and additional metadata. In our dataset, 99.43% of tweets are associated with a place, 0.04% come with a latitude/longitude pair only, and 0.53% bear no geolocation information whatsoever². Consequently, we chose to focus on tweets associated with a place as they cover almost the entire dataset. Of these, 91.77% contain places associated with an IT country code: we identified 34.8K unique places and computed the centroid of their bounding box, matching it with governmental data³ to generate choropleth maps for geographical analysis. The remaining data includes 8.16% tweets related to places with other country codes and 0.07% tweets representing the entirety of Italy. As well as geolocation information, tweets contain an ID, a user ID, a timestamp, the full text, a list of “entities” and other metadata. An “entity” is a character range in the full text labelled with a type (*url*, *user mention*, *hashtag*, *symbol* or *media*) and other metadata. Recognising the value of entity metadata for the downstream tokenisation process, we incorporated them into the full text as delimiter markers by selecting a unique pair of Unicode code points for each entity type from the Private Use Area in the Basic Multilingual Plane.

We tokenised our corpus with the spaCy v3.6.1 Italian tokeniser. Tweets present a number of challenges for a standard tokeniser due to the widespread use of Unicode, variable casing, white-spacing and punctuation marks. To address these issues, we took several steps, including replacing emojis with spaces, converting the text to lowercase, trimming sequences of white-space and extending the tokeniser’s infix matcher to identify sequences of commonly abused punctuation marks. Finally, our aforementioned entity annotation allowed us to wrap delimited regions in the text with spaces to help the tokeniser find their beginning; additionally, we defined a custom token matcher to catch any sequence bounded by our delimiter character pairs. Thanks to these adjustments, the tokeniser produced a minimal number of spurious tokens. Next, we filtered the output, eliminating tokens consisting of pure spaces, punctuation, numbers, and broken or non-existent handles (i.e., tokens beginning with @ but not marked as entities). We retained only hashtag entities. The process resulted in the extraction of 71.5M tokens (926K types).

To identify candidate neologisms, we employed two different strategies. The first is that adopted by [9, 10], which involves computing a measure of the degree to which the use of a token increases monotonically over time, and excluding tokens falling below a given positive threshold. The chosen measure, denoted as ρ_o , is the Spearman rank correlation coefficient between the daily occurrences of a token (normalised by the daily total token count) and the day number. We also computed the same measure on the daily users count ρ_u and allowed negative values because we noticed plausible patterns of new word use among them (e.g., an early peak followed by a lower plateau). Selecting forms matching any condition among $\rho_o > 0.2$, $\rho_o < -0.2$, $\rho_u > 0.2$ or $\rho_u < -0.2$, we obtained 6,737 candidate neologisms, which we defined as subset \mathcal{A} . In measuring monotonicity, this method might however still exclude other plausible patterns of new word use. We therefore employed an alternative, complementary strategy aimed at excluding usage patterns that we *do not* expect from emerging forms, such as accidental and sporadic phenomena (e.g., typos, inside jokes, etc.), excluded by setting lower bounds to the count of occurrences ($O > 9$) and unique users ($U > 9$); forms in use from the past or ending prematurely, excluded by setting lower bounds to the days of first ($A > 7$) and last ($Z > 351$) occurrence; and short-lived forms, excluded by establishing a lower bound to the usage lapse ($Z - A > 28$). The alternative subset, defined as \mathcal{B} , yielded 21,132 candidate neologisms (979 overlapping with \mathcal{A}). Combined, subsets \mathcal{A} and \mathcal{B} count 26,890 candidate neologisms (2.90% of the total extracted forms). From these, we automatically removed 15,366 using a lexicon of 514K Italian forms [19] and were thus left with 11,524 candidates between hashtags (3,391) and non-hashtags (8,133) for manual annotation.

The manual annotation process was conducted by two authors of the present paper. We used AntConc to look up the forms’ context and three online dictionaries for their attestation⁴, categorising candidates as either “innovative” or “non-innovative” and resolving sporadic disagreements through negotiation. Next, we grouped innovative forms into one or more categories of lexical creation based on an adjusted ONLI typology scheme.

Finally, we produced choropleth maps and developed a custom tool to visualise the distribution of our innovative forms across Italy.

4. RESULTS AND DISCUSSION

Methodologically, our approach avoids Grieve’s inherent bias towards monotonic patterns of emergence (by, for instance, accounting for forms that emerge early in the dataset but settle on a plateau lower than the initial peak), is conceptually simpler and computationally more efficient having benchmarked it to be 50x faster.

² This is possible because Twitter data can be redacted.

