



I dati linguistici

Metodologie
e strumenti della ricerca

a cura di
Caterina Cacioli,
Serena Carlamaria Crespi,
Stefano Miani, Barbara Patella,
Ersilia Russo, Carmelina Toscano



Studi e ricerche del Dipartimento di Lettere e Filosofia

direttore responsabile
Simone Magherini

direttore
Marco Biffi

Linguistica / 2



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DILEF
DIPARTIMENTO DI
LETTERE E FILOSOFIA

La collana «**Studi e ricerche del Dipartimento di Lettere e Filosofia**» dell'Università degli Studi di Firenze nasce, insieme a «DILEF. Rivista digitale del Dipartimento di Lettere e Filosofia», nel quadro delle attività condotte come Dipartimento di Eccellenza 2018-2022 sul Fondo assegnato dal Ministero dell'Istruzione, dell'Università e della Ricerca.

La collana si articola in quattro sezioni, che rispecchiano gli interessi e gli ambiti di studio delle rispettive sezioni dipartimentali: Antichità e Filologia, Filosofia, Letteratura italiana e Romanistica, Linguistica.

La pubblicazione in rete, in formato PDF, è ad accesso aperto; l'edizione a stampa è disponibile a pagamento.

Comitato direttivo

Benedetta Baldi, Giovanni Alberto Cecconi,
Simone Magherini, Mariagrazia Portera, Anna Rodolfi,
Salomé Vuelta García, Giovanni Zago

Comitato scientifico

Valentina Arena (University College, London)
Barbara Carnevali (École des Hautes Études en Sciences Sociales, Paris)
Mario Citroni (Scuola Normale Superiore di Pisa)
Matthias Heinz (Paris Lodron Universität Salzburg)
Susan Kozel (Università di Malmö)
Adam Ledgeway (University of Cambridge)
José María Micó (Università Pompeu Fabra, Barcellona)
Marco Petoletti (Università Cattolica di Milano)
Alessandro Polcri (Fordham University, NY)
Tommaso Raso (Universidade Federal del Minas Gerais)
Carole Talon-Hugon (Université de Nice-Sophia Antipolis)
Fabio Zinelli (École Pratique des Hautes Études, Paris)

I dati linguistici

Metodologie e strumenti della ricerca

a cura di

Caterina Cacioli,

Serena Carlamaria Crespi,

Stefano Miani, Barbara Patella,

Ersilia Russo, Carmelina Toscano

© 2024 Società Editrice Fiorentina, per la presente edizione
© 2024 The Authors, per i testi

via Aretina, 298 - 50136 Firenze
tel. 055 5532924
info@sefeditrice.it
edu.sefeditrice.it

E-ISSN 2974-6876
ISBN 978-88-6032-751-2
E-ISBN 978-88-6032-750-5
DOI 10.35948/DILEF/978-88-6032-750-5



La Collana è pubblicata ad Accesso Aperto con licenza Creative Commons
Licence CC-BY-NC-ND 4.0
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Crediti fotografici

© 2024 crediti dichiarati in didascalia.
Riproduzione in qualsiasi forma, intera o parziale, vietata

Progetto grafico e impaginazione
Francesco Sensoli

Copertina
Studio Grafico Norfini

Font
Alegreya ht e Alegreya Sans ht
(Juan Pablo del Peral, Huerta Tipográfica)

Indice

- VII *Introduzione dei curatori*
- I dati linguistici.
Metodologie e strumenti della ricerca
- 3 *La Marcatatura Differenziale dell'Oggetto in Calabria:
l'Area Lausberg*
Giusy Truncellito
- 21 *Sistemi di subordinazione nella Calabria meridionale:
il caso del verbo 'andare'*
Carmelina Toscano
- 45 *Fenomeni di contatto nel parlato di giovani abruzzesi.
Uno sguardo al corpus Nec Sine*
Valeria Garozzo, Enzo Santilli
- 71 *Manzoni postillatore della Crusca veronese: le unità fraseologiche*
Ersilia Russo
- 91 *Percorsi correttòri dal Fermo e Lucia alla Quarantana:
l'esempio dei proverbi glossati*
Irene Rumine
- 113 *Phraseological Research of the 16th-Century Slovenian Literary Language*
Eva Trivunović

- 131 *La "lingua delle donne" nella coscienza linguistico-letteraria di Sibilla Aleramo: realtà o costruzione sociale?*
Alessandra Rea
- 151 *Il lessico della psicanalisi in Italia nel primo Novecento: tra rimozioni, lacune lessicografiche e possibili sviluppi digitali*
Stefano Miani
- 171 *Esempi di interferenza tra lingua comune e lingua specialistica: note su alcuni termini artistico-architettonici e meccanici*
Matteo Mazzone
- 193 *Il "dato silente" in un corpus di parlato schizofrenico*
Simona Trillocco
- 215 *La raccolta e l'analisi dei dati linguistici in un'indagine su parlanti bilingui*
Yasmina Moussaid
- 229 *Tra lingue e risorse linguistiche: metodi di analisi contrastiva della semantica di verbi di azione*
Caterina Cacioli
- 251 *Cov-I-Cor: un corpus di italiano istituzionale relativo alla gestione dell'emergenza sanitaria*
Laura Occhipinti
- 269 *CORMIP (Corpus Multimodale dell'Italiano Parlato): questioni intorno al trattamento del dato linguistico multimodale*
Luca Lo Re
- 289 *Il dato del Paesaggio Linguistico: una proposta di classificazione con alcuni esempi dal mercato di S. Lorenzo (FI)*
Lorenzo Cambi
- 321 *Codifica XML-TEI: proposta di mark-up per i dizionari metodici*
Barbara Patella
- 339 *Indice dei nomi*

Introduzione dei curatori

Il presente volume raccoglie gli Atti del “Convegno dottorale internazionale LinUD”, un evento organizzato dai dottorandi dei cicli xxxv e xxxvi dei curricula di Linguistica e Umanistica Digitale, tenutosi a Firenze nei giorni 27, 28 e 29 giugno 2022 presso il Dipartimento di Lettere e Filosofia (DILEF) dell’Università degli Studi di Firenze.

LinUD è nato dall’idea di realizzare non solo un convegno accademico, ma anche e soprattutto un luogo di incontro interdisciplinare e di scambio scientifico, finalizzato a condividere le novità e i progetti che circolano nel campo della Linguistica e delle *Digital Humanities*, favorendo così un dialogo tra i partecipanti, sia dottorandi che giovani ricercatori.

Dopo esserci interrogati a lungo sul tema del convegno, la scelta conclusiva è ricaduta sui *dati linguistici*, poiché costituiscono un comune denominatore per le discipline umanistiche in questione e un oggetto esplorabile secondo molteplici sfaccettature, parametri e sistemi d’indagine: si pensi, ad esempio, alla raccolta dei dati, alla definizione di rappresentatività del campione oppure agli strumenti usati per l’estrappolazione e l’analisi dei dati; accanto a ciò – in un’ottica più squisitamente teorica – non sono da trascurare i criteri e le metodologie con cui essi vengono esaminati o ancora le modalità con cui possono essere condivisi e diventare fruibili (specie alla luce delle nuove tecnologie che consentono di accelerare i processi e le fasi di trattamento dei dati, nonché di renderli disponibili in formato digitale). In altre parole, si è lasciato spazio al dato linguistico sotto larghi punti di vista, nella

I dati linguistici. Metodologie e strumenti della ricerca

fattispecie ai modi in cui questo può essere elicitato, trattato e condiviso – a tal proposito, la diversità degli approcci riflette la vastità e la complessità degli studi linguistici e dimostra come metodi e modalità di indagine varino con il variare delle scelte e degli obiettivi di ogni singola ricerca.

A una molteplicità di analisi e di risorse corrispondono, dunque, molteplici prospettive di ricerca, in grado di stimolare nuove idee e produrre continui interrogativi. Ed è proprio in questo contesto che si inserisce LinUD, nato come occasione per caldeggiare la condivisione di pratiche, metodologie e modelli di ricerca – sia nel solco di tradizioni già consolidate sia nel panorama di frontiere sperimentali –, al fine di promuovere un'interazione proficua tra le diverse aree di ricerca. Crediamo, infatti, che idee innovative e stimolanti possano sgorgare anche dall'avvicinamento di ambiti di ricerca che di solito restano lontani fra loro, con l'auspicio di contrastare quell'isolamento che talvolta caratterizza il percorso delle singole ricerche col rischio di renderle aride: tutto ciò nell'ottica di far arricchire vicendevolmente le branche di un ventaglio composito come quello della Linguistica e dell'Umanistica digitale.

Con questa raccolta si intende offrire il contributo dei relatori che hanno preso parte alle giornate del convegno e che hanno permesso di toccare una vasta gamma di argomenti riguardanti i dati linguistici. Come anticipato, gli articoli proposti si muovono in ambiti eterogenei: si passa dal mutamento e dalla variazione nelle lingue a questioni di semantica, dai linguaggi specialistici alla fraseologia e alla lessicografia, fino ad arrivare all'analisi testuale, all'utilizzo dei corpora e ai linguaggi di mark-up.

A occuparsi di morfo-sintassi dialettale GIUSY TRUNCELLITO, che indaga la MDO (“Marca Differenziale dell'Oggetto”) nei dialetti calabresi settentrionali dell'area Lausberg, conducendo un'analisi micro-variazionista che tiene conto dei diversi parametri semantici, quali l'animità, la definitezza e la specificità dell'oggetto, col fine ultimo di chiarire quali siano i contesti in cui, nelle diverse varietà, è ammessa oppure esclusa la MDO; appartiene allo stesso filone di ricerca l'articolo di CARMELINA TOSCANO, che indaga il fenomeno della pseudo-co-

ordinazione (PseCo) nella Calabria meridionale attraverso un'analisi micro-variazionista, esaminando i contesti semantico-sintattici in cui la codifica della PseCo è esclusiva e in quali altri invece co-esiste con altre strutture di modo finito diffuse in questi stessi territori; l'obiettivo del lavoro è quello di individuare i contesti prototipici di codifica della PseCo e i fattori che governano la co-ricorrenza o l'alternanza tra questa e le altre costruzioni diffuse nella stessa area.

Spostandoci sul versante del rapporto tra lingua e dialetto, troviamo lo studio di VALERIA GAROZZO ed ENZO SANTILLI, il cui obiettivo consiste nell'individuare, categorizzare e analizzare i fenomeni di contatto italiano-dialetto: il lavoro dei due studiosi prende le mosse da un'analisi delle produzioni orali di parlanti bilingui che si avvicinano il più possibile al parlato spontaneo prototipico.

Alla fraseologia manzoniana sono dedicati due contributi: quello di ERSILIA RUSSO, che ricostruisce il ruolo rivestito dalla cosiddetta "Crusca veronese" nella formazione della fraseologia dei *Promessi sposi* (nella fattispecie, le postille apportate sul vocabolario testimoniano l'attenzione che Alessandro Manzoni ha riservato alle unità fraseologiche, in quanto tessere particolarmente rilevanti per la restituzione di una lingua "viva e vera") e quello di IRENE RUMINE, in cui si approfondisce l'aspetto paremiologico del romanzo manzoniano (in continuità con il clima romantico, i proverbi sono visti come un serbatoio di cultura popolare, in grado di conferire una molteplicità di toni e di registri al testo; in particolare, vengono messe in evidenza quelle paremie che, dal *Fermo e Lucia* alla *Quarantana*, sono segnalate da glosse metalinguistiche che ne attestano un uso consapevole da parte dell'autore). Sempre incentrato sulla fraseologia, ma in relazione alla lingua slovena, l'articolo di EVA TRIVUNOVIC, che presenta le problematicità dello studio diacronico della fraseologia, occupandosi nello specifico del materiale ricavato dal *Dizionario della lingua letteraria slovena del XVI secolo*: dall'analisi delle fonti del dizionario (53 testi di natura prettamente religiosa) si evincono due caratteristiche proprie delle unità fraseologiche, ossia la stabilità e l'espressività.

Focalizzato sulla lingua delle donne il contributo di ALESSANDRA REA, che indaga il rapporto che intercorre tra le scrittrici e la lingua,

I dati linguistici. Metodologie e strumenti della ricerca

intesa come costruzione sociale, studiando la percezione di Sibilla Aleramo riguardo a una presunta lingua delle donne; l'attenzione dell'autrice si concentra sull'analisi di più variabili, come la variazione diastratica, diafasica e di genere. L'analisi si sviluppa attraverso una griglia di riferimento contenente, da un lato, i tratti ritenuti caratterizzanti, dall'altro alcuni fenomeni linguistici attraverso cui è misurata la presenza di quei tratti.

Di linguaggi specialistici si è occupato STEFANO MIANI, in particolare della formazione del lessico italiano della psicanalisi nella prima metà del XX secolo; lo studio presenta il corpus di testi su cui sono stati condotti gli spogli lessicali, i criteri di digitalizzazione del materiale raccolto e alcuni dei dati quantitativi ricavati grazie all'utilizzo delle piattaforme *Sketch Engine* e *Nodegoat*.

Sull'interferenza tra lingua comune e lingua specialistica si concentra il saggio di MATTEO MAZZONE, in particolar modo sulla ridefinizione semantica in chiave tecnica di voci d'uso quotidiano, mediante l'analisi semantica di alcuni lemmi tratti dallo spoglio lessicale effettuato su quella che è oggi considerata come la seconda traduzione cinquecentesca del *De architectura* di Vitruvio, databile tra il 1504 e il 1511, e contenuta nel manoscritto Ottoboniano latino 1653 della Biblioteca Apostolica Vaticana.

Diverse ricerche ruotano, invece, intorno alla linguistica dei *corpora*, strumenti declinati in base alle diverse modalità di indagine, a partire da quella di SIMONA TRILLOCCO, che analizza e indaga il ruolo delle pause nel parlato schizofrenico, mediante l'analisi del *Corpus Italiano di Parlato Patologico Schizofrenico* (CIPPS) nel quadro teorico della "Teoria della Lingua in Atto", mostrando, mediante un confronto con i dati relativi al parlato normofasico, i primi risultati ottenuti.

YASMINA MOUSSAID analizza il bilinguismo di parlanti italo-arabofoni, mediante una ricerca che combina l'analisi delle informazioni ricavate da interviste semi-strutturate – rivolte a parlanti bilingui – con l'analisi di *corpora* bilingui, con l'intento di indagare alcune caratteristiche, come ad esempio le scelte linguistiche dei parlanti coinvolti, i fenomeni di contatto e ulteriori questioni di tipo sociolinguistico e socioculturale.

Nel suo saggio CATERINA CACIOLI propone una discussione dei metodi di analisi dei dati linguistici in ottica contrastiva e nel campo della semantica, focalizzandosi sull'uso di diverse fonti di dati – *corpora* e video-risorse linguistiche –; con un caso di studio sui verbi di azione in inglese, svedese e italiano si propone l'uso combinato di dati linguistici, mostrando i vantaggi dell'unione della componente visiva e d'uso nello studio dei verbi e delle azioni.

LAURA OCCHIPINTI si occupa di esplorare l'importanza della comunicazione istituzionale durante la pandemia da Covid-19, evidenziando la necessità di chiarezza nei messaggi ufficiali e introducendo, in relazione al soggetto trattato, il corpus Cov-I-Cor, appositamente creato per raccogliere i testi istituzionali prodotti in periodo di crisi e analizzare l'uso del linguaggio istituzionale durante l'emergenza sanitaria.

Ancora in relazione ai *corpora* il contributo di LUCA LO RE, il quale propone un metodo per sviluppare un corpus multimodale dell'italiano parlato, seguendo la tradizione dei *corpora* compilati dal gruppo di ricerca LABLITA dell'Università di Firenze e proponendo la creazione di un corpus pilota, CORMIP (acronimo di *Corpus Multimodale dell'Italiano Parlato*), attraverso cui testare un nuovo tipo di approccio ai dati e preparare il terreno per ricerche future e comprendere la lingua come un sistema complesso, multimodale, che utilizza gesti e linguaggio parlato per esprimere significati.

Di paesaggio linguistico parla LORENZO CAMBI, che presenta una classificazione del Paesaggio Linguistico (PL) basata sull'analisi di dati ed esempi provenienti dal mercato di San Lorenzo a Firenze; la proposta si distingue per un approccio sperimentale che integra metodologie quantitative e qualitative, enfatizzando l'importanza di coinvolgere attivamente gli attori e gli osservatori del PL, ed esaminando, al contempo, l'impatto delle politiche linguistiche e il ruolo sempre più rilevante del Paesaggio Linguistico Virtuale e dei legami tra spazio fisico e virtuale nell'ambito della società contemporanea.

A chiudere la raccolta il contributo di BARBARA PATELLA, nel quale si delineano – e si illustrano con una serie di esempi – i principali criteri adottati per marcare in XML-TEI i maggiori dizionari metodici

I dati linguistici. Metodologie e strumenti della ricerca

ottocenteschi della lingua italiana, al fine di mostrare le modalità con cui tali strumenti lessicografici (particolarmente rilevanti per la storia della lingua italiana e della lessicografia), da prodotti cartacei “statici”, siano stati trasformati in repertori digitali “dinamici” grazie a un lavoro di informatizzazione per renderli interrogabili (in funzione di una piattaforma online predisposta per consultare questo tipo di dati).

Chiudiamo la nostra premessa con l’augurio che questo volume possa essere una risorsa utile per gli studiosi, costituendo un pungolo per nuove riflessioni e nuovi spazi di ricerca.

Intendiamo rivolgere i nostri ringraziamenti a tutti i relatori che hanno partecipato con grande entusiasmo alle giornate del Convegno con i loro interventi, condividendo le loro ricerche e affidandone una parte a questo volume. Un sincero ringraziamento va poi ai relatori invitati che, con i loro contenuti, hanno impreziosito l’occasione di scambio accademico: il professor Alessandro Lenci e il professor Franco Fanciullo.

Un doveroso grazie al Dipartimento di Lettere e Filosofia dell’Università degli Studi di Firenze (DILEF), in particolare al professor Pierluigi Minari e alla professoressa Paola Manni che, nel 2022 – rispettivamente in qualità di direttore del Dipartimento e di coordinatrice del Dottorato di ricerca in “Filologia, Letteratura italiana, Linguistica” –, hanno reso possibile l’organizzazione del Convegno LinUD.

Concludiamo ringraziando i membri del Comitato scientifico, specialmente il professor Marco Biffi (referente del curriculum di Umanistica digitale) e la professoressa Monica Ballerini (referente del curriculum internazionale di Linguistica) per averci guidato e coordinato nella progettazione del Convegno, nonché durante le fasi organizzative, e infine per averci sostenuto nella realizzazione degli Atti raccolti in questo volume.

I dati linguistici.
Metodologie e strumenti della ricerca

La Marcatura Differenziale dell'Oggetto in Calabria: l'Area Lausberg

Giusy Truncellito

Introduzione

La Marcatura Differenziale dell'Oggetto (MDO d'ora in avanti), conosciuta anche come *accusativo preposizionale*, è il fenomeno morfosintattico per cui, in una lingua, alcune tipologie di oggetti diretti vengono marcate, a differenza di altri oggetti che non presentano la marca (cfr. BÁRÁNY 2018, p. 3; BOSSONG 1985).

È una tipologia di costruzione molto diffusa nelle lingue del mondo: «[t]he number of language groups and language families where phenomena describable as “differential object marking” can be found is very high [...]: Slavic, Armenian, Neo-Iranian, and Neo-Indoaryan inside the Indo-European phylum; and outside of it Finno-Ugric, Dravidian, Turkic, Mongolic, Tungusian, Tibeto-Burmese, Munda, Bantu, Pama-Nyungan, Micronesian, Uto-Aztec, Chicbcha, anche Tupi Guarani, among others» (BOSSONG 1991, p. 145).

Per quanto riguarda il panorama italo-romanzo, la MDO è attestata in italiano regionale del nord, in toscano (BENINCÀ 1986; BERRETTA 1989, IEMMOLO 2010, LOMBARDI VALLAURI-SUZUKI 2013), nonché nei dialetti toscani dell'isola d'Elba e della Corsica. È, infine, estremamente diffusa nelle varietà meridionali e meridionali estreme e in sardo (cfr. DE ANGELIS 2019; LEDGEWAY 2000; LEDGEWAY ET AL. 2019; MANZINI-SAVOIA 2005; ROHLFS 1969).

Tuttavia, la MDO presenta una enorme micro e macro-variazione, come affermato anche da Guardiano 2010 (p. 96), dal momento che «si

manifesta nella morfosintassi [...] ma è condizionata da fenomeni di carattere semantico o pragmatico-informazionale. È inoltre caratterizzata da una sensibile gamma di variazione sia interlinguistica sia intralinguistica».

In calabrese, come nelle altre varietà romanze e italo-romanze¹, la marca «coincide dovunque con la marca del dativo, normalmente un riflesso di AD > a, anche se non sono infrequenti riflessi di IN MEDIO (AD) 'in mezzo (a)' > *ma/me* nell'Italia centrale, nonché riflessi di DE+AB > *da* in determinati dialetti umbri» (LEDGEWAY 2019, p. 1) e nel dialetto gallosiculo di Nicosia (ROHLFS 1969, p. 8; 1984, 79).

È ormai noto che la presenza della MDO dipende da caratteristiche semantiche del referente (AISSEN 2003, p. 435), *in primis* l'animatezza (cfr. COMRIE 1989; AISSEN 2003; CROFT 2003), per cui:

umano > animato > inanimato.

L'animatezza sembra, infatti, il principale discrimine nella legittimazione della MDO, dal momento che, generalmente, gli oggetti marcati sono quelli che si trovano nelle posizioni più alte della gerarchia. È anche vero, tuttavia, che essa da sola non rappresenta un criterio sufficiente per determinare l'uso della marca differenziale, ma interagisce con altri parametri, quali la definitezza e la specificità (COMRIE 1989, p. 199), che AISSEN 2003 ha riassunto nella seguente gerarchia di definitezza che integra la specificità:

pronomi personale > nome proprio > SN definito > SN indefinito specifico > SN non specifico.

Dall'interazione tra animatezza, definitezza e specificità è stata successivamente messa a punto la *Extended Animacy Hierarchy* (cfr. SILVERSTEIN 1976; CROFT 2003, p. 130), fondamentale nello studio della MDO:

¹ Fa eccezione il rumeno con la marca *pe* < lat. (SU)PER.

pronomi personali di 1^a e 2^a persona < pronomi di 3^a persona < nomi propri e di parentela < nomi comuni umani < nomi comuni non umani animati < nomi comuni inanimati.

Metodologia e corpus

I dati analizzati in questo lavoro sono stati raccolti mediante la somministrazione di un questionario in sei località della Calabria settentrionale, tutte appartenenti alla cosiddetta *area Lausberg*, la quale prende il nome, com'è ben noto, dallo studioso – Heinrich Lausberg – che per primo si occupò di studiarla in modo approfondito ed è considerata una delle aree più conservative della Romania occidentale.

Per ogni località sono stati selezionati 6 informanti, suddivisi tenendo conto di sesso e fascia d'età².

Come illustrato nella Figura 1, le varietà analizzate sono sei: sul versante ionico Canna, Nocara, Alessandria del Carretto e Francavilla Marittima; su quello tirrenico Laino Borgo e Verbicaro.

Per l'analisi sono stati presi in considerazione i parametri semantici precedentemente menzionati (animatezza, definitezza e specificità, e, di conseguenza, la *Extended Animacy Hierarchy*), i quali sono fondamentali nella realizzazione del fenomeno della MDO, nonché parametri formali riguardanti la tipologia di nominali rappresentanti l'oggetto diretto marcato (es. teste D e sintagmi nominali complessi). Si procederà indicando, tuttavia, soltanto gli oggetti diretti che, almeno in una delle lingue esaminate, hanno dimostrato la possibilità di legittimare la MDO, mentre non sono tenute in considerazione tutte le altre tipologie di oggetto diretto (es. nomi comuni [+animati] con determinante esplicito; nomi comuni [+umani] con quantificatori universali e indefiniti; nomi comuni [-umani; +animati; -definiti]; oggetti con referente [-animato]).

² Nello specifico, sono stati intervistati un maschio e una femmina per la fascia da 18 a 40 anni, un maschio e una femmina per la fascia da 41 a 60 anni e un maschio e una femmina per la fascia dai 61 anni in su.

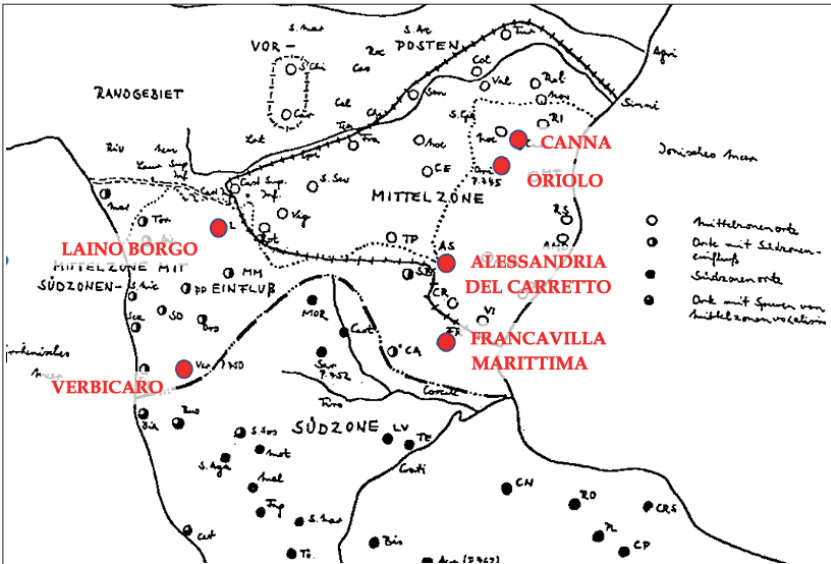


Figura 1 L'area Lausberg (Martino 1991, tav. v) e le località indagate

Pronomi

Com'è noto, in italiano colloquiale, Toscana e Italia settentrionale, nonché in varietà meridionali antiche come il napoletano, la distribuzione della MDO è effettuata in base a una scissione [\pm partecipanti al discorso] e non riguarda la 3^a persona (cfr. LEDGEWAY 2019).

In nessuna delle varietà analizzate, tuttavia, è presente una scissione pronominale, bensì ognuna delle sei varietà presenta obbligatorietà della marca differenziale con pronomi personali non solo quando questi si trovano in posizione di Topic o Focus (1c; 1f), ma anche in frasi non marcate (1a-b; 1d-e):

(1)

cann.

a. ma'ri:jə 'ɛ cca'mə:tə a 'mmi:jə / a 'tti:jə
 Maria avere.AUS.3SG chiamare.PPT MDO me / MDO te
 'Maria ha chiamato me / te'

La Marcatura Differenziale dell'Oggetto in Calabria

- b. ma'ri:jə 'tʃɛrkədə a 'ʝillə nɔn a 'ʝillə
 Maria cercare.PRS.3SG MDO lui non DOM lei
 'Maria cerca lui, non lei'
- c. a 'mmi:jə stu 'filmə mə 'fɛ 'rɪ:rə
 MDO me questo film mi fare.PRS.3SG ridere.INF
 'A me questo film fa ridere'
- verb.
- d. llə'ɣwiddʒə 'a 'vvistə a 'nnu.va
 Luigi avere.AUS.3SG vedere.PPT MDO noi
 'Luigi ha visto noi'
- e. 'ei ca'ma:tə a 'bbu.va 'nɔ:nə a 'ʝiddə
 avere.PRS.1SG chiamare.PPT MDO voi non DOM loro
 'Ho chiamato voi, non loro'
- f. a 'mmi:jə mə 'fa mbavə'rʷa 'su malə'tjembə
 MDO me mi fare.PRS.3SG spaventare.INF questo maltempo
 'Mi fa spaventare questo maltempo'

Altra tipologia di oggetti diretti obbligatoriamente marcati in tutte le varietà prese in esame è quella dei dimostrativi pronominali [+uma-ni]:

(2)

oriol.

- a. 'ei 'vistə a 'kkwistə 'kɔ:mə s_ɛ
 avere.AUS.2SG vedere.PPT MDO questo come si avere.AUS.3SG
 bbəs'tɔ:tə
 vestire.PPT
 'Hai visto questo come si è vestito?'

lain.

- b. 'ai sen'du:tə a 'kkwidda 'kɔ:mu 'parla
 avere.AUS.2SG sentire.PPT MDO quella come parlare.PRS.3SG
 'Hai sentito quella come parla?'

Giusy Truncellito

Anche nel caso dei pronomi interrogativi e indefiniti, è risultato cruciale il ruolo dell'animatezza o – per essere più specifici – dell'umanità. Con questa tipologia di pronomi (chi, qualcuno, nessuno, uno), la MDO è risultata obbligatoria per tutte le varietà, tranne le due più a nord (sul versante ionico Canna, su quello tirrenico Laino Borgo), per le quali, invece, la situazione appare un po' diversa: per il cannese, la marca risulta opzionale con i pronomi interrogativi e obbligatoria con i pronomi indefiniti (3a-b), mentre in lainesese risulta essere opzionale in entrambi i casi (3c-d):

(3)

cann.

- a. (a) (k)k^o 'ei sən'dʊ:tə
MDO chi avere.AUS.2SG sentire.PPT
'Chi hai sentito?'
b. ,nɔnn_u 'ɛddʒə sən'dʊ:tə a nnəf'fʊ:nə
non avere.AUS.1SG sentire.PPT MDO nessuno
'Non ho sentito nessuno'

lain.

- c. (a) '(k)ku 'ai 'visto
MDO chi avere.AUS.2SG vedere.PPT
'Chi hai visto?'
d. 'aʃʃu 'visto (a) aŋ'gu:nə nd_a 'cattsa
avere.AUS.1SG vedere.PPT MDO qualcuno in la piazza
'Ho visto qualcuno in piazza'

Nomi propri

Per quanto riguarda i nomi propri, la situazione che i dati ci restituiscono appare più variegata.

Nello specifico, con i nomi propri [+umani] (4a-b) e [-umani; +animati]³ (5a-b), tutte le varietà prevedono l'uso obbligatorio della marca, ma i dialetti di Canna (4a; 5a) e Laino Borgo (4b; 5b) si comportano

3 In particolare, si tratta di nomi di cani.

diversamente, dal momento che, in tali varietà, l'uso della MDO con questa tipologia di oggetti diretti è soltanto opzionale:

(4)

cann.

- a. ma'ri:jə 'ɛ ŋgun'drɛtə (a) (l)lu'iddzə ɛ (a) 'kkarlə
 Maria avere.AUS.3SG incontrare.PPT MDO Luigi e MDO Carlo
 nu 'mɛ:rə
 a il mare
 'Maria ha incontrato Luigi and Carlo al mare'

lain.

- b. ma'ri:a 'a ŋgon'dra:to (a) (l)lu'iddzi ɛ (a) 'kkarlo
 Maria avere.AUS.3SG incontrare.PPT MDO Luigi e MDO Carlo
 allu 'ma:re
 al mare
 'Maria ha incontrato Luigi and Carlo al mare'

(5)

cann.

- a. 'ɛ:nə tru'vɛ:tə (a) '(l)la:ki / (a) 'bbrəkkə'tʃillə
 avere.AUS.2PL trovare.PPT MDO Lucky / MDO Black
 ndu 'vuskə
 in il bosco
 'Hanno trovato Lucky / Black nel bosco'

lain.

- b. 'anno tru'va:to (a) '(l)la:ki / (a) '(r)rɛkksə
 avere.AUS.2PL trovare.PPT MDO Lucky / MDO Rex
 ndu 'vwɔʃkə
 in il bosco
 'Hanno trovato Lucky / Rex nel bosco'

La situazione si complica ulteriormente con i toponimi o nomi propri [-animati], i quali rappresentano «una sottoclasse di nomi propri che, sebbene apparentemente [-an.], si prestano facilmente all'antropomorfizzazione o, per lo meno, vanno interpretati come sottospecificati per il tratto [\pm an.] (LEDGEWAY 2019, p. 14). La MDO, fino ad oggi

Giusy Truncellito

registrata nel panorama italo-romanzo per le varietà della Corsica e dell'Isola d'Elba (cfr. ROHLFS 1969, p. 8) e per alcune varietà di calabrese meridionale estremo (ROHLFS 1971; DE ANGELIS 2019), è presente, seppure non obbligatoriamente con toponimi, in cinque delle varietà prese in esame nel presente lavoro – l'unico dialetto a non presentare la marca accusativa con toponimi è quello di Laino Borgo (6b).

(6)

oriol.

- a. a'wannə pas'sa:tə 'eddʒə vəsə'ta:tə (a) '(r)rɔ'mə
quest'anno passato avere.AUS.1SG visitare.PPT MDO Roma
'L'anno scorso ho visitato Roma'

lain.

- b. l'anno 'skɔrsə 'ajju visi'ta:tə Ø 'rɔ:ma
l'anno scorso avere.AUS.1SG visitare.PPT Ø Rome
'L'anno scorso ho visitato Roma'

Tuttavia, non tutti i toponimi presentano la MDO. È fatto noto, ad esempio, che alcuni nomi di luogo in passato fossero preceduti dall'articolo determinativo, che sembrerebbe, in casi del genere, bloccare la comparsa della MDO. Osservando, per esempio, la Figura 2 dall'«*Atlante Novissimo*» (ZATTA 1784), è possibile notare come alcuni toponimi fossero introdotti, all'epoca, da un determinante, usato ancora oggi nelle varietà dialettali come parte del nome proprio.

All'interno della mappa è possibile individuare, ad esempio, i toponimi per le località di Nocara, Canna e Oriolo, indicati rispettivamente come 'La Nucara', 'La Canna' e 'Oriuolo'; mentre nei primi due casi è impossibile che la MDO sia legittimata, nel terzo caso, anche se opzionalmente, l'uso della marca è previsto.

(7)

cann.

- a. 'eddʒə 'vistə Ø a 'kannə / Ø a nu'kɛ:rə / (a) rri'u:ə /
avere.AUS.1SG vedere.PPT Ø la Canna / Ø la Nocara / MDO Oriolo /



Figura 2 Calabria Citeriore (dettaglio), da Zatta 1784

(a) 'r(r)ɔ:mə
 MDO Roma
 'Ho visto Canna / Nocara / Oriolo / Roma'

oriol.

b. 'eddʒə 'vistə Øa mənnu'a:rə / Øa 'kannə / Øa 'rɔkkə /
 avere.AUS.1SG vedere.PPT Øla Amendolara / Øla Canna / Øla Rocca /
 a wə'sandrjə / (a) 'ssan 'dʒɔrdʒə
 la Alessandria / MDO San Giorgio
 'Ho visto Amendolara / Canna / Rocca / Alessandria / San Giorgio'

Ciò sarebbe, inoltre, riconducibile anche a questioni formali, dal momento che nella maggior parte delle varietà analizzate sono soltanto le teste D (come anche nel caso di 'San Giorgio') a prevedere la legittimazione della MDO, mentre i nominali complessi (pertanto, i sintagmi determinanti) non prevedono questa possibilità.

Nomi di parentela

I nomi di parentela, nella scala di animatezza, seguono i nomi propri e precedono i nomi comuni animati; ciononostante, la situazione nelle diverse località appare piuttosto omogenea. Cinque delle varietà analizzate, infatti, prevedono un uso obbligatorio della marca con i nomi di parentela singolari modificati da aggettivo possessivo tonico post-nominale, mentre solo il dialetto lainese presenta un uso della marca soltanto opzionale.

(8)

lain.

- a. 'mamma 'mɛ:a 'a cca'ma:to (a) ,(t)tsi:a 'tu:a
mamma mia avere.AUS.3SG chiamare.PPT MDO zia tua
'Mia mamma ha chiamato tua zia'
- b. 'stɔŋgə aspət'tann° (a) (m)muj'jɛ:rɔma 'nnandə a 'ka:sa
stare.AUS.1SG aspettare.GER DOM moglie+mia davanti la casa
'Sto aspettando mia moglie davanti casa'

I nomi di parentela in forma plurale, tuttavia, non ammettono in nessun caso e in nessuna delle varietà esaminate l'uso della marca, dal momento che questa tipologia di nominali è introdotta da articolo determinativo e – a livello formale – non costituisce una testa D, bensì un nominale complesso:

(9)

lain.

- a. 'mamma 'mɛ:a 'a cca'ma:to Ø i 'tsi:i 'tu:a
mamma mia avere.AUS.3SG chiamare.PPT Ø gli zii tuoi
'Mia mamma ha chiamato i tuoi zii'

Anche i nomi di parentela senza possessivo, essendo modificati da determinanti e costituendo, pertanto, dei nominali complessi, non ammettono un uso della marca (10a; 10c), ad eccezione di 'mamma', 'papà',

'nonno' e 'nonna', i quali si comportano esattamente come i nomi propri [+umani], identificando un referente in particolare – ed essendo, come questi ultimi, dei nomi nudi (*bare nouns*), per cui delle teste D (10b; 10d).

(10)

cann.

a. 'eddʒə ŋgun'drɛ:tə Ø i 'nɔnnə nd_a 'ci:sjə
 avere.AUS.1SG incontrare.PPT Ø i nonni in la chiesa
 'Ho incontrato i nonni in chiesa'

b. 'eddʒə pur'tɛ:tə (a) (m)mam'ma / (a) (p)pa'pa
 avere.AUS.1SG portare.PPT MDO mamma / MDO papà
 ad'du u 'mi:dəkə
 dove il medico
 'Ho portato mamma dal medico'

verb.

c. 'a:nə 'vistə Ø u 'fr'w:a:tə i ma'ri:a
 avere.AUS.1SG vedere.PPT Ø il fratello di Maria
 'Hanno visto il fratello di Maria'

d. 'ɛj pur't'w:a:tə a 'mmamma / a ppa'pa
 avere.AUS.1SG portare.PPT MDO mamma / MDO papà
 ad'du u 'mjɛ:dəkə
 dove il medico
 'Ho portato mamma dal medico'

Nominali complessi

Per quanto riguarda i nominali complessi, questi non legittimano la MDO con nomi comuni [+umani] nella maggior parte delle varietà indagate. In effetti, sono solo due le varietà che prevedono un uso – talvolta obbligatorio, talvolta opzionale – della marca con sintagmi complessi: si tratta delle due varietà situate più a sud, Francavilla Marittima sul lato ionico e Verbicaro su quello tirrenico. Mentre, però, il dialetto di Francavilla estende la marca soltanto a nominali [+umani] introdotti da

dimostrativo (11a), il dialetto di Verbicaro mostra, invece, un'estensione della MDO ancora più significativa, poiché non solo essa si estende (obbligatoriamente) ai nominali [+umani] con dimostrativo (11a-b), ma anche a nominali [+umani -definiti] con (12c-d) o senza lettura presupposizionale (12a-b) e modificati da articolo indeterminativo o quantificatori universali (12a-d). In questi ultimi casi, è proprio il tratto della specificità semantica a giocare un ruolo cruciale nella selezione della MDO. Ciò significa che la MDO viene legittimata quando il referente a cui l'oggetto diretto si riferisce è ben identificabile. Per esempio, in (12a), quando il parlante legittima marca accusativale fa riferimento a un medico specifico e ben identificabile e non ad uno qualunque.

(11)

francav.

- a. lu'iddʒə 'a 'ddʒa skun'dra:tə (a) 'sta waʎ'ʎə:nə
Luigi avere.AUS.3SG già incontrare.PPT MDO questa ragazza
'Luigi ha già incontrato questa ragazza'

verb.

- b. llə'ɣ'iddʒə 'a 'ddʒa mbrun'd'watə a 'sta
kwa'tra:ra
Luigi avere.AUS.3SG già incontrare.PPT MDO questa
ragazza
'Luigi ha già incontrato questa ragazza'

(12)

verb.

- a. ma'ri:a 'a cca'ma:tə (a) (n)nə 'mje:dəkə
Maria avere.AUS.3SG chiamare.PPT MDO un medico
'Maria ha chiamato un medico'
- b. 'karlə 'a 'bbistə (a) '(t)tʃertə pa'rendə alla 'missa
Carlo avere.AUS.3SG vedere.PPT MDO alcuni parenti alla messa
'Carlo ha visto alcuni parenti a messa'
- c. llə'ɣ'iddʒə 'a mbrun'd'watə (a) (n)nə krəsti'a:nə
Luigi avere.AUS.3SG incontrare.PPT MDO un signore

La Marcatura Differenziale dell'Oggetto in Calabria

ḳ'a 'kkanəʃ'fu:tə 'ji:erə
 che avere.AUS.3SG conoscere.PPT ieri
 'Luigi ha incontrato un signore che ha conosciuto ieri'

d. 'karlo 'a 'bbistə (a) '(t)tʃertə pa'rəndə ka 'nɔnnə
 Carlo avere.AUS.3SG vedere.PPT MOD alcuni parenti che non
 i sup'pɔrtədə
 li sopportare.PRS.3SG
 'Carlo ha visto alcuni parenti che non sopporta'

Considerazioni finali

In base a quanto fino ad ora osservato, è possibile suddividere le sei località indagate in cinque diversi gruppi (tabella 1), dalla più conservativa alla più innovativa:

	Gruppo 1 Laino Borgo	Gruppo 2 Canna	Gruppo 3 Oriolo	Gruppo 3 Alessandria del Carretto	Gruppo 4 Francavilla Marittima	Gruppo 5 Verbicaro
Pronomi 1 ^a /2 ^a persona	+	+	+	+	+	+
Pronomi 3 ^a persona	+	+	+	+	+	+
Pronomi dimostrativi	+	+	+	+	+	+
Pronomi interrogativi [+umani]	±	±	+	+	+	+
Pronomi indefiniti [+umani]	±	±	+	+	+	+
Nomi propri [+umani]	±	±	+	+	+	+
Nomi propri [-umani; +animati]	±	±	+	+	+	+
Nomi propri [-animati] (toponimi)	-	±	±	±	±	±
Nomi propri [-animati]	±	+	+	+	+	+

Giusy Truncellito

	Gruppo 1 Laino Borgo	Gruppo 2 Canna	Gruppo 3 Oriolo	Gruppo 3 Alessandria del Carretto	Gruppo 4 Franca villla Marittima	Gruppo 5 Verbicaro
Nominali complessi [+umani] con dimostrativi	-	-	-	-	±	±
Nominali complessi [+umani; -definiti; +presupposizionali] NPs	-	-	-	-	-	±
Nominali complessi [+umani] con quantificatori universali	-	-	-	-	-	±

Tabella 1 La MDO nelle località analizzate

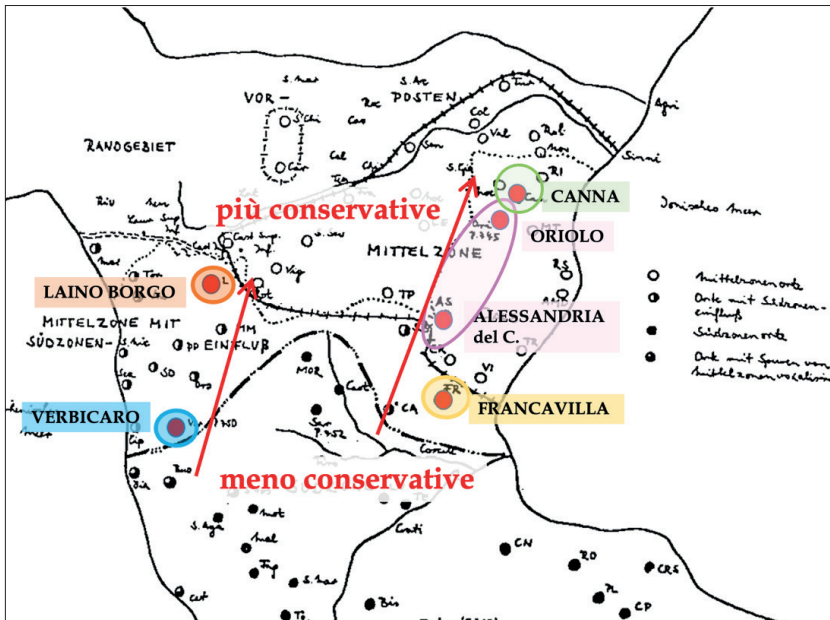


Figura 3 DOM e conservatività nell'area Lausberg

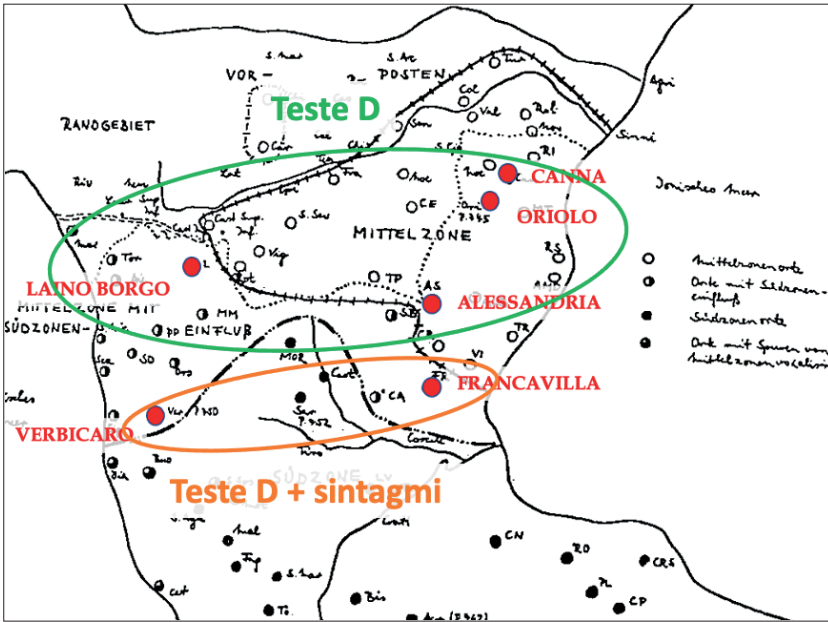


Figura 4 Tipi di nominali che legittimano la MDO nell'area Lausberg

Guardando attentamente la Figura 3, si nota come le varietà risultano essere più innovative (cioè estendono l'uso della MDO a una maggiore quantità di tipologie di oggetti diretti) man mano che si scende a sud, e diventano sempre più conservative man mano che ci si sposta al nord della regione, al confine con la Lucania e, pertanto, all'interno dell'area Lausberg, già conosciuta per la sua conservatività.

Osservando, infine, la tipologia di oggetti diretti che legittimano la MDO nei diversi dialetti (Figura 4), è possibile individuare come alcuni di essi – rappresentati dai gruppi 1, 2 e 3 – legittimino la marca soltanto con le teste D [+animate], come precedentemente accennato. Come afferma LEDGEWAY 2019 (p. 15), alcuni dialetti della Calabria nordorientale effettuano una scissione Testa-Sintagma, la quale raggiunge, in alcuni casi (es. nel dialetto di Amendolara, comune confinante con Oriolo) uno stadio estremo, con «l'estensione opzionale [...] o obbliga-

toria [...] dell'accusativo preposizionale ai pronominali anche [-an.], in modo che tutte le teste D legittimano l'accusativo preposizionale» (cfr. MANZINI-SAVOIA 2005, p. 510).

Le varietà più a sud, Verbicaro e Francavilla Marittima, considerate più innovative nella legittimazione della MDO – ossia i gruppi 4 e 5 –, invece, estendono l'uso della marca differenziale anche ad alcune tipologie di sintagmi (Figura 4), il cui referente è [+animato; ±definito; +specifico].

Bibliografia

- AISSEN 2003 = JUDITH AISSEN, *Differential object marking: iconicity vs. economy*, in «Natural language and linguistic theory», XXI, 3, 2003, pp. 435-483.
- BÁRÁNY 2018 = ANDRAS BÁRÁNY, *Person, case, and agreement. The morphosyntax of inverse agreement and global case splits*, Oxford, Oxford University Press, 2018.
- BENINCÀ 1986 = PAOLA BENINCÀ, *Il lato sinistro della frase italiana*, in «ATI Journal», 47, 1986, pp. 57-85.
- BERRETTA 1989 = MONICA BERRETTA, *Sulla presenza dell'accusativo preposizionale in italiano settentrionale: note tipologiche*, in «Vox Romanica», 48, 1989, pp. 13-37.
- BOSSONG 1985 = GEORG BOSSONG, *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen, Narr, 1985.
- BOSSONG 1985 = GEORG BOSSONG, *Differential Object Marking in Romance and Beyond*, in *New Analyses in Romance Linguistics*, Selected papers from the Linguistic Symposium on Romance Languages XVIII, Urbana-Champaign, 7-9 aprile 1988, 1985.
- COMRIE 1989 = BERNARD COMRIE, *Language, Universals and Linguistic Typology*, Chicago, The University of Chicago Press, 1989.
- CROFT 2003 = WILLIAM CROFT, *Typology and Universals*, Cambridge, Cambridge University Press, 2003.
- DE ANGELIS 2019 = ALESSANDRO DE ANGELIS, *Articolo espletivo e marcatura differenziale dell'Oggetto nel dialetto reggino di San Luca*, in «L'Italia Dialettale. Rivista di dialettologia italiana», LXXX, 2019, pp. 59-76.
- DIXON 1979 = ROBERT M.W. DIXON, *Ergativity*, Cambridge, Cambridge University Press, 1979.