³ <https://www.istat.it/it/archivio/222527>; <https://github.com/openpolis/geojson-italy>

⁴ <https://www.garzantilinguistica.it/>; <https://slengo.it/>; <https://www.treccani.it/vocabolario/>

The annotation of non-hashtag and hashtag forms differed with respect to innovation criteria and ONLI classification. Starting with non-hashtags forms, we excluded attestations, typographical errors caused by key proximity, popular terms (e.g., *bimbominchia*), established loanwords (e.g., *foliage*, *sponsorship*), adapted loanwords (e.g., *followo*, *crashare*), infrequently used foreign words (e.g., *smoothie*, *veggie*), acronyms (e.g., *PTSD*), regionalisms and regional variants (e.g., *annassero*, *ciolla*), gender-inclusive graphic variants (e.g., *cittadinə*), and nicknames (e.g., *pupone* for footballer Francesco Totti). The annotation resulted in 347 emerging forms, grouped into 14 categories of lexical creation. The most prolific categories are: orthographic variation used for emphasis (e.g., *pikkolo*), fun and sarcasm (e.g., *scienzah*), to shorten existing words or hide online conversation (e.g., *f4scist4*); univerbation (e.g., *stemmerde*); suffixation, used for intensification (e.g., *adorissimo*), pejorative connotation (e.g., *cinesata*) or augmentation (e.g., *soggettone*); loanwords, mostly borrowed from English (e.g., *reminder*); and portmanteau forms or blends, typically serving to amuse (e.g., *assurdistan*, *lettamaio*). Of these 347 forms, 22 are now part of the Zingarelli 2024 Italian monolingual dictionary, a 2023 publication superseding the previous edition with 250 new words and 750 new multi-word forms⁵. These 22 forms are obtained through suffixation, adapted and direct borrowing, prefixation, deonymic derivation, transcategorisation, and creation of portmanteau forms. Interestingly, forms resulting from orthographic variation, a common source of lexical innovation on social media, do not seem to make it into the dictionary, suggesting that spelling changes are not a strong criterion for lexicographic acceptance. Grammatically, most of our (now) institutionalised forms are nouns (12), with a few serving as both nouns and adjectives (4), followed by verbs (3) and adjectives (3). This suggests that, in the context of Twitter, nouns derived through suffixation are the most likely candidates for inclusion in dictionaries. As for hashtag annotation, to address biases introduced by forced univerbation and English dominance in detecting patterns of lexical creation, our analysis follows both objective and subjective criteria. We narrowed our selection by filtering out hashtags used by nine or fewer users, hashtags containing proper names (including but not limited to people, places, organisations, brands, sports teams, events, festivities, videogames, music bands, concerts, films, TV shows and political stances/movements); (combinations of) years, days of the week, times, and numbers; short-lived hashtags relating to a specific incident or time interval; univerbated hashtags with little to no probability of inclusion in lexical resources. The remaining hashtags were then categorised into single and univerbated forms. While the annotation of the former mirrored that of non-hashtag forms, in the annotation of univerbated hashtags (e.g., *#andratuttobene*, *#booklover*) we only considered those that intuitively seem likely to establish themselves as new non-hashtag forms in Italian social media communication or be acknowledged in authoritative lexical resources (e.g., *#avantitutta*, *#oldschool*). Of the emerging hashtags identified, 75% are loanwords.

We visualise forms belonging to subset \mathcal{A} as choropleth maps and forms subset \mathcal{A} and \mathcal{B} using a custom tool. The choropleth maps, shown in Figure 1, display the regional word instances per million tokens of forms from four different ONLI categories. The map of *gomblotto* shows widespread orthographic variation in almost all regions, particularly in Lombardy. Univerbated forms are thinly spread, with regional peaks like *miracomando*, which is particularly popular in Lombardy. Loanwords like *flexo* are prominent in the west but absent in the lower east, while *fattoni* is used in northern regions but is absent from the southern-east part of the country and the islands. Our custom tool, shown in Figure 2, generates a more accurate and interactive alternative visualisation from latitude/longitude pairs and timestamps for a geo-temporal analysis of the emerging forms. The tool currently takes a six-column CSV data file as input (*timestamp*, *user_id*, *tweet_id*, *latitude*, *longitude*, *word*) and allows users to sort one or multiple forms by time-period and/or geographical area.