La Marcatura Differenziale dell'Oggetto in Calabria

- GUARDIANO 2010 = CRISTINA GUARDIANO, *L'oggetto diretto preposizionale in siciliano. Una breve rassegna e qualche domanda*, in JACOPO GARZONIO, *Studi sui dialetti della Sicilia*, Padova, Unipress, 2010, pp. 83-101.
- IEMMOLO 2010 = GIORGIO IEMMOLO, *Topicality and differential object marking: Evidence from Romance and beyond*, in «Studies in Language», 34, 2, 2010, pp. 239-272.
- LEDGEWAY 2000 = ADAM LEDGEWAY, *A Comparative Syntax of the Dialects of Southern Italy: A Minimalist Approach*, Oxford, Blackwell, 2000.
- LEDGEWAY 2019 = ADAM LEDGEWAY, *L'accusativo preposizionale: parametri di variazione*, in *Dialetti: per parlare e per parlarne*, Sesto convegno internazionale di dialettologia, Progetto dell'Atlante Linguistico della Basilicata, Università degli Studi della Basilicata, 10-12 maggio 2019.
- LEDGEWAY ET AL. 2019 = ADAM LEDGEWAY, NORMA SCHIFANO, GIUSEPPINA SILVESTRI, *Differential Object Marking and the properties of D in the dialects of the extreme south of Italy*, in «Glossa: a journal of general linguistics», IV, 2019, art. 51.
- LOMBARDI VALLAURI-SUZUKI 2013 = EDOARDO LOMBARDI VALLAURI, SHINGO SUZUKI, *L'oggetto preposizionale in romeno e in italiano*, in «Studia Romanica», 46, 2013, pp. 11-19.
- MANZINI-SAVOIA 2005 = MARIA RITA MANZINI, LEONARDO MARIA SAVOIA, *I dialetti italiani e romanci. Morfosintassi generativa*, vol. 2, Alessandria, Edizioni Dell'Orso, 2005.
- MARTINO 1991 = PAOLO MARTINO, *L'«Area Lausberg». Isolamento e arcaicità*, Roma, Dipartimento di studi glottoantropologici dell'Università di Roma La Sapienza, 1991.
- ROHLFS 1969 = GERHARD ROHLFS, *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassi e formazione delle parole*, Torino, Einaudi, 1969.
- ROHLFS 1971 = GERHARD ROHLFS, *Autour de l'accusatif prépositionnel dans les langues romanes: concordances et discordances*, in «Revue de linguistique romane», 35, 1971, pp. 312-334.
- ROHLFS 1984 = GERHARD ROHLFS, *La Sicilia nei secoli. Profilo storico e linguistico*, Palermo, Sellerio, 1984.
- SILVERSTEIN 1976 = MICHAEL SILVERSTEIN, *Hierarchy of features and ergativity*, in *Grammatical categories in Australian languages*, a cura di Robert M.W. Dixon, Canberra, Australian Institute of Aboriginal Studies, 1976, pp. 112-171.
- ZATTA 1784 = ANTONIO ZATTA, *Atlante Novissimo illustrato ed accresciuto sulle osservazioni e scoperte fatte dai più celebri e più recenti geografi*, tomo III, Venezia, 1784.

Giusy Truncellito

Riassunto La Marcatura Differenziale dell'Oggetto (MDO) è un fenomeno ampiamente attestato e studiato nelle lingue del mondo. Per quanto riguarda la Calabria, tuttavia, gli studi più recenti si sono concentrati maggiormente sulla zona meridionale della regione. Lo scopo di questo intervento è analizzare la MDO in sei località in provincia di Cosenza, tutte appartenenti all'area Lausberg: Alessandria del Carretto, Canna, Oriolo, Francavilla Marittima, Laino Borgo, Verbicaro. I dati utilizzati sono stati ottenuti mediante la somministrazione di un questionario a parlanti nativi. Obiettivo del contributo è approfondire il tema della MDO nei dialetti della Calabria settentrionale dell'area Lausberg, tenendo conto dei parametri semantici di animatezza, definitezza e specificità, nonché delle caratteristiche formali degli oggetti diretti con i quali la marca è legittimata.

Abstract Differential Object Marking is a widely documented and studied phenomenon in the languages of the world. However, as far as the Italian region of Calabria is concerned, the most recent studies have focused more on the southern part of the region. The aim of this paper is to analyse DOM in six villages in the province of Cosenza, all belonging to the Lausberg area: Alessandria del Carretto, Canna, Oriolo, Francavilla Marittima, Laino Borgo, Verbicaro. The data used was collected through the administration of a questionnaire to native speakers. The main aim of the paper is to investigate DOM in the Northern Calabrian dialects of the Lausberg area, taking into account the semantic parameters of animacy, definiteness and specificity, as well as the formal characteristics of the direct objects with which DOM is licensed.

Sistemi di subordinazione nella Calabria meridionale: il caso del verbo ‘andare’

Carmelina Toscano

1. La pseudo-coordinazione e le strutture con il complementatore

È noto che nel Meridione estremo d'Italia, i costrutti all'infinito sono spesso sostituiti da strutture di modo finito, siano esse strutture con il complementatore¹ oppure pseudo-coordinazioni. Tale sostituzione è frequente con una classe ristretta di predicati, soprattutto modali e di movimento, e risulta sistematica quando il verbo principale è ‘andare’ (cfr. LEDGEWAY 2013).

- 1** Per “strutture con il complementatore” si fa riferimento a strutture in cui la dipendente presenta un verbo di modo finito, invariabilmente al presente indicativo, preceduto da un complementatore. Una frase del tipo ‘lo voglio vedere’, con dipendente di modo infinito e soggetto coreferente, in una varietà calabrese meridionale, come ad esempio quella di Melito di Porto Salvo, diventa: *vuliva*^{IMPRES.IND.1SG} *mu’u viju*^{PRES.IND.1SG}, letteralmente ‘volevo che lo vedo’. L’esempio dimostra come la dipendente di modo finito, al tempo presente, sia introdotta dal complementatore *mu*, corrispondente all’italiano ‘che’. Tali strutture sostituiscono le corrispettive all’infinito (‘lo voglio vedere’ > ‘voglio che lo vedo’) diffuse nella maggior parte delle varietà romanze occidentali. Il processo di sostituzione della dipendente all’infinito con una di modo finito preceduta da un complementatore è un fenomeno noto come “perdita dell’infinito”. Esso rappresenta uno dei calchi morfo-sintattici più evidenti che le varietà italo-romanze meridionali estreme devono all’azione dell’italo-greco, da cui sono state profondamente influenzate (cfr. DE ANGELIS 2013, 2016, 2017 a, b; LEDGEWAY 2007, 2013, 2016; FANCIULLO 1985; ROHLFS 1933, 1969, 1972).

Il verbo ‘andare’, in molte lingue romanze e non romanze, codifica costrutti pseudo-coordinativi. Si tratta di costrutti composti da due predicati di modo finito, i quali condividono il soggetto e presentano gli stessi tratti flessivi e temporali, esibendo dunque proprietà monofrasali. I due verbi possono essere coordinati per asindeto (cfr. 1.a), oppure, opzionalmente, collegati da un connettore: *e* (< ET) (cfr. 1.c) oppure *a* (l’origine, se da AC o da AD, è piuttosto discussa²) (cfr. 1.b). Sebbene tale costruzione formalmente sembri una struttura coordinata, a differenza di quest’ultima non codifica due eventi in successione ma un unico evento definito dal secondo verbo (=V2). Il primo verbo (=V1) può subire un processo di grammaticalizzazione in seguito al quale può apparire in una forma non flessa e invariabile e perdere il suo valore lessicale (cfr. 2 in cui *va* è una forma invariabile di 2SG usata in luogo di *iti* di 2PL). L’esito della sua desemantizzazione è ravvisabile nella possibilità del V1 di concatenarsi ad un V2 stativo. Tale possibilità, infatti, non sarebbe stata ammessa se ‘andare’ avesse mantenuto il suo significato di movimento perché quest’ultimo risulterebbe incompatibile con la semantica stativa del V2 (cfr. 2).

1.

a. *va* *stjinni* *li* *rrobbi* *vagnati*
 andare.IMP.2SG stendere.IMP.2SG *i* *vestiti* *bagnati*
 ‘*va*’ (e) stendi *i* *vestiti* *bagnati*’ (Delia [CL]; DI CARO 2019, p. 33)

2 La particella *a* è ricondotta, da alcuni studiosi, alla preposizione latina AD delle costruzioni infinitivali, le quali, nelle varietà meridionali estreme, spesso co-ricorrono con la pseudo-coordinazione (cfr. MANZINI-SAVOIA 2005, pp. 688-689) (cfr. es. *vaju*_{PRS.IND.1SG} *a mmanfu*_{PRS.IND.1SG} vs *vaju*_{PRS.IND.1SG} *a mmanfari*_{INF} [Modica – RG; MANZINI-LORUSSO-SAVOIA 2017, p. 19]). Tuttavia, c’è chi ritiene che l’elemento *a* della pseudo-coordinazione sia solo omofono ad *a* dei costrutti infinitivali e che, quindi, non derivi da AD ma dalla congiunzione latina AC (ASCOLI 1896, 1901; CARDINALETTI-GIUSTI 2001, 2003; ROHLFS 1969; SORRENTO 1950). Si consideri, però, che se si riconducesse *a* alla congiunzione AC significherebbe ammettere che *a* rappresenta l’unica continuazione di AC in italo-romanzo. Viceversa, se *a* fosse ricondotta alla preposizione latina AD, allora non si tratterebbe più di pseudo-coordinazione perché AD e i suoi continuatori romanzi introducono generalmente una dipendente infinitivale.

Sistemi di subordinazione nella Calabria meridionale

- b. u von a ffaʃʃəɲə
lo.CL andare.PRS.IND.3PL a fare.PRS.IND.3PL
lett. 'lo vanno a fanno'; 'vanno a farlo' (Putignano [BA];
MANZINI-LORUSSO-SAVOIA 2017, p. 17)
- c. sutta a la te finestra vegnu e staiu
sotto alla tua finestra venire.PRS.IND.1SG e stare.PRS.IND.1SG
lett. 'sotto alla tua finestra vengo e sto'; 'vengo a stare sotto la tua finestra'
(Reggio Calabria; ROHLFS 1969, p. 164)
2.
va stativi unni siti vui
andare.IMP.2SG stare.IMP.2PL dove essere.PRS.IND.2PL voi
lett. 'vai state dove siete voi' ('state/restate dove siete') (PITRÈ 1993 [1875], IV,
p. 150)

Nelle varietà italo-romanze meridionali estreme, gli enunciati di questo tipo possono essere codificati anche da altri costrutti di modo finito, ossia quelli introdotti da un complementatore³ (cfr. 3 per il salentino e 4 per il calabrese). La pseudo-coordinazione e le strutture con il complementatore sono costruzioni dalla genesi e dalle proprietà semantico-sintattiche molto diverse⁴ che co-esistono all'interno di uno

3 Il complementatore delle varietà meridionali estreme continua l'avverbio latino *MODO* e si realizza come *mu/u*, *mi/i*, *ma* in Sicilia e in Calabria; mentre, nel salentino, *cu* rappresenta la continuazione di *QUOD*. Per ulteriori approfondimenti sull'origine dei complementatori si veda DE ANGELIS 2016.

4 Le strutture con il COMP rappresentano l'esito di un lungo processo di interferenza tra il greco e il romanzo (cfr. almeno DE ANGELIS 2013, 2017a; LEDGEWAY 2013; ROHLFS 1933, 1969, 1972) e sono strutture bifrasali. Le strutture pseudo-coordinative sono, invece, monofrasali. Nei costrutti con il complementatore il clitico può collocarsi tra il V1 e il V2, mentre nelle pseudo-coordinazioni il clitico risale sempre al verbo matrice. Nei primi sono ammessi argomenti del V1 mentre non lo sono nelle pseudo-coordinazioni. Inoltre, nelle pseudo-coordinazioni esistono delle restrizioni sulla selezione del V1, il quale può essere anche ridotto fonologicamente, possibilità non prevista nelle strutture con il complementatore. Per esempi e ulteriori dettagli si veda CARDINALETTI-GIUSTI 2020.

Carmelina Toscano

stesso territorio e che, spesso, sembrerebbero intercambiabili tra loro (cfr. 3, 4).

3.

a. va e ppigghia a zzappa
andare.PRS.IND.3SG e prendere.PRS.IND.3SG la zappa

b. va mi pigghia a zzappa
andare.PRS.IND.3SG COMP prendere.PRS.IND.3SG la zappa
'va a prendere la zappa' (Melito di Porto Salvo [RC])

4.

a. lu vi'nia a ffa'f'ia
lo.CL venire.PF.IND.1SG a fare.PF.IND.1SG

b. vi'nia ku llu fattsu
venire.PF.IND.1SG COMP lo.CL fare.PRS.IND.1SG
'venni a farlo' (Mesagne [BR]; MANZINI-SAVOIA 2005, pp. 691-692)

Se la pseudo-coordinazione e le strutture con il complementatore sono state oggetto di studio privilegiato, lo stesso non può dirsi dell'analisi della loro co-esistenza e delle relative dinamiche di alternanza. Menzionata la presenza dei due costrutti all'interno della stessa area e la loro possibile intercambiabilità, nella Calabria centro-meridionale non sono stati mai indagati eventuali parametri di variazione o di selezione dell'uno o dell'altro costrutto. Recenti inchieste sul campo mettono, però, in luce delle nuove evidenze sia rispetto ai contesti in cui le due strutture possono o meno essere usate in maniera intercambiabile, sia rispetto a dei nuovi parametri di variazione e restrizione relativi alla pseudo-coordinazione. In base a questi ultimi, infatti, emerge come specifici fattori semantici inerenti all'animatezza del soggetto o alla classe di appartenenza del secondo verbo siano determinanti nella selezione dell'una o dell'altra struttura.

2. Il modo imperativo

2.1. Le varietà calabro-reggine settentrionali

Il modo imperativo è certamente quello in cui la pseudo-coordinazione risulta più produttiva: si riscontra, infatti, in tutte le varietà reggine sia alla 2SG che alla seconda persona plurale (=2PL). In particolare, nelle varietà calabro-reggine settentrionali⁵ la pseudo-coordinazione è attestata prevalentemente nella sua costruzione asindetica sia alla 2SG che alla 2PL.

5.

a. va cogghi a liva
andare.IMP.2SG raccogliere.IMP.2SG le olive
'vai a raccogliere le olive!' (Rizziconi [RC])

b. va mbuscati ddu liri
andare.IMP.2SG guadagnare.IMP.2SG+CL due lire
'vai a guadagnarti due lire!' (San Ferdinando [RC])

6.

a. jati curcativi
andare.IMP.2PL coricare.IMP.2PL
'andate a coricarvi!' (Feroletto della Chiesa [RC])

b. iti cercàtilu
andare.IMP.2PL cercare.IMP.2PL+CL
'andate a cercarlo!' (Rizziconi [RC])

5 Con "varietà calabro-reggine settentrionali" ci si riferisce alle varietà di San Ferdinando, Rizziconi, S. Ilario dello Jonio, Feroletto della Chiesa, Mammola e Siderno. Con "varietà calabro-reggine centro-meridionali" ci si riferisce all'insieme di varietà centrali e meridionali della provincia di Reggio. Quelle centrali sono composte dai dialetti di Castellace, Oppido Mamertina, Cosoleto, Santa Cristina e Delianuova. Nel gruppo delle varietà meridionali rientrano, invece, le varietà di Melito di Porto Salvo, San Lorenzo, Bova Marina, Bianco, Brancaleone, Cardeto e Roccaforte.

Nelle varietà calabro-reggine centro-meridionali, il tipo asindetico si riscontra esclusivamente alla 2SG (cfr. 7.a-b) e sempre in alternanza con il tipo dotato di connettore *e*⁶. Tale particella può essere esplicita (cfr. 7.a", b") oppure silente (cfr. 7.a', b'), ossia non realizzata in superficie ma presente nella struttura soggiacente, come dimostra la presenza del raddoppiamento fonosintattico in 7.a'-b' in un contesto che di per sé non lo innescherebbe (come dimostra la realizzazione in 7.a). Il raddoppiamento fonosintattico si giustifica, infatti, solo supponendo che la particella, non emersa in superficie, sia presente nella struttura soggiacente e ne lasci traccia nel raddoppiamento della consonante iniziale del secondo verbo che essa stessa innescherebbe.

7.

a. va **m**angia
andare.IMP.2SG mangiare.IMP.2SG

a'. va **mm**angia
andare.IMP.2SG mangiare.IMP.2SG

a". va e **mm**angia
andare.IMP.2SG e mangiare.IMP.2SG
'vai a mangiare!' (Melito di Porto Salvo [RC])

b. va **z**appa
andare.IMP.2SG zappare.IMP.2SG

b'. va **zz**appa
andare.IMP.2SG zappare.IMP.2SG

b". va e **zz**appa
andare.IMP.2SG e zappare.IMP.2SG
'vai a zappare!' (Delianuova [RC])

6 I dati AsiCa e AIS riportano comunque delle attestazioni della particella *a* nella Calabria settentrionale in provincia di Crotone.

Sistemi di subordinazione nella Calabria meridionale

A differenza di quanto si riscontra alla 2SG dell'imperativo, alla 2PL la giustapposizione del primo e del secondo verbo, nelle varietà centro-meridionali, non è ammessa in nessun caso. Infatti, il connettore *e* è sempre presente nella sua forma esplicita, altrimenti il costrutto risulterebbe agrammaticale (cfr. 8).

8.

- a. iti *(e) rrivatilu
 andare.IMP.2PL *(e) arrivare.IMP.2PL+CL
 'andate a raggiungerlo!' (Cosoletto [RC])
- b. iti *(e) ddurmiti
 andare.IMP.2PL *(e) dormire.IMP.2PL
 'andate a dormire!' (Santa Cristina [RC])

La variazione riscontrata al modo imperativo consente, dunque, di individuare diverse realizzazioni dello stesso costrutto a seconda della varietà considerata e della persona verbale coinvolta. Tale stato di cose, sintetizzato alla tabella 1, dimostra come le varietà calabro-reggine settentrionali siano le uniche a codificare sempre il costrutto asindetico, mentre le varietà calabro-reggine centro-meridionali lo ammettano esclusivamente alla 2SG e pur sempre in alternanza con il tipo dotato di particella, espressa o taciuta.

		Varietà calabro-reggine settentrionali	Varietà calabro-reggine centrali	Varietà calabro-reggine meridionali
IMPERATIVO	2SG	[V1+V2]	[V1+V2]	[V1+V2]
			[V1 (+e) V2]	[V1 (+e) V2]
			[V1 + e + V2]	[V1 + e + V2]
	2PL	[V1+V2]	[V1 + e + V2]	[V1 + e + V2]

Tabella 1

3. Il modo indicativo

3.1. L'indicativo presente

Se al modo imperativo è ammessa solo la pseudo-coordinazione, lo stesso non può dirsi del presente indicativo, dove si riscontrano anche i costrutti con il complementatore *(m)i/(m)u*. Data la co-esistenza della pseudo-coordinazione e della struttura con il complementatore *(m)i/(m)u* all'interno di uno stesso territorio (cfr. es. 3), spesso ci si è limitati ad osservare come l'una sembri alternarsi liberamente all'altra. Tuttavia, le due strutture non sono sempre intercambiabili; recenti inchieste che ho condotto sul campo rivelano una situazione linguistica ben più complessa che rende necessario considerare due parametri fondamentali: la variazione diatopica e la semantica del secondo verbo. A questo proposito, nelle varietà reggine settentrionali la struttura con il complementatore *(m)i/(m)u* è nettamente prevalente rispetto a quella pseudo-coordinativa codificata solo quando il verbo 'andare' è connesso ad un secondo verbo di attività (cfr. 9.a, b). Si osservi, inoltre, come la pseudo-coordinazione co-ricorra con la struttura con il complementatore *(m)i/(m)u* e da questo possa essere sempre sostituita (cfr. 9).

9.

a. vaju e ccattu u pani
andare.PRS.IND.ISG e comprare.PRS.IND.ISG il pane
'vado a comprare il pane' (S. Ilario dello Jonio)

a'. vaju (m)u ccattu u pani
andare.PRS.IND.ISG COMP comprare.PRS.IND.ISG il pane
'vado a comprare il pane' (S. Ilario dello Jonio)

b. chidu vaci e ccamina puru sutta
quello andare.PRS.IND.3SG e camminare.PRS.IND.3SG pure sotto

all'acqua

all'acqua

'quello va a camminare pure sotto la pioggia' (Mammola [RC])

Sistemi di subordinazione nella Calabria meridionale

b'. chidu vaci (m)u ccamina puru sutta
quello andare.PRS.IND.3SG COMP camminare.PRS.IND.3SG pure sotto

all'acqua

all'acqua

'quello va a camminare pure sotto la pioggia' (Mammola [RC])

Una particolarità delle varietà calabro-reggine settentrionali è la manifestazione di una giustapposizione del primo e secondo verbo esclusivamente alla 1PL del presente indicativo (cfr. 10). Tale caratteristica non si riscontra in tutte le altre persone, singolari e plurali, delle varietà reggine settentrionali e neppure nelle varietà più a Sud di Castellace che necessitano, invece, della particella esplicita per essere giudicate grammaticali.

10.

a. jamu putamu a vigna?
andare.PRS.IND.1PL potare.PRS.IND.1PL la vigna?
'andiamo a potare le viti?' (San Ferdinando [RC])

b. jamu cucinamu a pasta
andare.PRS.IND.1PL cucinare.PRS.IND.1PL la pasta
'andiamo a cucinare la pasta' (Feroletto [RC])

c. jamu guardamu comu puta?
andare.PRS.IND.1PL guardare.PRS.IND.1PL come potare.PRS.IND.3SG
'andiamo a guardare come pota?' (Rizziconi [RC])

Dell'alternanza strutturale sopra illustrata, nelle varietà reggine settentrionali non si riscontra traccia in nessun altro caso. In queste varietà, infatti, se il secondo verbo appartiene a una classe diversa di predicati (es. culminativi, risultativi o stativi), allora la pseudo-coordinazione non è ammessa e il costrutto con il complementatore *(m)i/(m)u* è l'unico ad essere codificato (cfr. 11.a-c).

Carmelina Toscano

11.

a. jati vui u svacantati a bumbola?
andare.PRS.IND.2PL voi COMP svuotare.PRS.IND.2PL la bombola?
'andate voi a svuotare la bombola?' (San Ferdinando [RC])

b. idi vannu (m)u ddormunu ansemi o
loro andare.PRS.IND.3PL COMP dormire.PRS.IND.3PL insieme al

figghiolu

bambino

'loro vanno a dormire insieme al bambino' (Siderno [RC])

c. vai u capisci sti mbrogghi
andare.PRS.IND.2SG COMP capire.PRS.IND.2SG queste bugie
'vai a capire queste bugie!' (Rizziconi [RC])

A differenza delle varietà reggine settentrionali, quelle centro-meridionali manifestano un uso molto più esteso della pseudo-coordinazione. Quest'ultima, infatti, si riscontra – pur sempre in alternanza con il costrutto dotato di complementatore *(m)i/(m)u* – sia se il V2 è un verbo di attività sia se il V2 appartiene alla classe dei predicati culminativi e risultativi (cfr. 12).

12.

a. vannu e ccogghiunu ddu cocci i racina
andare.PRS.IND.3PL e raccogliere.PRS.IND.3PL due chicchi di uva
'vanno a raccogliere due chicchi di uva' (Cosoletto [RC])

a'. vannu mu ccogghiunu ddu cocci i racina
andare.PRS.IND.3PL COMP raccogliere.PRS.IND.3PL due chicchi di uva
'vanno a raccogliere due chicchi di uva' (Cosoletto [RC])

b. vaju e nnettu ddu pisci pi stasira
andare.PRS.IND.1SG e pulire.PRS.IND.1SG due pesci per stasera
'vado a pulire due pesci per stasera' (Oppido Mamertina [RC])

Sistemi di subordinazione nella Calabria meridionale

b'. vaju mu nnettu ddu pisci pi stasira
 andare.PRS.IND.1SG COMP pulire.PRS.IND.1SG due pesci per stasera
 'vado a pulire due pesci per stasera' (Oppido Mamertina [RC])

c. non iti e u rrivati mi
 non andare.PRS.IND.2PL e lo.CL arrivare.PRS.IND.2PL COMP

avimu carica notizia i cchiù?
 avere.PRS.IND.2PL qualche notizia in più
 'non andate a raggiungerlo per avere qualche notizia in più?' (San Lorenzo [RC])

c'. non iti mu'u rrivati mi
 non andare.PRS.IND.2PL COMP+lo.CL arrivare.PRS.IND.2PL COMP

avimu
 avere.PRS.IND.2PL

carica notizia i cchiù?
 qualche notizia in più
 'non andate a raggiungerlo per avere qualche notizia in più?' (San Lorenzo [RC])

Le varietà centro-meridionali hanno finora mostrato un'identità di risultati, in opposizione alle varietà più settentrionali dove l'uso della pseudo-coordinazione risulta meno frequente e più circoscritto. Tale identità, però, non trova riscontro nei contesti in cui il costrutto presenta un secondo verbo stativo. Se, infatti, il V2 è un predicato stativo allora le varietà centrali codificano esclusivamente il costrutto con il complementatore *(m)i/(m)u* (cfr. 13), mentre le varietà più meridionali ammettono anche la pseudo-coordinazione (cfr. 14).

Carmelina Toscano

13.

- a. sicundu tia vaju mu criu a sti
secondo te andare.PRS.IND.1SG COMP credere.PRS.IND.1SG a queste

scemenzi?

scemenze?

‘secondo te credo a queste stupidaggini?’ (Delianuova [RC])

- b. si vannu mu si stannu fermi
se andare.PRS.IND.3PL COMP si stare.PRS.IND.3PL fermi

fannu cchiù bella figura

fare.PRS.IND.3PL più bella figura

‘se stanno fermi, fanno più bella figura’ (Cosoletto [RC])

14.

- a. sicundu tia vaju e criu a sti
secondo te andare.PRS.IND.1SG e credere.PRS.IND.1SG a queste

scemenzi?

scemenze?

‘secondo te credo a queste stupidaggini?’ (Brancaleone [RC])

- b. si vannu e si stannu fermi
se andare.PRS.IND.3PL e si stare.PRS.IND.3PL fermi

fannu cchiù bella figura

fare.PRS.IND.3PL più bella figura

‘se stanno fermi, fanno più bella figura’ (Cardeto [RC])

Si osservi, inoltre, che se il primo verbo si concatena a un secondo verbo stativo per esprimere un ordine perentorio allora la pseudo-coordinazione diventa prevalente nelle varietà meridionali della costa jonica e, presso alcuni parlanti, persino esclusiva (cfr. 15).

Sistemi di subordinazione nella Calabria meridionale

15.

- a. va statti settatu
 andare.IMP.2SG stare.IMP.2SG+CL sedere.PPT
 'stai seduto!' (Melito di Porto Salvo [RC])
- b. va ppensa chi nda ffari
 andare.IMP.2SG pensare.IMP.2SG che devi fare
 'pensa a cosa fare!' (Bianco [RC])
- c. va ccapisci sti trapuli
 andare.IMP.2SG capire.IMP.2SG questi intrallazzi
 'vai a capire questi intrallazzi!' (San Lorenzo [RC])

I dati offerti dalle varietà italo-romanze della provincia di Reggio Calabria rivelano, dunque, una forte micro-variazione, in base alla quale il costruito con il complementatore *(m)i/(m)u* risulta più frequente, più diffuso diatopicamente e senza alcun tipo di restrizione morfo-sintattica o semantica. Accanto a tale struttura, in alcuni contesti più di altri, viene codificata anche la pseudo-coordinazione. Quest'ultima, però, è soggetta a delle restrizioni sui V2 ammessi nel costruito. Di tali restrizioni non tutte le varietà esibiscono un'identità di risultati: come sintetizzato alla tabella 2, le varietà settentrionali ammettono solo i V2 di attività, quelle centrali pure V2 risultativi e cumulativi, quelle più meridionali anche predicati stativi.

INDICATIVO PRESENTE	[V1+V2] = 1PL [V1 + e + V2] COMP	[V1 + e + V2] COMP	[V1 + e + V2] COMP
V2 attività	COMP ± PseCo	COMP + PseCo	COMP + PseCo
V2 risultativi	COMP	COMP + PseCo	COMP + PseCo
V2 culminativi	COMP	COMP + PseCo	COMP + PseCo
V2 stativi	COMP	COMP	COMP + PseCo

Tabella 2

3.2. L'indicativo preterito

A differenza dell'indicativo presente ove la micro-variazione mostra usi e restrizioni diversi a seconda delle varietà considerate, al passato tutte le varietà reggine concordano nella codifica di un unico costrutto ossia quello con il complementatore *(m)i/(m)u* (cfr. 16.a, b, c). La pseudo-coordinazione, infatti, sembrerebbe non essere mai ammessa al passato in nessuna varietà calabro-reggina (cfr. 16.a', b', c').

16.

a. Isti u fui cu stu caddu?
 andare.PF.IND.2SG COMP correre.PRS.IND.2SG con questo caldo?
 'sei andato a correre con questo caldo?' (Oppido Mamertina [RC])

a'. *Isti e fujisti cu stu caddu?
 andare.PF.IND.2SG e correre.PF.IND.2SG con questo caldo?
 'sei andato a correre con questo caldo?' (Oppido Mamertina [RC])

b. chi isti u dici an giru?
 che andare.PF.IND.2SG COMP dire.PRS.IND.2SG in giro?
 'che cosa sei andato a dire in giro?' (Delianuova [RC])

b'. *chi isti e dicisti an giru?
 chi andare.PF.IND.2SG e dire.PF.IND.2SG in giro?
 'che cosa sei andato a dire in giro?' (Delianuova [RC])

c. iu u si ssetta propriu nta seggia
 andare.PF.IND.3SG COMP si sedere.PRS.IND.3SG proprio nella sedia

rrutta

rotta

'è andato a sedersi proprio nella sedia rotta' (Rizziconi [RC])

c'. *iu e si ssettau propriu nta seggia rrutta
 andare.PF.IND.3SG e si sedere.PF.IND.3SG proprio nella sedia rotta
 'andò a sedersi proprio nella sedia rotta' (Rizziconi [RC])

4. Animatezza del soggetto

Si è visto finora come la pseudo-coordinazione sia usata all'imperativo, mentre il costrutto con il complementatore $(m)i/(m)u$ sia codificato all'indicativo passato. Entrambe le strutture si riscontrano in co-ricorrenza solo all'indicativo presente, ma non sono sempre in variazione libera, dal momento che in specifici contesti, come abbiamo visto, mostrano distribuzione complementare (cfr. § 3 e tabella 3). L'unico caso in cui entrambe le strutture sembrerebbero ammesse liberamente in tutte le varietà è quando il V1 'andare' è unito a un V2 di attività. Tuttavia, ancora una volta, recenti inchieste che ho condotto sul campo rivelano come la codifica dell'uno o dell'altro costrutto non sia libera ma vincolata a parametri semantici inerenti ai tratti di animatezza e umanità del soggetto. Ad eccezione delle varietà calabro-reggine più settentrionali, dove la struttura con il complementatore $(m)i/(m)u$ prevale, i dati più significativi emergono proprio dalle varietà centrali e meridionali della provincia di Reggio Calabria. Entrambi i gruppi di varietà ammettono sia la pseudo-coordinazione che la struttura con il complementatore $(m)i/(m)u$ se il soggetto presenta i tratti [+animato] e [+umano] (cfr. 17).

17.

a. vabbò eu vaju e mmangiu
vabbè io andare.PRS.IND.1SG e mangiare.PRS.IND.1SG
'vabbè, io vado a mangiare' (Melito di Porto Salvo [RC])

b. imu e pputamu a vigna
andare.PRS.IND.1PL e potare.PRS.IND.1PL la vigna
'andiamo a potare le viti' (Delianuova [RC])

c. vannu u ccogghiunu ddu fica
andare.PRS.IND.3PL COMP raccogliere.PRS.IND.3PL due fichi
'vanno a raccogliere due fichi' (Cosoletto [RC])

Se, invece, il soggetto perde il tratto di umanità [-umano], allora le varietà reggine centrali non ammettono più la pseudo-coordinazione e codificano esclusivamente il costrutto con il complementatore

Carmelina Toscano

(*m*)/(*m*)*u* (cfr. 18) laddove le varietà calabro-reggine meridionali continuano, invece, a presentare pseudo-coordinazione (cfr. 19).

18.

- a. i cani vannu mu cercanu u
i cani andare.PRS.IND.3PL COMP cercare.PRS.IND.3PL COMP

mangianu
mangiare.PRS.IND.3PL
'i cani vanno a cercare da mangiare' (Castellace [RC])

- b. i cavaji vannu u si mangiunu u
i cavalli andare.PRS.IND.3PL COMP si.CL mangiare.PRS.IND.3PL il

frenu
fieno
'i cavalli vanno a mangiare il fieno' (Oppido Mamertina [RC])

19.

- a. i crapi vannu e ggiriunu nta ddi
le capre andare.PRS.IND.3PL e girare.PRS.IND.3PL in quelle

muntagni
montagne

chini d'erba
piene d'erba
'le capre pascolano sempre in quelle montagne piene di erba' (Melito di Porto Salvo [RC])

- b. non sacciu quali gaju vaci e
non sapere.PRS.IND.1SG quale gallo andare.PRS.IND.3SG e

bbecca
beccare.PRS.IND.3SG

i me jadini
le mie galline
'non so quale gallo va a beccare le mie galline' (Cardeto [RC])

La codifica del costrutto pseudo-coordinativo si arresta, invece, in tutte le varietà quando il soggetto perde il tratto dell'animatezza. In quest'ultimo caso, infatti, si riscontrano due diverse strategie nella codifica dell'enunciato. La prima è la struttura con il complementatore *(m)i/(m)u* (cfr. 20.a), l'altra, invece, riguarda la possibilità di riformulare l'enunciato eliminando il verbo 'andare'. Tale stato di cose lascia supporre che, molto probabilmente, il V1 conservi ancora il suo valore lessicale, per cui 'andare', in quanto verbo di movimento, è incompatibile con un soggetto inanimato e proprio per questo motivo è eliminato dalla struttura (es. 20.b).

20.

- a. peccatu chi tutta st'acqua (*va)
 peccato che tutta quest'acqua (*andare.PRS.IND.3SG)

cadi nto tumbinu
 cadere.PRS.IND.3SG nel tombino
 'peccato che tutta quest'acqua cade nel tombino' (Roccaforte [RC])

- b. non sacciu aundi vannu u si
 non sapere.PRS.IND.1SG dove andare.PRS.IND.3PL COMP si.CL

perdunu i me ugghi
 perdere.PRS.IND.3PL i miei aghi
 'non so dove vanno a finire i miei aghi' (Bianco [RC])

5. Conclusioni

La pseudo-coordinazione e la struttura con il complementatore *(m)i/(m)u* sono due strategie sintattiche ampiamente produttive nell'Italia meridionale estrema. Come abbiamo osservato, ambedue le strutture coesistono all'interno di uno stesso territorio e, spesso, sembrerebbero essere in concorrenza, intercambiabili tra di loro. Tuttavia, abbiamo mostrato che non sempre l'alternanza delle due costruzioni è ammessa, dal momento che, se in certe condizioni la loro selezione è libera, in

Carmelina Toscano

altri contesti è complementare e segue parametri ben precisi. In particolare, i dati da me raccolti nel corso del 2022 nella provincia di Reggio Calabria consentono di dimostrare come la pseudo-coordinazione sia codificata al modo imperativo e come il costruito con il complementatore $(m)i/(m)u$ sia usato in modo esclusivo all'indicativo presente. L'alternanza delle due costruzioni si verifica, quindi, solo all'indicativo presente: la struttura con il complementatore $(m)i/(m)u$ in questo caso è sempre usata, senza restrizioni di alcun tipo. Esistono, però, dei contesti in cui, più di altri, la struttura con il complementatore $(m)i/(m)u$ può co-ricorrere con la pseudo-coordinazione. Come si illustra nella tabella 3, le varietà calabro-reggine settentrionali ammettono la pseudo-coordinazione solo se il secondo è un verbo di attività, mentre in quelle centro-meridionali si riscontra anche con un V2 risultativo e culminativo. Solo le varietà più meridionali, specie quelle lungo la costa jonica della provincia di Reggio Calabria, codificano pseudo-coordinazione anche con i secondi verbi stativi.

	Varietà calabro-reggine settentrionali	Varietà calabro-reggine centrali	Varietà calabro-reggine meridionali
IMPERATIVO	[V1+V2]	[V1+V2] [V1 + e + V2]	[V1+V2] [V1 + e + V2]
INDICATIVO PRESENTE	[V1+V2] = 1PL [V1 + e + V2] COMP	[V1 + e + V2] COMP	[V1 + e + V2] COMP
V2 attività	COMP ± PseCo	COMP + PseCo	COMP + PseCo
V2 risultativi	COMP	COMP + PseCo	COMP + PseCo
V2 culminativi	COMP	COMP + PseCo	COMP + PseCo
V2 stativi	COMP	COMP	COMP + PseCo
IND. PRETERITO	COMP	COMP	COMP

Tabella 3

Si evince, pertanto, che il contesto prototipico in cui la pseudo-coordinazione si manifesta in co-ricorrenza con il costruito con il complementatore $(m)i/(m)u$ è all'indicativo presente, se il primo verbo

Sistemi di subordinazione nella Calabria meridionale

è 'andare' e il secondo è un verbo di attività. Queste, infatti, sono le uniche condizioni in cui tutte le varietà presentano la medesima alternanza strutturale. Entro queste condizioni, la pseudo-coordinazione e la struttura con il complementatore $(m)i/(m)u$ possono ricorrere in variazione libera o in distribuzione complementare a seconda dei tratti $[\pm \text{animato}]$ e $[\pm \text{umano}]$ (cfr. tabella 4). A questo proposito, ci sono alcune varietà, quali quelle centrali e settentrionali, che codificano pseudo-coordinazione solo se il soggetto ha tratti di umanità. Le varietà meridionali, invece, manifestano una maggiore estensione del costrutto: ammettono pseudo-coordinazione anche se il soggetto non è umano, purché esso sia animato. Se, infatti, il soggetto perde i tratti di animatezza la sola struttura ammessa è quella con il complementatore $(m)i/(m)u$.

	Varietà calabro- reggine settentrionali	Varietà calabro- reggine centrali	Varietà calabro- reggine meridionali
INDICATIVO			
PRESENTE	[+ ANIMATO]	[+ ANIMATO]	[+ ANIMATO]
V2 attività	[+ UMANO]	[+ UMANO]	[+ UMANO]
			[+ ANIMATO]
			[- UMANO]

Tabella 4

Queste nuove evidenze mettono, quindi, in luce due importanti novità sulla pseudo-coordinazione.

Oltre ai ben noti parametri di restrizione relativi ai modi e ai tempi in cui il costrutto può apparire, emergono, infatti, anche altri due parametri fondamentali finora trascurati in letteratura: la classe semantica di appartenenza dei V2 e i tratti di semantici del soggetto.

La seconda importante novità riguarda l'individuazione dei tratti prototipici, ossia in quali contesti è più probabile che la pseudo-coordinazione sia codificata. Come si illustra nella tabella 5, si devono verificare le seguenti condizioni: il primo verbo è 'andare' e il secondo un

Carmelina Toscano

verbo di attività; il modo è l'indicativo e il tempo un presente e, infine, il soggetto presenta i tratti [+ animato] e [+ umano].

	Varietà calabro- reggine settentrionali	Varietà calabro- reggine centrali	Varietà calabro- reggine meridionali
INDICATIVO PRESENTE V2 attività	COMP ± PseCo	COMP + PseCo	COMP + PseCo
	[+ AN, + UM]	[+ AN,+ UM]	[+ AN, + UM]

Tabella 5

Riferimenti bibliografici

- ASCOLI 1896 = GRAZIADIO ISAIA ASCOLI, *Un problema di sintassi comparata dialettale*, in «Archivio glottologico italiano» [ovvero: AGI], 14, 1896, pp. 453-468.
- ASCOLI 1901 = GRAZIADIO ISAIA ASCOLI, *Appendice all'articolo Un problema di sintassi comparata dialettale*, in «Archivio glottologico italiano» [ovvero: AGI], 15, 1901, pp. 221-225.
- CARDINALETTI-GIUSTI 2001 = ANNA CARDINALETTI, GIULIANA GIUSTI, *Semi-lexical Motion Verbs in Romance and Germanic*, in *Semi-lexical categories*, a cura di Norbert Corver e Henk Van Riemsdijk, Berlino, De Gruyter, 2001, pp. 371-414.
- CARDINALETTI-GIUSTI 2003 = ANNA CARDINALETTI, GIULIANA GIUSTI, *Motion Verbs as Functional Heads*, in *The Syntax of Italian Dialects*, a cura di Christina Tortora, New York, Oxford University Press, 2003, pp. 31-49.
- CARDINALETTI-GIUSTI 2020 = ANNA CARDINALETTI, GIULIANA GIUSTI, *Multiple Agreement in Southern Italian Dialects*, in *Linguistic Variations: Structure and Interpretation - Studies in Honor of M. Rita Manzini*, a cura di Ludovico Franco e Paolo Lorusso, Berlino, Mouton De Gruyter, 2020.
- DE ANGELIS 2013 = ALESSANDRO DE ANGELIS, *Strategie di complementazione frasale nell'estremo Meridione italiano*, Messina, SGB Edizioni, 2013.
- DE ANGELIS 2016 = ALESSANDRO DE ANGELIS, *Origini formali e funzionali della particella (m)i, (m)u, ma nell'area messinese e calabrese centro-meridionale*, Atti del IV Convegno Internazionale di Dialettologia, Potenza, Castelmezza-

Sistemi di subordinazione nella Calabria meridionale

- no, Lagopesole, 6-8 novembre 2014, a cura di Patrizia del Puente, Venosa, Osanna Edizioni, 2016, pp. 75-95.
- DE ANGELIS 2017a = ALESSANDRO DE ANGELIS, *Microvariazione, diacronia e interferenza: due case studies dall'Italia meridionale estrema*, in «Archivio glottologico italiano» [ovvero: AGI], 102, 1, 2017, pp. 40-69.
- DE ANGELIS 2017b = ALESSANDRO DE ANGELIS, *Between Greek and Romance: Competing complementation systems in Southern Italy*, in *Language and Identity in Multilingual Mediterranean settings. Challenges for Historical Sociolinguistics* («Trends in Linguistics. Studies and Monographs [TiLSM]», 310), a cura di Piera Molinelli, Berlino, Mouton de Gruyter, 2017, pp. 135-156.
- DI CARO 2019 = VINCENZO NICOLÒ DI CARO, *Multiple Agreement Constructions in Southern Italo-Romance The Syntax of Sicilian Pseudo-Coordination*, Tesi di Dottorato di Ricerca, Venezia, Università Ca' Foscari Venezia, 2019.
- FANCIULLO 1985 = FRANCO FANCIULLO, *Greek and Italian in Southern Italy*, in *Studies in Greek Linguistics. Proceedings of the 6th annual Meeting of the Department of Linguistics* (Thessaloniki, 22-24 April 1985), Aristotelian University of Thessaloniki, Thessaloniki, Kyriakidis, 1985, pp. 93-106 (poi in: FRANCO FANCIULLO, *Greco e italiano nell'Italia meridionale*, in *Fra Oriente e Occidente. Per una storia linguistica dell'Italia meridionale*, Pisa, ETS, 1996, pp. 31-40).
- LEDGEWAY 2007 = ADAM LEDGEWAY, *Diachrony of Finiteness: Subordination in the Dialects of Southern Italy*, in *Finiteness: Theoretical and Empirical Foundations*, a cura di Irina A. Nikolaeva, Oxford, Oxford University Press, 2007, pp. 335-365.
- LEDGEWAY 2013 = ADAM LEDGEWAY, *Greek disguised as Romance? The case of Southern Italy*, in *Proceedings of the 5th international conference on Greek dialects and linguistic theory*, a cura di Mark Janse et al., Patras, Laboratory of Modern Greek dialects, University of Patras, 2013, pp. 184-228.
- LEDGEWAY 2016 = ADAM LEDGEWAY, *Clausal complementation*, in *The Oxford guide to Romance languages*, a cura di Adam Ledgeway e Martin Maiden, Oxford, Oxford University Press, 2016, pp. 1013-1028.
- MANZINI-SAVOIA 2005 = RITA MANZINI, LEONARDO SAVOIA, *I dialetti Italiani e Romanci. Morfosintassi Generativa*, vol. 1, Alessandria, Edizioni dell'Orso, 2005.
- MANZINI-LORUSSO-SAVOIA 2017 = RITA MANZINI, PAOLO LORUSSO, LEONARDO SAVOIA, *A/bare finite complements in Southern Italian varieties: mono-clausal or bi-clausal syntax?*, in «Quaderni di Linguistica e Studi Orientali», 3, 2017, pp. 11-59.
- PITRÈ 1993 [1875] = GIUSEPPE PITRÈ, *Fiabe, novelle e racconti popolari siciliani*, voll. I-IV, Catania, Clio, 1993.

Carmelina Toscano

ROHLFS 1933 = GERHARD ROHLFS, *Scavi linguistici nella Magna Grecia*, Milano/Halle (Saale), Hoepli, 1933.

ROHLFS 1969 = GERHARD ROHLFS, *Grammatica storica della lingua italiana e dei suoi dialetti*, vol. 3, Torino, Einaudi, 1969.

ROHLFS 1972 [1958] = GERHARD ROHLFS, *La perdita dell'infinito nelle lingue balcaniche e nell'Italia meridionale*, in ID., *Studi e ricerche su lingua e dialetti d'Italia*. Firenze, Sansoni, 1972, pp. 318-332.

SORRENTO 1950 = LUIGI SORRENTO, *Sintassi romanza: ricerche e prospettive*, Milano, Cisalpino, 1950.

Riassunto Nelle varietà italo-romanze dell'estremo Sud Italia, se il verbo reggente è il verbo 'andare', si verifica una sistematica sostituzione delle dipendenti all'infinito con costrutti di modo finito. Essi possono essere costrutti con il complementatore (es. *vaju*^{PRS.IND.1SG} *mi viju*^{PRS.IND.1SG}, lett. 'vado che vedo') oppure pseudo-coordinazioni (es. *vaju*^{PRS.IND.1SG} [*a/e*] *viju*^{PRS.IND.1SG}, lett. 'vado a/e vedo'). Queste due costruzioni, dalla genesi e dalle proprietà morfosintattiche e semantiche diverse, coesistono all'interno di uno stesso territorio e, spesso, sembrano alternarsi liberamente. Se le singole strutture sono state oggetto di studio privilegiato, lo stesso non vale per i loro rapporti di coesistenza: quali siano le dinamiche e i parametri che governano la loro alternanza o co-ricorrenza resta ancora un'incognita. Con presente contributo si esamina la situazione offerta dalla Calabria meridionale, dove tale fenomeno di alternanza strutturale è molto accentuato ma è stato poco indagato. In particolare, dopo aver esaminato la distribuzione dei due costrutti nella Calabria reggina, si analizzeranno i relativi parametri di variazione. Si argomenterà, quindi, come la pseudo-coordinazione presenti delle restrizioni in termini di i) animatezza del soggetto; ii) predicati ammessi nel costrutto; iii) modi e tempi in cui la struttura può apparire. Si dimostrerà, pertanto, come la coesistenza delle dipendenti con il complementatore (*m*) *u/(m)* e la pseudo-coordinazione, lungi dall'essere in variazione libera, rivela, invece, una distribuzione complementare. Nuove e recenti evidenze consentono, infatti, di provare come la selezione dell'una o dell'altra struttura sia governata da parametri ben precisi, piuttosto che essere intercambiabile come, invece, è spesso stato ipotizzato in letteratura.

Abstract In the extreme southern Italo-romance varieties, whenever the first verb is 'go', the canonical infinitival structure that follows (e.g. 'I go to see') is replaced with finite mood constructions. These finite mood constructions can occur with the (*m*)*u/(m)* complementiser (e.g. *vaju mi viju* Lit. 'I go that I see') or with pseudo-coordination structures (e.g. *vaju [a/e] viju*, lit. 'I go and I see'). These two constructions, with different origins and different respective morphosyntactic and semantic properties, coexist within the same

Sistemi di subordinazione nella Calabria meridionale

geographical territory and often seem to alternate freely. If the individual structures have been the object of privileged study, respectively, the same does not apply to their relationships of coexistence; what the specific dynamics and parameters that govern their alternation or co-recurrence may be still remains unknown. This contribution investigates the distribution of these two constructions in southern Calabria, where their structural alternation is very pronounced. After having examined the distribution of the two constructs in Reggio Calabria, their parameters of variation will be analysed. It will, therefore, be argued that pseudo-coordination is subject to restrictions regarding: (i) subject animacy, (ii) the moods and the tenses with which it can occur, and (iii) the semantic criteria for the selection of the second verb. In conclusion, it will be shown that those structures with the complementiser *(m)u/(m)i* and those with pseudo-coordination are not in free variation but are, instead, in complementary distribution. Namely, new and recent evidence allows us to show how the selection of each structure is governed by very specific parameters, rather than being interchangeable as has often been posited up until now in the literature.

Fenomeni di contatto nel parlato di giovani abruzzesi. Uno sguardo al *corpus* Nec Sine

Valeria Garozzo, Enzo Santilli

Il presente contributo tratterà dei fenomeni di contatto fra italiano e dialetto rinvenuti nel parlato di giovani abruzzesi le cui produzioni sono raccolte in un *corpus* di recente creazione, il Nec Sine. Dopo aver presentato il *corpus* (§ 1), si procederà con l'illustrazione del quadro teorico alla base dell'individuazione e dell'analisi dei fenomeni di contatto (§ 2) e si daranno esempi di analisi di alcuni frammenti selezionati (§§ 3-4), per poi giungere alle conclusioni (§ 5)¹.

1. Il *corpus* Nec Sine

1.1. Le ragioni di un nuovo *corpus*

I dati analizzati per questo lavoro provengono dal parlato di giovani abruzzesi, le cui produzioni sono state raccolte per uno studio sull'italiano popolare che costituisce il nucleo della ricerca presentata in Santilli 2021².

- 1** La concezione e lo sviluppo di questo lavoro sono opera comune dell'autrice e dell'autore. A Valeria Garozzo si deve la stesura dei §§ 2-4, a Enzo Santilli la stesura del § 1, a entrambi la stesura del § 5. L'autrice e l'autore ringraziano Riccardo Regis e Massimo Cerruti per i preziosi suggerimenti concessi, che hanno contribuito a migliorare la versione finale di questo lavoro. Ogni eventuale imprecisione rimasta sarà invece, naturalmente, da imputare solo a chi scrive.
- 2** Nonostante il *corpus* presentato in queste righe sia stato originariamente pensato per raccogliere il parlato di semicoltori, fra gli informanti compaiono anche parlanti

L'esigenza di realizzare e mettere a disposizione un nuovo *corpus* è stata dettata da almeno tre ragioni. Innanzitutto, ancora fino a pochi anni fa, come faceva notare anche GAETANO BERRUTO 2014 (p. 284), si doveva «constatare una quasi totale assenza di studi specifici sul parlato di semicolti» dovuta fondamentalmente all'assenza di banche dati che ne rendessero disponibile le produzioni orali. Anche *corpora* già disponibili, come ad esempio il LIP (<http://badip.uni-graz.at/it/>), il CLIPS (www.clips.unina.it) e il C-ORAL-ROM (CRESTI-MONEGLIA 2005), grazie ai quali si potevano condurre pregevoli indagini sul parlato, non permettevano una immediata reperibilità delle informazioni sociodemografiche degli informanti. In secondo luogo, si lamentava un'oggettiva difficoltà di isolare nelle risorse disponibili sufficiente parlato di informanti giovani. È vero che, in anni recenti, questa lacuna è stata parzialmente colmata da risorse nuove, come ad esempio i *corpora* del gruppo KIParla (www.kiparla.it, vedi MAURI ET AL. 2019), grazie ai quali si sono potute dare nuove risposte su base empirica a domande che riguardavano le dinamiche di variazione che regolano il parlato contemporaneo delle giovani generazioni (si veda, ad esempio, BALLARÈ-CERRUTI-GORIA 2019). Infine, e questa è la terza ragione per cui si è deciso di realizzare un nuovo *corpus*, così come i *corpora* del gruppo KIParla si prefiggevano «di essere non solo una risorsa per lo studio del parlato in prospettiva linguistica ma anche un vero e proprio archivio di voci» (CERRUTI-BALLARÈ 2020, p. 176), anche la risorsa usata per questo lavoro ambisce a rendere finalmente accessibile agli studiosi il parlato di una comunità linguistica poco trattata finanche da un punto di vista teorico. I dialetti della Marsica, regione storico-culturale posta nel lembo occidentale d'Abruzzo da cui provengono i parlanti del *corpus*, non hanno mai potuto godere di studi approfonditi e specifici.

che hanno ottenuto il diploma liceale e/o frequentato alcuni anni di università. Volendo suggerire un'etichetta rappresentativa, in SANTILLI 2021, p. 39 si proponeva di categorizzare gli informanti del Nec Sine come «non dotti» – da intendere come non aventi necessariamente «una grande cultura, in senso assoluto o in un dato campo del sapere» (*ibidem*). È tuttavia un'etichetta può prestare il fianco ad alcune critiche di natura interpretativa (le stesse che riguardano, d'altronde, la definizione di *semicolto*), per cui non verrà riproposta in questa sede.

In attesa che tale mancanza venga colmata e si arrivi alla formazione di una vera e propria letteratura sull'argomento, la realizzazione di un *corpus* dedicata alle manifestazioni di tali varietà nell'oralità sarà utile ad avviare simili studi in una prospettiva *bottom-up*³. Al fine di sottolineare ulteriormente come per una comprensione completa e puntuale dello spazio linguistico italo-romanzo sia necessario non prescindere da alcuna realtà, al *corpus* è stato dato il nome *Nec Sine*, che richiama la massima attribuita allo storico Appiano di Alessandria: «*nec sine marsis nec contra marsos triumphari posse*».

1.2. Realizzazione e struttura del *corpus*

Il *corpus* *Nec Sine* è composto da due moduli: il primo modulo raccoglie il parlato di anziani aventi più di 65 anni di età nati e cresciuti in Marsica, il secondo modulo è invece dedicato al parlato di giovani marsicani che non avessero compiuto più di 30 anni al momento dell'intervista. Il modulo "anziani" è attualmente in fase di trascrizione, mentre il modulo "giovani" è già disponibile per la consultazione.

Il modulo "giovani" raccoglie 9 ore, 5 minuti e 58 secondi di parlato prodotto da 13 informanti (7 donne e 6 uomini) raccolto per mezzo di interviste semi-strutturate acquisite tramite registratore modello Zoom H4n-Pro e condotte fra i mesi di luglio e novembre 2020. Per partecipare all'intervista ognuna delle persone intervistate ha dovuto prendere visione dell'informativa sul progetto e firmare un modulo attraverso il quale acconsentiva al trattamento dei propri dati personali ai sensi del DLGS 196/2003 e a rendere disponibile il materiale audio fornito per la realizzazione di studi linguistici. Le domande presenti nella traccia dell'intervista ricalcavano in larga misura quelle sottopo-

3 Per una discussione sul problema della mancanza di trattazioni teoriche rilevanti aventi come oggetto i dialetti marsicani, e su alcuni tratti linguistici che li contraddistinguono, si rimanda a SANTILLI in pubblicazione.

ste agli informanti del *corpus* ParlaTO⁴. Anche se durante le conversazioni si è concessa assoluta libertà agli informanti e in alcuni casi è stato l'intervistatore stesso ad allontanarsi dalla traccia centrale, tre macro-argomenti principali hanno costituito il nucleo di ogni intervista: lavoro, considerazioni sulla zona di residenza, tempo libero. A questo primo gruppo di domande ne venivano generalmente aggiunte due ulteriori, relative a considerazioni personali sull'uso del dialetto e sulla vita durante il primo periodo di confinamento dovuto alla pandemia da Covid-19. Lo scheletro dell'intervista viene riportato in APPENDICE.

I file audio raccolti sono stati poi anonimizzati con aggiunta di rumore bianco tramite il software *Audacity* nelle parti in cui si potesse risalire alle identità degli informanti, e successivamente le registrazioni anonimizzate sono state trascritte. La maggior parte delle trascrizioni è avvenuta a mano, trasferendo su un foglio di videoscrittura il parlato registrato. Nei casi più complicati, ovvero quelli in cui più parlanti partecipavano a una stessa intervista eventualmente prendendo turno simultaneamente, si è reso necessario ricorrere all'utilizzo del software di trascrizione *Elan*. I file audio e i file di testo anonimizzati sono poi stati caricati in una cartella virtuale.

Al fine di restituire una resa ortografica quanto più aderente al parlato, nelle trascrizioni non è stato utilizzato alcun segno di punteggiatura a eccezione del punto interrogativo come segnalatore della presenza di una domanda, e sono state osservate poche altre convenzioni, parzialmente in linea con quanto indicato da JEFFERSON 2004, pp. 24-31 (alle righe 1-2 in tabella 1). Sono poi stati usati altri segni paragrafematici non esplicitamente trattati in JEFFERSON 2004 ma utili a una migliore fruibilità del *corpus* (3-4 in tabella 1) e, infine, per rendere conto di alcune caratteristiche fonetiche tipiche e ricorrenti nei dialetti marsicani, sono stati utilizzati i caratteri indicati alle righe 5-8 in tabella 1.

4 Si ringrazia Silvia Ballarè per avere fornito un modello da cui partire, poi adattato variamente a seconda dei soggetti intervistati.

Fenomeni di contatto nel parlato di giovani abruzzesi

	Simbolo	Significato
1	(testo)	il testo trascritto in quel punto è dubbio
2	-	la pronuncia della parola è stata interrotta prima della sua conclusione
3	'	troncamento di parola quando si tratta di verbi all'infinito (<i>canta', vede', fini'</i>) e all'indicativo presente (<i>so'</i> per 'sono'), articoli indeterminati (<i>n</i> per 'un', <i>'na</i> per 'una' seguito da consonante, <i>n'</i> per 'una' seguito da vocale), preposizioni (<i>co'</i> per <i>con</i> , <i>pe'</i> per 'per') e per le parole appartenenti a tutte le altre parti del discorso
4	#	inizio di un enunciato in dialetto o anche la sola presenza di una parola di origine dialettale
5	ə	vocale indistinta. Es.: <i>soldə</i> ['soldə] 'soldi'
6	ʃ	fricative postalveolari /ʃ/ e /ʒ/. Es.: <i>suʃta</i> ['suʃta] 'boria'
7	j	approssimante palatale /j/. Es.: ['bejjə] 'bello'
8	w	è usato per rendere conto della pronuncia del nesso it. <<qu>>, fortemente velare in Abruzzo. Es.: <i>qweʃtə</i> ['kweʃtə] 'questo'

Tabella 1 Convezioni ortografiche utilizzate per la trascrizione del *corpus* Nec Sine

Queste convenzioni ortografiche, integralmente applicate nella trascrizione del *corpus* originale, sono state parzialmente adottate anche per le trascrizioni riportate qui nei §§ 3-4, integrate da altre convenzioni di trascrizione per le quali si rimanda all'APPENDICE.

Il numero totale di *token* è stato conteggiato con il software #LancsBox 6.0 (BREZINA-WEILL-TESSIER-MCENERY 2021) ed è pari a 96.452 occorrenze se si tiene conto anche del parlato dell'intervistatore, 60.152 se si considera solo il parlato degli informanti. Degli informanti sono disponibili alcune informazioni sociobiografiche quali genere biologico, città di nascita e residenza, occupazione svolta al momento dell'intervista e livello di istruzione. Tali informazioni vengono suntuosamente riportate in tabella 2.

	Codice informante	Età al momento dell'intervista	Sesso	Professione	Titolo di studio	Luogo di nascita	Luogo di residenza	Presente nella registrazione
1	G01f	22	F	Barista	Dip. liceo linguistico	Avezzano (AQ)	Avezzano (AQ)	RECNS-G01
2	G02f	23	F	Studente	Dip. liceo artistico	Avezzano	Pescina (AQ)	RECNS-G01
3	G03f	23	F	Disoccupata	Dip. liceo linguistico	Avezzano	Pescina	RECNS-G01
4	G04m	27	M	Autotrasportatore	Dip. ist. tecn. agrario	Avezzano	Venere di Pescina (AQ)	RECNS-G02 RECNS-G03
5	G05m	23	M	Meccanico	Dip. scuola media inferiore	Avezzano	Venere di Pescina	RECNS-G03
6	G06f	25	F	Disoccupata / Influencer	Dip. ist. tecn. alberghiero	Avezzano	Venere di Pescina	RECNS-G04
7	G07f	22	F	Disoccupata	Dip. liceo magistrale	Avezzano	Pescina	RECNS-G05
8	G08m	19	M	Studente / Gamer	Dip. ist. tecn. informatico	Kielce (Polonia)	Pescina	RECNS-G06
9	G09m	27	M	Operaio	Dip. scuola media inferiore	Avezzano	Ortucchio	RECNS-G07
10	G10f	29	F	Barista / Addestratrice	Dip. liceo scientifico	Avezzano	Ortucchio	RECNS-G07
11	G11f	19	F	Commessa	Dip. ist. tecn.	Avezzano	Ortucchio	RECNS-G07 RECNS-G08
12	G12m	20	M	Studente	Dip. ist. tecn.	Avezzano	Pescina	RECNS-G09
13	G13m	29	M	Operaio	Dip. ist. tecn.	Avezzano	Venere di Pescina	RECNS-G10

Tabella 2 Caratteristiche sociodemografiche degli informanti del *corpus* Nec Sine

Il modulo “giovani” del *corpus* Nec Sine è ora disponibile, su richiesta, sulla piattaforma Sketch Engine, che ha anche fornito il servizio di *Pos-tagging* automatico.

2. Fenomeni di contatto tra italiano e dialetto: il quadro teorico di riferimento

Prima di passare all'analisi vera e propria dei frammenti selezionati dal modulo “giovani” del *corpus* Nec Sine, è necessario fare chiarezza su alcune questioni teoriche e terminologiche, data la molteplicità di definizioni e prospettive esistenti in letteratura per la descrizione dei fenomeni di contatto linguistico. La tabella 3⁵ propone un riepilogo sintetico dei fenomeni di contatto nel discorso presi in considerazione durante la ricerca, definendone le proprietà e i criteri considerati per la loro individuazione; in fondo alla tabella, per ogni fenomeno è indicato il numero di occorrenze riscontrate durante l'analisi del *corpus*.

Criteria e proprietà	Code switching	Enunciazione mistilingue
Categorie		
Stesso evento comunicativo	+	+
Livello	Interfrasale	Intrafrasale
Implicante una	Unità discorsiva	Unità morfosintattica
Ambito	Pragmatica	Morfosintassi
Funzionalità comunicativa	+	- +
Implicante bilinguismo	+	+
Occorrenze nel <i>corpus</i> Nec Sine	48	26

Tabella 3 Introduzione ai principali fenomeni di contatto nel discorso

Come indicato in tabella, il *code switching* e le enunciazioni mistilingue sono entrambi fenomeni caratterizzati dal manifestarsi all'interno di un unico evento comunicativo, ma differiscono sia sul piano sintattico che sul piano pragmatico: il *code switching* si manifesta a livello

5 La Tabella 3 è stata composta anche sulla base di schemi già presenti in: BERRUTO 1990 [2012], p. 281; BERRUTO 2001 [2012], p. 307.

interfrasale, ovvero tra un enunciato⁶ e un altro, coinvolgendo intere unità discorsive; le enunciazioni mistilingue, invece, si realizzano a livello intrafrasale, cioè dentro i confini di uno stesso enunciato, coinvolgendo unità morfosintattiche. Entrambi i fenomeni implicano il bilinguismo del parlante, ma le enunciazioni mistilingue rappresentano «il corrispettivo discorsivo dell'interferenza» (BERRUTO 1990 [2012], p. 288) e sono quindi tendenzialmente (ma non necessariamente) prive di funzionalità comunicativa. Le manifestazioni di *code switching*, al contrario, possono sempre essere ricondotte alla realizzazione di specifiche strategie conversazionali, risultando quindi marcate dal punto di vista della pragmatica del discorso.

3. Il *code switching*

Tra le molte alternative possibili, per la ricerca e l'analisi delle manifestazioni di *code switching* in questo lavoro è stata seguita la strada tracciata nei lavori di Giovanna Alfonzetti, definendo quindi il *code switching* tra italiano e dialetto come un «uso alterno e funzionale dei due codici all'interno dello stesso episodio comunicativo» (ALFONZETTI 1992, p. 16) che ha le sue realizzazioni prevalentemente al livello interfrasale. La parola chiave di questa definizione è senza dubbio l'aggettivo *funzionale*, che sottolinea la marcatezza comunicativa sempre riconoscibile nelle manifestazioni di *code switching*.

L'impianto teorico utilizzato da Alfonzetti è a sua volta il risultato di un'accurata integrazione di altri modelli, tra i quali spiccano per importanza i lavori di John Gumperz (dal quale è ripresa la cosiddetta

⁶ Si precisa che con l'etichetta di *enunciato* si identifica «un'unità linguistica interna al turno, prodotta da un singolo parlante, non interrotta da interventi verbali di altri locutori partecipanti allo stesso evento comunicato e delimitata da pause o cesure nella curva intonativa, in cui è riconoscibile un singolo nucleo illocutorio o una sequenza di segmenti riconducibili a una stessa mossa conversazionale» (CICCOLONE-DAL NEGRO 2021, p. 25).

teoria della contestualizzazione⁷) e, soprattutto, quelli di Peter Auer⁸, dal quale è ripreso il modello procedurale di conversazione bilingue che permette di coniugare l'analisi dei fenomeni di contatto linguistico in senso lato con i principi tratti dal campo dell'analisi della conversazione⁹. Dai lavori di Auer, inoltre, Alfonzetti riprende anche la suddivisione dei fenomeni di *code switching* in due macro-categorie: quando il passaggio da un codice all'altro serve a realizzare atti di negoziazione linguistica e la commutazione quindi segue la direzione delle competenze e delle preferenze dei parlanti coinvolti nell'interazione si parla di *code switching connesso ai partecipanti*, quando la commutazione serve a un unico parlante per organizzare al meglio il testo parlato e marcare la produzione di atti linguistici particolarmente significativi all'interno del proprio discorso, invece, si è in presenza di *code switching connesso al discorso*.

In virtù delle circostanze di staticità conversazionale in cui è stato raccolto il *corpus* Nec Sine, sono state rilevate solo 4 occorrenze di *code switching* connesso ai partecipanti. Molto più frequenti, invece, sono i casi di *code switching* connesso al discorso, con 44 occorrenze totali.

La tabella 4 contiene un riepilogo di tutte le funzioni pragmatiche associate in sede d'analisi alle manifestazioni di *code switching* connesso al discorso riscontrate nel *corpus*, elencate secondo il numero di occorrenze.

Guardando la tabella 4 è subito possibile farsi un'idea precisa di cosa s'intenda parlando di atti linguistici particolarmente significativi realizzabili attraverso il ricorso al *code switching*, e di conseguenza delle sue possibili funzioni pragmatiche. Nelle conversazioni indagate, infatti, il passaggio dall'italiano al dialetto viene sfruttato dai parlanti per la messa in atto di specifiche strategie conversazionali:

- 7 *Contextualization cue* nei termini di GUMPERZ 1982, p. 131: «any feature of linguistic form that contributes to the signalling of contextual presuppositions».
- 8 Si veda, in particolare, AUER 1984.
- 9 Tale proposito è esplicitato in AUER 1984, p. 6: «the model presented in the following chapters is meant as an attempt to flex the muscles of conversation analysis as much as possible».

17 volte il *code switching* è utilizzato per marcare l'inserimento di citazioni altrui all'interno di nuove formulazioni (funzione del 'discorso riportato'; § 3.1); 3 volte è utile per marcare delle glosse esplicative, da intendersi come frammenti di discorso in dialetto utili per fornire ulteriori specificazioni di qualcosa che è appena stato detto utilizzando l'italiano ('elaborazione'; § 3.2); 7 volte la giustapposizione tra italiano e dialetto marca la riproposizione più o meno equivalente dello stesso contenuto ('ripetizioni'); 1 volta il passaggio dall'italiano al dialetto serve al parlante per riprendere parte del discorso appena fatto, con intento autocorrettivo ('riformulazione'); 11 occorrenze di *code switching* sono utili in vari modi a supportare il racconto di una storia, ad esempio segnalando il raggiungimento del *climax* oppure la sua conclusione ('*storytelling*'). Infine, 5 volte il ricorso al *code switching* è utile per caricare di espressività specifici frammenti di discorso, rendendo più evidente l'intento scherzoso sotteso a tali formulazioni ('funzione ludico-espressiva').

N. occorrenze	Funzioni pragmatiche del <i>code switching</i> connesso al discorso nel corpus Nec Sine
17	Discorso riportato
11	<i>Storytelling</i>
7	Ripetizione
5	Ludica-espressiva
3	Elaborazione
1	Riformulazione
TOT. 44	

Tabella 4 Occorrenze di *code switching* connesso al discorso nel *corpus*

È bene sottolineare che le funzioni pragmatiche associabili al *code switching* connesso al discorso non si esauriscono in questo breve elenco, ma piuttosto fanno parte di un insieme aperto. La caratterizzazione dei fenomeni di *code switching* connesso al discorso, inoltre, dipende direttamente dallo specifico evento comunicativo all'interno del quale il passaggio da un codice all'altro si manifesta, ed è quindi possibile

solo attraverso una più ampia comprensione delle dinamiche conversazionali in corso. Per questa ragione, Auer definisce il *code switching* come «un fenomeno significativo a livello locale»¹⁰.

3.1. Il *code switching* connesso al discorso con funzione di ‘discorso riportato’

Poiché non possiamo dar conto, per ragioni di spazio, della totalità delle occorrenze di *code switching* connesso al discorso, ci limiteremo all’analisi di alcuni specifici frammenti ritenuti particolarmente significativi.

Il primo frammento selezionato rappresenta un esempio di *code switching* con funzione di ‘discorso riportato’. Si è in presenza di discorso riportato nel momento in cui «un locutore L riproduce, sulla catena verbale in cui realizza un proprio atto di enunciazione E, un altro atto di enunciazione E1, da ascriversi a una fonte L1, non necessariamente diversa da L» (MORATA GARAVELLI 1985, p. 21). All’interno del *corpus* Nec Sine, i passaggi dall’italiano al dialetto che coincidono con l’inserimento di discorso riportato sono all’incirca il 39% delle occorrenze totali di *code switching* connesso al discorso; tale funzione, generalmente indicata in letteratura semplicemente come un atto di “citazione”, è quindi quella più ricorrente nell’ambito del *code switching* connesso al discorso all’interno del Nec Sine, un dato che si presenta perfettamente in linea con gli altri studi sull’argomento a livello nazionale e internazionale¹¹.

La citazione prototipica è quella che permette a un parlante di riportare un discorso altrui in maniera diretta all’interno del proprio

10 La traduzione è degli autori. Nell’originale di AUER 1999, p. 310, si legge: «locally meaningful event».

11 Anche nella ricerca presentata in ALFONZETTI 1992, ad esempio, la “citazione” rappresenta la funzione con il maggior numero di occorrenze in relazione al *code switching* connesso al discorso; si veda ALFONZETTI 1992, p. 19, nota 7 anche per una panoramica della bibliografia internazionale relativa a studi sul *code switching* nei quali la funzione della “citazione” è ampiamente attestata.

atto enunciativo. Il ricorso al *code switching* permette di marcare in modo chiaro l'autonomia della sequenza citata, come si vede nel frammento riportato (Esempio 1), in cui la parlante G10f passa dall'italiano al dialetto per riferire le parole pronunciate da un'anziana signora, e il confine finale della citazione è individuabile proprio grazie al ritorno all'italiano.

Esempio 1

G10f: sì esatto infatti proprio una signora mentre loro facevano i lavori **diceva «guarda che bello mo #ecchə cə steva la nonna steva, c'ha nato i padrə, mo sta vənɪ' purə la la nəpote, quand'è bejjə quand'è bejjə»** perché poi so' tutte cose che comunque / anche perché s'è rimpopolato di ragazzi
{qua ci stava la nonna stava, c'è nato il padre, mo sta a venire (*scilicet*: sta venendo) pure la nipote, quanto è bello quanto è bello}

In questo specifico caso, è piuttosto plausibile che la parlante stia riportando fedelmente quello che ha sentito dire, e quindi che il codice scelto per il discorso riportato – il dialetto – sia davvero quello utilizzato dall'anziana signora intenta a osservare i lavori del cantiere, ma non è comunque possibile averne la certezza. Sarebbe ingenuo, infatti, pensare che l'utilizzo di citazioni serva sempre e soltanto per citare qualcosa di realmente detto da altri nel codice da loro realmente utilizzato, dato che la citazione marcata da *code switching* può anche essere del tutto fittizia o non rispecchiare il codice della reale enunciazione, assolvendo quindi funzioni diverse dal riportare semplicemente parti di discorsi altrui. Si consideri il frammento seguente (Esempio 2):

Esempio 2

I01: se tu vai a parla' con Mattarella non gli parli in dialetto
G02f: non XXX «**oh ciao comba' #cumma šta?**»
{oh ciao compare come stai?}

In questo esempio, siamo ancora in presenza di *code switching* con funzione di marcatura del discorso riportato, ma in un contesto totalmente diverso rispetto all'esempio precedente: nel momento in cui l'intervistatore IO1 afferma l'assurdità dell'ipotesi di rivolgersi in dialetto a Mattarella, GO2f interviene con una frase dialettale rivolta idealmente proprio al presidente della Repubblica, producendo una citazione fittizia. In questo caso, il *code switching* ha certamente anche una funzione ludico-espressiva: il passaggio dall'italiano al dialetto è utile anche a rendere più esplicito l'intento ludico che sta alla base della formulazione, essendo quella richiamata una situazione paradossale e irrealistica.

Nell'esempio che segue, infine, è riportato un caso di discorso riportato marcato da *code switching* dove il ricorso al sistema della citazione aiuta il parlante a raccontare al meglio un aneddoto personale (Esempio 3):

Esempio 3

GO9m: era passato uno / una quarantina d'anni, tutto abbronzato, un fisico da paura [...] io stavo vicino non c'ho detto niente poi c'ho dato la botta **c'ho #fattə «eh? tə ji si vistə quia?» ha fattə «chi quello?» fattə «che mə sta a pijjà purə pə culə i t'ho visto che l'hai guardato nən mə sta a di' də no» «eh ma è passato vicino»**
{fatto (*scilicet*: detto) «eh? te lo sei visto quello?»; ha fatto (*scilicet*: detto) [...] ho fatto (*scilicet*: detto) «che mi stai a prendere pure per il culo [...] non mi dire di no}

Nel discorso di GO9m, la giustapposizione tra italiano e dialetto si rivela funzionale per rendere al meglio il battibecco avvenuto tra lui e la fidanzata: il parlante attribuisce a sé stesso formulazioni interamente o quasi interamente dialettali, mentre citando le parole della fidanzata si esprime in italiano. Nelle ultime battute riportate, la successione dei turni di parola rimane scandita solo dal passaggio tra italiano e dialetto, mancando il *verbum dicendi*.

3.2. Il *code switching* connesso al discorso con funzione di ‘elaborazione’

Proseguiamo l’analisi con un caso (Esempio 4) di *code switching* connesso al discorso con funzione di ‘elaborazione’, selezionato poiché ritenuto particolarmente adeguato a rappresentare questa categoria:

Esempio 4

- IO1: ehm parliamo proprio appunto del / del lavoro in questa zona no? più più in generale ti pare che XXX, la Marsica, l’Abruzzo sia un buon posto dove lavorare rispetto magari ad altre realtà che hai sentito nominare?
- GO5m: oddio se l’aspettativa di vita è quella di fare il dottore, di fare l’imprenditore, sta’ tutti i giorni in giacca e cravatta no =
- IO1: mhmh
- GO5m: = sicuramente no / ci stanno lavori più umili appunto come bracciante agricolo il meccanico l’operaio l’autista
- IO1: mhmh
- GO5m: però lavoro ce n’è qua ce [n’è =
- IO1: [ok
- GO5m: **è che purtroppo nessuno accetta di farlo =**
- IO1: mh
- GO5m: = **che nessuno #cə jessə a Fucinə allə cinqwə la matina, nisciunə sə məttesə sopra nə camjə a stassə na sətimana forə**
{ci andrebbe a Fucino alle cinque del mattino, nessuno si metterebbe sopra un camion a starsi una settimana fuori}

Il frammento conversazionale qui riportato ha come codice base l’italiano. Per la formulazione dell’ultimo enunciato, tuttavia, il parlante GO5m ricorre al dialetto, utilizzando questo turno per specificare ulteriormente quanto intendeva dire con la sua formulazione precedente. L’enunciato in dialetto qui presente può dunque essere considerato una sorta di glossa di quanto lo precede: prima il parlante, utilizzando l’italiano, afferma che “nessuno accetta di fare” determinati lavori, poi passa al dialetto per le unità discorsive successive con lo scopo di precisare ulteriormente quali siano i lavori a cui si riferisce. Gumperz definisce simili unità discorsive – che anche nel suo *corpus* si presen-

tano marcate da *code switching* – come «qualifying constructions» (GUMPERZ 1982, p. 79), analizzando diversi esempi all'interno dei quali il messaggio principale viene trasmesso utilizzando un codice e il codice alternativo è usato per “qualificare” il messaggio in vari modi.

4. Enunciazioni mistilingui

Arriviamo adesso alle enunciazioni mistilingui, fenomeno spesso indicato in letteratura con la più problematica etichetta di *code mixing*¹². Secondo la definizione di BERRUTO (2001 [2012], p. 302), si parla di enunciazioni mistilingui in presenza di «enunciati frasali con costituenti appartenenti a due sistemi linguistici diversi, e quindi con il contributo di due grammatiche e due lessici nella formazione della frase»; il passaggio da un codice all'altro, a differenza di quanto succede in presenza del *code switching*, si realizza all'interno dei confini sintattici e informativi di un singolo enunciato. Si è detto che questo fenomeno rappresenta il corrispettivo discorsivo dell'interferenza, poiché le enunciazioni mistilingue spesso legate «a situazioni di incertezza nella scelta del codice, o addirittura a sovrapposizioni e ‘invasioni’ delle rispettive grammatiche» (BERRUTO 1990 [2012]: 282).

Seguendo la proposta di Regis 2005, si possono distinguere le enunciazioni mistilingui secondo l'ampiezza e il ruolo sintattico del segmento commutato. La tabella 5 sintetizza le occorrenze riscontrate nel *corpus*, suddivise nelle tre macro-categorie individuate da Regis e complete di specifiche sintattiche:

12 Si veda ad esempio ALFONZETTI 1992 (che si riferisce a questo fenomeno anche con l'espressione «switching intrafrasale», p. 21; p. 180ss.), oppure AUER 1999 – dove però l'etichetta di *code mixing* è attribuita non solo ai passaggi da un codice a un altro all'interno di una stessa preposizione, ma anche ad altri tipi di frammistione tra sistemi linguistici. Così come in effetti accade anche per l'etichetta *code switching*, sono molteplici le interpretazioni presenti in letteratura sulle caratteristiche e i limiti di questo fenomeno; in questo lavoro si è dunque preferito evitare questa denominazione, optando per l'etichetta più trasparente di *enunciazioni mistilingui*.

n. occorrenze	Categoria	Specifiche sintattiche
17	Monològa	
7	Sintagmatica	2 SN + 3 SV + 1 SAw + 1 SAgg
2	EM proposizionale	1 Prop. Limitativa + 1 Prop. Oggettiva

Tabella 5 Enunciazioni mistilingui nel *corpus* Nec Sine

Entrando nel dettaglio, si parla di *enunciazioni mistilingui monològhe* quando l'enunciato considerato contiene una sola parola commutata – come nel frammento che segue (Esempio 5), in cui l'unico elemento dialettale è il pronome soggetto di prima persona 'io':

Esempio 5

G09m: e là è stata l'unica cosa che dico sai che c'è se devo rianda' a una spiaggia / sicuramente a questa spiaggia qui è bellissima, però so che #**ji** mi bagno solamente i piedi
{io}

Si è invece in presenza di *enunciazioni mistilingui sintagmatiche* quando ad essere commutato è un sintagma intero, come nell'affermazione contenuta nel seguente (Esempio 6) dove l'unico contenuto dialettale è la locuzione avverbiale 'dopo un poco':

Esempio 6

G05m: è come il tappo della coca cola quando non avviti bene che la shakeri
[...] #**dopə nə pochə** la coca cola fa schifo
{dopo un poco}

Infine, ci sono casi in cui il passaggio da un codice a un altro coincide con il passaggio sintattico tra principale e subordinata: si parla allora di *enunciazioni mistilingue proposizionali*, come nel nostro ultimo frammento (Esempio 7):

Esempio 7

Go2f: perché ci sono dei termini specifici dialettali che danno dei concetti astratti che in italiano non riesci a spes / a spiegare tipo come fai a spiega' a uno #**chə te la sušta**?
{che tiene la susta (*scilicet*: che si comporta con altezzosità)}

In questo caso, il segmento dialettale non è un'unità discorsiva autonoma – come invece accade in presenza del *code switching* – ma è sintatticamente dipendente dalla proposizione principale, formulata in italiano.

5. Conclusioni

I risultati di questa breve disamina confermano tre tendenze della sociolinguistica contemporanea. Innanzitutto, è chiaro che uno studio che voglia trarre conclusioni forti su un determinato tema non può prescindere da un approccio *corpus based*. In un'era in cui i concetti stessi di lingua e varietà vengono messi in discussione da alcune correnti di pensiero¹³, è importante dare descrizioni degli effetti del contatto e della variazione appoggiandosi a solide basi empiriche. Da questo punto di vista, per quel che riguarda ad esempio gli studi variazionistici, alcuni risultati importanti sono già stati raggiunti: lavori come GHYSELEN-DE VOGELAER 2018, VILLENA-PONSODA-VIDA CASTRO 2020 e CERRUTI-VIETTI 2022 hanno dimostrato come sia possibile dare conto dell'esistenza di nozioni sociolinguistiche 'classiche' (nei casi citati, del concetto di *varietà*) traendo dati da *corpora* di piccole e medie dimensioni e analizzandoli con strumenti di indagine statistica. Nulla impedisce di avere lo stesso approccio per questioni relative al contatto fra italiano e dialetto: in questo studio, per ragioni di spazio, si è

¹³ Si fa riferimento, ad esempio, alla terza ondata degli studi variazionistici, di cui è data una descrizione (anche in riferimento alla prima e alla seconda ondata) in ECKERT 2014.

potuta condurre solo la prima fase, quella dell'analisi qualitativa, di un lavoro più complesso. Questa prima fase andrà completata in futuro dall'analisi quantitativa, che appoggiandosi a strumenti quali ad esempio l'analisi multivariata dovrà dimostrare se le conclusioni a cui si è preliminarmente giunti in questa sede daranno risposte convincenti anche da un punto di vista della rilevanza statistica.

In secondo luogo, per quanto riguarda il contatto fra italiano e dialetto, l'analisi condotta ha dimostrato una buona vitalità del codice dialettale, che stupisce se si pensa al “paradosso dell'osservatore” sintetizzato da LABOV 1972, p. 209:

the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.

Per studiare il discorso bilingue italiano-dialetto bisognerebbe infatti osservare il parlato spontaneo prototipico, che solo parzialmente può arrivare a coincidere con i dati raccolti durante un'intervista semi-strutturata. Nonostante questo importante impedimento, fenomeni di commutazione di codice sono stati riscontrati anche nel Nec Sine, alla cui raccolta gli informanti hanno partecipato consapevoli di essere intervistati: è probabile che l'affiorare di un numero tutt'altro che modesto di fenomeni di contatto connessi al discorso nel *corpus* dipenda dalle particolari condizioni che hanno fatto da sfondo alla conduzione delle interviste, caratterizzate ad esempio da una percezione di informalità e da rapporti pregressi tra il raccoglitore e le persone intervistate. Su questo, tuttavia, sarà necessario approfondire.

Infine, la terza tendenza riscontrata in queste pagine – anch'essa da confermare almeno da uno studio di natura quantitativa – riguarda la fenomenologia delle funzioni pragmatiche realizzabili attraverso il *code switching*: le manifestazioni di *code switching* riscontrate nel *corpus* sono infatti state associate alla realizzazione delle stesse funzioni pragmatiche individuate in GAROZZO 2021 (con un *corpus* raccolto in Sicilia) e in molta altra letteratura nazionale e internazionale; le stesse individuate da Gumperz e Auer in situazioni sociolinguisticamente

molto distanti dall'Italia contemporanea¹⁴. Ciò è perfettamente in linea con le posizioni di Auer¹⁵, che già negli anni '90 sosteneva l'esistenza di *conversational loci* particolarmente suscettibili a ospitare fenomeni di commutazione, indipendentemente dalle caratteristiche del repertorio sociolinguistico preso in considerazione.

Bibliografia

- ALFONZETTI 1992 = GIOVANNA ALFONZETTI, *Il discorso bilingue. Italiano e dialetto a Catania*, Milano, Franco Angeli, 1992.
- AUER 1984 = PETER AUER, *Bilingual conversation*, Amsterdam, Benjamin, 1984.
- AUER 1995 = PETER AUER, *The pragmatics of code-switching: a sequential approach*, in *One speaker, two languages. Cross-disciplinary perspectives on code-switching*, a cura di Lesley Milroy e Pieter Muysken, Cambridge, Cambridge University Press, 1995, pp. 45-67.
- AUER 1999 = PETER AUER, *From Code-Switching Via Language Mixing to Fuse Lects: Toward a Dynamic Typology of Bilingual Speech*, in «International Journal of Bilingualism», 3 (4), 1999, pp. 309-332.
- BALLARÈ-CERRUTI-GORIA 2019 = SILVIA BALLARÈ, MASSIMO CERRUTI, EUGENIO GORIA, *Variazione diastratica nel parlato di giovani: il caso delle costruzioni relative*, in *Le tendenze dell'italiano contemporaneo rivisitate*, Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6-8 settembre 2018), a cura di Bruno Moretti *et al.*, Milano, Officinavenutno, 2019, pp. 75-94.
- BERRUTO 1990 [2012] = GAETANO BERRUTO, *Italiano regionale, commutazione di codice e enunciati mistilingui*, in *Saggi di sociolinguistica e linguistica*, a cura di Giuliano Bernini *et. al.*, Alessandria, Edizioni dell'Orso, 2012, pp. 275-299 [già in *L'italiano regionale*, Atti del XVII Congresso della S.L.I. (Padova-Venezia, 14-16 settembre 1984), a cura di Michele Cortellazzo e Alberto Mioni, Roma, Bulzoni, 1990, pp. 105-130].

14 Lo studio di GUMPERZ 1982 si basa su un *corpus* di registrazioni di parlato spontaneo raccolto in diverse comunità bilingui, dove erano in contatto spagnolo e inglese, sloveno e tedesco e hindi e inglese. Per quanto riguarda AUER 1984, i suoi dati si riferiscono al parlato bilingue italiano-tedesco di una comunità di emigrati italiani a Costanza.

15 Si veda ad esempio AUER 1995, p. 120.

- BERRUTO 2001 [2012] = GAETANO BERRUTO, *Struttura dell'enunciazione mistilingue e contatti linguistici nell'Italia di Nord-Ovest (e altrove)*, in *Saggi di sociolinguistica e linguistica*, a cura di Giuliano Bernini et al., Alessandria, Edizioni dell'Orso, 2012, pp. 301-321 [già in *Italica-Raetica-Gallica. Studia linguarum litterarum artiumque in honorem Ricarda Liver*, a cura di Peter Wunderli, Iwar Werlen e Matthias Grünert, Tübingen, Francke, 2001, pp. 263-283].
- BERRUTO 2014 = GAETANO BERRUTO, *Esiste ancora l'italiano popolare? Una rivisitazione*, in *Dall'architettura della lingua italiana all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, a cura di Paul Danler e Christine Konecny, Frankfurt am Main, Peter Lang, 2014, pp. 277-290.
- BREZINA-WEILL-TESSIER-MCENERY ET AL. 2021 = VACLAV BREZINA, PIERRE WILL-TESSIER, ANTHONY MCENERY, #LancsBox v. 6.x. Scaricabile all'indirizzo: <http://corpora.lancs.ac.uk/lancsbox/download.php>
- CERRUTI-BALLARÈ 2020 = MASSIMO CERRUTI E SILVIA BALLARÈ, *ParlaTO: corpus del parlato di Torino*, in «Bollettino dell'Atlante Linguistico Italiano (BALI)», 44, 2020 (2021), pp. 171-196.
- CERRUTI-VIETTI 2022 = MASSIMO CERRUTI E ALESSANDRO VIETI, *Identifying language varieties: Coexisting standards in spoken Italian*, in *The coherence of linguistic communities: Orderly heterogeneity and social meaning*, a cura di Karen Beaman e Gregory R. Guy, London, Routledge, 2022, pp. 261-280.
- CICCOLONE-DAL NEGRO 2021 = SIMONE CICCOLONE E SILVIA DAL NEGRO, *Comunità bilingui e lingue in contatto. Uno studio sul parlato bilingue in Alto Adige*, Cesena-Bologna, Caissa Italia, 2021.
- CRESTI-MONEGLIA 2005 = EMANUELA CRESTI E MASSIMO MONEGLIA, *C-O-ral-Rom. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam-Philadelphia, John Benjamins, 2005.
- ECKERT 2014 = PENELOPE ECKERT, *Third Wave Variationism*, «Oxford Handbook Topics in Linguistics» (edizione on-line, 5 dicembre 2014). Disponibile on-line all'indirizzo: <https://academic.oup.com/edited-volume/42051/chapter/355823612>
- GAROZZO 2021 = VALERIA GAROZZO, *Italiano e dialetto nella conversazione. Un'indagine sulle manifestazioni del code switching nell'area ionica etnea*, in «Bollettino», 32, 2021, pp. 225-251.
- GHYSELEN-DE VOGELAER 2018 = ANNE-SOPHIE GHYSELEN E GUNTHER DE VOGELAER, *Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the 'Variety' Concept*, in «Frontiers in Psychology», 9, 2018. Disponibile on-line all'indirizzo: <https://www.frontiersin.org>, doi: 10.3389/fpsyg.2018.00385

Fenomeni di contatto nel parlato di giovani abruzzesi

- GUMPERZ 1982 = JOHN GUMPERZ, *Discourse strategies*, London, Cambridge University Press, 1982.
- JEFFERSON 2004 = GAIL JEFFERSON, *Glossary of transcript symbols with an introduction*, in *Conversation Analysis: Studies from the first generation*, a cura di Gene H. Lerner, Amsterdam-Philadelphia, John Benjamins, 2004, pp. 13-31.
- LABOV 1972 = WILLIAM LABOV, *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press, 1972.
- MAURI ET AL. 2019 = CATERINA MAURI, SILVIA BALLARÈ, EUGENIO GORIA, MASSIMO CERRUTI, FRANCESCO SURIANO, *KIParla Corpus: A New Resource for Spoken Italian*, in *CLiC-it 2019 – Italian Conference on Computational Linguistics*, Proceedings of the Sixth Italian Conference on Computational Linguistics, a cura di Raffaella Bernardi, Roberto Navigli, Giovanni Semeraro, 2019. Disponibile on-line all'indirizzo <http://ceur-ws.org/Vol-2481/paper45.pdf>
- MORTARA GARAVELLI 1985 = BICE MORTARA GARAVELLI, *La parola d'altri*, Palermo, Sellerio, 1985.
- REGIS 2005 = RICCARDO REGIS, *Appunti grammaticali sull'enunciazione mistilingue*, Muenchen, Lincom Europa, 2005.
- SANTILLI 2021 = ENZO SANTILLI, *Come fai a spiega' a uno chə te la sušta? Tratti di italiano popolare nel parlato di giovani non dotte/i della Marsica orientale*, Torino, Tesi di laurea magistrale non pubblicata, 2021.
- SANTILLI in pubblicazione = ENZO SANTILLI, *L'Italiano popolare nel parlato di giovani della Marsica orientale: evidenze dal corpus Nec Sine*, in «Rivista Italiana di Dialettologia (RID)», 47.
- VILLENA-PONSODA-VIDA CASTRO 2020 = JUAN ANDRÉS VILLENA-PONSODA E MATILDE VIDA CASTRO, *Variation, identity and indexicality in southern Spanish. On the emergence of a new variety in urban Andalusia*, in *Intermediate language varieties. Koinai and regional standards in Europe*, a cura di Massimo Cerruti e Stavroula Tsipplakou, Amsterdam-Philadelphia, 2020, pp. 149-182.

Appendice

Convenzioni di trascrizione

,	Brevissimo <i>break</i> intonativo senza pausa
...	Pausa più lunga, entro un turno
?	Intonazione interrogativa
!	Intonazione esclamativa
« »	Atti di citazione
/	Auto-interruzioni, auto-correzioni
[Parti pronunciate in sovrapposizione da due o più parlanti
[
=	Unisce parti del turno di uno stesso parlante, trascritte su righe diversi per consentire la trascrizione di un turno sovrapposto di un altro parlante
XXX	Elementi incerti, che all'ascolto non sono stati compresi / censura di dati sensibili riconducibili all'identità dell'informatore
[...]	
{testo}	
{(scilicet: testo)}	Omissioni volute di una o più parti
#	Traduzione italiana delle parti in dialetto
Grassetto	Aggiunge una spiegazione parentetica, oppure una parola omessa nel testo
	Evidenzia le parti in dialetto
	Evidenzia il fenomeno preso in esame all'interno di ciascun esempio

Traccia dell'intervista sottoposta alle/agli informanti del *corpus* Nec Sine

1. Che lavoro fai?
 - 1.1. Da quanto tempo?
 - 1.2. In cosa consiste?
 - 1.3. Lavori in proprio o sei un dipendente?
 - 1.4. Hai sempre fatto lo stesso lavoro?
 - 1.4.1. Cosa facevi prima?
 - 1.5. Ti piacerebbe andare avanti a fare lo stesso lavoro o vorresti cambiare?
 - 1.5.1. Cosa vorresti fare?
 - 1.6. Pensi che l'Abruzzo sia un buon posto per lavorare?
 - 1.6.1. Dove sarebbe meglio?
 - 1.6.2. Hai mai vissuto fuori dall'Abruzzo? Dove?
 - 1.6.3. Come ti sei trovato?
 - 1.7. Ti trasferiresti all'estero per lavorare?
 - 1.7.1. Pensi che all'estero ci sia più lavoro?
 - 1.7.2. Dove andresti?
2. In che zona dell'Abruzzo vivi?
 - 2.1. Hai sempre vissuto lì?
 - 2.2. Hai frequentato le scuole lì?
 - 2.3. I tuoi amici vivono (ancora) vicino a te?
 - 2.4. Ti piace la tua zona?
 - 2.4.1. Perché?
 - 2.4.2. Ci sono dei problemi? Come miglioreresti la tua zona?
 - 2.4.3. Trovi tutto nel tuo paese o devi spostarti spesso?
 - 2.4.4. Come ti muovi generalmente per spostarti? Quali mezzi usi?
 - 2.5. L'Abruzzo, la Marsica e il tuo paese sono cambiati molto negli ultimi anni? In bene o in male?
 - 2.5.1. Sei contento del tuo paese?
3. Cosa fai durante il tuo tempo libero?
 - 3.1. Attività sportiva:
 - 3.1.1. Fai lo stesso sport da molto tempo?
 - 3.1.2. Quante volte a settimana?
 - 3.2. Uscita:
 - 3.2.1. Che posti preferisci?
 - 3.2.2. Quante volte a settimana li frequenti?
 - 3.3. Musica:

- 3.3.1. Qual è il tuo genere preferito?
- 3.3.2. Vai a concerti?
- 3.4. Calcio:
 - 3.4.1. Che squadra tifi?
 - 3.4.2. Ti ricordi quando hai scelto la tua squadra?
 - 3.4.3. Ti piace andare allo stadio o preferisci guardare le partite in tv? Perché?
- 3.5. Cinema:
 - 3.5.1. Che tipo di film ti piacciono?
 - 3.5.2. Vai spesso al cinema?
 - 3.5.3. Quali sono le tue serie Tv preferite?
- 3.6. Che uso fai generalmente di internet e dei social network?
4. Come hai trascorso il *lockdown*?
 - 4.1. Pensi che la pandemia ci abbia reso davvero migliori come si credeva?
5. Cosa ne pensi del dialetto? Quando lo parli? Perché? E con chi?