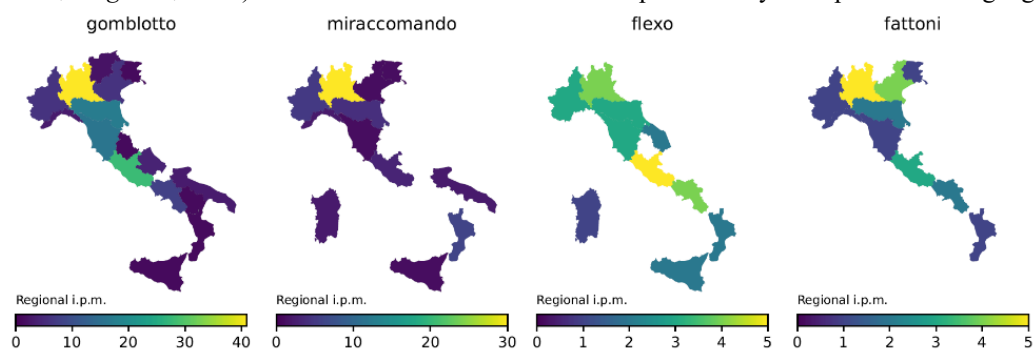


Figure 1. Choropleth maps of Italy showing regional instances per million tokens of *gomblotto*, *miracomando*, *flexo* and *fattoni* (forms taken from subset \mathcal{A})

⁵ <https://www.zanichelli.it/ricerca/prodotti/lo-zingarelli-2024>. This dictionary is generally regarded as the most up-to-date neologism lexicon for Italian.

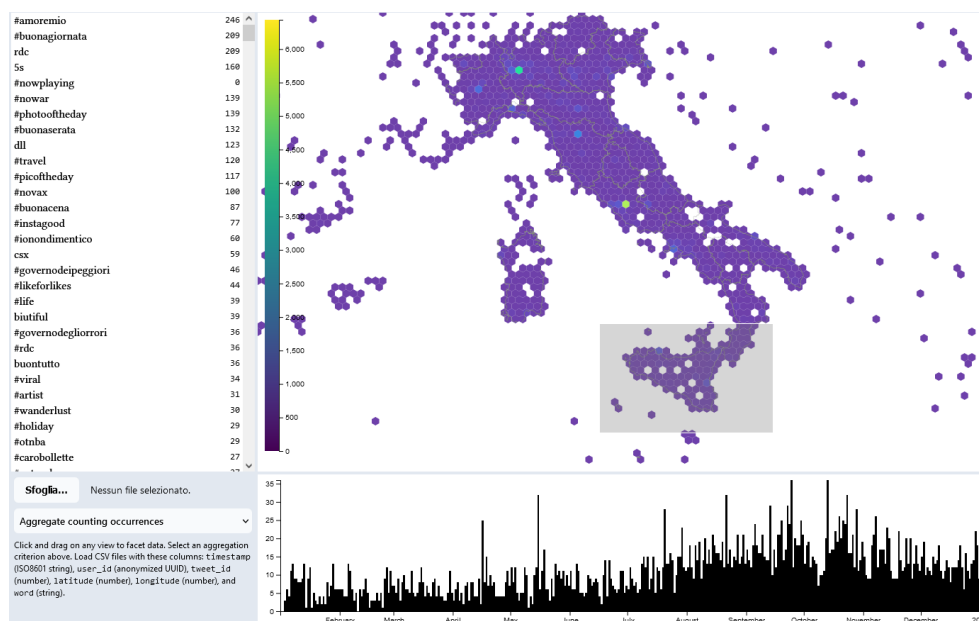


Figure 2. Screenshot of Frustum, our custom tool, developed to interactively visualise the geo-temporal behaviour of emerging forms in our dataset (forms taken from subset \mathcal{A} and \mathcal{B} combined)

5. CONCLUSION

This study investigates lexical innovation in Italian Twitter, revealing that the emergence of new words seems to be driven more by creativity, entertainment and a sense of belonging to a community rather than a need for novel terms to describe new concepts, situations or events. The 720 identified forms primarily serve functions related to irony, intensification and emphasis, and some coined expressions may extend beyond spoken discourse to online communication and media use. Overall, this study shows that Twitter and social media in general can be a reliable source for the study of lexical innovation, as they provide large amounts of data produced by ordinary speakers in their written and informal interactions. These data can effectively complement those traditionally used in the study of lexical innovation, such as data taken from the press. Furthermore, the study shows that, alongside traditional mechanisms of lexical innovation, Twitter interactions reveal a widespread use of spelling variation as a means of coinage of emerging forms, for it is capable of conveying different nuances of meaning.

While our one-year timeframe captures rapid linguistic phenomena on Twitter, it likely misses slower-spreading forms and, with them, other patterns of lexical creation. Follow-up work might, therefore, extend the analysis to additional timelines (and alternative micro-blogging platforms, given Twitter’s recent developments) focussing on the proliferation dynamics and potential institutionalisation of these candidate neologisms.

Our study contributes to the state of the art in the field with the largest study on Italian yet, an annotated dataset, a computationally-lighter and reproducible pipeline to automatically process micro-blogging posts for candidate neologisms, and a user-friendly visualisation tool to interactively browse geolocated data. The annotated datasets, the maps and the code resulting from this study are all available at <https://github.com/breviloquia-italica>.