Riassunto Nel repertorio linguistico dell'italiano contemporaneo, i dialetti italo-romanzi affiancano ancora l'italiano nei contesti di comunicazione informale, e le manifestazioni di fenomeni di contatto tra italiano e dialetto che possono essere rinvenute nelle produzioni linguistiche spontanee di parlanti bilingui sono di conseguenza molteplici. Per individuare, categorizzare e analizzare nel modo più appropriato tale fenomenologia, è di fondamentale importanza basare le ricerche su dati linguistici reali, focalizzandosi su produzioni orali di parlanti bilingui che si avvicinino il più possibile al parlato spontaneo prototipico. Sulla base di tale convinzione, il lavoro qui presentato estrae dati provenienti dal modulo "giovani" del Nec Sine, un *corpus* di recente composizione che raccoglie il parlato di tredici giovani informanti abruzzesi. Il *corpus* si compone di poco più di 9 ore di parlato, raccolto in interviste semi-strutturate, per un totale di circa 96.000 occorrenze. Ai fini di questo lavoro, sono state scandagliate tutte le occorrenze del *corpus* in cui i giovani parlanti intervistati hanno impiegato il dialetto, con l'obiettivo di individuare e categorizzare i fenomeni di contatto tra italiano e dialetto presenti nelle loro produzioni. Per l'analisi di tali produzioni sono state impiegate le due macrocategorie di *code switching* ed *enunciazione mistilingue*, discretizzando queste due modalità di discorso bilingue sia sulla base di criteri formali e sintattici sia attraverso criteri extralinguistici, in termini di funzionalità comunicativa. Dai risultati dell'analisi sarà evidente come, anche appoggiandosi a dati provenienti da *corpora* di piccole dimensioni, si possano dare dimostrazioni empiriche di teorie invalse nella sociolinguistica contemporanea.

Fenomeni di contatto nel parlato di giovani abruzzesi

Abstract In the linguistic repertoire of contemporary Italian, Italo-Romance dialects can still appear in usage along with Italian in informal communicative situations, hence, instances of contact phenomena are a common habit in the spontaneous speech of Italian speakers. In order to identify, categorize and analyse such set of phenomena in the most accurate way, it is of paramount importance to base research on real linguistic data, for example by focusing on spoken utterances of bilingual speakers that are as close as possible to natural spontaneous speech. By relying on this assumption, this work draws its data from the “young speakers” module of a recently created *corpus* called Nec Sine, which gathers little more than 9 hours of spoken data (ca. 96,000 tokens) by thirteen young informants from Abruzzo region (Italy). The structure of the Nec Sine *corpus* will be presented in detail in the first part of the paper, then the analysis will proceed to identify and categorize contact phenomena between Italian and dialect, which is the main goal of this work. This will be done, at first, by isolating all the occurrences in which the speakers used the dialect code. Subsequently, all the isolated occurrences will be analysed in terms of code-switching and code mixing, discerning these two modes of bilingual discourse on the basis of both syntactic and extralinguistic criteria, in terms of communicative functionality. From the analysis results it will be clear that, even relying on data drawn from small *corpora*, affirmed theories of contemporary sociolinguistics can be confirmed by means of empirical demonstrations.

Manzoni postillatore della Crusca veronese: le unità fraseologiche

Ersilia Russo

1. Lessicografia e fraseologia per il romanzo

Il *Vocabolario degli Accademici della Crusca* può essere considerato «uno dei più importanti ferri del mestiere»¹ di Alessandro Manzoni. In cerca di una lingua, lo scrittore adopera i testi lessicografici quali riferimenti «di carta e d'inchiostro»² cui rivolgersi quando la penna viene meno, perché viene meno la conoscenza della lingua. L'andirivieni tra testo e vocabolari è messo in atto a partire e soprattutto in occasione della revisione del primo manoscritto, il *Fermo e Lucia*, redatto in poco più di due anni tra il 24 aprile del 1821 e il 17 settembre del 1823. In questo periodo, la consapevolezza dell'autore muta considerevolmente: da questione marginale, l'apparato linguistico diventa uno dei nuclei teorici cruciali nella stesura del romanzo. La compagine della Prima minuta, dominata da «lo screziato, l'appezzato, il cangiante»³, viene difatti rifiutata in nome di una maggiore omogeneità formale, guidata dall'Uso e dal Bisogno⁴. Gli sforzi vengono di lì incanalati verso un'unica direzione, quella della toscanità scritta. Ma non quella appartenente al paradigma letterario tradizionale: sono, infatti, le opere teatrali,

1 ISELLA 2005, p. XII.

2 Nella lettera a Tommaso Grossi del 1824 (ARIETI 1986, t. I, p. 355, n. 198).

3 SLE *Lettera a Casanova*, p. 316, par. II.

4 I primi riferimenti alle due *auctoritates* della lingua scritta sono del 1818, all'interno del trattato inedito e incompiuto *I modi di dire irregolari*.

comiche, scientifiche, dal Cinque al Settecento, a dettare la norma e a formare un canone linguistico inedito. La tessitura verbale che Manzoni vuole intrecciare è già rivolta alla medietà dell'espressione e alla concretezza della parola in uso; le costruzioni sintattiche anomale per la scrittura e i tratti dell'oralità rinvenibili nelle opere del rusticismo fiorentino o negli scambi epistolari cinquecenteschi costituiscono la materia a cui attingere per la creazione di una lingua «viva e vera».

I vocabolari costituiscono quindi un appiglio sicuro nei casi, numerosi, di dubbio e incertezza. Il *Vocabolario della Crusca*, ma anche il *Vocabolario milanese-italiano* di Francesco Cherubini, il *Dictionnaire de l'Académie Française* o il *Dictionnaire des proverbes français* di La Mésangère si fanno interlocutori quotidiani, divenendo autorità in materia di scrittura anche nella ricerca fraseologica, che ben presto si innesta come il fulcro della questione linguistica. Attraverso la fraseologia, intesa come sovraordinato che include ogni caso di parole cooccorrenti – dalle collocazioni ai modi di dire e ai proverbi – è possibile trasmettere nello scritto la vividezza espressiva tipica del parlato⁵. Accanto ad altri espedienti morfo-sintattici e lessicali, essa assolve il compito di restituire vitalità a una lingua che Manzoni considera morta da tempo⁶. Per tali motivi, non è casuale che il numero delle unità fraseologiche – le unità minime della fraseologia – inizi a incrementare a partire proprio dalla revisione del *Fermo*, caratterizzando fortemente il testo della *Seconda minuta*⁷, e quindi dell'edizione definitiva del romanzo. A causa della loro fattura convenzionale, le strutture fraseologiche sono specifiche di un determinato idioma, e la globalità che caratterizza la loro semantica comporta una ricerca inesausta dell'espressione precisa, puntuale. È quanto emerge dalla ricostruzione della trafila diacronica di alcune varianti multiparola: *venire a taglio* della *Ventisettana* (nel capitolo XI un'occorrenza e nel capitolo XIX due), cioè 'in acconcio, a

⁵ Vedi KOESTERS-BERARDINI 2020, pp. 17-43.

⁶ Cfr. ARIETI 1986, t. I, p. 19, n. 11.

⁷ Cfr. RABONI 2008, p. 13.

proposito', viene sostituito da *venire in taglio* alla luce dell'indicazione della fiorentina Giovanna Feroci Luti, documentata su un foglietto⁸.

Prima delle inchieste rivolte ai parlanti toscani e fiorentini, l'indagine sulla fraseologia è sostenuta proprio dai vocabolari, i quali riportano una ricca e variegata casistica di modi idiomatici e proverbiali. In CHERUBINI 1814 è interessante notare la trasposizione di espressioni dialettali in forma italiana, corrispettiva dal punto di vista del significato ma non della forma: s.v. *dent* leggiamo la locuzione idiomatica milanese *Toccà gnanch on dent*, tradotto con 'Non toccar l'ugola' e glossato con «Dicesi di cosa di cui si è mangiato scarsamente»; oppure s.v. *can* «A man a man come fa i can. *To' qua e dà qua*. Si dice in maniera proverbiale quando si tratta con persone di poca fede». Altrettanto che il Cherubini, il *Vocabolario della Crusca* gode di un ricco patrimonio di elementi idiomatici e fraseologici, spesso rinvenibili nelle numerose esemplificazioni, derivate da fonti vicine al canone manzoniano, come le opere di Francesco Berni, di Michelangelo Buonarroti il Giovane, di Agnolo Firenzuola, di Anton Maria Salvini. Tuttavia, gli elementi fraseologici riportati dal vocabolario non soddisfano tutte le necessità espressive dello scrittore: per questo, numerose unità fraseologiche vengono appuntate sui margini dei sette volumi, conciate in modo da non lasciar vedere il testo⁹. Le postille arricchiscono le voci di nuove espressioni, per lo più derivate dalla lettura dei «libri toscani d'ogni secolo»¹⁰: le lettere di Annibal Caro, le commedie di Giovan Maria Cecchi, il *Malmantile racquistato* di Lorenzo Lippi, le *Lettere familiari* di Lorenzo Magalotti, le novelle di Franco Sacchetti, per citarne alcuni. Già VITALE 1992 nota che lo scrittore milanese

somministra, con costante legittimazione di esempi forniti dai consueti scrittori, nuove unità lessicali formate con vocaboli già esistenti nel vocabolario (le locuzioni così importanti per il Manzoni): sono postille, queste, che denotano nel Manzoni una coscienza terminologica e lessicale di grande scrittore e, più

8 SLII, p. 782, t. II.

9 Cfr. SLE *Appendice alla Relazione VI*, p. 234, par. 9.

10 *Ibidem*.

di tutto, il sollecito intento – riconoscibile anche nella postillatura già menzionata alla Proposta del Monti e a quella del dizionario milanese-italiano del Cherubini – di conquista della lingua toscana media e popolare, divenuta il fine costante della sua insistita lettura 'lessicografica' dei testi e dei dizionari¹¹.

Proprio per la pervasività del fenomeno, è apparso utile compiere uno studio sulle postille manzoniane alla *Crusca veronese* contenenti unità fraseologiche, in particolare idiomatiche, con lo scopo di far emergere il ruolo da esse ricoperto nella costituzione della lingua del romanzo. Nel presente contributo, a una digressione sull'esemplare manzoniano e sul rapporto dello scrittore col vocabolario, fa seguito un approfondimento sulle postille e sulla loro molteplice natura. In conclusione, si ricostruisce la vicenda redazionale di alcune postille particolarmente significative.

2. La *Crusca veronese* nella cultura dell'Ottocento

Il vocabolario posseduto e compulsato da Manzoni è il *Vocabolario degli Accademici della Crusca oltre le giunte fatteci finora, cresciuto d'assai migliaia di voci e modi de' classici, le più trovate da Veronesi* stampato a Verona presso Dionigi Ramanzini con data 1806, ma in realtà pubblicato tra il 1806 e il 1811, in sette volumi. L'edizione è voluta e curata dall'abate purista Antonio Cesari, che ne sostiene anche le spese. Non si tratta, infatti, di un'edizione licenziata dagli accademici cruscanti, ma di una ristampa della quarta impressione del vocabolario, la seconda veneziana di Francesco Pitteri (1763), ormai «uscita dai torchi da oltre un quarantennio»¹². A questa edizione, vengono aggiunti gli spogli del veronese padre Girolamo Lombardi, del classicista roveretano Clementino Vannetti, del veronese Paolo Zanotti, e del Cesari stesso¹³, che attingono «a

¹¹ VITALE 1992, pp. 184-185.

¹² DARDI 2003, p. 31.

¹³ In corso d'opera si aggiungono i contributi dei veneti Bartolomeo Bottari e Giuseppe Pederzani.

piene mani alla favolosa (e favoleggiata, come aveva mostrato il Cesariotti), incontaminata purezza della lingua del buon secolo»¹⁴. Il vocabolario si instaura come la conseguente deduzione lessicografica del purismo trecentesco, che propugna un integrale quanto anacronistico ritorno all'uso linguistico dell'aureo trecento, pure con alcune incursioni cinquecentesche.

In quanto «repertorio dell'italiano scritto tre-cinquecentesco di tipo fiorentino»¹⁵, l'edizione veronese diventa ben presto «uno strumento utilissimo per la pratica scrittoria e per l'unificazione linguistica dell'Italia colta»¹⁶. Difendendo una visione arcaizzante e puristica della lingua italiana, essa rischia, tuttavia, di ostacolare la naturale evoluzione delle strutture linguistiche e il loro arricchimento provocato da fattori endogeni ed esogeni. Dalla contrapposizione tra utilità pratica e arretratezza ideologica deriva il duplice atteggiamento degli scrittori e intellettuali di accettazione, legata alle necessità di scrittura, e di contestazione della sua «lacunosità, povertà, rigidità»¹⁷. Così la *Crusca veronese*, come è tradizionalmente denominata, diventa un «incontro obbligato»¹⁸ a prescindere dalle rispettive posizioni in fatto di lingua. Spesso annotata, le postille lasciate sui suoi volumi assolvono una duplice funzione: da una parte, arricchiscono il vocabolario di nuove voci, per lo più esterne al canone trecentesco; dall'altra, creano uno spazio in cui affermare le convinzioni linguistiche personali, solitamente contrarie a quelle dei compilatori. Il testo ha avuto illustri postillatori: oltre a Manzoni, si ricordano Foscolo, Leopardi, Monti e Tommaseo. Foscolo, in particolare, giudica il vocabolario «bello, e buono, ed utilissimo»¹⁹, ma cambia opinione in un secondo momento²⁰.

¹⁴ DARDI 2003, p. 31.

¹⁵ NENCIONI 1986, p. 286.

¹⁶ *Ibidem*.

¹⁷ *Ibidem*.

¹⁸ *Ibidem*.

¹⁹ CARLI 1952, pp. 137-38.

²⁰ CARLI 1954, p. 142. Cfr. DARDI 2003, p. 33, n. 3.

Monti, invece, si scaglia subito contro la proposta anacronistica dei puristi, «sfogando il proprio disappunto nei vivagni della propria copia della *Crusca* ‘veronese’, quindi pubblicamente con quattro dialoghi satirici pubblicati tra “Il Poligrafo” e la “Biblioteca Italiana”²¹, fino alla pubblicazione, tra il 1817 e il 1826, della *Proposta di alcune correzioni ed aggiunte al Vocabolario della Crusca*²².

3. La postillatura manzoniana

I sette volumi della *Crusca veronese* appartenuti e postillati da Manzoni sono conservati nella Sala Manzoniana della Biblioteca Nazionale Braidense con segnatura MANZ. 16. 0205-0211 dal 1994. Prima di allora, sono appartenuti a due collezionisti privati milanesi (Emilio Brusa per i tomi I, III-VII e Giovanni Treccani degli Alfieri per il tomo II), i quali hanno concesso la realizzazione della prima edizione critica, curata da Dante Isella, nel 1964. Oggi, sul portale *Manzoni Online*, è liberamente consultabile la loro completa digitalizzazione²³.

Da un punto di vista materiale, è possibile notare che le pagine del vocabolario riversino in uno «stato di grave usura [...] e diciamo pure di sporcizia depositata sugli angoli utilizzati per sfogliarle»²⁴, a dimostrazione di una consultazione quotidiana dell’opera, considerata alla stregua di un diario su cui tenere traccia delle conquiste linguistiche acquisite *in fieri*. I volumi sono «crivellati»²⁵ di annotazioni, contenenti riferimenti delle letture dei testi toscani di epoca antica e moderna, confronti puntuali con il milanese, con il francese e con il latino, giudizi e domande di vario tenore, spesso retoriche²⁶. La fittezza delle po-

²¹ DARDI 2003, p. 34.

²² Cfr. BONSI 2018.

²³ All’indirizzo alessandromanzoni.org.

²⁴ DANZI 1998, p. 59.

²⁵ NENCIONI 1986, p. 284.

²⁶ I riferimenti alle lingue altre (tra cui anche il latino) possono essere considerati, alla luce di una «subordinazione intenzionale e istituzionale al toscano» (CORTI

stille genera una sensazione «conturbante»²⁷ in chi si avvicina a quelle carte per la prima volta. «È l'impressione», scrive Corti, «di una sublime pedanteria, lievemente imparentata col nevrastenico»²⁸. Le ragioni che muovono lo scrittore a intraprendere «un indomito corpo a corpo con le parole»²⁹ così capillare può essere ricondotta all'«ansia di assoluto»³⁰ tipica della sua disposizione mentale, all'aspirazione all'ordine, alla razionalità, alla sistematicità del metodo, nei quali risiede la verità di ogni operazione morale e intellettuale. Annotare i testi con postille verbali e non verbali è, infatti, un «processo congeniale alla mente del Manzoni»³¹, che in questo modo intraprende un «dialogo impegnato con i suoi autori, pronto al consenso o al dissenso»³². In questo senso, le postille rendono manifesta la continua vigilanza di Manzoni sul modo altrui di lavorare. Sulla *Crusca veronese*, confluisce, dunque, una massa poderosa di testi toscani, popolareggianti e berneschi, novellistici e cronistici, nonché scientifici ed epistolari, che fissano un nuovo modello di scrittura, in cui si manifestano l'ideale e il programma di lingua “media” per la propria prosa letteraria. La postillatura manzoniana alla *Crusca veronese*, condotta dallo scrittore milanese parallelamente alla revisione del romanzo, appare senza dubbio funzionale alla formazione della sua lingua³³. DANZI 1998 sottolinea «l'alto coefficiente di 'strumentalità' dei materiali lessicali della Crusca, cioè la loro 'capacità di servizio' (Nencioni) nei confronti della lingua del romanzo»³⁴.

1969, p. 157) come «noto dal quale ci si muove per chiarire l'ignoto: l'italiano» (FORTI 1965, p. 312).

27 CORTI 1969, p. 149.

28 Ivi, pp. 149-150.

29 ISELLA 2005, p. XII.

30 Ivi, p. XXV.

31 CORTI 1969, p. 150.

32 ISELLA 2005, p. XXI.

33 Laddove, invece, quella condotta sui vocabolari del Cherubini è legata a un lavoro per lo più successivo e collettivo, e avente ragioni prettamente sociolinguistiche (cfr. DANZI 2001, FERRARI 2017).

34 DANZI 1998, p. 56.

Secondo una prospettiva contenutistica, dalle postille trapela un sentimento, comune alla cultura lombarda, di anticlassicismo e antipurismo. Ne consegue una chiara avversione per gli arcaismi fonetici, lessicali e sintattici, in forza di un principio di chiarezza di matrice razional-illuministica. Proprio in virtù di essa, «l'arcaismo non assume dignità di testimonianza storica, ma appare il doppiamente grottesco e superfluo d'una forma vivente, buono solo a ingenerare confusione e incertezza nell'uso della lingua»³⁵. Lo strumento linguistico, invece, deve essere efficiente e moderno, conforme agli usi della società in cui è adottato. A questo principio si lega l'incessante ricerca di coerenza logica, che si esprime tramite osservazioni morfologiche e sintattiche rivolte agli stessi estensori del vocabolario³⁶. Manzoni, inoltre, rifiuta strenuamente le voci sinonimiche e i doppioni linguistici, in nome di una chiara distinzione dei significati³⁷. Emerge, infine, un interesse precoce per l'uso "normale", "medio", della lingua e della parlata toscana: nei passi annotati si nota la volontà di perseguire un tono quotidiano, colloquiale, rifiutando la frase tipica, il gioco di parole, l'uso espressionistico o popolare.

Alla complessa stratificazione delle postille, Dante Isella ha cercato di mettere ordine individuando una triplice categorizzazione tipologica, cui corrisponde una parziale cronologia genetica. Su prove materiali, quali differenze di inchiostro e mutamenti del *ductus*, è in realtà pressoché impossibile ricostruire la diacronia di un'annotazione così frammentaria. Alcuni elementi utili alla datazione sono i rimandi cronologici alle edizioni citate e i riferimenti che caratterizzano l'evoluzione del pensiero linguistico manzoniano. La prima tipologia comprende le postille riferite alle giunte veronesi. Di natura anticlassica e antipuristica, esse esprimono un impetuoso dissenso verso i metodi e le scelte dei compilatori, pronti al recupero di qualsiasi forma de-

³⁵ DARDI 1998, p. 55.

³⁶ Cfr. s.v. alto, in ISELLA 2005, p. 19.

³⁷ Celebre, in questo senso, l'intervento manzoniano alla voce panna del *Prontuario* di Giacinto Carena. Nel vocabolario, vedi il verbo sintagmatico *andare dietro* in ISELLA 2005, p. 28.

sumibile dai testi antichi, a stampa o anche manoscritti, spesso con lezioni frutto di errori di copista o di lettura. È il caso della postilla a *bugnone*: «A veder questo bugnone, vien subito in testa che possa esser fallo di copista, per *buscione*, e infatti, a questa voce, trovi un es.º colle stesse parole a un dipresso»³⁸ e a *dilanato* «E se fosse un granchio del copista? o dell'autore? sarebbe lingua? anzi sarebbe mai stata?»³⁹. Viceversa, «[d]ove l'arcaismo fonetico o fonetico-lessicale è riconosciuto come storicamente effettivo, viene combattuto in forza di un principio di chiarezza e di semplificazione»⁴⁰; Manzoni commenta così la voce *comata*: «E questa è ricchezza di lingua, aver due parole a significare una medesima idea? sarebbe confusione se fosse nella lingua, ma qui è da credere sia uno strafalcione di F. Giordano, o del copista. E fosse pure una volta questa voce stata nella lingua, sarebbe ora da metterla nel dimenticatojo – È da credere che quei valentuomini che compilarono il Vocab. avranno avuta contezza di questa e di molte altre, e le avranno a bella posta tralasciate; e se la cosa è così, hanno fatto bene»⁴¹. Ancora, «[l]'arcaismo sintattico, lessicale o semantico è combattuto in forza del principio dell'uso»⁴², come in *aver luogo*⁴³, e sono respinte le forme occasionali, con variazioni fonetiche arbitrarie, probabilmente oggetto di errori tipografici. Tali annotazioni si distinguono per l'impulsività della scrittura, che denota l'immediatezza della reazione e la volontà di prendere le distanze da quanto riportato nel testo; tant'è che talvolta queste possono essere censurate in un secondo momento con fitte inchiostature. I termini *ante quem* per la datazione del primo gruppo di postille sono individuabili nella prima *Introduzione* al *Fermo* e nella *Proposta* montiana, postillata da Manzoni; anzi, più nello specifico, nei *Dialoghetti*.

³⁸ ISELLA 2005, p. 70.

³⁹ Ivi, p. 162.

⁴⁰ CORTI 1969, p. 155.

⁴¹ ISELLA 2005, p. 103.

⁴² CORTI 1969, p. 155.

⁴³ «Si usa? come lo sapete? perchè il Cavalca l'ha usato una volta? E perchè l'ha usato *alla latina*? traducendo? È questa l'idea dell'Uso?» (ISELLA 2005, p. 53).

Il secondo gruppo, che costituisce il nucleo maggiore, comprende le postille che arricchiscono l'esemplificazione del vocabolario. I nuovi lemmi hanno la funzione di precisare il nucleo del significato della voce a cui appartengono, segnalare nuovi significati o sfumature di questi. In esso, confluisce il bottino delle molte letture condotte negli anni di scrittura del romanzo. Il canone letterario che ne emerge è rivoluzionario; quasi non compaiono i poeti dell'alta tradizione lirica, mentre abbondano gli accademici del rusticismo fiorentino (Michelangelo Buonarroti il Giovane, con la *Fiera* e la *Tancia*, Lorenzo Lippi del *Malmantile* e commentatori linguistici: Salvini, Minucci e Biscioni), gli scrittori «comici» (Francesco D'Ambra, Giovan Maria Cecchi, il Lasca), i berneschi; fra i prosatori, i comici (Giovan Battista Gelli, Agnolo Firenzuola, Giovanni Battista Fagiuoli), i novellieri (Franco Sacchetti); i cronisti; i volgarizzatori storici (Niccolò Machiavelli, Dino Compagni, Giovanni Villani); i trattatisti (Benedetto Varchi, Leonardo Salviati); gli iniziatori della prosa scientifica (Galileo Galilei, Lorenzo Magalotti); gli scrittori di lettere (Annibale Caro) e di memorie (Benvenuto Cellini)⁴⁴. L'immissione del patrimonio letterario toscano tra le pagine della *Crusca* avviene prevalentemente nel periodo precedente e contemporaneo alla revisione del primo manoscritto del romanzo, in preparazione della sua stampa, tra il '23 e il '25. In questo lasso temporale vengono appuntate molte voci «toscano-milanesi»⁴⁵, che presentano una concordanza tra la forma milanese e quella toscana: una felice coincidenza che porta a constatare una comunanza linguistica che supera i limiti territoriali⁴⁶. Sono presenti anche alcune aggiunte tarde, successive al 1827, collocabili cronologicamente grazie alla data di pubblicazione dei testi citati, come l'*editio maior* del Vocabolario del Cherubini (1839-42) o le lettere inedite del Caro del '27; oppure postille di locuzioni toscane che entrano soltanto nell'edizione definitiva dei *Promessi sposi* «come

⁴⁴ Cfr. CARTAGO 2013.

⁴⁵ Cfr. ARIETI 1986, p. 380, n. 220.

⁴⁶ Come avviene nella postilla alla voce *panno*, dove l'espressione idiomatica *bianco come panno curato* tratta dal Minucci diventa *come panno lavato* secondo l'uso milanese e toscano (ISELLA 2005, p. 62).

correzione alla lezione della Ventisettana, al tempo della quale Manzoni sembra ignorarle»⁴⁷. Sempre successive al 1827, sono le postille che fanno riferimento all'uso di Firenze, «del quale si intende specificamente chiedere conto»⁴⁸ o che «recano i segni di una già acquisita e piena conoscenza dell'uso tosco-fiorentino, della quale Manzoni deve essere entrato in possesso dopo la sua andata a Firenze»⁴⁹.

Infine, all'ultimo gruppo appartengono le osservazioni rivolte agli stessi compilatori cruscanti: alla definizione della voce, alla sua struttura, alle corrispondenze latine, all'interpretazione e alla congruenza degli esempi con la definizione. Tali commenti rispecchiano interessi più propriamente lessicografici, riferibili allo stesso piano cronologico della prima categoria.

4. Le postille fraseologiche: alcuni casi di studio

Situate a metà tra lessico e sintassi, le unità fraseologiche si costituiscono come la chiave di volta per la creazione di una lingua «viva e vera». «[Q]ue' modi di dire composti di più vocaboli, e che hanno un loro significato particolare e determinato»⁵⁰, secondo la definizione manzoniana, diventano essenziali per arricchire il tessuto narrativo del romanzo. Tra le postille del secondo gruppo, numerose coinvolgono unità fraseologiche, scoperte o riscoperte attraverso le recenti letture. Di seguito viene fornita una serie di esempi di unità fraseologiche contenute nelle annotazioni alla *Crusca veronese*, secondo una doppia casistica: postille che commentano locuzioni registrate nel vocabolario

⁴⁷ VITALE 1992, p. 202.

⁴⁸ *Ibidem*. Alla voce *insino*, Manzoni commenta «Insognarsi una cosa in milanese vale sognarla; il Magal. lett. fam. 21. par. 1. pag. 351, ha: chi vi ha detto che voi non ve l'insognate? Verificare se il vocabolo sia veramente usitato in Firenze» (ISELLA 2005, p. 292).

⁴⁹ VITALE 1992, p. 202. S.v. *Fonda* «Ora in Firenze si dice: piena.» (ISELLA 2005, p. 249).

⁵⁰ SLI *Della lingua italiana* v, p. 399, par. 353.

(*dare noia, fare fico*) e postille che aggiungono nuove locuzioni, derivate dai testi toscani (*avere le lune, tremare come una foglia*).

Dare noia

Alla voce *dare*, il vocabolario registra la locuzione verbale *dare noia*, accompagnata dal corrispettivo monorematico *noiare* e da una ricostruzione etimologica che fa risalire l'espressione al latino *molestia afficere*. Secondo il GRADIT s.v. *noia*, la locuzione significa 'dare fastidio, disturbare'. La postilla manzoniana aggiunge la glossa sottintesa nella voce lessicografica: «Dar noia, vale anche: essere ostacolo, e importunare»⁵¹. Appare significativo l'utilizzo della forma apocopata del verbo. A questa specificazione semantica, segue la citazione dall'*Ercolano* di Benedetto Varchi, uno dei testi di lingua postillato da Manzoni: «Varch. Erc. 1. 134. e' ne pizzica (del furbesco) ... ma che noia dà, o qual mia colpa? – È frequente presso i comici. Verificare se sia dell'uso vivente»⁵². Alla citazione, succede una marca di frequenza: l'espressione è tipica presso gli scrittori comici. Dopodiché, viene invocata la verifica dell'uso vivo, concreto, dei parlanti. «L'autentica ossessione del Manzoni per il vivente», scrive FORTI 1965, «si rivela anche indirettamente nella attenzione particolare che egli concede alle locuzioni proverbiali e ai commentatori toscani dei testi di lingua, che gli avvicinano cronologicamente il testo e gli documentano la vitalità nel tempo di certe espressioni»⁵³.

La locuzione è assente nel *Fermo e Lucia*, mentre inizia ad essere inserita ne *Gli sposi promessi*, per essere confermata fino alla *Quarantana*; l'ammissione nel testo è sicuramente correlata alla testimonianza varchiana riportata sulla *Crusca veronese*:

⁵¹ ISELLA 2005, p. 147.

⁵² *Ibidem*.

⁵³ FORTI 1965, p. 316.

- «continuava tra sè: – che noia mi dà costei! Liberiamocene. –» (Sp II XX 45);
- «con tutto ciò, non si può dire che noia gli desse quella vista.» (Sp II XXX 6);
- «Levami un po' quel lume dinanzi, che m'accieca... mi dà una noia...!» (Sp II XXXIII 3);
- «la luce del giorno già inoltrato gli dava noia, quanto quella della candela, la sera avanti» (Sp II XXXIII 10).

In quattro casi, la locuzione verbale viene aggiunta solo nella *Quarantana*:

- «Quel che Dio vuole, – rispondeva ai pensieri che gli davan più noia: – quel che Dio vuole» (Q XVII 26), sostituisce «che più imperversavano» (Sp e V II XVII 26)
- «lì non c'era politica: era proprio vero che gli dava noia d'aver i suoi anni» (Q XXIX 20)
- «gli dà noia il bene stare» (Q XXIII 59) sostituisce «ha male di troppo bene» (Sp e V II XXIII 59)
- «Piuttosto, quel che mi dà noia...» (Ps40 XXIX 23) «quel che mi dà fastidio» (Sp e V II XXIX 23)

Un'occorrenza entrata negli *Sposi promessi* cade nelle redazioni successive: «gli dava non so qual noia inesplicabile» (Sp II XVII 15) è sostituita nella *Ventisettana* con «aveva pel suo orecchio non so che di odioso» (V II XVII 15).

Fare fico

La locuzione verbale *fare fico* è registrata sulla *Crusca veronese s.v. fare*, con il significato di «Dare in nulla», dal latino *frigere; indoli, aut expectationi non respondere* e dal greco *καταψύχεται*⁵⁴. Con un segno di rappic-

⁵⁴ ISELLA 2005, p. 223.

co accanto al sottolemma, la postilla indicante la locuzione e la fonte linguistica è riportata sul margine superiore: «Fare fico. Salv. Pr. Tosc. Cic. III. p. 419. (517) Andò... (Crasso) ... Che ne avvenne? Ebbe la rotta, fece fico»⁵⁵. Si tratta del volume delle *Prose toscane* di Antonio Maria Salvini, stampato a Venezia nel 1734. Sull'esemplare manzoniano, l'espressione *fare fico* a p. 419 è sottolineata a penna: si concretizza in tal modo il cortocircuito tra la compulsazione dei testi di lingua e la trascrizione dei modi più interessanti e utili sul vocabolario cruscante. Sul margine destro invece, all'altezza della voce lessicografica, troviamo specifiche ulteriori, forse apportate in un secondo momento, essendo la penna diversa da quella della nota precedente: «disusato: ora si dice a Firenze come forse per tutta Italia: far fiasco»⁵⁶. Anche in questo caso, Manzoni è attento all'uso concreto dell'unità fraseologica: *fare fico* appare come un'espressione antiquata, sostituita ormai dalla locuzione *fare fiasco*, fiorentina e probabilmente italiana. Entrambe le espressioni non sono impiegate nel romanzo.

Avere le lune

L'espressione idiomatica *avere le lune* è aggiunta da Manzoni alla voce lessicografia *luna*, di cui era sprovvista. Si tratta di un caso in cui lo scrittore, tramite la lettura dei testi di lingua, incontra espressioni a lui congeniali e calzanti, ma che fino a quel momento non ha rinvenuto nella lingua letteraria. In questo caso, sono le annotazioni alla *Fiera* di Salvini a costituire la fonte di lingua viva e agile: «Aver le lune, cioè le paturne; e d'un bisbetico si dice: aver la luna a rovescio. Salv. not. Fier. pag. 386. col. 1»⁵⁷. Sul volume manzoniano, il passo è sottolineato e rilevato da una croce uncinata. La certificazione salviniana è tanto più fortunata in quanto serve a constatare la corrispondenza dell'espressione, avvertita come milanese, con il tosco-fiorentino; e infatti

⁵⁵ *Ibidem.*

⁵⁶ *Ibidem.*

⁵⁷ ISELLA 2005, p. 313.

Manzoni annota sul margine inferiore della pagina «Benedetto il Salvini che mi somministra un esempio toscano di questo modo di dire tanto usitato in Lombardia»⁵⁸. L'espressione entra nel romanzo, però, solo nella *Ventisettana*, dove sostituisce la polirematica «è già tanto di mal umore per quel maledetto Casale» (Sp II XIII 63), secondo la forma estesa «ha già tanto le lune a rovescio per quel maladetto Casale» (V II XIII 63), reso al singolare nella *Quarantana* «ha già tanto la luna a rovescio per quel maledetto Casale» (Q XIII 63).

Tremare come una foglia

Il caso del paragone standardizzato *tremare come una foglia* testimonia la tensione manzoniana alla sistematicità del processo di ricerca linguistica, per cui le note stesse sono collegate da rimandi intratestuali. Alla voce *tremare* del vocabolario, sottolemma *Tremare verga a verga o come una verga*, cioè 'eccessivamente', Manzoni registra la fonte letteraria «Noi diciamo tremar come una foglia. Salv. not. alla Fiera, p. 377, col. 1. – Nè più nè meno diciamo noi milanesi»⁵⁹. Seguono due righe pesantemente cassate con la penna. Alla citazione succede una considerazione sociolinguistica, che sottolinea la concordanza tra toscano e milanese. Manzoni riscrive la locuzione sotto il lemma *verga*, rimandando alla voce *tremare* («V. Tremare») per la definizione. Nel romanzo, il paragone viene inserito nella *Ventisettana*: «Il garzoncello trema come una foglia» (V I VIII 43).

5. Conclusioni

Come abbiamo visto negli esempi sopra riportati, le unità fraseologiche derivate dai testi toscani che confluiscono nelle postille alla *Crusca veronese* documentano la necessità da parte di Manzoni di circoscrive-

⁵⁸ *Ibidem*.

⁵⁹ *Ivi*, p. 550.

re un canone linguistico nuovo per il romanzo; un canone che diventa precocemente un programma di lingua “media” e colloquiale, per quanto ancora ricercata sui testi scritti. Le postille «riflettono lo sforzo del romanziere per raggiungere una lingua italiana diversa da quella della tradizione letteraria, perché arricchita, in prospettiva, di un nuovo e moderno fattore, la dimensione sociale della comunicazione linguistica»⁶⁰. Il vocabolario può essere così considerato un «indispensabile reagente»⁶¹ per la ricostruzione del pensiero linguistico manzoniano e dell'evoluzione linguistica dei *Promessi sposi*, nascondendo il «segreto dell'officina del romanzo»⁶², il laboratorio dove, attraverso una ricerca minuziosa, sistematica, durevole, mettere a punto una lingua non immaginata esclusivamente per la letteratura, i cui sviluppi influenzeranno la stabilizzazione della lingua italiana moderna anche per quanto riguarda il patrimonio idiomatico e fraseologico.

Bibliografia primaria

FL = *Fermo e Lucia: prima minuta (1821-1823)*, a cura di Barbara Colli, Paola Italia, Giulia Raboni, Milano, Casa del Manzoni, 2007.

Sp = *Gli Sposi promessi: seconda minuta (1823-1827)*, a cura di Barbara Colli, Giulia Raboni, Milano, Casa del Manzoni, 2012.

V = *I promessi sposi (1827)*, a cura di Salvatore Silvano Nigro, Milano, Mondadori, 2002.

Q = *I promessi sposi (1840) – Storia della colonna infame*, a cura di Salvatore Silvano Nigro, Ermanno Paccagnini, Milano, Mondadori, 2002.

Arieti 1986 = *Tutte le lettere*, a cura di Cesare Arieti, 3 tomi, Milano, Adelphi, 1986.

ISELLA 1964 = ALESSANDRO MANZONI, *Postille al vocabolario della Crusca nell'edizione veronese*, a cura di Dante Isella, Milano-Napoli, Ricciardi, 1964.

ISELLA 2005 = ALESSANDRO MANZONI, *Postille al Vocabolario della Crusca nell'edizione veronese*, a cura di Dante Isella, in *Edizione nazionale ed europea delle*

⁶⁰ DANZI 1998, p. 59

⁶¹ CORTI 1969, p. 159.

⁶² ISELLA 2005, p. XIII.

Manzoni postillatore della Crusca veronese

- opere di Alessandro Manzoni*, diretta da Giancarlo Vigorelli, vol. 24, Milano, Casa del Manzoni, 2005.
- SLE = *Scritti linguistici editi*, a cura di Angelo Stella e Maurizio Vitale, in *Edizione nazionale europea delle opere di Alessandro Manzoni*, diretta da Giancarlo Vigorelli, vol. 19, Milano, Casa del Manzoni, 2000.
- SLI = *Scritti linguistici inediti*, a cura di Angelo Stella, Maurizio Vitale, in *Edizione nazionale ed europea delle opere di Alessandro Manzoni*, diretta da Giancarlo Vigorelli, vol. 17, Milano, Casa del Manzoni, 2000.
- SLII = *Scritti linguistici inediti*, a cura di Angelo Stella, Maurizio Vitale, in *Edizione nazionale ed europea delle opere di Alessandro Manzoni*, diretta da Giancarlo Vigorelli, vol. 18, Milano, Casa del Manzoni, 2000.

Bibliografia secondaria

- BONSI 2018 = CLAUDIA BONSI, *Liste lessicali e nascita di un lessicografo: il caso di Vincenzo Monti*, in «Prassi Ecdotiche della Modernità Letteraria», 3, 2018, pp. 369-383.
- CARLI 1952 = PLINIO CARLI, *Epistolario*, Volume 2 (luglio 1804-dicembre 1808) in *Edizione Nazionale delle opere di Ugo Foscolo*, vol. 15, Firenze, Le Monnier, 1952.
- CARLI 1954 = PLINIO CARLI, *Epistolario*, Volume 4 (1812-1813) in *Edizione Nazionale delle opere di Ugo Foscolo*, vol. 17, Firenze, Le Monnier, 1954.
- CARTAGO 2013 = GABRIELLA CARTAGO, *Un laboratorio di italiano venturo: postille manzoniane ai testi di lingua*, Milano, Centro nazionale studi manzoniani, 2013.
- CHERUBINI 1814 = FRANCESCO CHERUBINI, *Vocabolario milanese-italiano*, Milano, dalla Stamperia Reale, 1814.
- CORTI 1969 = MARIA CORTI, *Uno scrittore in cerca della lingua*, in *Metodi e fantasmi*, Milano, Feltrinelli, 1969, pp. 143-59 (ora in *Nuovi metodi e fantasmi*, ivi, 2001), già in «L'Approdo Letterario», XXVII, 1964, pp. 3-20.
- DANZI 1988 = ALESSANDRO MANZONI, *Le postille del Manzoni al Vocabolario della Crusca: con una appendice di altri documenti*, a cura di Luca Danzi, Milano, Biblioteca Nazionale Braidense, 1988.
- DANZI 2001 = LUCA DANZI, *Lingua nazionale lessicografia milanese. Manzoni e Cherubini*, Alessandria, Edizioni dell'Orso, 2001.
- DARDI 1988 = ANDREA DARDI, *Per la datazione di alcune postille del Manzoni alla Crusca Veronese*, in «Studi linguistici italiani», XIV, 1988, pp. 50-65.

- DARDI 2003 = *La Crusca veronese del Cesari*, in *La Crusca nell'Ottocento*, a cura di Elisabetta Benucci, Andrea Dardi, Massimo Fanfani, Firenze, Società Editrice Fiorentina, 2003, pp. 31-36.
- FERRARI 2017 = JACOPO FERRARI, *Sulle postille inedite manzoniane alla seconda edizione del Vocabolario milanese italiano di Francesco Cherubini*, in «Studi italiani di linguistica teorica e applicata (Centro Interfacoltà di Linguistica Teorica e Applicata. Università di Bologna)», XLVI, 1, 2017, pp. 95-108.
- FORTI 1965 = FIORENZO FORTI, *Fra le carte dei poeti*, Milano-Napoli, Ricciardi, 1965, pp. 304-320, già in «Giornale Storico della Letteratura Italiana», CXXXI, 1954, pp. 352-385.
- GRADIT = TULLIO DE MAURO, *Grande Dizionario italiano dell'uso*, Torino, UTET, 1999-2000.
- KOESTERS-BERARDINI 2020 = SABINE KOESTERS GENSINI, ANDREA BERARDINI, *Si dice in molti modi. Fraseologia e traduzioni nel Visconte dimezzato di Italo Calvino*, Roma, Sapienza Università, 2020.
- NENCIONI 1986 = GIOVANNI NENCIONI, A. *Manzoni e l'Accademia della Crusca*, in «Atti e Memorie dell'Arcadia Letteraria Italiana», serie III, vol. VIII, 1983-1985, f. 2-3, pp. 1-29 dell'estratto; poi in «Nuova Antologia», CXXI, 1986, f. 2157, pp. 279-302.
- RABONI 2008 = GIULIA RABONI, *La scrittura purgata. Sulla cronologia della Seconda minuta dei "Promessi sposi"*, in «Filologia italiana», 5, 2008, pp. 191-208.
- VITALE 1992 = MAURIZIO VITALE, *Le postille del Manzoni al «Vocabolario della Crusca» nell'edizione Veronese*, in *Studi di storia della lingua italiana*, Milano, LED (Edizioni universitarie di Lettere Economia Diritto), 1992, pp. 181-204, già in «Annali Manzoniani», n.s., I, 1990, pp. 175-196.

Riassunto L'attenzione manzoniana nei confronti delle unità fraseologiche – le unità minime della fraseologia – confluisce, oltre che negli scritti linguistici e nell'epistolario, anche nell'operazione sistematica di postillatura del *Vocabolario degli Accademici della Crusca*, nell'edizione curata dall'abate Cesari del 1806. L'esemplare manzoniano, conservato presso la Sala Manzoni della Biblioteca Nazionale Braidense, costituisce uno dei principali «ferri del mestiere» dello scrittore per la creazione del tessuto linguistico dei *Promessi sposi*. In particolare, il suo ruolo nella definizione della lingua del romanzo diventa cruciale nel momento di revisione della Prima minuta, in vista della stampa. Tra il 1823 e il 1825, Manzoni si dedica allo studio del vocabolario, dialogando con le sue voci attraverso una fitta rete di note, che ne arricchiscono il patrimonio lessicale e fraseologico attraverso i rimandi alle letture dei testi toscani svolte in quel periodo. Tra le postille,

Manzoni postillatore della Crusca veronese

di cui Dante Isella ha proposto una classificazione per cronologia e tipologia, numerosi sono i riferimenti alle unità fraseologiche. Tali strutture diventano essenziali a partire dalla Seconda minuta, perché capaci di veicolare nella scrittura un maggiore senso di oralità e di medietà. Scopo del contributo sarà, dunque, quello di portare alla luce la ricerca manzoniana svolta a livello fraseologico attraverso le tracce lasciate sulle pagine della *Crusca veronese*, in cui è possibile ricostruire l'innovativo canone letterario utilizzato per dare forma alla lingua della seconda stesura del romanzo.

Abstract Manzoni's interest in phraseological units – the minimal units of phraseology – concerns not only his linguistic writings and letters, but also the systematic annotation of *Vocabolario degli Accademici della Crusca*, edited by Antonio Cesari in 1806. Manzoni's volume, preserved at Sala Manzoniana of the Biblioteca Nazionale Braidense, is one of the writer's main «tools of the trade» in creating the linguistic fabric of *The Betrothed*. In particular, its role in defining the language of the novel becomes crucial at the time of revision of the first manuscript, in preparation for printing. Between 1823 and 1825, Manzoni devoted himself to the study of the vocabulary, dialoguing with its entries through a dense network of notes, which enriched its lexical and phraseological heritage through references to the readings of Tuscan texts carried out during that period. Among the notes, for which Dante Isella proposed a classification by chronology and type, there are numerous references to phraseological units. Such structures become essential from the second manuscript onward, as they are capable of conveying a greater sense of orality and common use in writing. Therefore, the contribution aims to bring to light the research carried out by Manzoni at phraseological level through the traces left on the pages of the *Crusca Veronese*, in which it's possible to reconstruct the innovative literary canon that shaped the language of the second draft of the novel.

Percorsi correttòri dal *Fermo e Lucia* alla Quarantana: l'esempio dei proverbi glossati

Irene Rumine

1. Il proverbio nei postillati e negli scritti linguistici manzoniani

Gli scritti linguistici manzoniani, raccolti e pubblicati nell'Edizione Nazionale ed Europea delle Opere del Manzoni, raccolgono numerosi esempi di espressioni fraseologiche e proverbiali. Dalla *Seconda Introduzione* al *Fermo e Lucia* (1823) all'incompiuto trattato *Della lingua italiana* (1830-1859), l'Autore parla di «frasi», «locuzioni», «modi di dire» e altre simili nozioni con cui si riferisce ai modi espressivi, vivi nell'uso parlato e scritto¹. Ciò che non si riscontra negli scritti linguistici è, tut-

1 Si veda la *Seconda Introduzione* al *Fermo e Lucia*, dove il lessema «frasi» ricorre più volte nel binomio con «parole» (cfr. *SL I*, pp. 17-28, in particolare pp. 20-21). Subito dopo la stesura del *Fermo*, tra il 1823 e il 1824, nell'esigenza di definire il concetto di uso, Manzoni affronta lo studio della fraseologia negli incompiuti e frammentari *Modi di dire irregolari*, dove legittima modi o maniere di dire contrari alle prescrizioni grammaticali, ma giustificati dall'«Uso» o dal «Bisogno» (cfr. *SL I*, pp. 39-46, e *SL II*, pp. 39-69). Agli stessi anni risale anche il cd. *Libro «d'avanzo»*, il distrutto lavoro sulla lingua di cui rimangono quattro frammenti, il quarto dei quali documenta l'interesse manzoniano per la «fraseologia moderna», riscontrata negli innovativi autori toscani del Sei-Settecento, come il Salvini, il Redi, il Galilei (cfr. *SL I*, pp. 29-37). Sull'opportunità di spogliare i testi toscani antichi quali riserve di «vocaboli e modi», ma pur sempre da sottoporre alla regola dell'uso vivo ed effettivo, Manzoni insiste in due minute di una lettera all'abate Cesari, composte di ritorno dal viaggio in Toscana nel 1827 (cfr. *SL II*, pp. 47-67). Ancora, nel saggio sul *Sistema del Padre Cesari*, progettato intorno al 1830 ma anch'esso solo abbozzato, si fa menzione delle «locuzioni» (cfr. *ivi*, pp. 139-220, a p. 200). In due minute non concluse

tavia, una teorizzazione del *proverbio* e del suo trattamento linguistico che permetta di considerarlo un elemento autonomo dalla *locuzione*, e gli appunti ed elenchi lessicali manzoniani registrano spesso indistintamente gli uni e le altre. L'interesse di Manzoni per il proverbio, altrettanto significativo quanto lo studio delle locuzioni, muove dalla medesima istanza linguistica, fondata sulla regola dell'uso, di ricercare per i personaggi del romanzo un «parlato con larga escursione di registro»²: nelle quattro redazioni del romanzo sono presenti sia proverbi latini, pronunciati da ecclesiastici o personaggi di alto rango civile, sia proverbi diffusi nella lingua d'uso dei personaggi popolari. Già nel 1818, in una postilla a un passo della *Lezione sui proverbi toscani del Cecchi* di Luigi Fiacchi in cui si menziona il Cesarotti del *Saggio sulla filosofia delle lingue*, Manzoni riconduce i «modi proverbiali» all'arbitrio

di una lettera al Tommaseo, della fine del 1830, Manzoni ribadisce la tesi dell'uso effettivo e corrente come principale riferimento anche nelle compilazioni lessicografiche, tornando a parlare di «locuzioni», «espressioni», «modi di dire», «modi» (cfr. *SL I*, pp. 69-94). La sistemazione della propria riflessione sulla lingua è avviata con la stesura del trattato *Della lingua italiana*, intrapresa nel 1830 e che prosegue a intervalli per circa trent'anni, fino al 1859. Già nella *Prima redazione*, Manzoni affronta l'argomento delle «locuzioni parlate popolari», annotando alcuni esempi di «locuzioni popolari, familiari, plebee, se si vuole, o se pur si vuole, di *riboboli*» (cfr. *SL II*, pp. 250-55, a pp. 252 e 253). In un frammento del cap. II della *Quarta redazione* si legge, sempre a proposito delle locuzioni: «non vogliam certo prender l'impegno di definirle a parte a parte, quando ciò non occorre. / I francesi le chiamano molto propriamente *frasi fatte* (a); presso di noi non trovo ch'elle abbiano una propria special denominazione. Talvolta sono dette *idiotismi*; ma non tutte da tutti: chè un tal termine è inteso in troppo diversi sensi; e del resto ha per tutti un senso più esteso, poichè comprende anche vocaboli e forme grammaticali. Un senso più generico hanno ugualmente i termini *locuzioni*, *dizioni*, *modi di dire*, *proprietà*. Tuttavia noi le chiameremo, per brevità, *locuzioni* senza più, sottintendendo: composte di più vocaboli, e ricevute in una lingua qualunque» (*Frammento 10*, §§ 3-7, in *SL II*, pp. 628-29). Cfr. anche la *Quinta redazione*, iniziata nel 1843, dove l'Autore si sofferma distesamente, con spiegazioni ed esempi, su quelle che definisce «locuzioni composte di più vocaboli, e che hanno una loro forma determinata, e un senso ugualmente determinato, il quale però non resulterebbe naturalmente dal concorso de' vocaboli medesimi» (cap. II, § 128, in *SL I*, p. 444).

2 GIOVANNI NENCIONI, *L'incessante itinerario di una 'concezione democratica' della lingua*, Premessa a *SL I*, pp. XIII-XXVII, a p. XVIII.

dell'uso, rilevando che la loro caratteristica è quella di avere una «significazione [...] speciale, di rappresentare insomma un complesso d'idee», e che tale «significazione» non deriva che dall'uso³. L'impiego del proverbio in Manzoni⁴, inoltre, risente probabilmente del gusto degli scrittori del Romanticismo per le ricerche folkloriche, etnografiche e dialettologiche⁵ e anticipa l'attitudine, poi consolidatasi in Italia nella seconda metà dell'Ottocento, a studiare le tradizioni popolari anche attraverso la ricezione letteraria e lessicografica dei proverbi⁶ e la loro catalogazione in veri e propri repertori⁷. È significativo che, pubbli-

- 3 LUIGI FIACCHI, *Dei proverbi toscani. Lezione di Luigi Fiacchi detta nell'Accademia della Crusca il dì 30 novembre 1813. Con la dichiarazione de' proverbi di Gio. Maria Cecchi testo di lingua citato dagli accademici della Crusca*, Firenze, dalla stamperia Piatti, 1818, p. 7. Sul postillato cfr. portale *Manzoni Online*, scheda di Mariarosa Bricchi (www.alessandromanzoni.org/biblioteca/eseemplari/2832/postille).
- 4 In quanto mezzo comunicativo efficace, il proverbio è adoperato da Manzoni non solo nella scrittura narrativa, ma anche nell'esposizione teorica delle proprie tesi linguistiche, come, ad es., quando confuta il *Sistema del Padre Cesari*: «Sarebbe certo cosa più corta il rimettere il lettore al sistema stesso; ma dice il proverbio: chi vuol far vada, e chi non vuol, mandi» (*Frammento 5*, in *SL II*, p. 214). Ma si vedano anche le lettere di Manzoni, come quella a Ruggero Bonghi del 1868, intorno al *De vulgari eloquio di Dante*, che inizia con un modo proverbiale, ricorrente anche altrove: «Dico a socera perché nora intenda» (cfr. *Manzoni Online*, www.alessandromanzoni.org/manoscritti/717).
- 5 Può aver funzionato da stimolo alla raccolta manzoniana dei proverbi la stessa passione per la poesia popolare dell'amico Fauriel, che giunto in Italia negli anni 1823-1825, si dedica alla composizione del «Discours préliminaire» degli *Chants populaires de la Grèce moderne* (cfr. RAIMONDI 2000, pp. XIII-XLI). A proposito del carattere “popolare” della poesia romantica e nei *Promessi sposi*, cfr. CIANFAGLIONI 2006, pp. 21-40.
- 6 Nella tradizione letteraria italiana, i *Canti popolari toscani corsi, illirici, greci* (1841-1842) di Niccolò Tommaseo raccolgono molti proverbi tratti dall'uso vivo del popolo «nel suo aspetto più genuino, vero e autentico» (cfr. TOMMASEO/BALSAMO CRIVELLI 1920, II, p. 132). Sul legame tra i *Canti popolari greci* e il modello degli *Chants de Fauriel*, si veda TOMMASEO/MAIOLINI 2017, *Introduzione* (cfr. anche MARTINELLI 2020). Buona parte dei proverbi introdotti nei *Canti popolari* sono registrati dal lessicografo dalmata nel *Dizionario dei Sinonimi* e nel *Dizionario della lingua italiana*.
- 7 Si consideri la *Raccolta di proverbi toscani* di Giuseppe Giusti, che nella *Prefazione* avverte: «per proverbio intendo quel dettato, che chiude una sentenza, un precetto,

cata nel 1823 la terza edizione del *Dictionnaire des proverbes français* di Pierre de La Mésangère, l'Autore se ne procuri di lì a poco una copia, che postillerà nel 1823-24, arrivando fino alla lettera M e inserendo in alcune postille dei rimandi al toscano della *Crusca veronese* («Cr.») e al milanese («mil.»)⁸. A documentare la ricerca paremiologica manzoniana in quegli anni è, infatti, soprattutto la postillatura del vocabolario del Cesari⁹ e di quello del Cherubini, alla quale risultano parzialmente complementari e integrativi gli spogli dalla stessa *Crusca* e dagli scrittori comici toscani, nonché le postille a Plauto e Terenzio, durante la riscrittura del *Fermo*¹⁰.

-
- un avvertimento qualunque, ed escludo da questa raccolta certi altri detti [...] che si dicono proverbi e che i raccoglitori registrano per proverbi, ma mi pare a tutto rigore che debbano chiamarsi o modi di dire o modi proverbiali» (per i manoscritti A e B della raccolta e le successive edizioni del 1853 e del 1871, curate da Gino Capponi, cfr. la banca dati *Proverbi italiani*, www.proverbi-italiani.org).
- 8 Sul volume postillato del Mésangère si vedano gli studi di EVANGELISTI 2001-2003, che ha, peraltro, rilevato una coincidenza fra alcuni dei proverbi del dizionario francese sui quali ricade l'attenzione di Manzoni e quelli registrati in SEBASTIANO PAULI, *Modi di dire ricercati nella loro origine*, Venezia, Simone Occhi, 1740, di cui l'Autore era in possesso. Cfr. poi CIANFAGLIONI 2006, pp. 43-55; e, da ultimo, GHIRARDI 2018, secondo cui «è difficile stabilire con esattezza come Manzoni, per quanto assidui fossero i suoi contatti con il mondo culturale parigino, sia entrato in possesso della terza ed. (datata 1823) dell'opera» (ivi, p. 209). La studiosa ha curato la trascrizione delle 106 postille al Mésangère, in *Manzoni Online* (www.alessandromanzoni.org/biblioteca/esemplari/10415).
- 9 L'esemplare postillato da Manzoni è il *Vocabolario degli Accademici della Crusca oltre le giunte fatteci finora, cresciuto d'assai migliaia di voci e modi de' classici, le piu trovate da veronesi*, Verona, dalla stamperia di Dionigi Ramanzini, 1806, voll. I-VII [d'ora in avanti: *Cr. ver.*], conservato presso la Biblioteca Nazionale Braidense, con segn. MANZ. 16. 0205-0211.
- 10 Un elenco di espressioni radunato negli spogli dalla *Crusca* reca il titolo di *Modi proverbiali* (cfr. *SL I*, pp. 11-12, § 5). Tra le centinaia di postille manzoniane al commediografo latino, una a un passo della *Casina* II III («Bibe, es, disperde rem») riporta la traduzione italiana del proverbio latino: «Mangia, bevi, butta via» (cfr. BASSI 1932, p. 239). Una postilla alla *Mostellaria* registra il proverbio *a buon intenditor poche parole*, che traduce il latino «Dictum sapientis sat est» (cfr. BASSI 1932, p. 255), e un'altra a un passo del *Trinummus* («Non aetate, verum ingenio adipiscitur sapientia») riporta il proverbio *chi non nasce savio non si farà mai* (cfr. ivi, p. 264). Per la datazione delle postille al *corpus* plautino, nell'ed. in tre volumi del 1788, si veda ora

Dopo l'edizione Ventisettana, e in funzione della sua revisione, qualche proverbio è annotato da Manzoni negli appunti lessicali milanesi del 1827 (es. «Uomo avvertito è mezzo salvo»)¹¹, nelle richieste di avallo al Cioni e al Niccolini delle scelte linguistiche della Ventisettana (ad es., al quesito sulla forma toscana di «*Dura più un carro rotto che un nuovo*» i due rispondono: «“Basta più una conca fessa che una salda”, Crusca»)¹². Il Libri, incaricato nel 1830 di verificare la corrispondenza toscana anche di alcuni proverbi milanesi, assenti nel romanzo, conferma, ad esempio, l'uso toscano di *chi sprezza ama* e di *aspetta bue che l'erba cresce*¹³, e lo esclude per *can ringhioso* e *non forzoso, guai alla sua pelle*¹⁴.

Nel 1835, gli spogli del Grossi per la *Risposta* alle critiche di Michele Ponza, che preparano la stesura del *Sentir messa*, registrano alcuni proverbi attestati negli scrittori della tradizione toscana, soprattutto comica: dal Lippi del *Malmantile* con le note del Minucci (es. «*Chi non ha*

MARTINELLI 2017, pp. 54-55 (cfr. anche DANZI 2001, pp. 204-10). La stessa ha curato la trascrizione delle postille in *Manzoni Online*, www.alessandromanzoni.org/biblioteca/eseemplari/10427 (si vedano anche le altre due ed. delle commedie plautine postillate da Manzoni, quella parigina del 1830-32 e quella del *Théâtre de Plaute*, con traduzione francese a fronte, entrambe riprodotte nel portale). Un proverbio si legge anche in una postilla a TERENCE, *Phormio* I I («Ut semper aliquid addant divitioribus!»), nell'ed. volgarizzata in fiorentino e annotata dal Cesari: al margine della nota «*colmar il sacco a' ricchi*. Questo bel proverbio abbiám noi Lombardi, per dir, che la roba corre sempre dietro a' ricchi», Manzoni postilla «e i toscani: la roba va alla roba. Monos.[ini] p. 325» (cfr. BASSI 1932, p. 273). L'esemplare postillato da Manzoni, *Le sei commedie di Terenzio recate in volgar fiorentino da Antonio Cesari con note postoci innanzi un ragionamento cioè Difesa dello stil comico fiorentino*, è riprodotto in *Manzoni Online* (www.alessandromanzoni.org/biblioteca/eseemplari/3566).

11 Cfr. *SL II*, p. 77, n. 1. Il proverbio in questione entra nel romanzo dalla Seconda minuta, ma è lasciato in sospenso («Uomo avvertito... ella c'intende», in *SP I I* 32), facendo pensare che si tratti di un proverbio comune e comprensibile a tutti. In *FL I I* 32-33, si leggeva al suo posto: «era nostro dovere d'avvisarla e l'abbiamo fatto».

12 Cfr. *ivi*, p. 86, n. 112.

13 Cfr. *ivi*, p. 111, n. 4, e p. 135, n. 382.

14 Cfr. *ivi*, p. 125, n. 205. Altri proverbi milanesi sottoposti al Libri, nelle inchieste del 1830, si leggono *ivi*, p. 133, n. 178, e p. 135, n. 387.

cervello / Abbia gambe)¹⁵, al Cecchi della *Dote* (es. «Chi vuol far vadia; e chi non vuol far mandi») ¹⁶, della *Moglie* (es. «Ogni troppo è troppo») ¹⁷, dei *Dissimili* (es. «il diavolo non è brutto come e' si dipinge») ¹⁸, al Buonarroto della *Fiera* e della *Tancia* annotate dal Salvini (es. «*a caval donato non gli guardare in bocca*: proverbio») ¹⁹.

Per la verifica dell'uso fiorentino della Quarantana, alcuni proverbi sono oggetto dei quesiti manzoniani, datati 1839-1842, a Giovanna Feroci Luti (es. «Dimmi *con chi tratti* o *chi tratti* o *chi pratici* e ti dirò chi sei») ²⁰, mentre un appunto successivo di Emilia Luti reca scritto: «Il proverbio toscano corrispondente al milanese *tegnì de cunt i guggià* non lo so, scriverò a Firenze» ²¹.

Pubblicata l'edizione definitiva del romanzo, la ricerca paremiologica diventa funzionale al progetto di realizzazione del *Vocabolario dell'uso fiorentino*, del 1856. Appunti lessicali e fraseologici registrano, ora con una certa sistematicità, anche proverbi, alcuni dei quali già postillati da Manzoni nella *Crusca veronese* e presenti nel romanzo. Nel *Saggio di vocabolario* in collaborazione col Capponi si legge, ad esempio, *can che abbaia non morde*, già usato nel *Fermo* ²², e il modo proverbiale *dico a socera perché nora*

¹⁵ *SL II*, p. 395, n. 202.

¹⁶ Ivi, p. 410, n. 418. Il proverbio è impiegato da Manzoni nelle richiamate bozze del *Sistema del Padre Cesari*.

¹⁷ Ivi, p. 413, n. 499.

¹⁸ Ivi, p. 416, n. 561. Esattamente in tale forma il proverbio è impiegato nel romanzo da Agnese, in *Q III 10* (il proverbio è già in *FL I III 14*, nella dicitura: «il diavolo non è mai brutto come si dipinge»).

¹⁹ Ivi, p. 445, n. 1101. Varie altre espressioni attestate nella *Fiera* e nella *Tancia* sono qualificate dal Salvini come proverbi o modi proverbiali: cfr., ad esempio, ivi, p. 438, nn. 1012 e 1023; p. 442, n. 1073; p. 443, n. 1079; p. 444, n. 1089.

²⁰ Cfr. *SL II*, p. 790, 17, n. 8.

²¹ Ivi, p. 813, 7. Cfr. CHERUBINI 1839-1856, II (1840), s. v. *guggiàda*, che registra *tegnì a man i guggiad e trà-via i remissegj*. Si vedano anche, tra i quesiti alle Luti, le note ai proverbi: «Il buon vino non vuol frasche» e «Nelle scatole foderate di piombo, il tabacco non si risecca», da considerare postille alle voci, rispettivamente, *frasca* e *fodràa* di CHERUBINI 1839-1856, I (1839) (cfr. *SL II*, p. 785, 10).

²² Cfr. *SL II*, p. 952, n. 49. Il proverbio è introdotto in *FL I I 74*: «cane che abbaia non morde». In *SP I I 76* è modificato con la dicitura «Eh! se questi cani dovessero mor-

intenda, assente nel romanzo, ma che ritorna in una postilla al *Supplemento a' vocabolarj* del Gherardini e, con l'ordine invertito dei componenti, in una postilla alla seconda edizione del Cherubini²³. Numerosi proverbi sono annotati dal Matteucci nelle *Maniere di dire fiorentine* (es. *chi vuol far l'altrui mestiere fa la zuppa nel paniere; fammi fattore un anno, se son povero mio danno; il pazzo fa la festa e il savio se la gode; finché uno ha i denti in bocca, non si sa ciò che gli tocca*)²⁴, e, tra quelli impiegati nel romanzo, *il lupo lascia il pelo, ma il vizio, mai*²⁵ e *una mano lava l'altra e tutte e due lavano il viso*²⁶.

Qualche altro proverbio è oggetto delle inchieste viareggine (es. *dove vai tu? Le son cipolle e quanto il cento? Vo a Firenze*)²⁷, oppure è raccolto nel materiale per la «revista» della seconda edizione del Cherubini (es. *vuna addoss e l'altra al foss*)²⁸, nonché in alcuni appunti sparsi di Manzoni, di datazione e luogo incerti (es. «La farina del diavol la va tutta in Crusca». 'Quel che va di ruffa in raffa, se ne va di buffa in baffa')²⁹.

dere tutte le volte che abbajano!», poi confermata, con qualche modifica, in V I 1 29 («e guai se questi cani dovessero mordere tutte le volte che abbaiano!»), e in Q I 76. Cfr. anche la postilla al *Mésangère*, s. v. *chien*: «Can che abbaia poco morde».

23 Cfr. *SL II*, p. 989, n. 809. La postilla al *Supplemento a' vocabolarj italiani proposto da Giovanni Gherardini*, Milano, dalla stamperia di Gius. Bernardoni di Gio., 1852-1857, voll. I-VI, v (1857), s. v. *suocera*, reca scritto: «Dico a socera perché nora intenda. Modo proverbiale» (cfr. *Manzoni Online*, scheda di M. Bricchi, www.alessandromanzoni.org/biblioteca/eseemplari/7467). Per la postilla al *Vocabolario milanese*, cfr. CHERUBINI 1839-1856, III (1841), s. v. *noèra*: «Dico a nora perché socera intenda. fior.», cfr. *Manzoni Online*, scheda di Jacopo Ferrari, (www.alessandromanzoni.org/biblioteca/eseemplari/10451).

24 Quelli citati sono siglati dal Matteucci come «Proverbio» (cfr. *SL II*, p. 997, s. v. *fare v.*, nn. 44, 56 e 100, e p. 1000, s. v. *sapere v.*, n. 8). Ma molti altri proverbi sono raccolti nelle *Maniere di dire fiorentine* (cfr. *ivi*, s. v. *stare in senso proprio*, n. 13; *Locuzioni diverse*, nn. 9, 10, 18, 35, 44, 59, 61, 68, 69, 74, 75, 77, 78, 80, 82, 85, 95, 98, 102, 103).

25 Cfr. *SL II*, p. 1004, *Locuzioni diverse*, n. 60.

26 Cfr. *ivi*, n. 67.

27 Cfr. *SL II*, p. 959, n. 2.

28 Cfr. *ivi*, p. 1011, § 10, n. 108.

29 Si tratta della forma milanese di un noto proverbio che Manzoni compara con le dizioni veneziana, bolognese e toscana (cfr. *ivi*, pp. 1047-48, *Appunti sparsi*, § 2). Il proverbio *quel che va di ruffa in raffa, se ne va di buffa in baffa* è già annotato da Manzoni in una postilla al *Mésangère*, al lemma *diable*, e nel romanzo è richiamato dalla locuzio-

2. Il proverbio nel romanzo di Manzoni

L'insistito impiego del proverbio nei *Promessi sposi* non si spiega, tuttavia, soltanto con la volontà di conformare la lingua del romanzo all'uso vivo, restituendo al discorso dei personaggi o alla voce del narratore uno degli elementi più espressivi del parlato. Il proverbio serve a dare ulteriore realismo alla storia narrata, soprattutto quando è inserito nei dialoghi o nei monologhi. Che sia pronunciato dal narratore o dai personaggi, sia noto o frutto dell'inventiva di Manzoni, il proverbio costituisce, inoltre, un «momento di pausa, di riflessione e di distacco dell'autore dalla sua opera»³⁰, divenendo un espediente privilegiato dell'ironia manzoniana. Ciò è tanto più evidente quando è dichiarato attraverso delle glosse metalinguistiche o esplicative, incentrate il più delle volte, ma non sempre, su un *verbum dicendi* e introdotte da formule del tipo «come dice il proverbio», «ciò che l'uom dice in proverbio», «lo dice anche il proverbio», «si dice in proverbio», o «secondo quel proverbio», «è un proverbio», e simili³¹. A partire da tali proverbi dichiarati e corredati di glossa, si ripercorrono qui i passaggi correttori che sugli stessi intervengono, dal *Fermo* alla Quarantana.

2.1 I proverbi glossati nel *Fermo*

Il «composto indigesto di frasi un po' lombarde, un po' toscane, un po' francesi, un po' anche latine; di frasi che non appartengono a nessu-

ne idiomatica *fare a ruffa ruffa*, 'fare a chi ruba di più' (cfr. *SP II XIV 45*: «fanno a ruffa ruffa», poi in *V* e in *Q*). Negli *Appunti sparsi* si vedano, inoltre, ancora nel raffronto tra le forme milanesi, veneziane, bolognesi e toscane, i proverbi «Chi no risega no rosega», «Morto un papa se ghe ne fa un altro», «Chi va pian va san» (in *SL II*, p. 1048, § 4).

30 Cfr. CIANFAGLIONI 2006, pp. 133-34.

31 Ci sono, tuttavia, anche dei casi di proverbi non dichiarati ma accompagnati da formule metalinguistiche, del tipo «come si dice», «suol dirsi» (sull'uso di tali formule nel romanzo si rinvia a ANTONELLI 2008, che ha notato «una precisa corrispondenza fra tipologia dell'osservazione metalinguistica e formule usate per introdurla»: *ivi*, p. 142).

na di queste categorie, ma sono cavate per analogia e per estensione o dall'una o dall'altra di esse»³², si riflette, nel *Fermo*, anche nella scelta dei proverbi dichiarati (otto, più una «frase proverbiale lombarda»). Due, infatti, sono latini: *vox populi vox Dei* (che ritorna alla fine del romanzo, tradotto in italiano, in perifrasi e in inciso)³³, e *si Romae fueris, romano vivito more*³⁴. Gli altri, di antica origine e d'uso comune, sono in genere prelevati dal dialetto milanese o lombardo: lo è, dichiaratamente, la frase proverbiale *meglio perderlo che trovarlo*³⁵, ma anche il proverbio *una mano lava l'altra*, in riferimento al quale il personaggio del birro, che lo adopera per estorcere informazioni a Fermo, dice: «è un proverbio che l'avrete anche nel vostro paese»³⁶. Quest'ultimo ha, infatti, il suo equivalente milanese in *ona mano lava l'oltra e tutt'e dó laven el mostacc*³⁷, così come *ambasciator non porta pena* e *voce del popolo, voce di Dio* hanno corrispondenza nelle rispettive varianti dialettali registrate nel Che-

32 Dalla seconda 'Introduzione' al '*Fermo e Lucia*', § 26 (in *SL I*, pp. 23-24).

33 *FL I* v 53, e *IV III* 22. Il detto latino trova una prima attestazione nella tradizione medievale, in Alcuino di York (cfr. TOSI 1991, p. 3, num. 1). Su tale proverbio e su *ambasciator non porta pena*, cfr. RUMINE 2022.

34 *FL II VIII* 13 (prima, al posto del passo «ricordandosi di quel proverbio [etc.]» si leggeva: «risoluto di stare alla legge», per cui cfr. *FL II VIII* 14, *Apparato*, p. 248). Il proverbio latino è citazione virgiliana da *Eneide V* v 231: «si Romae fueris romano vivito more, si fueris alibi, vivito sicut ibi».

35 L'espressione ha corrispondenza nel milanese *l'è mei perdel che trovall* (cfr. CHERUBINI 1839-1856, *IV* [1843], s. v. *trovà*). Tale frase accentua la distanza ironica tra Autore e personaggio: con essa, infatti, il curato, al cospetto di Federico Borromeo, commenta tra sé l'evangelico *perierat et inventus est*, pronunciato dall'arcivescovo per riferirsi all'innominato convertito.

36 *FL III VII* 52. Il proverbio si legge già ivi, *II VII* 71, senza glossa metalinguistica: «Bene tu avrai da questi informazioni, e ajuti al caso. Una mano lava l'altra, e le due il viso. Coraggio, e prudenza: comprare e non vendere; andare e tornare» (ad adoperarlo è don Rodrigo nel dialogo con il Griso).

37 L'equivalente milanese di questo proverbio di origine classica (lat. *manus manum lavat*) è registrato in CHERUBINI 1839-1856, *III* (1841), s. v. *man* (cfr. anche *V* [1856], s. v. *man*, dove si legge *ona man con l'altra se lava la faccia*).

rubini³⁸, mentre *dimmi con chi tratti e ti dirò chi sei* si legge nella colonna delle forme milanesi, tra i più tardi quesiti di Manzoni alla Feroci Luti, sopra ricordati. Sebbene lo spoglio dagli autori e dalla *Crusca*, collocabile nell'intensissimo periodo tra il 1823 e il 1827, sia nella fase di redazione del *Fermo* ancora da avviare³⁹, alcuni di tali proverbi trovano una più o meno esatta corrispondenza nelle forme attestate nella tradizione fiorentina: dalle commedie del Cecchi⁴⁰, del Firenzuola⁴¹ e del Lasca⁴² alle *Prose toscane* del Salvini⁴³. I restanti proverbi, *chi pensa*

- 38** Il proverbio *ambasciator non porta pena*, che richiama l'antico adagio *legatus non ceditur, neque violatur*, ha esatta corrispondenza nel milanese *imbassador no porta penna* registrato in CHERUBINI 1839-1856, III (1841), s. v. *penna* (ma già in CHERUBINI 1814, s. v. *mess*, si legge, come traduce della forma italiana, *mess mandaa no porta penna*). Per la diffusione milanese del proverbio *vox populi vox Dei*, cfr. CHERUBINI 1814, s. v. *popel*: «*Vox popel o popul, vox Dei*» (lo stesso si legge in CHERUBINI 1839-1856, III [1841], s. v. *popol o popel*).
- 39** Qualche traccia della lingua e dello stile comico fiorentino si intravede negli ultimi capitoli della Prima minuta e nella seconda *Introduzione* al romanzo, ma la presenza della tradizione letteraria toscana è preponderante a partire dalla Seconda minuta, redatta dal 1823 al 1825. Si veda, a proposito, DANZI 2001, pp. 195-243, a p. 209.
- 40** Cfr. GIOVAN MARIA CECCHI, *La dote* IV III: «sarebbe meglio 'l perderlo che smarrirlo»; e *Lo sviato* I II: «L'una man lava l'altra. / [C.] E tutte e due Lavano poi il mostaccio». *Lo sviato* non è presente tra le venti commedie raccolte nel *Teatro comico fiorentino delle più rare commedie citate da' sig. Accademici della Crusca diviso in tomi sei*, il volume postillato da Manzoni, contenente oltre alle commedie del Cecchi, quelle del Lasca, dell'Ambra, del Salviati e del Buonarroti (cfr. *Manzoni Online*, scheda di S. Ghirardi, www.alessandromanzoni.org/biblioteca/eseemplari/6229).
- 41** Cfr. AGNOLO FIRENZUOLA, *La Trinuzia* II VI: «Voce di popol, voce del signore». Sui volumi delle *Opere di messer Agnolo Firenzuola fiorentino* postillati da Manzoni negli anni 1823-1827, si veda *Manzoni Online*, scheda di S. Ghirardi (www.alessandromanzoni.org/biblioteca/eseemplari/7894).
- 42** Cfr. ANTON FRANCESCO GRAZZINI, detto Il Lasca, *La spiritata* IV I: «l'una mano lava l'altra; e le due il viso»; e *La gelosia, Prologo agli uomini*: «Voce di popolo, voce di Dio».
- 43** Cfr. ANTON MARIA SALVINI, *Prose toscane recitate nell'Accademia della Crusca* II XVIII: «l'ambasciatore, come si dice in proverbio, non porta pena» (una precedente attestazione cinquecentesca del proverbio è nel poema dell'Ariosto [ed. 1532], XXIV 109: «Per dir, ch'ambasciator pena non porta»). Il proverbio si ritrova anche nella trattatistica storica seicentesca, leggendosi nei *Consigli cavallereschi* (*Consiglio Secondo*,

male pensa una sola volta e a questo mondo, niente per niente, sono probabilmente ricavati per analogia o per estensione dal milanese, in assenza di una tradizione, ma ritenuti comprensibili a ogni parlante italiano⁴⁴: *a questo mondo, niente per niente*, come dice Tonio, «è un proverbio che lo sa anche Gervaso, lo sanno anche quelli che non sanno dire il *Credo*»⁴⁵. Non solo rilevanti per la veste linguistica, le glosse ai proverbi del *Fermo* si caratterizzano anche per il contenuto, perché contengono osservazioni sulla natura del proverbio, che il narratore riferisce non senza ironia: così, «i proverbi sono voce di Dio secondo quell'altro proverbio = vox populi vox Dei»⁴⁶ e «[d]immi con chi tratti e ti dirò chi sei è un proverbio; e | come tutti i proverbj, non solo è infallibile, ma ha anche la facoltà di rendere infallibile l'applicazione che ne fa chi lo cita»⁴⁷.

2.2 I proverbi glossati nella Seconda minuta

Radicalmente mutato è il quadro dei proverbi dichiarati nel «rifacimento» costituito dalla Seconda minuta. Sono, infatti, eliminati il proverbio della latinità classica *si Romae fueris, romano vivito more*⁴⁸, i due proverbi probabilmente diffusi nel milanese ma non attestati nella tradizione toscana, *a questo mondo, niente per niente* e *chi pensa male pensa una sola volta*, e la frase proverbiale «lombarda» *meglio perderlo che trovarlo*. Sono invece confermati, ma senza glossa, i proverbi *una mano*

7) del milanese Francesco Birago, accertata fonte del romanzo (cfr. NUNNARI 2013, pp. 259-60, nota 23).

44 Nella seconda *Introduzione* al *Fermo* Manzoni ammette di aver adoperato alcune frasi lombarde, quando fossero intelligibili a ogni parlante italiano e a lui non fosse nota la variante comune (cfr. §§ 24-25, in *SL I*, p. 23).

45 *FL I* vi 75.

46 *FL I* v 53. E si veda anche ivi, IV III 22: «facendo eco alla voce del popolo, la quale in questo caso – se è lecito fare una eccezione ad un proverbio – non era certamente voce di Dio».

47 *FL III* ix 41.

48 Il proverbio latino *vox populi, vox Dei*, invece, è reintrodotta nel capitolo finale, ma senza glossa (cfr. *SP III* xxxviii 41), e confermato in *V* e in *Q*.

lava l'altra, ora completato con la dicitura «e le due il viso»⁴⁹, *dimmi con chi tratti, e ti dirò chi sei*⁵⁰, e *voce del popolo voce di Dio*, quest'ultimo riformulato con un'interrogativa parentetica⁵¹. Un'interessante modifica riguarda la glossa che introduce il proverbio *ambasciator non porta pena*: «e senza andar tanto a cercare, lo dice anche il proverbio: ambasciator non porta pena. E i proverbii, signor conte, sono la sapienza del genere umano»⁵². Sono aggiunti cinque nuovi proverbi glossati: *chi è in difetto è in sospetto, fare della necessità virtù, l'abito non fa il monaco*, al quale si accompagna *il lupo lascia il pelo, ma non il vizio*, e il proverbio, coniato da Manzoni e attribuito all'inventiva dell'anonimo seicentista, «chi vuole aver molti in aiuto, faccia di non averne bisogno»⁵³. È introdotto altresì un «modo proverbiale» che riformula il proverbio *basta più una conca fessa che una salda* («sono, per servirci di un modo proverbiale, di quelle conche fesse che bastano un pezzo»)⁵⁴. Infine, nei fogli della Seconda minuta non confluiti nel testo si leggono due altre espressioni corredate di glossa: *non si può cantare e portar la croce* e *menare la danza*⁵⁵.

I proverbi introdotti *ex novo* in questa fase toscano-libresca di scrittura del romanzo, sono verificati sui vocabolari e sugli spogli dagli scrittori toscani. Il più antico *fare della necessità virtù* è registrato nel

49 SP II XIV 17. La scelta è confermata in V, mentre in Q la forma è nuovamente modificata, secondo l'uso fiorentino, in «una mano lava l'altra, e tutt'e due il viso» (XIV 17).

50 SP III XXV 29-30. Nella stessa forma il proverbio passa in V, mentre in Q è modificato, su suggerimento della Feroci Luti, in «Dimmi chi pratici, e ti dirò chi sei» (XXV 30).

51 Cfr. SP III XXXI 32: «facendo eco alla voce del popolo, (era essa, anche in questo caso, voce di Dio?)». La forma è identica in V e, con l'eliminazione del pronome, passa in Q XXXI 31.

52 SP I v 35.

53 Ivi, III XXV 39.

54 Ivi, II XIX 31.

55 Il proverbio *non si può cantare e portar la croce* è introdotto in V II XVI 47, ma senza glossa metalinguistica («non si può mica cantare e portar la croce»). L'espressione *menare la danza* (cfr. SP II XIX 21: «mena la danza, come si dice in proverbio»), invece, non entra nel romanzo.

Flos italicae linguae del Monosini⁵⁶ e nel *Lexicon* del Forcellini, postillato, ma in anni più tardi, da Manzoni⁵⁷, ma manca nel Cherubini, e anche *l'abito non fa il monaco*, di origine classica, è presente nel Monosini⁵⁸ e trova corrispondenza nel milanese *l'è minga el vestii che cunta*⁵⁹. Altri tre proverbi, di cui il Cherubini registra la variante milanese, sono attestati nella tradizione toscana: nel Cecchi e nel Fagioli si leggono, rispettivamente, *non si può cantare e portar la croce*⁶⁰ e *il lupo lascia il pelo ma non il vizio*⁶¹, mentre la *Crusca veronese* registra, ma senza esempi, *basta più una conca fessa che una salda*⁶². Solo di un proverbio Manzoni circoscrive

- 56** Il proverbio traduce il latino *facere de necessitate virtutem* ed è registrato in *Cr. ver.*, s. v. *necessità, necessitate, e necessitate*, § I, e s. v. *virtù, virtude, e virtute*, § IV.
- 57** Cfr. FORCELLINI 1827-1831, III, s. v. *necessitas*. Le postille manzoniane al Forcellini sono state pubblicate da MARTINELLI 1994, secondo la quale «[t]erminus post quem è l'uscita a dispense del *Lexicon* dal 1827 al 1830» (ivi, p. 45) (cfr. DANZI 2001, pp. 204-10). L'esemplare postillato da Manzoni, nell'ed. in quattro volumi del 1827-1831, è riprodotto in *Manzoni Online* (www.alessandromanzoni.org/biblioteca/esemplari/10502 e www.alessandromanzoni.org/biblioteca/esemplari/10501).
- 58** Il proverbio, che equivale al latino *barba non facit philosophum*, è registrato in *Cr. ver.*, s. v. *abito*, § I, e s. v. *monaco*, § I.
- 59** Il proverbio è registrato in CHERUBINI 1839-1856, IV (1843), s. v. *vestii*, con l'equivalente italiano «*La vesta non fa il monaco*».
- 60** Il proverbio (corrispondente al lat. *simul flare, atque sorbere difficile est*) è attestato in CECCHI, *La dote* (ed. 1585), IV II: «e non si può portar la croce / E cantare» (il proverbio non compare invece nell'ed. del *Teatro comico fiorentino*). Ha corrispondenza milanese in *se pò minga cantà e portà la cros* (cfr. CHERUBINI 1839-1856, II [1839], s. v. *cantà*; ma vedi anche s. v. *cròs*).
- 61** Il proverbio (dal lat. *lupus pilum mutat, non mentem*) ha come corrispondente milanese *el lóff el perd el pèl ma minga el vizzi* (cfr. CHERUBINI 1839-1856, II [1840], s. v. *lóff*). Nella forma di SP (col verbo *lascia*) si legge in FAGIUOLI, *Lavaro punito* I XI: «il lupo lascia il pelo, ma il vizio mai», e *Rime piacevoli* II XV: «Contro tai mascherati animalacci / Che il pelo lascian sì, ma non il vizio» (ma si vedano anche CECCHI, *Gli incantesimi* I III: «benché il lupo muti il pelo, e non muta natura», e *Le cedole* I II: «Il lupo muta il pelo / Ma non il vizio»). È da precisare però che *Le cedole* del Cecchi non figurano nella biblioteca di Manzoni.
- 62** Cfr. *Cr. ver.*, s. v. *conca*, § VII. CHERUBINI 1814, s. v. *carr*, registra come traduce milanese *tirà là pussee on carr rott che on carr nœuf*, la forma che si legge nei già richiamati appunti milanesi del 1827. Nella forma toscana registrata nella *Crusca*, il pro-

con la glossa la diffusione locale: «Chi è in difetto è in sospetto, dice il proverbio milanese», pur trattandosi di un proverbio comune⁶³.

2.3 I proverbi glossati nella Ventisetтана

Meno sostanziose sono le correzioni ai proverbi e alle rispettive glosse nel passaggio dalla Seconda minuta alla prima edizione del romanzo, le quali intervengono sul proverbio *il lupo muta il pelo, ma non il vizio*, con la sostituzione dell'elemento verbale (*lascia > muta*)⁶⁴, e su quello attribuito all'anonimo, che viene modificato dalla forma affermativa a quella interrogativa: «volete aver molti in aiuto? Fate di non averne

verbo è annotato da Manzoni anche in una postilla al *Mésangère*, s. v. *long-temps* («Basta più una conca fessa che una salda»), per cui cfr. GHIRARDI 2018, p. 226-27.

63 SP I VIII 58. CHERUBINI 1814, s. v. *pattell*, il proverbio italiano si legge come traduce del milanese *el sa come el sta in di pattij*, e in CHERUBINI 1839-1856, I (1839), s. v. *difett*, è registrato l'esatto corrispondente milanese *chi è in difett è in sospett*. Il proverbio non è però solo milanese. Sebbene non sia registrato nella *Crusca* del Cesari, né riscontrato da Manzoni nella tradizione, è, tuttavia, attestato nella raccolta cinquecentesca di proverbi toscani del Serdonati (cfr. *Proverbi italiani*, <https://proverbi-italiani.org/>). È registrato nel *Dizionario universale critico enciclopedico della lingua italiana* di Francesco d'Alberti di Villanova, Lucca, Maescandoli, 1797-1805, voll. I-VI, II (1797), s. v. *difetto* (cfr. LEI, XIX, 671, s. v. *defectus*), ed entra poi nel TOMMASEO-BELLINI 1861-1879, s. v. *difetto*, § VI (e s. v. *sospetto*, § XIV), e nella *V Crusca*, s. v. *difetto*, § XXXII, che non riporta esempi dalla letteratura (cfr. *Lessicografia della Crusca in rete*, www.new.lessicografia.it). La variante veneziana del proverbio si legge nel *Dizionario del dialetto veneziano* di Giuseppe Boerio, Venezia, Santini, 1829, s. v. *difeto* (*che xe in difeto xe in sospeto*), per cui cfr. ZOLLI 1989, p. 404. La glossa diatopica, nel romanzo, svolge in tal caso una funzione mimetica e rappresentativa della realtà popolare seicentesca (si ricordi quanto dichiara Manzoni nell'*Introduzione* alla Seconda minuta, ribadendo di aver introdotto anche nel rifacimento del romanzo qualche modo lombardo, «non solo intelligibile ad ogni lettore italiano, ma il più proprio ad esprimere italianamente il concetto che si voleva esprimere»: SP, *Introduzione*, § 13).

64 Cfr. VII XIX 16. La variante del proverbio col verbo *muta* è attestata, come si è visto, nel Cecchi.

bisogno»⁶⁵. È, inoltre, eliminato il «modo proverbiale», che tuttavia lascia traccia nel romanzo nella locuzione *essere una conca fessa*⁶⁶.

2.4 I proverbi glossati nella Quarantana

Nella Quarantana sono confermati tutti i proverbi glossati della Ventisettesima, ma con una nuova sostituzione del verbo nei proverbi *il lupo cambia il pelo, ma non il vizio (muta > cambia)*⁶⁷ e «volete aver molti in aiuto? cercate di non averne bisogno» (*fate > cercate*)⁶⁸, e un'ulteriore modifica del proverbio *far di necessità virtù (della > di)* e della glossa che lo introduce, dove all'uso tradizionale dell'impersonale «ciò che l'uom dice» si sostituisce il più diffuso «si dice»⁶⁹. È aggiunta, inoltre, una glossa diacronica a un proverbio già presente nella narrazione: «come dice un antico proverbio, del senno di poi ne son piene le fosse»⁷⁰, un proverbio in realtà non più antico di altri presenti nella narrazione, e della cui diffusione toscana Manzoni trova conferma, ancora una volta, nel Lasca⁷¹ e nel Cecchi⁷².

⁶⁵ Cfr. V III xxv 39.

⁶⁶ La locuzione *essere una conca fessa*, già presente in SP III xxxviii 9, è quindi confermata in V e poi in Q.

⁶⁷ Cfr. Q XIX 16. Nelle *Maniere di dire fiorentine* del Matteucci, sopra richiamate, è annotata la forma, ancora diversa, *il lupo perde il pelo, ma il vizio, mai*, con la corrispondenza del verbo al milanese *perd*.

⁶⁸ Cfr. Q xxv 39.

⁶⁹ Cfr. Q x 71-72.

⁷⁰ Q xxiv 80.

⁷¹ Cfr. LASCA, *La sibilla* V VIII: «Del senno di poi ne son piene le fosse» (si veda ANTONELLI 2008, p. 164, nota 83).

⁷² Cfr. CECCHI, *La dote* III I: «del senno di poi ne son piene le fosse». Sul proverbio in questione, cfr. FANFANI 2021.

3. Osservazioni conclusive

La ricognizione delle modifiche ai proverbi dichiarati e corredati di glossa traccia il percorso correttorio seguito da uno scrittore in cerca di una forma e di una lingua per il suo romanzo⁷³. Le correzioni ai proverbi dichiarati, nel passaggio dalla Prima alla Seconda minuta, si collocano all'interno di una modifica più ampia che interessa, non ancora singoli lessemi, ma porzioni di testo: sono eliminate o corrette, in particolare, le glosse del *Fermo* che contengono considerazioni sul valore del proverbio e sulla sua diffusione locale. Percorsa quindi la fase toscano-milanese che porta alla Ventisettona, dove le poche modifiche ai proverbi glossati investono l'aspetto lessicale e consistono sostanzialmente nella sostituzione di singoli elementi, Manzoni approda alla tappa definitiva della Quarantana, avendo rilevato in molti casi l'universalità dei proverbi adoperati e verificato, ove possibile, la loro conformità al fiorentino⁷⁴.

73 Le modifiche ai proverbi glossati sono schematizzate nella tabella in *Appendice*.

74 La glossa diatopica a *chi è in difetto è sospetto*, che sopravvive ancora in Q, non sembra potersi spiegare come un'eccezione alla soluzione linguistica del fiorentino, ma risponde più probabilmente a quelle esigenze di vero storico richieste dal romanzo. Il proverbio in questione, come si è detto, non è, infatti, solo milanese, ma «proprio perché dice una verità universale, a M.[anzoni] piace dichiararlo di diffusione geograficamente ristretta, quasi valesse solo per i milanesi» (POGGI SALANI in Q VIII 43, nota 65).

Appendice

Prima minuta	Seconda minuta	Ventisettana	Quarantana
secondo quel proverbio= ella m'insegna che i proverbi sono voce di Dio secondo quell'altro proverbio= vox populi vox Dei= quel proverbio che dice: ambasciator non porta pena (I v 53)	lo dice anche il proverbio: ambasciatore non porta pena: e i proverbi sono la sapienza del genere umano (I v 35)	lo dice anche il proverbio: ambasciator non porta pena. E i proverbi, signor conte, sono la sapienza del genere umano (I v 35)	lo dice anche il proverbio: ambasciator non porta pena. E, i proverbi, signor conte, sono la sapienza del genere umano (v 35)
a questo mondo, niente per niente: è un proverbio che lo sa anche Gervaso, lo sanno anche quelli che non sanno dire il Credo (I VI 75)	[senza glossa] vox populi, vox Dei (III XXXVIII 41)	[senza glossa] vox populi, vox Dei (III XXXVIII 41)	[senza glossa] vox populi, vox Dei (XXXVIII 41)
-	-	-	-
-	Chi è in difetto è in sospetto, dice il proverbio milanese (I VIII 58)	Chi è in difetto è in sospetto, dice il proverbio milanese (I VIII 43)	Chi è in difetto è in sospetto, dice il proverbio milanese (VIII 43)
-	fare realmente e in effetto, ciò che l'uomo dice in proverbio, della necessità virtù (I X 71-72)	fare realmente e in effetto, ciò che l'uomo dice in proverbio, della necessità virtù (I X 71-72)	far realmente e in effetto, ciò che si dice in proverbio, di necessità virtù (X 71-72)
quell'indegno proverbio che dice, chi pensa male pensa una volta sola (II IV 39)	-	-	-

Prima minuta	Seconda minuta	Ventisetтана	Quarantana
ricordandosi di quel proverbio si Romae fueris, romano vivito more (II VIII 13)	-	-	-
-	Non si può, dice il proverbio, cantare e portar la croce (*II XIX 15, espunto da SP)	[senza glossa] non si può mica cantare e portar la croce (II XVI 47) (< SP II XVI 47: non si può mica far due fatti in una volta)	[senza glossa] non si può cantare e portar la croce (XVI 47)
-	come dice il proverbio... l'abito non fa il monaco (II XIX 15)	come dice il proverbio... l'abito non fa il monaco (II XIX 15)	come dice il proverbio, l'abito non fa il monaco (XIX 15)
-	Il proverbio non veniva a taglio esattamente, ma il conte lo aveva citato in sostituzione d'un altro che gli passava in mente: il lupo lascia il pelo, ma non il vizio (II XIX 16)	Il proverbio non veniva a taglio esattamente; ma il conte lo aveva citato in sostituzione d'un altro che gli passava in mente: il lupo muta il pelo, ma non il vizio (II XIX 16)	Il proverbio non veniva in taglio esattamente; ma il conte l'aveva sostituito in fretta a un altro che gli era venuto sulla punta della lingua: il lupo cambia il pelo, ma non il vizio (XIX 16)
-	mena la danza, come si dice in proverbio (*II XIX 21, espunto da SP)	-	-
-	sono, per servirci di un modo proverbiale, di quelle conche fesse che bastano un pezzo (II XIX 31)	[senza glossa] sono una conca fessa (III XXXVIII 9)	[senza glossa] sono una conca fessa (XXXVIII 9)

Percorsi correttòri dal *Fermo e Lucia* alla Quarantana

Prima minuta	Seconda minuta	Ventisettana	Quarantana
rispose, con una frase proverbiale lombarda: meglio perderlo che trovarlo (III I 55)	-	-	-
Una mano lava l'altra, è un proverbio che l'avrete anche nel vostro paese; quale è il vostro paese? non per cercare i fatti vostri... (III VII 52)	[senza glossa] una mano lava l'altra, e le due il viso (II XIV 17)	[senza glossa] una mano lava l'altra, e le due il viso (II XIV 17)	[senza glossa] una mano lava l'altra, e tutt'e due lavano il viso (XIV 17)
-	[senza glossa] Ma del senno di poi ne son piene le fosse (II XXIV 80)	[senza glossa] Ma del senno di poi ne son piene le fosse (II XXIV 80)	come dice un antico proverbio, del senno di poi ne son piene le fosse (XXIV 80)
Dimmi con chi tratti e ti dirò chi sei è un proverbio; e come tutti i proverbj, non solo è infallibile, ma ha anche la facoltà di rendere infallibile l'applicazione che ne fa chi lo cita (III IX 41)	[senza glossa] dimmi con chi tratti, e ti dirò chi sei (III XXV 29-30)	[senza glossa] Dimmi con chi tratti, e ti dirò chi sei (III XXV 29-30)	[senza glossa] Dimmi chi pratici, e ti dirò chi sei (XXV 30)
-	Su di che il nostro anonimo stimò bene di formare un proverbio: chi vuole aver molti in aiuto, faccia di non averne bisogno (III XXV 39)	Sul qual fatto, il nostro anonimo stimò bene di formare un proverbio: volete aver molti in aiuto? Fate di non averne bisogno (III XXV 39)	Sul qual fatto, il nostro anonimo credé bene di formare un proverbio: volete aver molti in aiuto? Cercate di non averne bisogno un proverbio (XXV 39)

Prima minuta	Seconda minuta	Ventisettana	Quarantana
facendo eco alla voce del popolo, la quale in questo caso – se è lecito fare una eccezione ad un proverbio – non era certamente voce di Dio (IV III 22)	[senza glossa] voce del popolo, (era essa, anche in questo caso, voce di Dio?) (III xxxi 32)	[senza glossa] voce del popolo, (era essa, anche in questo caso, voce di Dio?) (III xxxi 32)	[senza glossa] voce del popolo (era, anche in questo caso, voce di Dio?) (xxxii 32)

Riferimenti bibliografici

- ANTONELLI 2008 = GIUSEPPE ANTONELLI, *Le glosse metalinguistiche dei «Promessi sposi»*, in «Studi di Lessicografia Italiana», xxv, 2008, pp. 141-78.
- BASSI 1932 = DOMENICO BASSI, *Postille inedite di Alessandro Manzoni a Plauto e a Terenzio*, in «Aevum», VI, 1932, pp. 225-74.
- CHERUBINI 1814 = FRANCESCO CHERUBINI, *Vocabolario milanese-italiano*, Milano, Stamperia reale, 1814, tt. I-II.
- CHERUBINI 1839-1856 = FRANCESCO CHERUBINI, *Vocabolario milanese-italiano*, Milano, Imperial Regia Stamperia, 1839-1843, voll. I-IV, più un vol. postumo curato da Giuseppe Villa, Milano, Società tipografica de' Classici italiani, 1856.
- CIANFAGLIONI 2006 = CLAUDIO CIANFAGLIONI, *Vox populi vox Dei? Proverbi e locuzioni idiomatiche nei «Promessi sposi»*, San Martino delle Scale, Abadir, 2006.
- DANZI 2001 = LUCA DANZI, *Lingua nazionale lessicografia milanese. Manzoni e Cherubini*, Alessandria, Edizioni dell'Orso, 2001.
- EVANGELISTI 2001-2003 = SIMONE EVANGELISTI, *Un postillato inedito manzoniano*, in «Annali Manzoniani», n.s., IV-V, 2001-2003, pp. 309-12.
- FANFANI 2021 = MASSIMO FANFANI, *Senno del poi o senno di poi?*, in «Italiano digitale», XVI, 1, 2021, pp. 20-3.
- FORCELLINI 1827-1831 = *Totius Latinitatis lexicon consilio et cura Jacobi Facciolati opera et studio Aegidii Forcellini*, Patavii, typis Seminarii, voll. I-IV, 1827-1831.
- GHIRARDI 2018 = SABINA GHIRARDI, *Le postille manzoniane al Dictionnaire des proverbes français di Pierre de la Mésangère*, in «Prassi ecdotiche della Modernità Letteraria», III, 2018, pp. 205-32.