REFERENCES

- [1] Adamo, Giovanni, and Valeria Della Valle, eds. ‘Neologismi quotidiani. Un dizionario a cavallo del Millennio’. In *Lessico intellettuale europeo*, Vol. 95. Olschki, 2003.
- [2] Adamo, Giovanni, and Valeria Dalle Valle. ‘Osservatorio neologico della lingua italiana. Lessico e parole Nuove ‘italiano’. In *I*. Roma: ILIESI-CNR, 2019. <http://www.iliesi.cnr.it/scheda.php?id=253&cl=1/TS>.
- [3] Adamo, Giovanni, and Valeria Della Valle, eds. ‘Innovazione lessicale e terminologie specialistiche’, *Lessico Intellettuale Europeo*, Vol. 92. Olschki, 2003.
- [4] Adamo, Giovanni, and Valeria Della Valle. ‘Neologismi (Parole nuove dai giornali 2008-2018)’. In *Istituto dell’Enciclopedia Treccani*. Roma, 2018.
- [5] Brasolin, Paolo, Greta H. Franzini, and Stefania Spina. “‘Ti Blocco Perché Sei Un Trollazzo’. Lexical Innovation in Contemporary Italian in a Large Twitter Corpus.” Edited by Federico Boschetti, Gianluca E. Leboni, Bernardo Magnini, and Nicole Novielli. *Proceedings of the 9th Italian Conference on Computational Linguistics CLiC-It 2023*: Venice, Italy, November

- [6] Brigadir, Igor. 'Advanced Search On Twitter', 2023. <https://github.com/igorbrigadir/twitter-advanced-search>
- [7] Croft, William. *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Education, 2000.
- [8] Fischer, Roswitha. *Lexical Change in Present-Day English: A Corpus-Based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*. Vol. 17. *Language in Performance*. Tübingen: G. Narr, 1998.
- [9] Grieve, Jack, Nini Andrea, and Diansheng Guo. 'Analyzing Lexical Emergence in Modern American English Online'. *English Language & Linguistics* 21, no. 1 (2016): 99–127.
- [10] Grieve, Jack, Andrea Nini, and Diansheng Guo. 'Mapping Lexical Innovation on American Social Media'. *Journal of English Linguistics* 46, no. 4 (1 December 2018): 293–319. <https://doi.org/10.1177/0075424218793191>
- [11] Ježek, Elisabetta. *The Lexicon: An Introduction*. Oxford Textbooks in Linguistics. Oxford, New York: Oxford University Press, 2016.
- [12] Kerremans, Daphné. *A Web of New Word*. Frankfurt am Mai: Peter Lang, 2015.
- [13] Kershaw, Daniel, Matthew Rowe, and Patrick Stacey. 'Towards Modelling Language Innovation Acceptance in Online Social Networks.' In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 553–562. WSDM '16. New York, NY, USA: Association for Computing Machinery, 2016. <https://doi.org/10.1145/2835776.2835784>
- [14] Labov, William. *Principles of Linguistic Change*, Volume 2: Social Factors. Vol. 2–3. Oxford: Wiley-Blackwell, 2001.
- [15] Laitinen, Mikko, Masoud Fatemi, and Jonas Lundberg. 'Size Matters: Digital Social Networks and Language Change'. *Frontiers in Artificial Intelligence*, 46, 3 (20 July 2020). <https://doi.org/10.3389/frai.2020.00046>
- [16] Marri, Fabio. 'The Neologisms inside and Outside the Recent Repertoires'. *Quaderns d'Italia* 23, no. 11 (5 December 2018). <https://revistes.uab.cat/quadernsitalia/article/view/v23-marri>
- [17] Rodríguez Arrizabalaga, Beatriz. 'Social Networks: A Source of Lexical Innovation and Creativity in Contemporary Spanish'. *Languages*, 138, 6, no. 3 (September 2021): 1–22. <https://doi.org/10.3390/languages6030138>
- [18] Spina, Stefania. *Fiumi di parole. Discorso e grammatica delle conversazioni scritte in Twitter*. Aracne, 2019.
- [19] Spina, Stefania. 'Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione'. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-It 2014*, edited by Roberto Basili, Lenci Alessandro, and Bernardo Magnini, 1:354–359. Pisa: Pisa University Press, 2014.
- [20] Tarrade, Louise, Jean-Philippe Magué, and Jean-Pierre Chevrot. 'Detecting and Categorising Lexical Innovations in a Corpus of Tweets.' *Psychology of Language and Communication* 26, no. 1 (1 January 2022): 319–329. <https://doi.org/10.2478/plc-2022-15>
- [21] Würschinger, Quirin. 'Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter.' *Frontiers in Artificial Intelligence*, 648583, 4 (1 November 2021). <https://doi.org/10.3389/frai.2021.648583>
- [22] Zolli, Paolo. *Come nascono le parole italiane*. Milano: Rizzoli, 1989.