- LEI 1979- = *Lessico Etimologico Italiano*, fondato da Max Pfister, diretto da Wolfgang Schweickard e Elton Prifti, Wiesbaden, L. Reichert, 1979-.
Lessicografia della Crusca in rete (<http://new.lessicografia.it>).
Manzoni Online (www.alessandromanzoni.org).
- MANZONI/MARTINELLI 2022 = *I Promessi sposi. Ventisettana* [ovvero: V], edizione critica diretta da Dante Isella, a cura di Donatella Martinelli, Milano, Casa del Manzoni, 2022.
- MANZONI/COLLI-ITALIA-RABONI 2006 = Alessandro Manzoni, *I Promessi sposi*, edizione critica diretta da Dante Isella, vol. I: *Fermo e Lucia. Prima minuta (1821-1823)* [ovvero: FL], a cura di Barbara Colli, Paola Italia, Giulia Raboni, Milano, Casa del Manzoni, 2006.
- MANZONI/COLLI-RABONI 2012 = ALESSANDRO MANZONI, *I Promessi sposi*, edizione critica diretta da Dante Isella, vol. II: *Gli sposi promessi. Seconda minuta (1823-1827)* [ovvero: SP], a cura di Barbara Colli e Giulia Raboni, Milano, Casa del Manzoni, 2012.
- MANZONI/POGGI SALANI 2013 = ALESSANDRO MANZONI, *I promessi sposi (1840-1842)* [ovvero: Q], a cura di Teresa Poggi Salani, Milano, Centro Nazionale Studi Manzoni, 2013.
- MARTINELLI 1994 = DONATELLA MARTINELLI, *Postille inedite del Manzoni al «Lexicon» del Forcellini*, in «Annali manzoniani», n.s., II, 1994, pp. 35-78.
- MARTINELLI 2017 = DONATELLA MARTINELLI, *L'edizione digitale delle postille manzoniane a Plauto: problemi ecdotici*, in «Ecdotica», XIV, 2017, pp. 48-88.
- MARTINELLI 2020 = DONATELLA MARTINELLI, *Per la lettura dei «Canti popolari greci» tradotti da Tommaseo. In margine a una nuova edizione*, in «Italiano LinguaDue», I, 2020, pp. 863-72.
- NUNNARI 2013 = TANO NUNNARI, *Le fonti storiche dei «Promessi sposi»*, Milano, Centro Nazionale Studi Manzoni, 2013.
Proverbi italiani (www.proverbi-italiani.org).
- RAIMONDI 2000 = EZIO RAIMONDI, *Un'amicizia europea*, premessa all'edizione del *Carteggio Manzoni-Fauriel*, a cura di Irene Botta, Milano, Centro Nazionale Studi Manzoni, 2000.
- RUMINE 2022 = IRENE RUMINE, *Proverbi e tradizione popolare nei capitoli V e XXXVIII dei Promessi sposi*, in «Studi linguistici italiani», I, 2022, pp. 84-103.
- STELLA-VITALE 2000a = *Scritti linguistici inediti I* [ovvero: SL I], a cura di Angelo Stella e Maurizio Vitale, con Premessa di Giovanni Nencioni, Milano, Centro Nazionale Studi Manzoni, 2000.
- STELLA-VITALE 2000b = *Scritti linguistici inediti II* [ovvero: SL II], a cura di Angelo Stella e Maurizio Vitale, Milano, Centro Nazionale Studi Manzoni, tt. I-II, 2000.

Irene Rumine

TOMMASEO/BALSAMO CRIVELLI 1920 = Niccolò Tommaseo, *Dell'Italia*, a cura di Gustavo Balsamo Crivelli, Torino, Utet, voll. I-II, 1920.

TOMMASEO/MAIOLINI 2017 = NICCOLÒ TOMMASEO, *Canti popolari greci*, a cura di Elena Maiolini, Parma, Guanda, 2017.

TOMMASEO-BELLINI 1861-1879 = NICCOLÒ TOMMASEO-BERNARDO BELLINI, *Dizionario della lingua italiana*, Torino, Unione tipografico-editrice, voll. I-IV, 1861-1879.

TOSI 1991 = RENZO TOSI, *Dizionario delle sentenze latine e greche*, Milano, Rizzoli, 1991.

ZOLLI 1989 = PAOLO ZOLLI, *Il sostrato linguistico della "Quarantana"*, in *Dialettologia e varia linguistica per Manlio Cortelazzo*, a cura di GianLuigi Borgato e Alberto Zamboni, Padova, Unipress, 1989, pp. 401-19.

Riassunto Il contributo si propone di esaminare, nell'ambito della riflessione manzoniana sulla lingua e sulla fraseologia, l'interesse di Manzoni per il proverbio, come documentato nei postillati e negli scritti linguistici. Tale interesse si inserisce nella ricerca della lingua per il romanzo, per poi estendersi alla dimensione sociale dell'individuazione di una lingua nazionale, promossa anche attraverso la realizzazione di un vocabolario dell'uso fiorentino. Soffermandosi quindi sui proverbi dichiarati e corredati di glossa esplicativa nelle quattro redazioni del romanzo, si ripercorrono le correzioni intervenute sugli stessi, dalla Prima minuta all'edizione definitiva.

Abstract This paper aims to analyse, in the context of Manzoni's study on language and phraseology, the interest of the Author in proverbs, as shown particularly by volumes that he annotated and by his linguistic writings. This interest is part of the research, at first, of novel's language and, later, of a national language, also promoted by the project of a dictionary of Florentine spoken language. Focusing on proverbs declared and accompanied by a gloss in the drafts and editions of the novel, the linguistic corrections occurred on them are retraced, from *Fermo e Lucia* to the *Promessi sposi's* final edition.

Phraseological Research of the 16th-Century Slovenian Literary Language

Eva Trivunović

Introduction

In modern standard Slovenian language, the majority of phraseological research is based on large corpora that offer a wide array of tools and options for optimizing searching and sorting results (e.g. GANTAR 2006; GANTAR 2007; KRŽIŠNIK 2013; METERC 2019). In addition, the importance of spoken language and the survey method (particularly of native speakers) have been given more prominence (e. g. METERC 2019, pp. 42-43; cf. KRŽIŠNIK 1996, pp. 136-37). However, in phraseology research of past periods of Slovenian literary language (specifically the 16th century), the research methods differ significantly because those methods (corpus approach and survey methods) are not applicable. Also, although there are many works dedicated to the Slovenian literary language of the 16th century, only a few deal with phraseology. This is why the methods for historical and diachronic phraseological research of the Slovenian language are still being developed and refined. The pursuit for the optimal way of research is still ongoing and many basic questions remain (at least partly) unanswered. This article¹

1 The article was written in the context of the research programme *IMAGE – WORD – KNOWLEDGE. The Transmission and Transformation of Ideas on the Territory Between the Eastern Alps and the North Adriatic 1400-1800 (P6-0437)*. The author acknowledges the project was financially supported by the Slovenian Research Agency from the state budget.

brings attention to the much underresearched field of phraseology of the Slovenian literary language of 16th century, presents sources, methods and specific characteristics of historical phraseological research based on actual language use in written texts rather than drawing data exclusively from older dictionaries.

1. Fixed Multi-Word Units and Phraseological Units

In Slovenian linguistics the term fixed multi-word units (or fixed MWUs) is usually used as a hypernym for different types of fixed MWUs, the common feature of which being the fact that they are a part of the speakers' mental lexicon as a whole and therefore constitute a single lexeme (VIDOVIČ MUHA 2013, p. 109). Unlike free MWUs, which are formed spontaneously, fixed MWUs are recalled by speakers from memory as a whole (*ibidem*).

One of the types of fixed MWUs are phrasemes or phraseological units (or PUs). In the beginning of Slovenian phraseological research in 1970s, there were many different definitions of the basic unit of phraseology and many terminological inconsistencies and variants (KRŽIŠNIK 2013, pp. 16-21), but in the past decade or so these have mostly been resolved. The most common definition in Slovenian phraseology nowadays defines PUs as a fixed MWUs with the following features: relative stability in form (as some variation is anticipated), partially or fully figurative (idiomatic) meaning (not motivated by the sum of its components; the degree of their idiomaticity ranges from more or less transparent to completely opaque), expressiveness and, as all MWUs they are an essential part of a human vocabulary, the lexicon, they are reproducible and lexicalized (GLIHA KOMAC ET AL. 2015, pp. 56-57). It is important to note that with this broader definition of PU the majority of paremiological units (such as sayings and proverbs) also fall within the scope of phraseology.

Modern phraseology increasingly emphasizes that PUs are not as concrete units as traditional lexicography presents them (KRŽIŠNIK 1996, p. 133; MOON 1996, p. 246; GANTAR 2003, p. 212; JESENŠEK-ULČNIK

2014, p. 282; GEERAERT *ET AL.* 2017, p. 87; HANKS 2017, p. 95; METERC 2019, pp. 34-35). «The misconception that idioms are frozen is probably due to the fact that much of the literature on idioms and collocations is based on data derived via introspection» (FELLBAUM 2016, p. 419). Especially with the use of the corpus approach, it has been shown that PUs are much more unstable as corpus material shows a great diversity and variability of these language units (JESENŠEK-ULČNIK 2014, p. 282; FELLBAUM 2016, pp. 412, 419). In modern conceptions variability of idioms is understood as their inherent property and not an error (JESENŠEK-ULČNIK 2014, p. 285). Stability is still considered a defining feature of PUs, but the actual use indicates greater freedom in choosing individual components of the idiom, which calls into question not only the previously registered Slovene phraseology in dictionaries, but also the notion of phraseological stability and stability in general (GANTAR 2003, p. 212; cf. GANTAR 2007, pp. 99-101).

2. 16th-Century Slovenian Literary Language

2.1. Sources and Historical Background

In the second half of the 16th century Slovenian language was established as a literary language as a result of two European movements: Renaissance humanism and the Reformation (HERRITY 2016, p. 4). «The linguistic principle of the Reformation was that the liturgical language should be intelligible to all, and this gave rise to a stimulus to create a literary language based on the vernacular for proselytising purposes» (HERRITY 2016, p. 5). In 1550 the first Slovenian printed books (*Catechismus* and *Abecedarium vnd der klein Catechismus* both authored by Primož Trubar) were published, marking the beginning of the Slovenian literary language. Until 1603 a total of 53 books were published in the Slovenian language (MERŠE 2011, p. 7). The authors were Protestant reformers whose main goal was religious reform, and thus most of the works are of a religious nature: Bible translations, catechisms, postils (a collection of Bible passages and sermons about them), Agenda etc.

The most prominent and comprehensive among them is Jurij Dalmatin's full translation of the Bible in 1584, which remained the only full Slovenian Bible translation for 200 years (GUTHRIE 1984, p. 74). In addition to religious texts there were also some linguistic works published: primers, the first grammar of the Slovenian language (written in Latin, Adam Bohorič: *Arcticae horulae fuccifivae*)² and two multilingual dictionaries that include Slovenian lexemes (HIERONYMUS MEGISER 1592: *DICTIONARIVM QVATVOR LINGVARVM* and Hieronymus Megiser 1603: *Thefaurus Polyglottus*)³.

These 53 Protestant works represent the material for the *Dictionary of the 16th-Century Slovenian Literary Language*, the first volume of which (A-D) was published in 2021 (SSKJ16) and was converted in 2022 into digital form and published at the Fran dictionary portal⁴ (eSSKJ16)⁵. The source material constitutes a paper corpora composed of paper slips with text fragments excerpted from the 16th-century Slovenian Protestants' texts. These paper slips make up a complete list of all occurrences of all words from all 16th-century Slovenian printed Protestant texts (which means that every occurrence of every word is listed in alphabetical order and can be analysed), which contains approximately three million alphabetically arranged paper slips (MERŠE 2011, p. 7). The most extensive text is Dalmatin's Bible, which represents about one third of this paper corpus. Paper slips are in A6 format, with the lemma written in the top left corner, the main part is the word in context and below is information about the exact location of the word

- 2 Most of the works are written in the Bohorič alphabet *bohoričica*, named after Adam Bohorič, who codified it in his grammar. The Bohorič alphabet was in use until it was replaced by Gaj's Latin alphabet *gajica* in the early to mid-19th century (HERRITY 2016, p. 10).
- 3 Although published in 1603 (and thus in 17th century) it is commonly included in research about 16th century, because it contains much of the vocabulary that was in use in the 16th century (MERŠE 2011; AHAČIČ ET AL. 2021).
- 4 <https://fran.si/>
- 5 Instead of waiting to finish and publish the whole second volume (E–J) in approx. 2028, it is planned that new dictionary entries will be added each year to this online version of the dictionary.

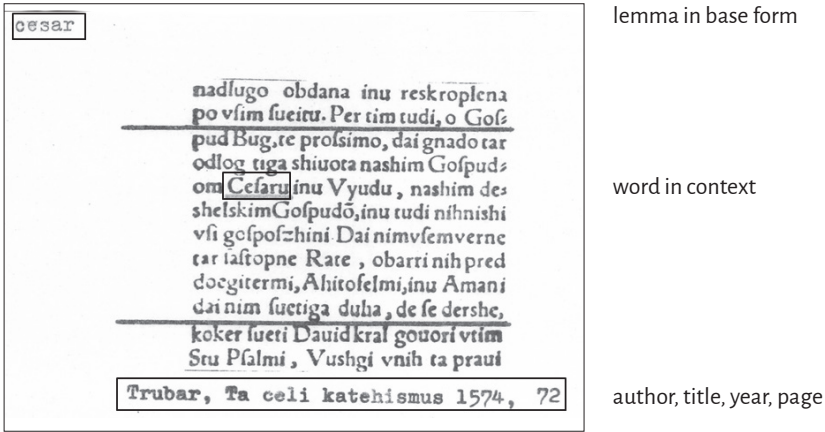


Image 1 A quotation slip of the lemma *cesar* (*caesar*) from the paper corpus

in the text, work, author and year, so that the word can be checked in the primary source if necessary.

Having a complete list of occurrences for each lemma has proved to be a great advantage as it gives us a thorough overview of the material. Because of this, paper concordances are very reliable, but the analysis is time-consuming and prone to human errors in sorting the language material. In phraseological research this means that in order to conduct a thorough analysis, you have to analyse all occurrences of all words that make up the PU, sometimes even the whole word family (families).

For this reason, there have been efforts to digitize this material to make it more accessible and convenient and the first version of corpus (Korpus 16)⁶ is available. It contains 45 of the 53 source books. The books were transcribed by a company and transcriptions were then revised by students and researchers (AHAČIČ ET. AL. 2021, p. 8). This produced a better final result than scanning and optical character recognition, but there are still errors in transcription due to the length and complexity of the texts (*ibidem*). The corpus is searchable, but not

⁶ <https://fran.si/korpus16>

yet lemmatized. The search is conducted character-by-character with some most common spelling variations taken into account. For example: if you enter the letter *a* in the search box, it finds the letter *a* with different superscript marks: *à, á, â, ã*; the entry *nj* finds results *n, nj, jn, jnj, inj* and the entry *š* finds *sh, ſh, ſsh, fſh, ſsh*⁷. Because of the inconsistent orthography and many diverse word forms the search unfortunately rarely covers all relevant phenomena. The corpus it is not yet lemmatized, which decreases its reliability and usefulness for phraseological studies (as multi-word search is not available) in comparison with the paper corpus.

2.2 Language of 16th-Century Texts

The Protestant reformers were well aware of the vast dialectal diversity of Slovenian and strived towards a written language understood by all Slovenians. The language was based on the central dialect without its more local peculiarities, but every author slightly modified it with influences from their own or other dialects (HERRITY 2016, p. 5; GREENBERG 2006, p. 10).

In historical linguistics it is crucial to be aware that the language of past eras is very different from modern language. This is most apparent in the use of different alphabet, spelling (sound-letter relation), word forms and many dialectal features, but it affects all linguistic levels, including denotative and connotative meanings, which are essential for phraseological research. In connection to the connotative meaning, the word *baba* ‘woman’, for example, is always used pejoratively in the modern Slovenian language⁸, but in 16th-century texts neutral uses can be observed as well. Similarly, difference in denotative meaning can be

⁷ <https://fran.si/korpus16/o-portalu?page=Search>

⁸ Modern Slovenian words and their meanings and use are from *Dictionary of the Slovenian Standard Language, 2nd and 3rd Edition* (SSKJ2, eSSKJ) and the information about 16th-century Slovenian literary language is from eSSKJ16 unless stated otherwise.

observed with the adjective *čemern*, which today means ‘gloomy, sullen’, but in 16th-century Slovenian literary language it is attested with a very different meaning: ‘poisonous’ (LEGAN RAVNIKAR 2020, p. 277). Another example is the word *črka*. In modern Slovenian it denotes 1. ‘symbols used to write a language, letter’ and 2. ‘what represents the external side of what is said’. In 16th century Slovenian it had three different meanings of which only the last one is partially equivalent to the first modern meaning: 1. ‘a short line’, 2. ‘a graphic superscript mark indicating a particular pronunciation or accent of the sound(s) marked by a letter’ and 3. ‘character used when writing text: a line, period, comma, letter’.

These examples also illustrate the fact that the researcher’s interpretations can sometimes be biased by his or her linguistic intuition derived from the modern language. As such linguists cannot rely on the linguistic intuition that they have as native speakers of the language more than 450 years after texts were produced, because it can often be misleading (LEGAN RAVNIKAR 2015, p. 451; LEGAN RAVNIKAR 2020, p. 267). Conclusions must be drawn based on the data from the texts themselves and not on assumptions. When trying to describe a linguistic reality of the past, which often differs from the modern reality both in terms of expression and meaning, we often do not have a precise understanding (JELOVŠEK 2018, p. 464). In historical lexicology there is always «inevitably a degree of subjectivity in the interpretation of problematic quotations» (HAWKE 2016, p. 179). Hawke continues that not having sufficient context can lead to misinterpretation. How much context is needed, varies greatly: from a complete sentence to a paragraph (*ibidem*) and in some cases even the entire opus of an author (LEGAN RAVNIKAR 2017, p. 41). Some of the meanings in 16th-century Slovenian cannot be unambiguously determined even with the overview of the very broad context, all occurrences of the lexeme and comparison with the original text in foreign language (LEGAN RAVNIKAR 2015, p. 454).

The functional needs of the Slovenian literary language of the 16th century were different than they are today. Polysemy is more common as lexemes were more semantically burdened (LEGAN RAVNIKAR

2017, p. 48). In translating religious texts translators often faced lexical gaps in Slovenian, and in order to resolve them, new meanings were added to existing borrowed and domestic lexemes as this was easier than creating new words and believed to be more appropriate than loan translations or uncritical borrowings from foreign languages (*ibidem*). Most of the Slovenian texts in the 16th century are translations or adaptations of foreign (mostly German) texts. An important method used to determine meaning of lexemes is a comparison of the Slovenian translation with the original text (JELOVŠEK 2018, p. 467; LEGAN RAVNIKAR 2015, p. 451; LEGAN RAVNIKAR 2017, pp. 41-42; LEGAN RAVNIKAR 2020, pp. 267, 270) and comparison of foreign language equivalent in historical dictionaries of other languages, most often German or other Slavic languages (JELOVŠEK 2018, p. 467; LEGAN RAVNIKAR 2020, p. 267).

Additionally, the knowledge we have at our disposal today is far broader than the knowledge the writers in 16th century had. It is not uncommon that the modern meaning of the lexeme (for example in modern biblical translations) and the one from the 16th century do not match, because today's perception of a given denotatum differs from that of the 16th century (LEGAN RAVNIKAR 2020, p. 279). Having at least some basic knowledge and understanding of the broader socio-historical, cultural and ideological context in which the texts originated is a necessity when dealing with centuries old texts (LEGAN RAVNIKAR 2015, p. 451). For the 16th century, mainly knowledge of the fields of theology, church and cultural history is required (LEGAN RAVNIKAR 2017, p. 42), as well as knowledge of religious Protestant movements in the 16th century, Old and New Testament events etc. (LEGAN RAVNIKAR 2020, p. 270).

2.3 Phraseology of the 16th-Century Slovenian Literary Language

Researching 16th-century phraseology is challenging because phraseologization is a long and gradual process (AHAČIČ ET AL. 2020, p. 74) and PUs historically represent the end point in development (VRBINC

1997, p. 140, following COWIE 1983). In 16th-century Slovenian literary language these processes were just beginning and it is essential to make conclusions based on the features that a MWU exhibits in the 16th century Slovenian (AHAČIČ *ET AL.* 2020, p. 74). In the source material of the 16th-century Slovenian literary language some modern PUs are non-phraseological MWUs that only later developed into PUs (KRŽIŠNIK 1986, p. 441).

It is also important to note that most texts from the 16th century are of a religious nature and there is not a lot of diversity in text types, as mentioned above. This is not necessarily a reflection of the general language use in the 16th century or even in older periods of language history (LEGAN RAVNIKAR 2017, p. 39). Today it is established that many PUs are more often used in spoken than in written communication (METERC 2017, p. 186), and from the past periods only written sources are preserved. For this reason, it is not possible to say with absolute certainty which PUs really were in use. Therefore, if a PU is not attested in the preserved sources, this does not necessarily mean that it was not in use in that period in other contexts or in spoken language.

The problem of distinguishing between different types of MWUs is a universal problem, but it gets significantly more difficult trying to do so in texts hundreds of years old for reasons discussed earlier (cf. 2.2). When distinguishing between different types of MWUs, the main guidelines are the defining features of the different types of MWUs, on the basis of which the criteria are formed. But although this may sound easy in theory, it is far more difficult when dealing with actual language use. At the semantic level, it is especially difficult to distinguish between PUs and (non-phraseological or non-terminological) fixed MWUs as the boundaries are not always unambiguously definable, neither at the level of motivation nor expressiveness (JAKOP-KRŽIŠNIK 2015, p. 422). Nowadays it has become more and more obvious that phraseological theory is the most complex and inconsistent when it comes to defining features (GANTAR 2007, p. 79). The two main problem areas of phraseology of the 16th century Slovenian further presented below are: (in)stability and expressiveness.

2.3.1 Stability

The revised defining feature relative stability (cf. 1) is even further loosened in 16th century Slovenian as the variation across all linguistic levels is far greater than in modern Slovenian due to the fact that written norm was not yet established. What can be defined as stable in a language whose main feature is its instability? And how much variation is still acceptable for something to be considered a fixed MWU or PU? In case of phraseology not all variation is equally significant. The orthographic and phonetic variations are not as significant for phraseological research because they do not influence the (denotative and/or connotative) meanings or usages of the PU. More important to note are word-formational, lexical and syntactic variations and limitations of use⁹. Some examples are presented below in order to illustrate the many types of variations that occur in the 16th-century Slovenian literary language¹⁰.

a) word-formational variants (separated by /):

- variation in aspect (perfective and imperfective verb forms):
dati/dajati kaj na dan [lit. give something to the day] 'to make

- 9 There are some grammatical changes or transformations which also affect some PUs that are (at least in Slovenian phraseology) not considered a variant but form of usage. They can affect one component (e.g. the negation of verb in Slovenian demands a change of the case of the noun/noun clause from accusative to genitive which is reflected in a different ending: *mlatiti prazno slamo* [lit. to beat empty straw] 'to talk without any content' → *ne mlatiti prazne slame* [lit. to not beat empty straw]) or the PU as a whole. The latter has three types: passivization (→ *prazna slama se mlati* [lit. empty straw is beaten]), nominalization (→ *mlatenje prazne slame* [lit. the beating of empty straw]) and relativization (→ *prazna slama, ki jo (kdo) mlati* [lit. empty straw, that someone is beating]) (all based on KRŽIŠNIK 2018). It is also important to note that PUs are highly anomalous lexical units, which means: not all PUs with a similar surface structure allow the same transformations (GANTAR 2007, p. 72). The example used here is a one of very few PUs that can undergo all of the transformations listed above.
- 10 The examples are either from eSSKJ16 or based on the author's own analysis of paper concordances.

something known', closest English PU with similar components and meaning: *see the light of day*

- diminutive and non-diminutive form: *zgubljena ovčica/ovca* [lit. the lost (little) sheep] 'a person without a goal or meaning in life'

b) lexical variants (separated by /):

- *žalostna/revna dolina* [lit. sad/poor valley] 'a place of suffering, especially earthly existence as opposed to afterlife'

c) syntactic variants (separated by //)

- variation in part of speech (noun + noun in genitive // adjective + noun): *dolg krvi // krvavi dolg* [lit. debt of blood // bloody debt] 'guilt, responsibility before God for spilled blood'
- variation in word order: *vinograd božji // božji vinograd* [lit. vineyard of God // God's vineyard] 'a community of people who believe in God'
- variation in verb valency (verb + accusative // verb + instrumental): *biti (celo) dejano za koga/kaj // biti (celo) dejano s kom/čim* [lit. to be given (whole) for who/what // to be given (whole) with who/what] 'expresses that the existence or the duration of who/what has come to an end'

d) combination of different types of variants

- variation in preposition and a lexical variant: *držati na uzdi/vbrzdah* [lit. to hold on bridle/in bit] 'to control, to restrain'
- lexical variant and variation in word order: *po desnici ino levici // z levico ino desnico*, [lit. on the right hand and the left hand // with the left hand and right hand] 'in every way, by all means'

Sometimes, especially concerning paremiological units, the degree of variation is so high that it is impossible to set a base from that would cover all occurrences of the paremiological unit in the source material.

An example of such unit are the four occurrences of the same PU that is derived from the Bible passage: «You can see the speck in your friend's eye. But you don't notice the log in your own eye.» (Luke 6,41; similar also Matthew 7,3-5)

- «Inu bersh v'drusih ozhéh bijl, kakòr v'foih trame vidio» (KRELJ 1567, p. VIb)
- «drugim Trohu videti Vozheh, febi pak Bruuna nepozhutiti» (JURIČIČ 1578, p. II,181a)
- «drugim eno Troho vide vnyh Oku, febi pak Bruuna vlastnem Oku nepozhute» (JURIČIČ 1578, p. II,108)
- «Ti vidilsh eno troho v'eniga drusiga okej, ampak v'tvoim okej trama nepozhutilsh» (DALMATIN 1584: p. III,78)

The next relevant question connected to stability concerns frequency: how many occurrences does it take to consider a MWU a fixed MWU, or in our case a PU, as only usage in different contexts in which the meaning of the phrase as a whole is preserved (AHAČIČ ET AL. 2020, p. 75) can truly prove its presence and prevalence in a language. In modern lexicography five occurrences in the Slovene reference corpus *Gigafida*¹¹ are often used as a frequency threshold for inclusion in the eSSKJ (METERC 2019, p. 35), but sometimes even less is considered enough if the examples are very convincing or if we have other sources indicating its stability and frequent use (*ibidem*). In the corpus of the 16th century Slovenian language five occurrences is too high of a threshold for many PUs. If there is only one occurrence of a MWU, it is most often classified as a metaphor (cf. TRIVUNOVIĆ 2022, p. 37). In some cases, however, even one occurrence can be enough to confirm the presence of a PU in the 16th century Slovenian. Sometimes its context provides information on its general use in the Slovenian language of that time. Text introducers such as *kakor pravimo* 'as we say' are strong

¹¹ Corpus *Gigafida* is available in multiple versions (*Gigafida 1.0*, *Gigafida 1.1*, *Gigafida 1.1 deduplicirana*; *Gigafida 2.0*, *Gigafida 2.0 deduplicirani*). The updated *Gigafida 2.0* was published in 2019 (<https://clarin.si/noske/>).

indicators that something was a fixed expression (AHAČIČ *ET AL.* 2020, p. 75.) and together with idiomatic meaning and expressive function of a unit in the text we can be more certain that we are dealing with a PU even with as little as one occurrence.

2.3.2 Expressiveness

PUs differ from other fixed MWUs as they carry an additional connotative meaning and are expressive in use, but these properties as a criterion are very subjective and therefore challenging. Usually an expressive language unit (such as PUs) has a neutral variant or homonym in the language system, but in the thematically and quantitatively limited corpus of 16th century Slovenian that is often not the case.

There are three basic factors that can help with discerning expressive function and connotative meaning: context, culture and dictionary labels, but in the 16th-century Slovenian language all of them are to a certain extent problematic. Not all the texts are preserved in whole and the context is often difficult to understand as it deals with theological or philosophical issues that are difficult to be fully grasped, which makes it harder to not only determine denotative meanings but also connotative ones as well. Expressiveness is closely linked to culture and as mentioned before (cf. 2.2): the socio-historical, cultural and ideological context in which the texts were written differs significantly from today. In some cases, a good indication that a lexical unit was used expressively in past periods are the dictionaries written at that time and the use of labels. Unfortunately, the 16th-century multilingual dictionaries do not provide such additional information.

Conclusion

The distinction between different types of MWUs in historical linguistics is a demanding task and there are no definitive answers or perfect quick solutions, but a key guideline is to always take into account all of

the defining features rather than just one. Furthermore, it is important to examine and analyse all the components that make up the PU to truly accurately determine its features and try to take into account the extralinguistic circumstances as well. One must not be discouraged just because it requires hard and tedious work. 16th century is an important milestone in Slovenian literary language development and it often surprises researchers with its richness, diversity and timeless relevance.

References

- AHAČIČ ET AL. 2020 = KOZMA AHAČIČ, METOD ČEPAR, ALENKA JELOVŠEK, ANDREJA LEGAN RAVNIKAR, MAJDA MERŠE, JOŽICA NARAT, FRANCE NOVAK, EVA TRIVUNOVIĆ, *SSKJ16: Priročnik za sestavljanje in vpisovanje gesel*, edited by Kozma Ahačič, Ljubljana, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, 2020.
- AHAČIČ ET AL. 2021 = KOZMA AHAČIČ ET AL., *Slovar slovenskega knjižnega jezika 16. stoletja. A-D*, Ljubljana, Založba ZRC, ZRC SAZU, 2021.
- DALMATIN 1584 = JURIJ DALMATIN, *BIBLIA*, Wittenberg, 1584.
- eSSKJ = eSSKJ: *Slovar slovenskega knjižnega jezika 2016–*, www.fran.si.
- FELLBAUM 2016 = CHRISTIANE FELLBAUM, *The Treatment of Multi-word Units in Lexicography*, in *The Oxford Handbook of Lexicography*, edited by Philip Durkin, Oxford University Press, 2016, pp. 411-425.
- GANTAR 2003 = POLONA GANTAR, *Stalnost in spremenljivost frazema v slovarju*, in *Współczesna polska i słoweńska sytuacja językowa (Sodobni jezikovni položaj na Poljskem in v Sloveniji)*, edited by Stanisław Gajda and Ada Vidovič Muha, Opole, Uniwersytet Opolski, Instytut Filologii Polskiej, [Ljubljana], Univerza v Ljubljani, Filozofska fakulteta, 2003, pp. 209-223.
- GANTAR 2006 = POLONA GANTAR, *Korpusni pristop v frazeologiji in slovarske aplikacije*, in *Slovensko jezikoslovje danes*, edited by Ada Vidovič Muha, Ljubljana, Slavistično društvo Slovenije, 2006, pp. 151-163.
- GANTAR 2007 = POLONA GANTAR, *Stalne besedne zveze v slovenščini: korpusni pristop*, Ljubljana, Založba ZRC, ZRC SAZU, 2007.
- GEERAERT-NEWMAN-BAAYEN 2017 = KRISTINA GEERAERT, R. HARALD BAAYEN, JOHN NEWMAN, *Understanding idiomatic variation*, in *Proceedings of the 13th Workshop on Multiword Expressions*, Valencia, Association for Computational Linguistics, 2017, pp. 80-90.

- GLIHA KOMAC ET AL. 2015 = NATAŠA GLIHA KOMAC ET AL., *Koncept novega razlagalnega slovarja slovenskega knjižnega jezika*, Ljubljana, Založba ZRC, ZRC SAZU, 2015.
- GREENBERG 2006 = MARC LELAND GREENBERG, *A Short Reference Grammar of Standard Slovene*, Kansas, University of Kansas, 2006.
- GUTHRIE 1984 = DONALD GUTHRIE, *Besedila in prevodi*, in *Svetopisemski vodnik*, main authors David Alexander and Pat Alexander, Koper, Ognjišče, 1984, pp. 69-80.
- HANKS-EL MAROOF-OAKES 2017 = PATRICK HANKS, ISMAIL EL MAAROUF, MICHAEL OAKES, *Flexibility of Multi-Word Expressions and Corpus Pattern Analysis*, in *Multiword Expressions: Insights from a Multilingual Perspective*, edited by Manfred Sailer and Stella Markantonatou, Berlin, Language Science Press, 2017, pp. 93-119.
- HAWKE 2016 = ANDREW HAWKE, *Quotation Evidence and Definitions*, in *The Oxford Handbook of Lexicography*, edited by Philip Durkin, Oxford University Press, 2016, pp. 176-202.
- HERRITY 2016 = PETER HERRITY, *Slovene: a comprehensive grammar*, London, New York, Routledge, 2016.
- JAKOP-KRŽIŠNIK 2015 = NATAŠA JAKOP, ERIKA KRŽIŠNIK, *Konceptualna izhodišča za vključitev frazeologije v splošni enojezični slovar slovenskega jezika*, in *Slovnica in slovar - aktualni jezikovni opis*, edited by Mojca Smolej, Ljubljana, Znanstvena založba Filozofske fakultete, Obdobja 34, part 1, 2015, pp. 417-424.
- JELOVŠEK 2018 = ALENKA JELOVŠEK, *Slovar slovenskega knjižnega jezika 16. stoletja: pomenske razlage in zgodovinska stvarnost*, in *Slovanská lexikografie počátkem 21. století: sborník příspěvků z mezinárodní konference, Praha 20.-22. 4. 2016*, Praha, Slovanský ústav AV ČR, 2018, pp. 463-470.
- JESENŠEK-ULČNIK 2014 = VIDA JESENŠEK, NATALIJA ULČNIK, *Spletni frazeološko-paremiološki portal: redakcijska vprašanja ob slovenskem jezikovnem gradivu*, in *Več glav več ve: frazeologija in paremiologija v slovarju in vsakdanji rabi*, edited by Vida Jesenšek and Saša Babič, Maribor, Oddelek za germanistiko Filozofske fakultete Univerze v Mariboru, Ljubljana, Inštitut za slovensko narodopisje ZRC SAZU, 2014, pp. 276-292.
- JURIČIČ 1578 = JURIJ JURIČIČ, *POSTILLA*, Ljubljana, 1578.
- Korpus 16: Korpus besedil slovenskih protestantskih piscev 16. stoletja*. Version 1.0, www.fran.si/korpus16.
- KRELJ 1567 = SEBASTIJAN KRELJ, *POSTILLA SLOVENSKA*, Regensburg, 1567.
- KRŽIŠNIK 1986 = ERIKA KRŽIŠNIK, *Poskus razvrstitve stalnih besednih zvez v Trubarjevi Cerkovni ordningi*, in *16. stoletje v slovenskem jeziku, književnosti in kulturi: mednarodni simpozij v Ljubljani od 27.do 29. junija 1984*, edited by

- Breda Pogorelec with collaboration of Jože Koruza, Ljubljana, Filozofska fakulteta, *Obdobja* 6, 1986, pp. 435-445.
- KRŽIŠNIK 1996 = ERIKA KRŽIŠNIK, *Norma v frazeologiji in odstopi od nje v besedilih*, in «Slavistična revija», 44, 2, 1996, pp. 133-154.
- KRŽIŠNIK 2013 = ERIKA KRŽIŠNIK, *Moderna frazeološka veda v slovenistiki*, in *Frazeološka simfonija: Sodobni pogledi na frazeologijo*, edited by Nataša Jakop and Mateja Jemec Tomazin, Ljubljana, Založba ZRC, ZRC SAZU, 2013, pp. 15-26.
- KRŽIŠNIK 2018 = ERIKA KRŽIŠNIK, *Skladnja v frazeologiji med sistemom in rabo*, in «Jezik in slovstvo», 63, 2-3, 2018, pp. 35-47.
- LEGAN RAVNIKAR 2015 = ANDREJA LEGAN RAVNIKAR, *Iz slovaropisne delavnice: ugotavljanje in strukturiranje pomenov pri večpomenskih leksemih v zgodovinskem slovarju*, in *Slovnica in slovar – aktualni jezikovni opis*, edited by Mojca Smolej, Ljubljana, Znanstvena založba Filozofske fakultete, *Obdobja* 34, part 2, 2015, pp. 449-456.
- LEGAN RAVNIKAR 2017 = ANDREJA LEGAN RAVNIKAR, *K problematiki vpliva stičnega jezika - nemščine na semantične spremembe in stilno vrednost najstarejše slovenske knjižne leksike (16. stoletje)*, in «Slovenski jezik - Slovene linguistic studies», 11, 2017, pp. 35-53.
- LEGAN RAVNIKAR 2019 = ANDREJA LEGAN RAVNIKAR, *Neslovnični kvalifikatorji in kvalifikatorska pojasnila v zgodovinskem slovaropisju*, in «Slavia Centralis», 12 (1), 2019, pp. 158-168.
- LEGAN RAVNIKAR 2020 = ANDREJA LEGAN RAVNIKAR, *Pomenske lastnosti knjižne leksike v zgodovinskem slovarju SSKJ16*, in *Slovensko jezikoslovje, književnost in poučevanje slovenščine*, edited by Marko Jesenšek, Maribor, Univerzitetna založba Univerze, Mednarodna knjižna zbirka Zora, 2020, pp. 263-286.
- MERŠE 2011 = MAJDA MERŠE, *Uvod*, in *Besedje slovenskega knjižnega jezika 16. stoletja*, Kozma Ahačič *et al.*, Ljubljana, Založba ZRC, ZRC SAZU, Zbirka Slovarji, 2011, pp. 7-16.
- METERC 2017 = MATEJ METERC, *Paremiološki optimum: najbolj poznani in pogosti pregovori ter sorodne paremije v slovenščini*, Založba ZRC, ZRC SAZU, Zbirka Lingua Slovenica, 10, 2017.
- METERC 2019 = MATEJ METERC, *Analiza frazeološke variantnosti za slovarski prikaz v eSSKJ-ju in SPP-ju*, in «Jezikoslovni zapiski», 25, 2, 2019, pp. 33-45.
- MOON 1996 = ROSAMUND MOON, *Data, Description and Idioms in Corpus Lexicography*, in *Euralex 1996 Proceedings II*, edited by Martin Gellerstam *et al.*, Göteborg University, Department of Swedish, 1996, pp. 245-256.
- SSKJ2 = *Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja*, www.fran.si.

TRIVUNOVIĆ 2020 = EVA TRIVUNOVIĆ, *Tipi stalnih besednih zvez v Slovarju slovenskega knjižnega jezika 16. stoletja*, in *Slovensko jezikoslovje, književnost in poučevanje slovenščine*, edited by Marko Jesenšek, Maribor, Univerzitetna založba Univerze, Mednarodna knjižna zbirka Zora, 2020, pp. 323-355.

VIDOVIČ MUHA 2013 = ADA VIDOVIČ MUHA, *Frazem med besedo in stalno besedno zvezo*, in *Frazeološka simfonija: Sodobni pogledi na frazeologijo*, edited by Nataša Jakop and Mateja Jemec Tomazin, Ljubljana, Založba ZRC, ZRC SAZU, 2013, pp. 109-118.

VRBINC 1997 = MARJETA VRBINC, *Frazeološke enote: definicija in razvrstitev*, in «Vestnik», 31, 1-2, 1997, pp. 127-161.

Abstract The paper deals with phraseology of the Slovenian literary language of the 16th century and tries to fill the gap in this field of linguistic research. Firstly, definition and types of multi-word units are presented, one type of them being phraseological units. Secondly, sources, methods and some linguistic features of 16th century Slovenian literary language are presented. Finally, some phraseological dilemmas of dealing with centuries old texts are presented; main focus is on the relativization of the definitional features of phraseological units: relative stability and expressiveness.

La “lingua delle donne” nella coscienza linguistico-letteraria di Sibilla Aleramo: realtà o costruzione sociale?

Alessandra Rea

1. Introduzione

A partire dagli anni Settanta, quando la diffusione del movimento femminista era incipiente, il dibattito sull'innegabile assenza del soggetto donna nel canone letterario ha consentito la nascita di istituzioni – come la Società Italiana delle Letterate – e di archivi delle donne che hanno compiuto un intenso lavoro di scavo documentaristico. Se osserviamo la produzione in volgare del XIII e del XIV secolo, la donna è dominante nel ruolo di destinataria, ma non in quello di scrittrice. La situazione è paradossale solo in apparenza, perché la rappresentazione del femminile nella lirica duecentesca è una figura simbolica e concepita da menti maschili. L'apparente divario tra la presenza ingente del femminile nella letteratura e la sua esigua rilevanza nella vita quotidiana era stato evidenziato da Virginia Woolf:

Imaginatively she is of the highest importance; practically she is completely insignificant. She pervades poetry from cover to cover; she is all but absent from history. She dominates the lives of kings and conquerors in fiction; in fact she was the slave of any boy whose parents forced a ring upon her finger. Some of the most inspired words, some of the most profound thoughts in literature fall from her lips; in real life she could hardly read, could scarcely spell, and was the property of her husband. It was certainly an odd monster that one made up by reading the historians first and the poets afterwards - a worm winged like an eagle [...] (WOOLF 1935, p. 66)

Non è questa la sede per percorrere il flusso, di rilevante portata, della Letteratura italiana e rintracciarvi quelli che – riadattando il titolo di un'opera di Virginia Woolf – si potrebbero definire i “segni sul muro”, ma sembra fondamentale evidenziare che il problema dell'assenza del femminile come soggetto della letteratura non riguarda la realtà storica in sé, quanto i criteri con cui quella realtà è stata selezionata e tramandata (STORINI 2016, p. 15).

2. Un'assenza che pone alcuni interrogativi

Una produzione letteraria femminile, se pur mediata da mani maschili e poco conosciuta, è sempre esistita (vedi: PLEBANI 2019). Solo nel primo Ottocento, però, le scritture delle donne iniziarono a circolare, a tal punto che Virginia Woolf durante la sua sistematica ricerca afferma: «Here, then, one had reached the early nineteenth century. And here, for the first time, I found several shelves given up entirely to the works of women» (WOOLF 1935, p. 99). Alla fase iniziale del fenomeno appartengono per lo più nomi inglesi, francesi e tedeschi, tra cui: Jane Austen, le sorelle Brontë, Mary Shelley, George Sand, Elizabeth Barrett Browning, George Eliot. Si tratta di scrittrici che con le loro opere hanno rappresentato fenomeni difficili da ignorare, grazie all'attenzione posta al valore espressivo e alla capacità di rifuggire da stilemi programmati. Nel corso del diciannovesimo secolo il giornalismo femminile e la letteratura delle donne crebbero anche in Italia, dando vita a un caleidoscopio di scritture, che non riguardava più soltanto le produzioni private (lettere, diari, memorie). Neera, Marchesa Colombi, Matilde Serao, Ada Negri, Contessa Lara, Sibilla Aleramo, sono alcune delle scrittrici che, attraverso la strada del giornalismo, varcarono la soglia delle Lettere presentando al pubblico i volti della femminilità che da sempre erano stati occultati, il desiderio di evasione da uno spazio – fisico e mentale – angusto, le condizioni di ristrettezza economica in cui la maggior parte delle donne che si accostava ai nuovi mestieri viveva. Le protagoniste erano operaie, maestre, prostitute, monache, mondine, viaggiatrici, ecc., accomunate dal desiderio di percorrere strade diverse per

raggiungere nuovi orizzonti, prima inimmaginabili. Spesso, le stesse scriventi provenivano da situazioni economiche familiari non agiate, ma grazie alla scolarizzazione molte di loro conseguirono il diploma di maestra e si cimentarono nella scrittura giornalistica, nei romanzi, nella pubblicistica educativa e di intrattenimento.

Nonostante la quantità ingente di testi prodotta dalle donne a cavaliere dei due secoli, difficilmente le loro penne riuscirono a «sconfinare dal recinto della letteratura cosiddetta “di consumo”» (FRESU 2019, p. 99), in cui furono relegate anche le opere con scopi pedagogici. Per quanto riguarda la stesura di forme più libere, invece, la pluralità dei percorsi intrapresi dalle scrittrici non fu immediatamente evidente, perché «le prime scritture femminili destinate alla stampa non sono quasi mai scritture dell’io; è l’esito di una generale reticenza all’auto-rappresentazione [...] nel segno di una auto-cancellazione che solo a fine secolo mostra i primi segni di un cedimento» (VILLANI 2018, p. 50). Inoltre, alcune di loro usarono nomi maschili o pseudonimi per firmare e occultare l’identità femminile, che ancora faticava a essere accettata dalle cautele imposte dalla società borghese.

Il quadro esposto sin qui dimostra che storicamente le donne hanno scritto anche in un passato più remoto rispetto a quello presentato dalla storia della letteratura *mainstream*. Non si pretende di aver dato spazio a tutte le voci femminili esistenti, né di aver menzionato tutti i generi sperimentati dalle scriventi. Si spera di aver fornito, per lo meno, un’idea della notevole quantità di scritture sommerse e di aver suscitato una curiosità tale da intraprendere un cammino alla ricerca di altre tracce cadute nell’oblio. Se le donne hanno sempre scritto, però, viene da chiedersi in che modo questo fiume – irriducibile a un singolo rivolo – sia stato soffocato nel proprio alveo. Inoltre, è doveroso domandarsi se sia sufficiente affermare l’esistenza di “una” scrittura femminile, là dove sembra celarsi una *variopinta pluralità*. Osservando la vasta produzione di queste scriventi – che Benedetto Croce indicò con l’espressione «infinito pulviscolo» (CROCE 1940, vol. VI, p. 185) – sembra di trovarsi nella situazione

[...] in cui i parametri diafasici, quali la tipologia testuale (sussidiario, novella edificante, *pièce*, manuale comportamentale, galateo, racconto sentimentale-

le-psicologico, romanzo rosa), la funzione (alfabetizzare, educare, intrattenere), l'utenza (infantile, giovanile, adulta, maschile e femminile), devono aver costituito, insieme al *background* culturale delle autrici, un significativo condizionamento delle scelte linguistiche adottate. (FRESU 2019, p. 106)

La quasi totale assenza di autrici nel canone letterario ha contribuito alla sedimentazione di logori stereotipi riguardo alla subalternità della produzione femminile rispetto a quella maschile. Inoltre, là dove le scritture delle donne hanno trovato angoli di spazio per sopravvivere sono state relegate a una letteratura non ritenuta di primo grado. Ciò è avvenuto anche perché le disamine di taglio linguistico su testi di mano femminile sono state condotte principalmente nell'ambito della scrittura privata, dove il registro adottato è tendenzialmente più basso rispetto a quello impiegato nelle opere pensate per la pubblicazione. Rita Fresu ha parlato di una «miopia» (FRESU 2016, p. 18), a causa della quale non è stato possibile tracciare un *continuum* tra le epoche, tra le tipologie testuali, tra le varietà diastraticamente distanti. Pertanto, si ritiene interessante analizzare in prospettiva linguistica un'opera nata con ambizioni letterarie che prende le distanze dalla «varietà (come detto non letteraria) di tipo “medio”, osservata per lo più attraverso un genere testuale strutturalmente ben definito (e precipuamente considerato “femminile”) come quello epistolare» (FRESU 2016, p. 17). Come accennato poc'anzi, la produzione delle autrici in Italia tra diciannovesimo e ventesimo secolo è caratterizzata da una *variopinta pluralità* di voci che non è possibile analizzare in questa sede. A partire dalla riflessione di Francesca Sanvitale (SANVITALE 1995, p. IV), si è deciso di condurre un'analisi in prospettiva linguistica e di genere del romanzo di Sibilla Aleramo *Una donna* (1906). Davanti al successo di questa scrittrice la “cancellazione ufficiale” delle autrici dal canone letterario (vedi: CRISPINO 2015 e STORINI 2016) sembra aver subito un'inibizione e il suo pensiero ha aperto la strada alle riflessioni riguardo ai ruoli che la società e la storia hanno scolpito per donne e uomini. Con *Una donna* Sibilla Aleramo mette in scena la graduale conquista di una consapevolezza piena di sé da parte della protagonista, giungendo alla realizzazione di un'opera che ha le sembianze del romanzo di formazione,

in cui il confine tra realtà e mito di sé è quasi impercettibile. L'opera è il frutto di una prima stagione femminista che si manifestò fra la fine dell'Ottocento e gli inizi del Novecento e rappresenta «un documento, un manifesto, l'attestato di un martirio consapevole e vissuto al di fuori di ogni retorica» (POZZATO 1980, p. 47).

3. Metodo

Il rapporto che intercorre tra le scrittrici e la lingua sarà osservato analizzando la percezione di Sibilla Aleramo riguardo a una presunta “lingua delle donne” e il confine fra la lingua dei personaggi e quella della scrittrice. Anzitutto, era plausibile per l'autrice pensare a una “lingua delle donne” diversa da quella degli uomini per la mera variabile *sex*? L'attenzione è stata rivolta all'intreccio di più variabili, come ad esempio la variazione diastratica, diafasica e di *genere*. La ricerca prende avvio dai dati raccolti ed elaborati nei sondaggi sociolinguistici di Monica Berretta e Rita Fresu: le indagini condotte in ambito sociolinguistico sulla presunta variazione correlata con il *sex* hanno dimostrato che quando le donne non sono imbrigliate in reti sociali fitte o, ancor meglio, sono in movimento verso l'alto delle gerarchie sociali non ricorrono a fenomeni linguistici considerati prototipici (vedi: GIACALONE RAMAT 1979; MILROY 1980; LABOV 1998). A partire dai tratti ritenuti più caratterizzanti della varietà femminile, fra cui la *politeness*, Monica Berretta nel 1983 ha condotto una ricerca che ha come oggetto lo studio della lingua delle donne in qualità di immagine sociale e non di sistema oggettivamente esistente. Infatti, nel suo lavoro la studiosa specifica che

un'indagine puramente linguistica del comportamento verbale femminile non porta a risultati soddisfacenti e generalizzabili: qualsiasi ricerca approfondita non può che correlare il comportamento verbale al ruolo rivestito dalle donne nelle interazioni o in genere nella cultura di appartenenza. (BERRETTA 1983, p. 217)

L'indagine è stata condotta da Berretta somministrando al campione di studenti universitari selezionato due diversi tipi di intervista: un

questionario (proposto a 174 persone) di domande strutturate, o semistrutturate, e una prova di identificazione (proposta a 172 persone) di testi maschili e femminili, dopo la quale si chiedeva di motivare la scelta dell'attribuzione. Si ritengono interessanti le riflessioni ottenute dall'incrocio delle due modalità di intervista, con cui Berretta ha commentato i risultati ottenuti. In particolare, per quanto riguarda la risposta alla prima domanda del questionario – “secondo lei c'è differenza fra uomini e donne nel modo di parlare?” – si deve notare come l'80% degli/delle intervistati abbia risposto affermativamente.

Complessivamente, dai dati osservati si può dedurre che le differenze evidenziate dal campione sono la conseguenza di paradigmi socialmente costruiti e correlati inevitabilmente con le dimensioni diastratiche e diafasiche. A distanza di circa vent'anni i questionari usati da Monica Berretta sono stati somministrati, con alcune innovazioni, da Rita Fresu a un campione prevalentemente giovanile romano e cagliaritano, con il medesimo fine di osservare alcune dinamiche sulla percezione della lingua che la comunità di parlanti ha in relazione al genere (FRESU 2008a, pp. 129-163). Lo studio è partito soprattutto dall'osservazione dei dati ricavati dall'indagine sociolinguistica del 1983 e dalla conseguente rilevazione di una diffusa rappresentazione negativa della lingua femminile, da cui è emerso

un orizzonte di attese ben definito da parte della comunità di parlanti nei confronti del comportamento linguistico attribuito ai generi [...] In altre parole ci si aspetta che uomini e donne, in quanto tali, si esprimano ricorrendo a precise strategie, si servano di determinate espressioni e ne evitino altre. (FRESU 2008b, pp. 92-93)

Le risposte e i criteri con cui il campione intervistato dalle due linguiste ha attribuito i testi a uomini e a donne sono stati il punto di partenza per sviluppare una griglia di riferimento per lo studio qui presentato, contenente da un lato i tratti ritenuti caratterizzanti per la lingua delle donne dall'opinione comune, dall'altro alcuni fenomeni linguistici attraverso i quali è misurata la presenza di quei tratti nel romanzo esaminato:

Tratti ritenuti caratterizzanti per la lingua delle donne	Fenomeni da analizzare nei testi scelti
Tendenza alla conservatività	<ul style="list-style-type: none"> – Soluzioni arcaizzanti – Tratti regionali – Forestierismi
Strategia del garbo (<i>politeness</i>)	<ul style="list-style-type: none"> – Uso di alterati, in particolare diminutivi e vezzeggiativi – Impiego di espressioni affettive e di formule fatiche – Sistema allocutivo (in particolare, pronomi, appellativi, apposizioni ed epiteti) – Interiezioni, esclamazioni
Inclinazione alla distensione del discorso	<ul style="list-style-type: none"> – Aggettivazione ricca – Strutture binarie e ternarie – Reduplicazioni lessicali e/o sintattiche e strutture foderate
Attenzione ai dettagli	<ul style="list-style-type: none"> – Varietà lessicale – Capacità di astrazione – Uso di crononimi
Selezione di argomenti specifici	<ul style="list-style-type: none"> – Sfera degli affetti e dei sentimenti – Sfera domestica e familiare – Curiosità riguardanti particolari privati e osservazioni dettagliate su altre persone

Tabella 1 Griglia di riferimento per l'analisi linguistica

Nel presente contributo saranno osservati e commentati i dati relativi al tratto *strategia del garbo (politeness)*, in quanto si ritengono quelli più adatti a illustrare la metodologia e gli strumenti di ricerca impiegati nello studio effettuato. A partire dalle discussioni dei dati ricavati dalle ricerche sopra citate, si è deciso di rintracciare nel romanzo di Sibilla Aleramo quei fenomeni che più di altri sono stati chiamati in causa dagli intervistati e dalle intervistate nei sondaggi di Fresu e Berretta. Per rintracciare gli eventuali segni della *strategia del garbo* la ricerca è stata indirizzata verso l'uso di *diminutivi e vezzeggiativi*, l'impiego di

espressioni affettive e formule fatiche e, soprattutto nell'analisi del *sistema allocutivo*, il ricorso a *interiezioni ed esclamazioni*.

Per contestualizzare nell'opera l'uso di alcuni tratti rinvenuti durante lo spoglio dei dialoghi e di capire quanto le variabili relative ai personaggi abbiano influenzato determinate scelte linguistiche, è stato impiegato GATE. Si tratta di un software *open-source* per lo sviluppo di sistemi *Information Extraction* (IE) proposto dall'Università di Sheffield: «[...] i processi di IE [sono] basati sul *Natural Language Processing* (NLP), inteso come un insieme di tecniche e strumenti dedicati all'analisi e all'elaborazione del linguaggio usato comunemente dalle persone in forma scritta e orale» (GALLERANI 2015, p. 2). Grazie a questa ulteriore analisi del testo è possibile confrontare i fenomeni linguistici tratti dai dialoghi con l'*usus scribendi* della scrittrice nelle parti narrate. Per svolgere questo tipo di ricerca è stato anzitutto necessario reperire il romanzo in formato *TXT* (vedi: FUSI 2011), scaricato da *Liber Liber* nell'edizione *Una donna*, Firenze, R. Bemporad & Figlio, Editori, MCMXXI³.

Prima di caricare il file nel software scelto per l'AAT è stato necessario servirsi di un ulteriore *tool*, così da fornire a GATE un testo già suddiviso in parti del discorso. GATE, infatti, è ben strutturato per dividere in categorie grammaticali testi in lingua inglese ma la stessa precisione di attribuzione non si ottiene per l'italiano. Dunque, si è fatto ricorso a *Linguistic Annotation Pipeline*, elaborato dall'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa.

Le annotazioni applicate al testo scaricabile fanno parte del TANL (*Text Analytics and Natural Language processing*) e sono state concepite a partire dall'ISST (*Italian Syntactic-Semantic Treebank*), un *corpus* articolato sui livelli ortografico, morfo-sintattico, sintattico e semantico. Caricando il testo *annotato* su GATE, il software riconosce gli *annotation types* relativi alle categorie linguistiche dei livelli sopra citati e permette di formulare richieste specifiche per condurre l'analisi testuale. È stato adottato l'approccio *rule based*, che è fondato proprio sulle annotazioni e sulle informazioni di cui esse sono portatrici.

Pur non trattandosi di un'analisi relativa alla lingua letteraria dell'epoca, è sembrato interessante cercare di delineare alcune tendenze

linguistiche del *modus scribendi* dell'autrice, così da rintracciare i confini (eventuali) tra la lingua della voce narrante e quella dei personaggi. Anche questo lavoro di analisi automatica del testo – complementare allo spoglio tradizionale – è stato condotto con l'obiettivo principale di capire quale fosse, e se ci fosse, la percezione di una lingua delle donne da attribuire ai personaggi femminili.

4. Risultati¹

Fenomeni linguistici	Personaggi	Esempi
Uso di alterati, in particolare diminutivi e vezzeggiativi	Anziane del paese	Demonietto
	Marito di Sibilla	Gli amici vantano il tuo ingegno, mi dicono che ho una sposina invidiabile
	Madre del collega di Sibilla con cui lavora per la rivista <i>Mulier</i>	E tu, piccina, hai subito un poco il suo fascino?
	Direttrice di <i>Mulier</i>	Cappellini
Impiego di espressioni affettive e formule fatiche	Sibilla	Figlio mio, figlio mio...; figliuolo mio
	Madre del collega di Sibilla con cui lavora per la rivista <i>Mulier</i>	Cara; e tu, piccina...;
	Direttrice di <i>Mulier</i>	Caro Perugino (soprannome che dà a Sibilla)
	Disegnatrice norvegese	Mia piccola
	Suocera di Sibilla	Ah, figlio mio
	Madre di Sibilla	Poveri figli miei (in una lettera)

¹ Da qui in poi si è fatto ricorso al **grassetto** per evidenziare i fenomeni linguistici oggetto di interesse nelle parole e nelle espressioni riportate nella tabella. Nel commento le frasi e i periodi tratti dal romanzo sono stati scritti in *corsivo*, mentre è stato usato il **grassetto corsivo** per indicare specifiche parole o suffissi all'interno dei passi riportati.

Alessandra Rea

Fenomeni linguistici	Personaggi	Esempi	
Sistema allocutivo (in particolare, pronomi, appellativi, apposizioni ed epiteti)	Padre di Sibilla	Devo dire dunque che sei una civetta ?	
	Madre di Sibilla	È vero, dica , che lei accompagna mio marito a passeggio la notte dalla parte del fiume? (uso dell'allocutivo lei per rivolgersi a un notaio, amico del marito)	
	Suoceri di Sibilla	La voce narrante racconta come i suoceri si rivolgessero a lei con l'allocutivo voi .	
	Forestiero	La voce narrante indica il momento in cui il forestiero iniziò a darle del tu .	
	Suocero di Sibilla	Signora baronessa	
	Anziana donna di servizio	Ah, signorina mia (cui segue l'impiego dell'allocutivo lei per rivolgersi a Sibilla)	
	Donna non identificata	Da due ore, signora	
	Suocera di Sibilla	Povera donna , non sapete quante altre sono nel caso vostro (uso dell'allocutivo voi)	
	Interiezioni, esclamazioni	Padre di Sibilla	Finalmente respiro!; Quella donna impazzisce!; Taci, tu!; Esci!
		Sibilla	Oh, se veramente il babbo ci portasse pel mondo!
Madre di Sibilla		Ah, se tu avessi avuto un bimbo!; Perdonatemi, perdonatemi...	
Marito di Sibilla		Vivi! Per nostro figlio!; Fa' quello che vuoi!	
Anziana donna di servizio		Ah, signorina mia, fossi trent'anni di meno! Chi sa che avreste fatto di me!	
Disegnatrice norvegese		Ah che noia, mia piccola, che noia!; Era brutta, sai!	

Fenomeni linguistici	Personaggi	Esempi
Interiezioni, esclamazioni	Amica anziana di Sibilla	Agire! Questa è la vera propaganda!
	Attrice	Avessi un quarto di secolo in meno! Io l'imporrei!
	Suocera di Sibilla	Il paradiso! State qui come una regina!
	Cognata di Sibilla	Ah, voglio vederlo padrone del paese, un giorno!

Tabella 2 Strategia del garbo (*politeness*)

Relativamente all'uso di *alterati* sono state trovate nei dialoghi quattro occorrenze tra vezzeggiativi e diminutivi: *demonietto*, *sposina*, *piccina*, *cappellini*. Si ritiene interessante la collocazione co-testuale del diminutivo impiegato dal marito di Sibilla, rivolgendosi proprio a lei: *Gli amici vantano il tuo ingegno, mi dicono che ho una **sposina** invidiabile*. Il passo appena riportato si può associare alla riflessione di Robin Lakoff, secondo cui tratti ritenuti caratterizzanti della cosiddetta “lingua delle donne” sarebbero usati dagli uomini solo a scopo ironico e parodistico (LAKOFF 1975, pp. 8-18). *Demonietto*, invece, è adoperato dalle anziane del paese nei confronti della protagonista per sottolineare le differenze di pensiero e di comportamento tra la ragazza e i ricordi della loro gioventù. Gli ultimi due diminutivi rinvenuti sono attribuiti a personaggi femminili: *E tu, **piccina**, hai subito un poco il suo fascino?*; *caro Perugino, con un po' di garbo puoi far l'elogio tanto dello struzzo provveditore dei **cappellini**, quanto di Sant'Antonio protettore del matrimonio*. Nel primo caso il diminutivo è usato con finalità affettive, mentre nel secondo la scelta cela intenzioni ironiche.

Mediante l'analisi automatica del testo è stato possibile ampliare lo sguardo all'intero *corpus* testuale, così da osservare la distribuzione di diminutivi e vezzeggiativi nell'*usus scribendi* della scrittrice. Come spiegato nel paragrafo 3, i risultati delle indagini sociolinguistiche menzionate mostrano la presenza di un'aspettativa dei/delle parlanti che tra i tratti “tipicamente femminili” della lingua usata delle donne prevede il

ricorso a diminutivi e vezzeggiativi. Su un totale di 121.794 *token* (LENCI-MONTEMAGNI-PIRRELLI 2020, p. 102) sono state rinvenute le seguenti occorrenze di alterati affettivi o attenuativi (oltre ai quattro inseriti nella tabella 2): *pezzetto*; *giovinetto* x2; *viaggetti*; *figliuollette*; *stanzuccia*; *piccina*; *sorellina* x2; *donnina*; *padroncina*; *personcina*; *ragazzina*; *mussolina*; *sorelline* x3; *stanzine*; *fratellini* x5; *piccini*; *saletta* x2; *giornaletto*; *letticciuoli*; *corpellini*; *canzonette*; *caminetto*; *bocchetta*; *boccuccia* x3; *lettino* x2; *appartamento*; *piccino* x24; *libriccino*; *corpicino*; *nipotino*; *avvocatino* x3; *cuoricino*; *ragazzina*; *testina* x4; *creaturina* x2; *donnina*; *faccina*; *piccolina*; *gambette*; *alberghetto*; *figliuolletto*; *stanzette* x2; *articoletto*; *vecchietta*; *casetta* x2; *tavolino*; *cervellino*; *ometto*; *vocetta*; *letticciuolo* x2; *foglietto*; *labruzzo*; *stanzetta* x2; *cantuccio*; *quartierino*; *nasino*; *visino*; *tavolino* x2; *stanzino*; *madonnina*; *manina* x3; *vocina*; *mammine*; *cappellini*, *ditini*. La presenza di diminutivi e vezzeggiativi è considerevole, ma contestualizzando i contesti in cui appaiono si nota che molti sono usati dalla protagonista per parlare del figlio. Inoltre, nella prosa filotoscana del XIX secolo l'uso dei valutativi era molto diffuso (SERIANNI 2020).

Sarebbe interessante confrontare tali risultati con quelli provenienti da analisi analoghe svolte su opere coeve di produzione maschile. Come ricorda Rita Fresu, infatti:

[...] diversi studi, per lo più condotti sul parlato, hanno confutato il maggiore utilizzo di forme attenuative, soprattutto di diminutivi, e degli alterati in genere, nelle produzioni femminili, dimostrando una sostanziale equità d'uso da parte di uomini e donne, in linea con una tendenza che da qualche decennio sta mettendo in luce l'infondatezza delle opinioni attribuite alla lingua delle donne. (FRESU 2008c, p. 21)

Le espressioni affettive e le formule allocutive riportate nella tabella 2 provengono da voci esclusivamente femminili e sono emerse principalmente nei contesti in cui le donne si rivolgono ai propri figli. Anche in questo caso – come per la sezione precedente – è necessario osservare il contesto in cui tali espressioni appaiono, così da non correre il rischio di ricondurle esclusivamente alla variabile *gender*.

Nella sezione della tabella 2 dedicata al sistema allocutivo emerge l'apposizione attribuita dal padre di Sibilla alla propria moglie, men-

tre «indugiava dinanzi allo specchio, dubbiosa della sua toeletta che non indossava da molto tempo: *Devo dire dunque che sei una civetta?*». L'epiteto assegnato al personaggio femminile fa parte dell'insieme di metafore zoomorfe che nella storia della lingua italiana sono state attribuite quasi esclusivamente alle donne e indica il tentativo di attirare l'attenzione (Fusco 2012, pp. 70-76). Restando nell'ambito di appellativi, apposizioni ed epiteti attribuiti alle donne, si notano un'occorrenza di *signorina mia* e due di *signora*. La prima è usata da un'anziana per rivolgersi alla protagonista, mentre le occorrenze della seconda forma citata dell'appellativo sono impiegate dal suocero di Sibilla e da una donna non identificata. Il primo ricorre a *Signora* per rivolgersi alla figlia – che «aveva un temperamento imperioso ed egoista, freddo e lunatico» –, aggiungendo anche l'apposizione *baronessa*. Indubbiamente, si tratta di uno di quei casi in cui il suffisso *-essa* «è pronto ad essere riutilizzato e a risprigionare la sua carica negativa» (SABATINI 1987, p. 116).

Interiezioni ed esclamazioni nei dialoghi sono sempre usate da donne – fatta eccezione per *Finalmente respiro!* del padre di Sibilla – quando hanno tono ottativo o enfatico: *Ah, se tu avessi avuto un bimbo; Ah, signorina mia, fossi trent'anni di meno! Chi sa che avreste fatto di me!; Ah che noia, mia piccola, che noia! Ah, voglio vederlo padrone del paese, un giorno! ecc.* Al contrario, le esclamazioni attribuite agli uomini assumono nel contesto del romanzo un tono imperativo e non garbato: *Quella donna impazzisce!; Taci, tu!; Esci!; Vivi! Per nostro figlio!; Fa' quello che vuoi!*

Nel corpus analizzato in toto le formule in cui confluiscono *interiezioni ed esclamazioni* – considerate dall'opinione comune un indizio di coinvolgimento emotivo da parte del/della mittente – rilevate da GATE sono 40, cui bisogna sottrarre quelle ottenute dallo spoglio dei dialoghi per volgere lo sguardo all'*usus scribendi* di Sibilla Aleramo. Si riportano alcuni esempi: *non mi amava, ah, certo che non mi amava più!; il figlio lo seguirebbe, oh, dovunque!; mi amava, oh!; Mha...quando sarei grande!; Oh la mia bella adolescenza selvaggia!; Ah, che davvero non sapevo nulla, in fondo, della vita, per aver troppo ed esclusivamente contemplato mio padre!; Oh, la perfetta letizia di quell'estate!; Ah, come di fronte alla fine cade ogni speranza di sfidare e vincere l'Ignoto!*

5. Discussione

Durante l'effettuazione dell'analisi automatica del testo sono stati trovati numerosi *alterati affettivi* o *attenuativi* che, secondo i campioni intervistati da Monica Berretta e da Rita Fresu (cfr. § 3), rappresentano indizi di coinvolgimento emotivo, tipici della lingua impiegata dalle donne. Come evidenziato nel paragrafo precedente, la frequenza assoluta di vezzeggiativi e diminutivi è considerevole, ma è necessario calcolare la frequenza relativa delle occorrenze raccolte per ottenere un dato oggettivo, adatto a rispondere alle istanze che hanno guidato lo studio qui presentato. Dal calcolo della frequenza relativa si ottiene il risultato infinitesimale di 0,009. Dunque, la presenza dei *valutativi* non sembra poter esser considerata una marca di *politeness* nella presunta lingua delle donne, soprattutto se analizzata nel quadro più generale della prosa fiorentina del XIX secolo.

L'uso di *espressioni affettive* e di *formule allocutive* è stato analizzato soltanto nei dialoghi, dove l'interazione tra personaggi – anche appartenenti a diverse estrazioni sociali – consente di capire la variazione dei rapporti interpersonali al mutare degli scenari scelti dalle scrittrici. I dati raccolti provengono solo da personaggi femminili, che costituendo la schiacciante maggioranza, rischiano di non far luce su altre variabili. Si notino, infatti, i sentimenti di affetto, amicizia, conoscenza, amore genitoriale che veicolano tali espressioni e formule.

Sul piano dei pronomi allocutivi impiegati si può affermare che in *Una donna* l'uso del pronome *voi* rappresenta l'eccezione. A tal proposito, la voce narrante scrive: «e mi sorrideva un po' confusa dandomi del *voi*². Anche mio suocero non riusciva a dirmi *tu*». Anche in un altro passo la protagonista nota – e fa notare a chi legge – il passaggio dal *voi* al *tu*, associandolo a un'evoluzione dei rapporti interpersonali. Si tratta del momento in cui il “forestiero” che si è invaghito di lei le inizia a dare del *tu*: «Si risolse, troncò le dimostrazioni, mi prese le mani, ravvivò gli occhi, mi disse che mi amava, ch'io pure l'amavo, che saremmo stati felici presto; mi dava del *tu*».

² Si riferisce alla suocera.

Il pronome allocutivo *lei*, invece, è usato dalla madre di Sibilla per rivolgersi a un amico notaio del marito, che era solito trascorrere serate a casa loro e dall’anziana donna di servizio per parlare con la protagonista.

Per quanto riguarda l’ultima sezione della tabella 2, *interiezioni ed esclamazioni*, sono state individuate 40 occorrenze. Si ribadiscono le ripercussioni delle scelte linguistiche sull’*atto illocutivo* e su quello *perlocutivo*, con conseguenti differenze nella rappresentazione comportamentale dei personaggi maschili e femminili. Le mittenti cercano di essere garbate e di non imporsi al destinatario, con il prevedibile risultato di apparire costantemente vulnerabili e insicure. Al contrario, i personaggi maschili ricorrono al modo verbale imperativo per rivolgersi a donne che ricoprono il ruolo di madri o mogli, testimoniando il retroterra patriarcale vigente nel panorama storico e sociale che fa da sfondo al romanzo.

6. Conclusioni

Come già specificato nel terzo paragrafo, il presente contributo è stato elaborato a partire da una ricerca più ampia che ha previsto l’analisi anche di altri tratti ritenuti prototipici della cosiddetta lingua delle donne all’interno del romanzo di Sibilla Aleramo. Pertanto, si ritiene opportuno collocare i risultati qui presentati e commentati all’interno di una cornice che prende in considerazione anche i dati ricavati dalla misurazione degli altri tratti (cfr. tabella 1). Come è stato scritto nell’introduzione (cfr. § 1), la prolifica produzione femminile elaborata tra l’Otto e il Novecento si presenta ancora come un terreno di indagine poco esplorato, soprattutto se considerata in relazione alla questione della lingua, ai processi che hanno portato all’unità linguistica nazionale e alla partecipazione delle scrittrici – se pur attraverso modalità e canali differenti – alla graduale formazione di una coscienza letteraria che pone al centro le donne, i volti della loro fatica, l’insoddisfazione per la disuguaglianza di genere, l’intraprendenza, il sogno di libertà intellettuale. Attraverso l’analisi linguistica dei passi

dialogici e l'esame dei profili delineati da Sibilla Aleramo sono state evidenziate le eventuali relazioni tra le scelte linguistiche della scrittrice e le rappresentazioni femminili da lei tratteggiate, così da esplorare la sua sensibilità circa gli stereotipi di genere e la sua volontà di metterli in discussione. Pur non trattandosi di uno studio linguistico in prospettiva della lingua letteraria dell'epoca, mediante l'analisi automatica del testo è stato possibile volgere lo sguardo anche all'*usus scribendi* dell'autrice e inserire i fenomeni linguistici più ricorrenti nel quadro linguistico-letterario di riferimento. Per la restituzione di un esaustivo e attendibile sfondo teorico l'analisi di un solo romanzo non è sufficiente; pertanto, sarebbe opportuno condurre un esame sistematico dei tratti ritenuti prototipici della lingua delle donne su un campione di testi – di mano femminile e maschile – rappresentativo, con l'obiettivo di delineare un *continuum* tra la percezione del *genere* negli scrittori e nelle scrittrici e gli usi linguistici scelti per far parlare i propri personaggi.

Dall'analisi qui presentata complessivamente è emersa l'assenza della percezione di una "lingua delle donne" da parte dell'autrice, ma si rileva la consapevolezza riguardo alla correlazione tra ruolo sociale, retroterra culturale e lingua dei personaggi. Infatti, la protagonista – *alter ego* di Sibilla Aleramo – è la figura femminile che più delle altre cerca di divincolarsi dal contesto che la società ha attribuito alle donne e al contempo quella che ricorre meno ai tratti ritenuti prototipici della "lingua femminile", se non nei discorsi diretti in cui si rivolge affettuosamente al figlio e ai sentimenti che nutre verso di lui. Nella coscienza letteraria della scrittrice, il fattore che influisce sulle scelte linguistiche non è il *sex*, biologicamente determinato, bensì il *genere*, socialmente e storicamente determinato, insieme alle sue intersezioni con altre varianti. È possibile, infatti, trovare nel romanzo due rappresentazioni del femminile: da un lato le anziane donne del paese che vivono pensando in maniera ossessiva al matrimonio, panacea di un destino che altrimenti sarebbe certamente segnato dalla miseria e dall'esclusione sociale; dall'altro il ritratto che Sibilla Aleramo fa di sé. Un ritratto che nel corso del libro diventa sempre più consapevole, combattivo, libero. Ugualmente libera e lontana dalle pieghe degli stereotipi è l'amica an-

ziana della protagonista di *Una donna* che esclama: «bisogna riformare la coscienza dell'uomo, creare quella donna».

Inoltre, si ritiene che non si possa parlare di una specifica lingua delle donne riscontrabile nell'*usus scribendi* dell'autrice. Anche là dove le occorrenze dei fenomeni linguistici analizzati presentano una frequenza assoluta considerevole, il calcolo della frequenza relativa conduce a risultati infinitesimali. La presenza cospicua di determinati fenomeni, quali i valutativi diminutivi e vezzeggiativi, risulta in linea con la prosa letteraria del tempo e, dunque, non denota alcuna eccezionalità legata al sesso della scrivente.

Nella prospettiva di ricerche future si ritiene indispensabile l'analisi della produzione letteraria, maschile e femminile, coeva a quella di Sibilla Aleramo, così da poter ampliare il quadro metodologico-teorico e confrontare la presenza dei tratti ritenuti prototipici per la “lingua delle donne” alla luce dell'intersezione di più variabili, come quelle diastratiche, diafasiche, il *genere*, la tipologia testuale. Al contempo, sarebbe interessante indagare, attraverso l'analisi dei passi dialogici, la consapevolezza degli scrittori e delle scrittrici rispetto al ruolo che il *genere* e le variabili ad esso connesse ricoprono nelle scelte linguistiche attribuite a personaggi di diversa estrazione sociale e con retroterra culturali differenti.

Bibliografia

- BERRETTA 1983 = MONICA BERRETTA, *Per una retorica popolare del linguaggio femminile, ovvero: la lingua delle donne come costruzione sociale*, in *Comunicare nella vita quotidiana*, a cura di Franca Orletti, Bologna, il Mulino, 1983.
- CRISPINO 2015 = *Oltre canonone. Generi, genealogie, tradizioni*, a cura di Crispino Anna Maria, Guidonia (RM), Iacobelli Editore, 2015.
- CROCE 1940 = BENEDETTO CROCE, *La letteratura della nuova Italia. Saggi critici*, VI vol., Bari, Laterza, 1940.
- FRESU 2008a = RITA FRESU, «*Gli uomini parlano delle donne, le donne parlano degli uomini*». *Indagine sociolinguistica in un campione giovanile di area romana e cagliaritano in Lingua italiana del Novecento-scritture private, nuovi linguaggi, gender-*, Roma, Edizioni Nuova Cultura, 2008, pp. 129-163.

Alessandra Rea

- FRESU 2008b = RITA FRESU, *Il gender nella storia linguistica italiana (1988-2008)*, in «Bollettino di italianistica. Rivista di critica, storia letteraria, filologia e linguistica», n.s., anno V, n. 1, 2008, pp. 86-111.
- FRESU 2008c = RITA FRESU, «*la mia testa è un po' mattatuccia*». *Gli alterati nelle lettere di S. Gemma Galgani*, in *Lingua italiana del Novecento-scritture private, nuovi linguaggi, gender-*, Roma, Edizioni Nuova Cultura, 2008, pp. 11-22.
- FRESU 2016 = RITA FRESU, *L'infinito pulviscolo: tipologia linguistica della (para)letteratura femminile in Italia fra Otto e Novecento*, Milano, Franco Angeli, 2016.
- FRESU 2019 = RITA FRESU, *Donne, lingua, letteratura in Italia*, in *Rivista internazionale di testi e studi*, Pisa-Roma, Fabrizio Serra Editore, 1, 2019.
- FUSCO 2012 = FABIANA FUSCO, *La lingua e il femminile nella lessicografia italiana – tra stereotipi e (in)visibilità*, Alessandria, Edizioni dell'Orso, 2012.
- FUSI 2011 = DANIELE FUSI, *Informatica per le scienze umane*, vol. 1, Roma, Nuova cultura, 2011.
- GALLERANI 2015 = ROBERTO GALLERANI, *Natural Language Processing (NLP) e Information Extraction (IE)*, in *Quaderni-appunti digitali*, <https://www.slideshare.net/rgallerani/natural-language-processing-nlp-e-information-extraction-ie>.
- GIACALONE RAMAT 1979 = ANNA GIACALONE RAMAT, *Lingua, dialetto e comportamento linguistico. La situazione di Gressoney*, Aosta, Tipo-Offset Musumeci, 1979.
- LABOV 1998 = WILLIAM LABOV, *Resolving the Gender Paradox in the Study of Linguistic Change*, in *Linguistica storica e sociolinguistica*, Atti del Convegno della Società Italiana di Glottologia, Roma, 22-24 ottobre 1998, testi raccolti a cura di Palmira Cipriano, Rita d'Avino, Paolo Di Giovine, Roma, editrice "IL CALAMO", MM, 1998, pp. 37-38.
- LAKOFF 1975 = ROBIN LAKOFF, *Language and woman's place*, New York, Harper&Row, 1975.
- LENCI-MONTEMAGNI-PIRRELLI 2020 = ALESSANDRO LENCÌ, SIMONETTA MONTEMAGNI, VITO PIRRELLI, *Testo e computer-Elementi di linguistica computazionale*, Roma, Carocci editore, 2020.
- MILROY 1980 = LESLEY MILROY, *Language and social networks*, Blackwell, Oxford, 1980.
- PLEBANI 2019 = TIZIANA PLEBANI, *Le scritture delle donne in Europa. Pratiche quotidiane e ambizioni letterarie (secoli XIII-XX)*, Roma, Carocci editore, 2019.
- POZZATO 1980 = MARIA PIA POZZATO, *I romanzi e le prose di Sibilla Aleramo*, in *Sibilla Aleramo*, a cura di Maria Pia Pozzato e Isabella Pezzini, Scandicci, La Nuova Italia Editrice, 1980.
- SABATINI 1987 = ALMA SABATINI, *Il sessismo nella lingua italiana*, Roma, Presidenza del Consiglio dei Ministri, 1987.

SANVITALE 1995 = FRANCESCA SANVITALE, *Le scrittrici dell'Ottocento. Da Eleonora De Fonseca Pimental a Matilde Serao*, Roma, Istituto Poligrafico e Zecca dello Stato, 1995.

SERIANNI 2013 = LUCA SERIANNI, *Storia dell'italiano nell'Ottocento*, Bologna, il Mulino, 2013.

STORINI 2016 = MONICA CRISTINA STORINI, *Il secchio di Duchamp. Usi e riusi della scrittura femminile in Italia dalla fine dell'Ottocento al terzo millennio*, Pisa, Pacini Editore, 2016.

VILLANI 2018 = PAOLA VILLANI, *Ritratti di signore. I galatei femminili nell'Italia belle époque e il caso Serao*, Milano, Franco Angeli, 2018.

WOOLF 1935 = VIRGINIA WOOLF, *A room of one's own*, London, Hogarth press, 1935.

Riassunto Nell'immaginario collettivo della cultura occidentale è stata spesso recriminata alle donne una tendenza alla ridondanza espressiva, mentre il silenzio e la capacità di non far parlare di sé sono state ritenute caratteristiche apprezzabili nel sesso femminile. L'esistenza stessa, sin dall'antichità, del concetto di una lingua femminile è emblematica, soprattutto se la si accosta all'assenza della nozione di lingua maschile. In questa prospettiva, dunque, la presunta lingua delle donne rappresenta la varietà *marcata*, rispetto alla norma di cui sono latori gli uomini. Tali considerazioni antropologiche e sociolinguistiche hanno avuto ricadute anche sul giudizio circa la produzione letteraria femminile, esclusa o relegata ai margini del canone. Lo scopo del contributo è quello di affrontare il tema riguardante il rapporto che intercorre tra le scrittrici e la lingua, come costruzione sociale. In particolare, si è deciso di indagare sulla percezione di Sibilla Aleramo riguardo a una presunta lingua delle donne. Anzitutto, era plausibile per la scrittrice pensare a una “lingua delle donne” diversa da quella degli uomini per la mera variabile sesso? Per rispondere a questo interrogativo, l'attenzione è stata rivolta all'intercetto di più variabili, come ad esempio la variazione diastratica, diafasica e di *genere*. L'analisi è stata sviluppata attraverso una griglia di riferimento contenente da un lato i tratti ritenuti caratterizzanti per la lingua delle donne (tendenza alla conservatività, ricorso alla strategia del garbo, inclinazione alla distensione del discorso), dall'altro alcuni fenomeni linguistici (soluzioni arcaizzanti, tratti regionali, uso di valutativi, impiego di espressioni affettive e di formule fatiche) attraverso i quali è misurata la presenza di quei tratti nei dialoghi esaminati.

Abstract Since ancient times women have been considered as derivative of men or subordinate to them. This cultural vision impacts the expectations regarding their language and their literary production. The aim of this paper is to analyse the correlation

Alessandra Rea

between women writers and language, as social construct strongly correlated to diastratic and diaphasic factors. This study will examine Sibilla Aleramo's awareness about the existence of a "women's language" in their novel *Una donna*. The research starts from sociolinguistic surveys, which have revealed the existence of a set of expectations held by a community of speakers regarding the linguistic behaviours attributed to *gender*. The analysis applies to the novel a set of prototypical features of "women's language", according to speakers' expectation, and some linguistic phenomena (divided on the bases of different linguistic levels, like morpho-phonological, syntactic, and pragmatic-textual) for each one of these traits. The linguistic examination has been performed with GATE, an open-source software used for the development of Information Extraction system. Thanks to the data obtained it is possible to compare the occurrences of each prototypical feature with all the others linguistic possibilities available to express the same concept. The relative frequency of the linguistic phenomena demonstrates that they depend on an intersection of variables like *gender*, diastratic and diaphasic factors. It is very important to focus on the difference between *sex* and *gender*. As a matter of fact, existing literature on the subject has proven there are no differences in masculine and feminine language dependent on *sex*, being a biological factor. Instead, there can be differences depending on *gender*, in the sense of a factor built on social and historical phenomena.

Il lessico della psicanalisi in Italia nel primo Novecento: tra rimozioni, lacune lessicografiche e possibili sviluppi digitali

Stefano Miani

Le parole erano in origine degli incanti e la parola ha conservato ancora molto della sua antica forza magica. Con la parola un uomo può rendere il suo simile felice oppure spingerlo alla disperazione, con la parola il maestro trasmette la sua scienza agli scolari, con la parola l'oratore trascina a sé l'uditorio e ne determina i giudizi e le decisioni. La parola suscita degli affetti ed è il mezzo comune col quale gli uomini influiscono l'uno sull'altro.

(SIGMUND FREUD, *Introduzione allo studio della psicoanalisi*, nella traduzione di Edoardo Weiss del 1922)

La terminologia psicoanalitica è uno dei problemi della psicoanalisi.

(*Avvertenza generale alle Opere di Sigmund Freud*, Torino, Bollati Boringhieri, [1967-1980])

Nel 1964, in un articolo pubblicato sulla rivista «Lingua Nostra», l'italianista Michel David (1924-2020) auspicava che qualcuno volesse «studiare la psicanalisi e la sua storia per mezzo del linguaggio»:

quali sono le parole della psicanalisi passate nell'uso volgare, nell'uso colto, nell'uso genericamente tecnico, o rimaste nell'uso specializzato degli analisti di oggi? e quali parole erano già in uso, in tutte queste categorie di parlanti, prima della guerra del 1914, poi in ognuno dei due decenni interbellici? (DAVID 1964, p. 19).

Abbiamo deciso di indagare la formazione del vocabolario italiano della psicanalisi partendo dall'auspicio di David e dalla constatazione dell'assenza di un lavoro che studi sistematicamente, da un punto di vista storico-linguistico, la formazione in Italia di questo linguaggio specialistico, importante sia in sé sia per le sue ricadute sulla lingua comune e sulla letteratura¹. Il ricorso ai repertori lessicografici è tutt'altro che soddisfacente e l'uso da parte dei lessicografi di fonti non specifiche (si pensi che *transfert* è registrato da *GDLI* al 1981 con *Vino al vino* di Mario Soldati), ha la conseguenza storicamente distorta di far credere che una terminologia che si è formata nel complesso entro i primi due decenni del Novecento si sia formata nell'arco di quasi un secolo.

Nel presente intervento ci concentreremo sulla descrizione del *corpus* di testi su cui verranno effettuati gli spogli linguistici ed esporremo alcuni dati quantitativi ottenuti dalla sua digitalizzazione.

1. Un linguaggio specialistico senza specialisti?

È stato osservato che, perché si possa parlare di *linguaggio specialistico*², per prima cosa devono esserci degli specialisti che utilizzino tra loro una determinata terminologia (CORTELAZZO 1994). Se dovessimo seguire rigidamente questo assunto potremmo concludere che, man-

- ¹ I notevolissimi contributi di David (due volumi e due brevi articoli, che, pur aggiornati successivamente, risalgono agli anni Sessanta) esprimono un approccio alla materia sostanzialmente storico-culturale e letterario, con escursioni linguistiche occasionali e limitate. Oltre a questi, si contano studi linguistici su un paio di parole (CORTELAZZO 1981 e PELILLO 2012).
- ² Per una disamina del «pulviscolo di denominazioni» (GUALDO 2016, p. 372) – *linguaggio specialistico, lingua speciale, lingua settoriale* –, rimandiamo a GUALDO-TELVE 2011 (§ 1.1). Noi parliamo di *lessico* – seguendo BECCARIA 2004 (s.v.) –, in quanto ci riferiamo esclusivamente a quelle parole e locuzioni che formano la terminologia psicanalitica. Segnaliamo, inoltre, che utilizziamo la variante *psicanalisi* e non, quella più “ortodossa”, *psicoanalisi* (così parliamo di *psicanalisti* e usiamo l'agg. *psicanalitico*).

cando in Italia veri psicanalisti, tutto il periodo precedente all'opera di Edoardo Weiss (1889-1970) sarebbe dovuto uscire dalla nostra analisi, essendo questi l'unico che potesse vantare tale titolo nel nostro Paese tra i due conflitti mondiali e avendo iniziato a pubblicare solo a partire dai primi anni Venti del XX secolo. Avremmo, quindi, potuto studiare la terminologia presente nei suoi *Elementi di psicoanalisi* (WEISS 1931) o nella «Rivista italiana di psicoanalisi», da lui diretta e attiva tra il 1932 e il 1934. Già questo lavoro avrebbe contribuito a correggere, ma solo in parte, le distorsioni cronologiche della lessicografia di cui abbiamo già parlato. Ma, grazie all'opera di oscuri (talvolta insospettabili) pionieri o curiosi divulgatori, gran parte della terminologia psicanalitica era già stata tradotta in italiano prima del 1918.

Una delle peculiarità della psicanalisi, infatti, è che mentre di solito i linguaggi specialistici si formano nel tempo con un lento processo di accumulo, la terminologia psicanalitica si forma e si coagula nel giro di pochi decenni. Questo lessico venne plasmato dalla straordinaria capacità espressiva di Freud (tanto che il suo nome fu proposto per il Nobel per la letteratura!), che, compiendo una scelta “galileiana”, al tecnicismo (magari coniato *ex novo* su base greca o latina) preferì «parole normali, non tecniche, parole della lingua d'uso» (RANCHETTI 2002, p. 5). Ma è facile rilevare che Freud non ha esitato a operare travasi lessicali³ (basti pensare al meccanismo della *sublimazione*, dal ted. *Sublimierung*, termine proprio della fisica) e a utilizzare parole coniate da altri scienziati (come è noto il termine *Es* è di Groddeck).

Questo lessico *sui generis*, composito ed evocativo, ma considerato spesso fin troppo vago, insieme al problema della infalsificabilità dei risultati della ricerca psicanalitica⁴, ha contribuito a porre in dubbio la scientificità della disciplina stessa.

Dal momento che, pur con vistose eccezioni (basterà accennare alla seconda topica freudiana – *Io*, *Es* e *Super Io* – teorizzata solo nei primi anni Venti), nel periodo in cui in Italia si inizia a parlare di psicanalisi (1907-1909) questa possiede già una sua terminologia e che si tratta,

³ Utilizzo *travaso* al posto di *transfert* seguendo GUALDO 2021, pp. 130-131.

⁴ Su questo argomento rimandiamo a CAROTENUTO 2002², § 1.

quindi, di tradurla in italiano, abbiamo ritenuto di poter procedere con una indagine che si sviluppasse sul doppio binario dello specialismo e della divulgazione (con relativa banalizzazione). Questo ci ha permesso di scoprire che non sempre le parole che oggi sono considerate tecnicismi psicanalitici hanno seguito una trafila regolare: talvolta, infatti, scelte di non specialisti sono state, quasi *ob torto collo*, accettate dagli stessi psicanalisti⁵.

2. Definizione dell'arco cronologico

Nei primi quattro decenni del XX secolo la psicoanalisi in Italia coincide, in pratica, con l'opera del solo Freud e presenta molte lacune e ritardi rispetto ai progressi e alle diversificazioni già in atto nei Paesi dell'area tedesca o anglosassone. Ancora intorno al 1915, infatti, la disciplina è identificata quasi esclusivamente con *L'interpretazione dei sogni*, la cui prima edizione tedesca risale al 1899⁶, e le traduzioni italiane di libri di Freud nella prima metà del secolo sono poche e non comprendono opere capitali come *La psicopatologia della vita quotidiana* e la stessa *Interpretazione dei sogni*⁷. Almeno fino all'«esplosione anarchica» (DAVID 1990, p. 143)⁸ del secondo dopoguerra, gli sparuti psicanalisti

⁵ Cfr. MIANI 2022.

⁶ Il volume *Traumdeutung*, infatti, «fu pubblicato in realtà il 4 novembre 1899, ma l'editore vi appose la data dell'anno successivo» (FREUD 1971, p. XXI).

⁷ La prima monografia di Freud tradotta in italiano è *Sulla psicoanalisi* (FREUD 1915).

⁸ Michel David, nel suo imprescindibile volume *La psicoanalisi nella cultura italiana* suddivide in tre fasi la penetrazione della psicoanalisi nella cultura italiana: a) un primo volgarizzamento ad alto livello (1908-1915) seguito dal b) periodo tra le due guerre mondiali in cui si inasprisce la polemica tra oppositori (in primo luogo l'ambiente idealista, la Chiesa cattolica e il regime fascista) e i pochi fautori della nuova disciplina per giungere, infine, nel secondo dopoguerra, a quello che David definisce c) «esplosione di libertà anarchica», seguita da una corposa opera di traduzione e divulgazione dei freudiani, junghiani e americani (DAVID 1990, p. 143). Per un'analisi generale dei rapporti tra psicoanalisi e ideologie politiche (fascismo, marxismo, cattolicesimo) cfr. *ivi*. pp. 29-141. Per un quadro generale della situa-

che, tra mille difficoltà e diffidenze, riuscirono a operare nel nostro Paese osservarono una stretta ortodossia freudiana, con le partigianerie e omissioni di altri autori che questa posizione comportava⁹.

Abbiamo circoscritto, quindi, l'ambito cronologico dei nostri spogli al periodo che va dal 1899, data con cui di norma si fa nascere la psicanalisi¹⁰, al 1939. Questo anno, infatti, non segna solo lo scoppio della seconda guerra mondiale e la morte di Freud, fuggito a Londra a seguito dell'*Anschluss*, ma è l'anno in cui Edoardo Weiss emigra negli Stati Uniti a causa della legislazione antisemita messa in piedi dal regime fascista a partire dal 1938. Con l'uscita di scena di Weiss, centro dello sparuto gruppo di membri della *Società Italiana di Psicoanalisi*, molti dei quali, essendo ebrei, costretti alla diaspora o al silenzio, cala un velo di silenzio sulla disciplina che sarà destinato a perdurare fino al secondo dopoguerra¹¹.

Essendo ogni scelta cronologica, come è ovvio, arbitraria, per alcuni termini non abbiamo rispettato tale criterio, operando incursioni sia negli anni precedenti sia in quelli successivi; vedremo tuttavia che gran parte del lessico psicanalitico italiano si è formato entro il 1918, stabilizzandosi non più tardi dei primi anni Trenta.

zione tra il 1915 e il 1945 cfr. *ivi*, pp. 162-218. Per un'analisi dettagliata dei rapporti tra idealismo crociano e psicanalisi cfr. DAVID 1963. Per la tardiva e faticosa penetrazione in Italia di Jung cfr. CAROTENUTO 1977. Questi testi sono stati una fonte bibliografica importante per la costruzione del *corpus*.

- 9 È tuttavia possibile riscontrare un marginale, ma comunque presente, apporto terminologico anche da parte di altri psicanalisti "eterodossi" come Alfred Adler e Carl Gustav Jung o Eugen Bleuler. Nonostante i tre già prima del primo conflitto mondiale si fossero distaccati dal Freud furono spesso considerati esponenti, magari non "ortodossi", della psicanalisi.
- 10 Anche se «l'atto di nascita della psicoanalisi» è individuato in una lettera di Freud al collega e confidente Wilhelm Fliess, datata 21 settembre 1897 (cfr. MUSATTI 1991, p. x). Il 1899 è l'anno di pubblicazione della *Traumdeutung*.
- 11 La constatazione che anche per profili storici di una disciplina affine alla psicanalisi come la psicologia sia stato scelto come discriminare la fine della seconda guerra mondiale, poiché l'influenza, metodologica e, non ultimo, linguistica anglo-americana fu così ingente da stravolgerne le fondamenta, ci ha convinto della bontà della scelta operata. Cfr. CIMINO-DAZZI 1998 e MARHABA 2003.

3. Tipologie di fonti che compongono il corpus

La tipologia delle fonti raccolte (figura 1) è molto eterogenea. In gran parte le abbiamo individuate grazie all'imponente mole di scritti indicati in DAVID 1990. A queste sono stati aggiunti articoli individuati dallo spoglio degli indici delle principali riviste mediche del tempo e, non ultimo, da ricerche mirate di termini psicanalitici negli archivi on line di alcuni quotidiani italiani («La Stampa» e il «Corriere della Sera») e su *Google libri*. Si è cercato di costruire un *corpus* quanto più rappresentativo di ciò che è stato pubblicato sulla psicanalisi in Italia nel nostro periodo di riferimento.

Il *corpus* così ottenuto è composto per lo più da articoli provenienti da riviste psichiatriche e psicologiche. Dato poi che, come abbiamo detto, fino alla pubblicazione della «Rivista italiana di psicoanalisi», non esiste un organo ufficiale della psicanalisi in Italia (nonostante il meritorio «Archivio generale di neurologia, psichiatria e psicoanalisi» diretto da Marco Levi Bianchini [1875-1961], su cui ritorneremo), quasi tutti i contributi che si occupano della nuova disciplina sono disseminati tra riviste psicologiche, psichiatriche, filosofiche, antropologiche e, nell'ultimo periodo, anche giuridiche. Sebbene poche e, con rare eccezioni, di scarso contenuto scientifico, sono importanti alcuni interventi su quotidiani come il «Corriere della Sera» e «La Stampa» o su riviste non scientifiche come la «Nuova Antologia», la «Civiltà Cattolica», «La Voce» o la «Fiera letteraria». Questi scritti non hanno solo il merito di proporre a un pubblico di non addetti la nuovissima disciplina, ma ne diffondono, certo banalizzandolo, il lessico nel momento stesso in cui questo va formandosi (ponendo le basi per una certa confusione e causando molti futuri fraintendimenti). Una situazione, insomma, peculiare.

Il *corpus* non contiene però solo articoli, ma, grazie soprattutto all'instancabile opera dello psichiatra Marco Levi Bianchini, che sarà ben presto affiancato dal più competente Edoardo Weiss, anche traduzioni di opere di Freud e di altri psicanalisti (Otto Rank, Anna Freud, Marie Bonaparte). Sono presenti, inoltre, altre monografie, piccoli manuali, pubblicati da case editrici importanti come Fratelli Bocca a

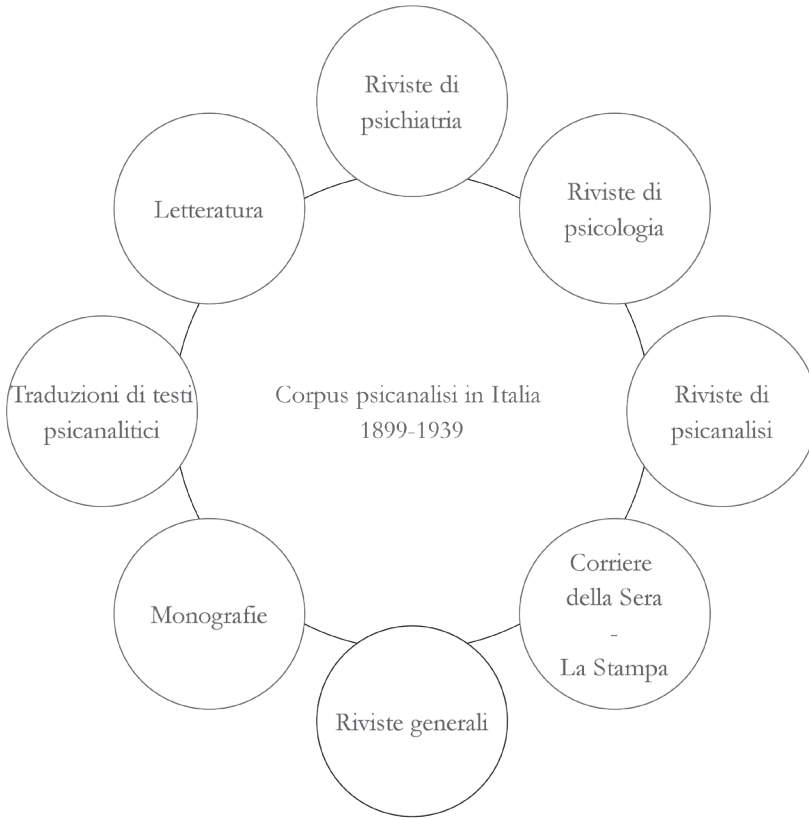


Figura 1 Tipologia di fonti raccolte

Torino, Hoepli a Milano e Zanichelli a Bologna, che, pur con infiniti limiti, lacune e fraintendimenti, hanno il merito di tentare la divulgazione al grande pubblico che vuole farsi un'idea della psicanalisi. Una terza tipologia di monografia è rappresentata dai contributi originali. Tuttavia, vista l'esiguità dei membri della *Società psicoanalitica italiana* e la scarsa qualità di questi (eccetto Weiss, nessuno di loro si era formato a Vienna e si era sottoposto a un'analisi), le opere che rientrano in questa categoria sono poche e, inutile dirlo, i contributi innovativi

italiani di riferimento nel nostro periodo, se si eccettua qualche studio di Weiss sull'agorafobia, sono praticamente inesistenti.

Sempre esigue, ma di non poco conto, le opere letterarie che accolgono, diffondendolo, il lessico psicanalitico. È soprattutto dopo la sua diffusione in Francia a partire dal primo dopoguerra, che è possibile assistere a una, seppur modestissima, presenza di lessico psicanalitico in autori come Giuseppe Ungaretti, Giovanni Papini e, soprattutto, Riccardo Bacchelli che nel romanzo *Oggi, domani e mai* (1932) descriverà una seduta psicanalitica. Interessanti, inoltre, a cavallo tra gli anni Venti e Trenta, i primitivi tentativi di applicare i concetti della psicanalisi alla critica letteraria (TISSI 1929) e all'arte in generale (ANGELUCCI 1929).

4. Il «corpus» digitalizzato

Il materiale raccolto è stato digitalizzato, dapprima in formato PDF, poi, mediante processo OCR, in file di testo (.txt) e, dopo una tokenizzazione tramite Python, è stato caricato in formato XML, in locale, sulla piattaforma *NoSketch Engine* e, on line, sulla piattaforma *Sketch Engine*¹².

Ciò che presentiamo è ancora un *work in progress*.

Il *corpus* è attualmente composto (figura 2) da 460 documenti, per un totale di oltre 5 milioni di parole. Ogni documento è stato marcato in modo da poterne rilevare l'autore, l'anno di pubblicazione, la tipologia e, se è un articolo, la rivista su cui è stato pubblicato. In questo modo è stato possibile estrarre dati quantitativi che riteniamo salienti.

Possiamo innanzitutto osservare (figura 3)¹³ che le 63 monografie raccolte, con oltre 3 milioni di parole, rappresentano il 52% del *corpus*, soprattutto a causa del fatto che i 397 articoli sono per lo più brevi note

¹² Cfr. KILGARIFF ET AL. 2004.

¹³ Come è ovvio l'unità di misura per valutare il “peso”, che chiameremo *estensione*, all'interno del *corpus* di un documento sarà il numero di parole. Perciò avvertiamo, per evitare equivoci, che gli aerogrammi, che suddividono le fonti tra monografie o riviste o tra tipologie di riviste, dove non chiaramente indicato, si basano sull'*estensione* e non sul numero dei documenti.

Corpus della Psicanalisi in Italia (1899-1939) ?

Creato da Stefano Miani

Counts		General info		Lexicon sizes	
Tokens	6,309,347	Language	Italian	word	190,514
Words	5,289,909	Encoding	UTF-8	tag	52
Documents	460	Compiled	01/14/2023 12:02:02	lemma	126,946

Structures and attributes

doc 460

Figura 2 Maschera di NoSketch Engine con le Informazioni sul «Corpus» della psicanalisi in Italia (1899-1939)

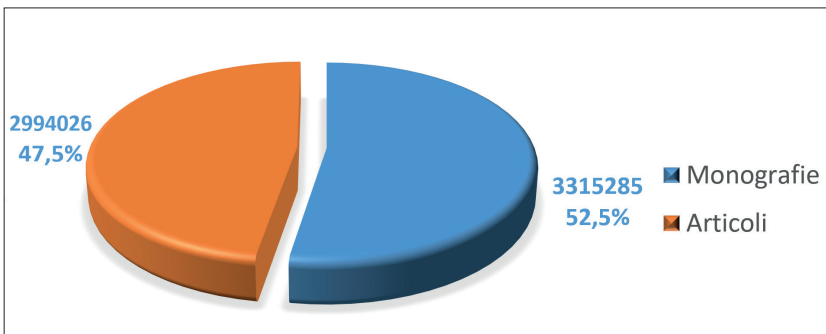


Figura 3 Suddivisione del «corpus» in monografie e articoli. Sopra la percentuale è indicato il numero di parole

e recensioni. In alcuni casi si è scelto di digitalizzare intere opere (così ad esempio per alcune monografie del neuropsichiatra e psicologo Sante De Sanctis [1862-1935]¹⁴), in altri, si è scelto invece di non includere la monografia completa, ma di inserire esclusivamente un capitolo o un passo specifico che trattassero di psicanalisi.

¹⁴ In particolare ci riferiamo a DE SANCTIS 1925 e 1929-1930.

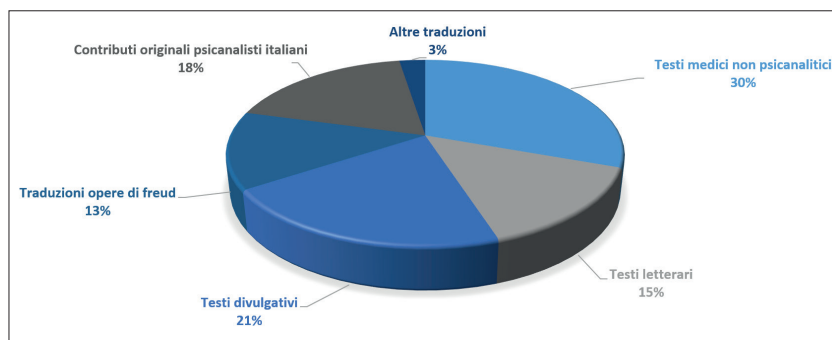


Figura 4 Tipologie di monografie all'interno del «corpus»

Per i motivi di cui sopra, se guardiamo alla composizione delle monografie (figura 4) possiamo osservare come il 30% (stiamo sempre parlando di numero di parole e non di pubblicazioni) appartenga a testi etichettati come «testi medici non psicanalitici», in quanto si tratta di opere, soprattutto trattati di psichiatria e di neuropsichiatria, in cui solo incidentalmente si parla di psicanalisi, spesso per attaccarla.

Sotto l'etichetta «testi divulgativi» abbiamo incluso anche due corposi volumi di Enrico Morselli intitolati *La psicanalisi* (MORSELLI 1926) che, con oltre 400.000 parole, rappresentano la fonte più estesa del nostro corpus (7,7 % del totale). Questi non possono rientrare nella categoria di testi psicanalitici, in quanto l'autore non era uno psicanalista ma ne era, anzi, un grande detrattore. Nonostante l'avversione dello stesso Freud per quest'opera¹⁵, i due volumi di Morselli sono importantissimi, in quanto furono «le fonti di quasi tutti i critici italiani di Freud che scrissero dopo il 1925 [...] e forse anche per chi parteggiò a favore della psicoanalisi» (DAVID 1990, p. 175).

Come è stato già detto, le opere originali di psicanalisti italiani contribuiscono solo per il 18% al totale delle monografie ed è importante osservare come le tre edizioni degli *Elementi di psicoanalisi* di Weiss rappresentino da sole il 30% di questa categoria.

¹⁵ Cfr. WEISS 1971, pp. 72-73.

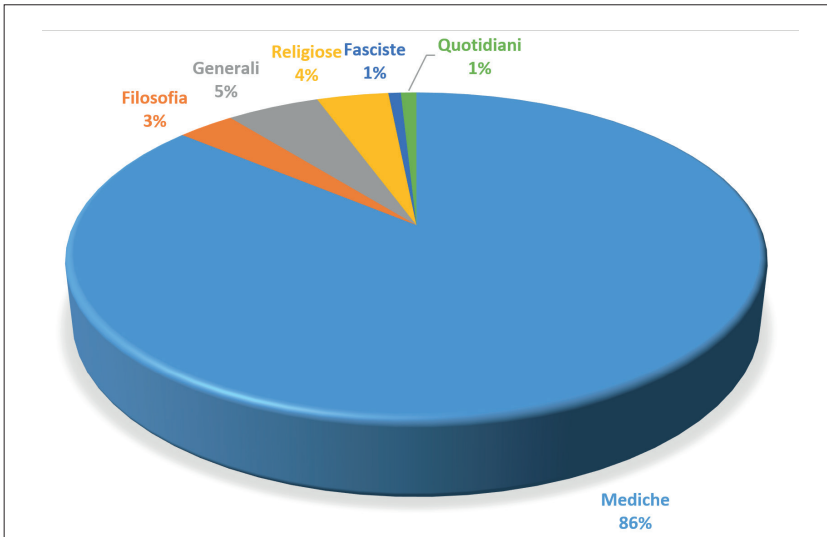


Figura 5 Suddivisione tipologie riviste del «corpus» per numero di parole

I testi letterari sono pochi. Se escludiamo la *Coscienza di Zeno* o qualche lirica di Saba (terminologicamente comunque poco importanti, se non per la divulgazione del termine stesso *psico-analisi*), possiamo segnalare due romanzi di Riccardo Bacchelli, *La città degli amanti* (1929) e *Oggi, domani e mai* (1932) in cui l'autore, critico e sprezzante nei confronti della psicanalisi, ne utilizza, banalizzandola ma divulgandola, la terminologia (per esempio *Es*, *inconscio*, *preconscio*, *repressione*). Abbiamo, inoltre, una traduzione italiana del 1931 del volume *L'anima che guarisce* di Stefan Zweig (1881-1942), entusiasta propugnatore della psicanalisi, la cui traduzione in italiano vide la collaborazione di Weiss.

Passando alle riviste (397 documenti), analizzando il peso delle varie tipologie ci rendiamo conto che, come prevedibile, le pubblicazioni che abbiamo ricompreso sotto l'etichetta «mediche» (284 documenti) rappresentano, per numero di parole, l'86% di quanto pubblicato (figura 5). Se di queste, poi, andiamo ad analizzare nello specifico quali ambiti disciplinari hanno maggiormente scritto di psicoanalisi (figura

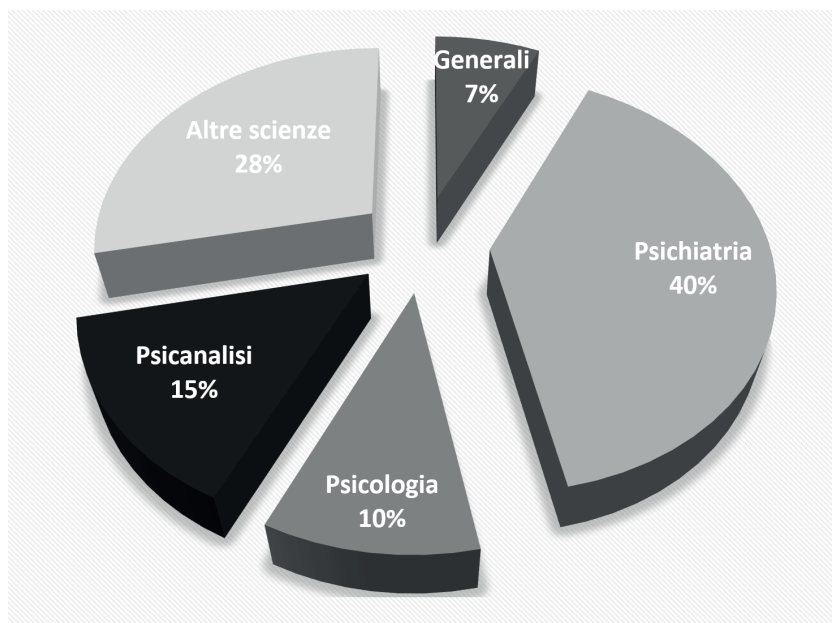


Figura 6 Suddivisione per branca di scienza delle riviste mediche

6) ci rendiamo ben conto che la psichiatria, nel bene e nel male, è stata la scienza che più si è occupata di psicoanalisi nel nostro periodo di riferimento, anche perché tra le riviste che abbiamo incluso in questa categoria c'è l'«Archivio» di Levi Bianchini che, per tutti gli anni Venti, e fino all'avvio della «Rivista italiana di psicoanalisi», si è occupato costantemente di psicanalisi, con ben 74 articoli (che da soli ne rappresentano – in termini assoluti di numero di parole – il 49% di quanto è stato scritto sulle riviste psichiatriche).

È importante evidenziare che a un'analisi qualitativa, rispetto alla massa di articoli delle riviste afferenti alla categoria «mediche», risultano spesso più interessanti le attestazioni presenti nelle riviste di altre categorie: gli articoli di riviste «religiose» (per es. «La civiltà cattolica»), «fasciste» («La difesa della razza») o filosofiche («La critica») hanno spesso dato luogo a diatribe e a censure che hanno avuto l'effetto di

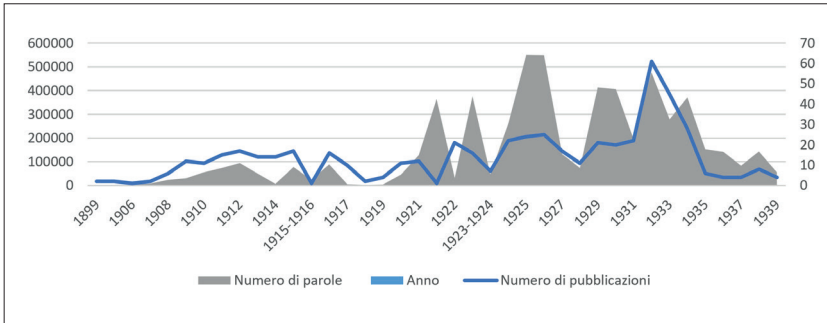


Figura 7 Numero di pubblicazioni e numero di parole per anno (1899-1939)

diffondere ancor più, al di fuori degli ambienti specialistici, la terminologia psicanalitica.

Da un punto di vista diacronico, il grafico combinato (figura 7) ci mostra la distribuzione del numero di parole e del numero di pubblicazioni, entrambe suddivise per anno. Possiamo così confermare quanto già detto: le prime pubblicazioni erano per lo più brevi note e recensioni. Il picco intorno al 1926 è causato dai già citati volumi di Morselli e dalle opere di De Sanctis comprese integralmente nel *corpus*. Si può notare come la weissiana «Rivista italiana di psicoanalisi» avesse prodotto una discreta mole di articoli (42 nel 1933) e che la sua prematura chiusura abbia rapidamente portato al silenzio.

L'organizzazione delle fonti in un database digitale permette, infine, di studiare e visualizzare a colpo d'occhio, la distribuzione geografica delle pubblicazioni, per cercare di capire le vie attraverso cui si è diffusa in Italia la psicanalisi. Per questo scopo è stata usata la piattaforma *Nodegoat*¹⁶, che permette di creare un database di fonti geolocalizzate (nel nostro caso mediante i luoghi di pubblicazioni delle riviste e delle monografie) e di visualizzare determinati rapporti relazionali tra queste. Nella figura 8 è possibile osservare l'immagine della situazione delle pubblicazioni in Italia al 1919. I cerchi indicano il luogo di

¹⁶ Cfr. BEE-KESSELS 2013.

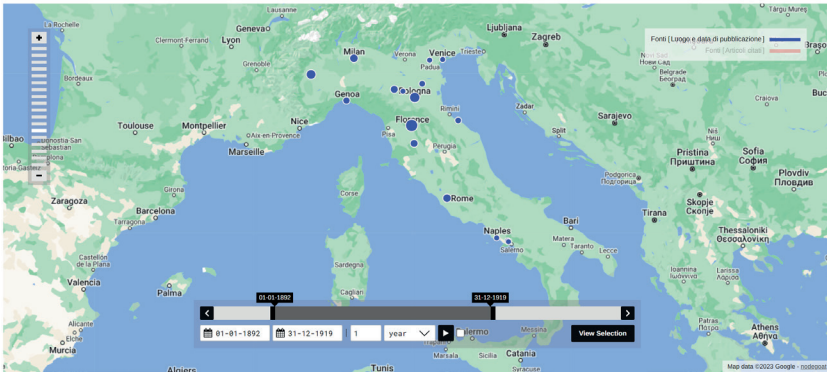


Figura 8 Geolocalizzazione al 31 12 1919 delle fonti del «corpus»

pubblicazione e maggiore è il diametro della circonferenza, maggiore è il numero delle fonti. Ci si accorge così che di psicanalisi si scriveva maggiormente nell'Italia centrale, in particolare a Firenze (che, oltre ad Andrea De Sarlo e il suo laboratorio di psicologia, poteva vantare la pionieristica rivista «Psiche») a Bologna (dove veniva pubblicata l'importante «Rivista di psicologia») e a Regio Emilia (dove veniva pubblicata la «Rivista sperimentale di Freniatria», la rivista ufficiale della società psichiatrica italiana).

Questo lavoro permette di correggere qualche “falso mito”, per esempio la centralità di Trieste. Questa città, infatti, fino al 1918 appartenente all'impero Austro-Ungarico, ha sì dato i natali a Weiss e ha visto l'entusiasmo di Saba (curato da Weiss) o il distaccato sarcasmo di Svevo (il cui cognato fu anch'esso curato da Weiss), ma, soprattutto negli ambienti medici, fu così ostile alla disciplina da costringere lo stesso Weiss a trasferirsi a Roma¹⁷.

17 Per gli anni triestini di Weiss cfr. *CORSA 2018*, mentre per il periodo romano, per approfondimenti sulla *Società psicoanalitica italiana* e per le attenzioni che il regime fascista dedicò alla psicoanalisi cfr. *CORSA 2017*.

5. Possibili sviluppi futuri

Per quanto riguarda i rapporti tra la lessicografia e la psicanalisi, come abbiamo già detto, questo lavoro può servire a correggere alcune “sviste” e “rimozioni” da parte di repertori troppo “letterariocentrici”. Il letterariocentrismo può essere anche interessante per vedere il grado di penetrazione nella lingua non specialistica di certa terminologia, ma da un punto di vista delle attestazioni crea non pochi problemi. Spesso si ignorano autori importanti, psichiatri, psicologi, medici, che per primi hanno contribuito a creare, mediante prestiti, calchi o il conio di neologismi, la terminologia psicanalitica (lo stesso discorso si potrebbe fare per la psichiatria). O questi autori non sono citati o sono usati solo occasionalmente (per esempio il *GDLI* inizia a citare Morselli solo a partire dalla lettera P, come se giunti a *psichiatria*, *psicologia* e *psicoanalisi* gli autori si fossero resi conto della necessità di spogliare psichiatri, psicologi e psicanalisti¹⁸. Qual è la conseguenza? Come detto, molte delle attestazioni presenti nel *GDLI*, ma anche, in parte, nel *GRADIT*, sono basate su fonti non specifiche e fanno sì che una terminologia che, come speriamo di dimostrare al termine di questo lavoro, si era formata già nei primi tre decenni del secolo, sia “snocciolata” per quasi tutto il secolo.

Riportiamo un esempio che mostra quanto sia possibile retrodatare quanto proposto da *GDLI* (figura 9) e corretto solo in parte, per esempio dal *GRADIT* (figura 10).

Naturalmente questa è solo la parte quantitativa del lavoro. Come detto, la maggiore ambizione è ricostruire la storia delle principali parole della psicanalisi in Italia. Questo lavoro non solo permette di riscoprire grandi figure che rischiano l'oblio (basti pensare che personaggi di primo piano come Roberto Assagioli o lo stesso Edoardo Weiss non hanno una scheda sul *Dizionario Biografico degli Italiani*), ma anche di ricreare un microcosmo di rapporti, stroncature e diatribe, piccole e grandi tragedie, che sono andate creandosi in un periodo particolarmente travagliato della storia d'Italia. La psicanalisi è per prima cosa il

¹⁸ Nel Supplemento 2004 e 2009 questo viene corretto solo in parte.

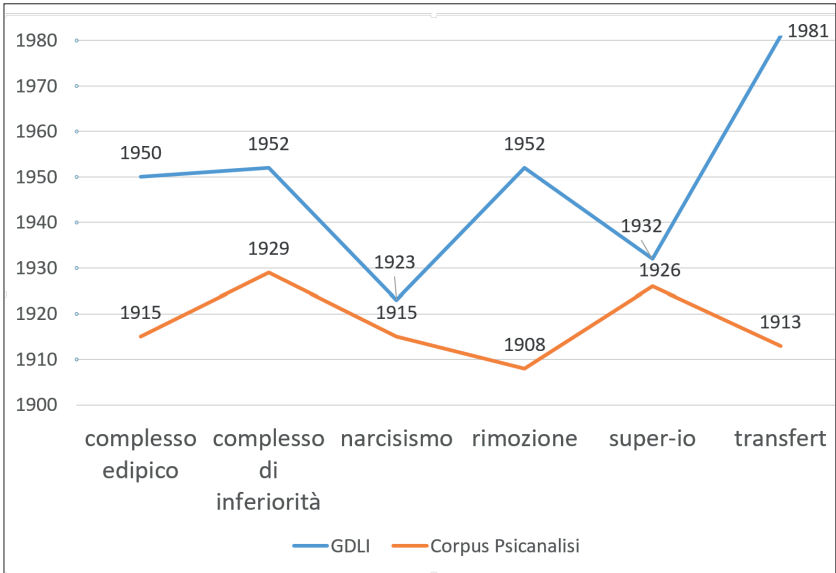


Figura 9 Confronto delle attestazioni di alcuni termini psicanalitici fornite dal GDLI con quelle estratte dal nostro «corpus»

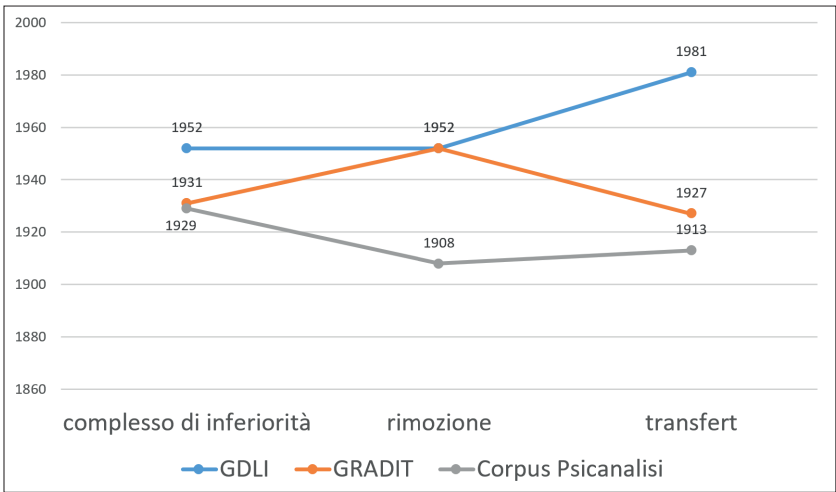


Figura 10 Confronto delle attestazioni di alcuni termini psicanalitici fornite da GDLI e GRADIT con quelle estratte dal nostro «corpus»

suo linguaggio (mezzo per definire e per operare) ed è interessante vedere come proprio a causa di alcune sue parole (*Libido*, spesso in mala fede confusa con *libidine*, è solo il caso più eclatante) sia stata attaccata da un fuoco incrociato che ha visto spesso un fronte trasversale (cattolici, idealisti e fascisti) cospirare contro la sua piena affermazione nel nostro Paese, con conseguenze ancora tangibili.

Un ulteriore auspicio è quello di poter rendere consultabile e pubblico il *corpus* che abbiamo costruito, che riteniamo utile non solo per ulteriori studi linguistici, ma anche per gli storici e per gli stessi psicanalisti interessati a scavare nel proprio passato, per ritrovare precursori spesso indebitamente “rimossi”.

Bibliografia

- ANGELUCCI 1929 = Arnaldo Angelucci, *Psicanalisi e sublimazione nell'arte*, Napoli, Casa editrice libraria Vittorio Idelson, 1929.
- BECCARIA 2004 = GIAN LUIGI BECCARIA, *Dizionario di linguistica e di filologia, metrica e retorica*, Torino, Einaudi, 2004.
- BREE-KESSELS 2013 = PIM VAN BREE-GEERT KESSELS, *Nodegoat: a web-based data management, network analysis & visualisation environment*, 2013, <<http://nodegoat.net>>.
- CAROTENUTO 1977 = ALDO CAROTENUTO, *Jung e la cultura italiana. Dal 1903 a oggi, una storia delle idee, dei libri e delle riviste che hanno contribuito a diffondere in Italia il pensiero di uno dei più originali psicologi contemporanei*, Roma, Astro-labio, 1977.
- CAROTENUTO 2002² = ALDO CAROTENUTO, *Breve Storia della psicanalisi*, Milano, Bompiani, 2002².
- CIMINO-DAZZI 1998 = *La psicologia in Italia. I protagonisti, i problemi scientifici, filosofici e istituzionali (1870-1945)*, a cura di Guido Cimino e Nino Dazzi, Milano, LED, 1998.
- CORSA 2017 = RITA CORSA, *Vanda Shrenger Weiss. La prima psicoanalista in Italia. La psicoanalisi a Roma in epoca fascista*, Roma, Alpes, 2017.
- CORSA 2018 = RITA CORSA, *Edoardo Weiss a Trieste con Freud. Alle origini della psicoanalisi italiana. Le ricerche di Nathan, Bartol e Veneziani*, Roma, Alpes, 2018.
- DAVID 1963 = MICHEL DAVID, *L'idealismo crociano e la psicoanalisi*, in «Rivista di psicoanalisi», IX, 1963, pp. 189-234.

Stefano Miani

- DAVID 1964 = MICHEL DAVID, *La lingua della psicanalisi nella lessicografia italiana*, in «Lingua nostra», XXV, 1964, pp. 19-24.
- DAVID 1990 = MICHEL DAVID, *La psicoanalisi nella cultura italiana*, Torino, Bollati Boringhieri, 1990 [1966¹].
- DE SANCTIS 1925 = SANTE DE SANCTIS, *Neuropsichiatria infantile. Patologia e diagnostica*, Roma, Stock, 1925.
- DE SANCTIS 1929-1930 = SANTE DE SANCTIS, *Psicologia sperimentale*, Roma, Stock, 2 voll., 1929-1930.
- FREUD 1915 = SIGMUND FREUD, *Sulla psicoanalisi. Cinque conferenze tenute nel settembre 1909 alla Clark University di Worcester Mass. in occasione del 20° anniversario di fondazione. Prima traduzione italiana sulla seconda edizione tedesca del 1912 del prof. M. Levi Bianchini*, Nocera Superiore, Il Manicomio, 1915.
- FREUD 1971 = SIGMUND FREUD, *Opere 1899. L'interpretazione dei sogni*, Torino, Bollati Boringhieri, 1971.
- GUALDO 2016 = RICCARDO GUALDO, *Linguaggi specialistici e settoriali*, in LUBELLO 2016, pp. 371-395.
- GUALDO 2021 = *Introduzione ai linguaggi specialistici*, Roma, Carocci, 2021.
- GUALDO-TELVE 2011 = RICCARDO GUALDO, STEFANO TELVE, *Linguaggi specialistici dell'italiano*, Roma, Carocci, 2011.
- KILGARRIFF ET AL. 2004 = ADAM KILGARRIFF, PAVEL RYCHLÝ, PAVEL SMRZ, DAVID TUGWELL, *The Sketch Engine*, in *Proceedings of the XI Euralex International Congress*, July 6-10, 2004, a cura di G. Williams e S. Vessier, Lorient, France, 2004, pp. 105-116.
- LUBELLO 2016 = *Manuale di Linguistica italiana*, a cura di Sergio Lubello, Berlin/Boston, De Gruyter, 2016.
- MARHABA 2003 = SADI MARHABA, *Lineamenti della psicologia italiana (1870-1945)*, Firenze, Giunti, 2003.
- MIANI 2022 = STEFANO MIANI, *Per uno studio della terminologia della psicanalisi in Italia: atti mancati e lapsus freudiano*, in «Lingua Nostra», LXXXIII, 2022, pp. 110-18.
- MUSATTI 1991 = CESARE MUSATTI, *Mia sorella gemella la psicoanalisi*, Pordenone, Edizioni Studio Tesi, 1991.
- RANCHETTI 2002 = MICHELE RANCHETTI, *Le difficili origini della psicoanalisi*, in «Psicoterapia e scienze umane», XXXVI, 2002, 2, pp. 5-9.
- TISSI 1929 = SILVIO TISSI, *La psicanalisi scienza dell'Io o del mistero-problema psichico con saggi di analisi psichica su drammi di Pirandello, Shakespeare, Ibsen, Tolstoj e Shaw*, Milano, Hoepli, 1929.
- WEISS 1931 = EDOARDO WEISS, *Elementi di psicoanalisi*, Milano, Hoepli, 1931.

WEISS 1971 = EDOARDO WEISS, *Sigmund Freud come consulente*, Roma, Astrolabio, 1971.

Riassunto L'intervento descrive lo stato dell'arte (alla fine del secondo anno) del progetto di dottorato *Formazione del lessico italiano della psicanalisi*. La ricerca nasce dalla constatazione dell'assenza di un tale lavoro nel panorama degli studi storico-linguistici italiani e dalla verifica di imprecisioni e lacune nei principali repertori lessicografici dovuti, per lo più, al ricorso a fonti letterarie o, comunque, non specifiche, che rischia di diffrangere lungo tutto l'arco del secolo una terminologia formatasi, in gran parte, entro i primi quattro decenni del Novecento. Viene presentato il *corpus* di testi, monografie e articoli di argomento psicanalitico pubblicati in Italia tra il 1899 e il 1939, su cui sono stati condotti gli spogli lessicali. Vengono esposti i criteri di digitalizzazione del materiale raccolto e discussi alcuni dei dati quantitativi (certo ancora provvisori) che, grazie alla trattazione con *Sketch Engine* e *Nodegoat*, è già possibile ricavare.

Abstract The speech describes the state of the art (at the end of the second year) of the doctoral project *Formazione del lessico italiano della psicanalisi* (*Formation of the Italian lexicon of psychoanalysis*). The research arises from the observation of the absence of such work in the panorama of Italian historical-linguistic studies and from the verification of inaccuracies and gaps in the main lexicographic repertoires due, mostly, to the use of literary or, in any case, non-specific sources, which it risks diffracting throughout the century a terminology formed, largely, within the first four decades of the twentieth century. The corpus of texts, monographs and articles on psychoanalytic topics published in Italy between 1899 and 1939, on which the lexical searches were conducted, is presented. The criteria for digitizing the collected material are exposed and some of the quantitative data (certainly still provisional) which, thanks to the treatment with *Sketch Engine* and *Nodegoat*, is already possible to obtain, are discussed.

Esempi di interferenza tra lingua comune e lingua specialistica: note su alcuni termini artistico-architettonici e meccanici

Matteo Mazzone

1. Introduzione

Sin dalla prima metà degli anni Settanta la letteratura scientifica ha offerto, sempre più sistematicamente, vari contributi allo studio e alla definizione delle differenze che intercorrono tra lingue specialistiche e lingue settoriali¹: da monografie ad atti di convegno, sino ad articoli e a saggi *ad hoc* volti ad analizzare le caratteristiche che hanno portato alla formazione di terminologie relative a specifiche varietà di lingua (come, ad esempio, la lingua dell'architettura, quella della matematica, quella della botanica, ecc.). Certo è che tra le due varietà esiste un discrimine netto, riguardante «l'alto grado di specializzazione»² delle lingue specialistiche o speciali³, che manca invece a quelle settoriali. La specializzazione del lessico tecnico delle prime e la conseguente difficoltà interpretativa e comprensiva per chi non appartenga alla cerchia di coloro che si occupano dell'ambito a cui si riferisce tale nomenclatura

1 Sulle lingue specialistiche e su quelle settoriali, si rimanda in generale a BECCARIA 1973; GOTTI 1991; CORTELAZZO 1994; SOBRERO-MIGLIETTA 2006. Da ultimo, si vedano i lavori di VISCONTI-MANFREDINI-COVERI 2020, GUALDO-TELVE 2021 e GUALDO 2021.

2 Cfr. SOBRERO-MIGLIETTA 2006, p. 137.

3 Di "lingua speciale" parla, per primo, Giacomo Devoto: cfr. DEVOTO 1939a, pp. 17-21; ID. 1939b, pp. 114-121. Per approfondimenti sulle diverse denominazioni, da *lingue speciali* a *tecnoletti* o *microlingue*, rimando a CAVAGNOLI 2007, pp. 13-17.

ra costituiscono il divario più forte rispetto al lessico delle lingue settoriali, che risulta invece facilmente comprensibile anche ai non addetti ai lavori.

Di seguito riporto la definizione che Michele A. Cortelazzo fornisce per *lingua speciale*, riprendendo alcune delle considerazioni già espresse da Gaetano Berruto nel 1974⁴:

[...] varietà funzionale di una lingua naturale, dipendente da un settore di conoscenze o da una sfera di attività specialistici, utilizzata, nella sua interezza, da un gruppo di parlanti più ristretto della totalità dei parlanti la lingua di cui quella speciale è una varietà, per soddisfare i bisogni comunicativi (in primo luogo quelli referenziali) di quel settore specialistico; la lingua speciale è costituita a livello lessicale da una serie di corrispondenze aggiuntive rispetto a quelle generali e comuni della lingua e a quello morfosintattico da un insieme di selezioni, ricorrenti con regolarità, all'interno dell'inventario di forme disponibili nella lingua⁵.

La definizione evidenzia quali siano gli aspetti essenziali delle lingue specialistiche e quali siano i processi lessicali che determinano lo sviluppo e la formazione del loro vocabolario tecnico. Esse sono considerate come una varietà funzionale e diafasica, utilizzata da un gruppo ristretto di parlanti, perlopiù tecnici, esperti del settore o addetti ai lavori. Esse, inoltre, si costituiscono attraverso procedimenti basati sulla risemantizzazione (talvolta detta anche *specializzazione*, *tecnificazione* o *terminologizzazione*⁶) di parole o di espressioni appartenenti alla lingua comune⁷, alle quali vengono assegnate informazioni semanti-

⁴ Cfr. BERRUTO 1974, p. 68.

⁵ Cfr. CORTELAZZO 1994, p. 8.

⁶ Cfr. *ivi*, p. 121.

⁷ Altri sono i procedimenti utilizzati per la creazione di lessici specialistici, come la metafora lessicale, il *transfert* o travaso lessicale, l'uso di prestiti e calchi da lingue straniere, la creazione di neoformazioni derivazionali o compositivi con elementi greci o latini e, infine, le neologie (cfr. CORTELAZZO 1994, pp. 13-14 e GUALDO 2021, pp. 115-167). Inoltre, arricchiscono il lessico delle lingue specialistiche i cosid-

che aggiuntive, volte a designare in maniera univoca nozioni e concetti propri di quel particolare ambito tecnico.

In questo articolo mi concentrerò sull'aspetto relativo all'interferenza tra lingua comune e lingua specialistica, ossia sulla rideterminazione semantica di voci d'uso quotidiano in chiave tecnica per la denominazione di peculiari elementi artistico-architettonici e meccanici. Nello specifico, analizzerò alcune forme tratte dallo spoglio lessicale⁸ che ho effettuato da quella che, a oggi, si può considerare come la seconda traduzione del *De architectura* di Vitruvio⁹, contenuta nel manoscritto *Ottoboniano latino 1653*¹⁰ e databile tra il 1504 e il 1511¹¹.

detti *tecnicismi collaterali*, per i quali si rimanda più dettagliatamente a SERIANNI 2005, pp. 127-159.

- 8 Le voci sono ricavate dalla lista degli oltre 930 lemmi che costituiscono il *Glossario* della terminologia artistico-architettonica, meccanica e idraulica della traduzione ottoboniana e rientra nel mio lavoro di tesi dottorale discussa nell'aprile 2022 presso l'Università degli Studi di Firenze (cfr. MAZZONE 2018-2021, voll. II-III). La lista è l'esito di una preliminare realizzazione di una versione elettronica della traduzione, processata attraverso lo strumento di analisi *StanfordNLP* (<https://stanfordnlp.github.io/stanfordnlp/>), con il quale è stato possibile tokenizzare il testo ed estrapolare una tabella contenente l'elenco delle forme affiancate dalla loro singola frequenza.
- 9 Per una ricognizione sulle traduzioni quattro-cinquecentesche del trattato vitruviano, cfr. BIFFI 2005, pp. 151-153. Sulla precisazione circa il primato di Francesco di Giorgio Martini relativo alla sua traduzione organica del *De architectura*, si rimanda a BIFFI 2022, pp. 26-27 e relative note.
- 10 Il manoscritto, consultabile in rete all'indirizzo https://digi.vatlib.it/view/MSS_Ott.lat.1653, è stato oggetto della mia tesi di laurea magistrale in "Filologia moderna" discussa presso l'Università degli Studi di Firenze (cfr. MAZZONE 2017-2018). L'edizione del testo, da me curata, è di prossima pubblicazione. Per alcune notizie sul metodo di traduzione adottato dall'estensore dell'*Ottoboniano* e per un primo sondaggio sul lessico artistico e architettonico in esso contenuto, rimando a MAZZONE 2022a e ID. 2022b, oltre alle analisi e alle conclusioni già espresse in BIFFI 2009, pp. 75-78.
- 11 La datazione del manoscritto si deve agli studi di BIFFI (1999, pp. 60-61 e nota 96; ivi, p. 81 e nota 164; ID. 2009, p. 74 e ss.). Si veda anche MAZZONE 2021, pp. 39-40, nota 10.

Il volgarizzamento, scritto in fiorentino quattro-cinquecentesco¹², è conservato presso la Biblioteca Apostolica Vaticana di Roma¹³ ed è tutt'oggi da ritenere adespoto¹⁴.

2. Interferenza tra lingua comune e lingua specialistica: alcuni casi tratti dalla traduzione ottoboniana

La risemantizzazione di voci appartenenti alla lingua comune è uno dei processi di creazione di lessico tecnico adottati da coloro che per primi, tra la fine del Quattrocento e per tutto il Cinquecento (agli albori, dunque, della formazione di una nomenclatura dell'architettura¹⁵ e

12 La fiorentinità linguistica è stata dimostrata da BIFFI (2009, pp. 73-74 e relative note) e da me confermata attraverso uno studio linguistico sistematico, che ha preso in esame non solo i tratti tipici del fiorentino argenteo, ma anche quelli più generalmente fiorentini e che lo contraddistinguono dagli altri volgari della Toscana (cfr. MAZZONE 2021).

13 Di questa traduzione esiste una copia contenuta nel manoscritto *Italien 472* (Parigi, Biblioteca Nazionale di Francia). Il rapporto tra i due testi è stato indagato da Marco Biffi, il quale ha dimostrato che il ms. *Italien* è *codex descriptus* dell'*Ottoboniano* (cfr. BIFFI 1999, p. 110 e nota 270; ID. 2009, pp. 72-73 e nota 33; per altri dati che confermano la parentela tra i due testi, rimando a MAZZONE 2018-2021, vol. I, pp. XXVI-LVII). Il codice, digitalizzato sul sito *Gallica* all'indirizzo <https://bit.ly/3qF4FdT>, è stato oggetto di una parte del lavoro di tesi dottorale, che ne ha previsto l'analisi, l'edizione e il commento linguistico. Esso è databile, per perizia paleografica, non oltre gli anni '40 del Cinquecento, e la lingua si caratterizza per un registro di base toscana con inserti linguistici dell'area umbra e mediana (ivi, vol. I, pp. 318-546).

14 Per alcune notizie codicologiche, cfr. MAZZONE 2021, pp. 41-43.

15 Si deve a Giovanni Nencioni l'inizio di uno studio sistematico sulla formazione della lingua dell'architettura (cfr. NENCIONI 1995, poi ID. 2000). In tempi più recenti, meritano di essere ricordati alcuni scritti di Marco Biffi dedicati all'analisi della formazione del lessico architettonico italiano e alla contestualizzazione storica dell'opera di Vitruvio in rapporto alle traduzioni del *De Architectura* (si vedano, almeno, BIFFI 1999; ID. 2001; ID. 2003; ID. 2005; ID. 2006; ID. 2017; ID. 2019; ID. 2022). La natura specialistica della lingua dell'architettura e la sua posizione nello spazio linguistico sono state studiate da Patrizia Bellucci, che ha esaminato

della meccanica applicata¹⁶), si sono imbattuti nell'oggettiva difficoltà di costruire e di determinare quale fosse la tipologia di lingua specialistica di stampo sovraregionale che potesse essere al tempo stesso funzionale e comprensibile per tutti. È questa una delle sfide più importanti che impegna i primi traduttori di testi tecnici – compreso l'enigmatico *De architectura* di Vitruvio¹⁷ – i quali, spesso illetterati o poco familiari con il latino, solo successivamente imparato da autodidatti, appartengono a quello che Carlo Maccagni ha definito come lo «strato culturale intermedio»¹⁸ composto da tecnici, esperti e operai del settore. Lo scopo non è solo quello di rispondere all'istanza teorico-esegetica di comprensione della fonte latina di partenza, mediante la translitterazione passiva e/o la traduzione interpretativa del testo vitruviano (il cosiddetto *sapere che*, ossia l'insieme delle nozioni astratte e concettuali), ma, soprattutto, quello di acquisire esperienza e perfezionare la competenza pratica relativa alla sfera produttiva (il cosiddetto *sapere come*, basato sul conseguimento di determinate abilità operative).

È dovuto proprio al perseguimento di questo secondo fine lo stretto legame che intercorre tra lingua comune e lingua specialistica. All'interno delle numerose precisazioni semasiologiche – costituite da glosse e annotazioni di lunghezza variabile che spiegano i latinismi e

la lingua dell'architettura con particolare riferimento a quella contemporanea. Le conclusioni a cui giunge la studiosa sono altresì valide ed estendibili in diacronia (cfr. BELLUCCI 1997).

- 16** I più significativi studi relativi alla nascita del lessico della meccanica applicata risalgono ai primi anni Ottanta e si devono a Paola Manni (cfr. MANNI 1980 ed EAD. 1981). Il problema relativo all'importanza circa l'individuazione di un *corpus* di testi tecnici cinquecenteschi – considerato come un ricco serbatoio a fondamento della terminologia della meccanica applicata – era stato già affrontato, alla fine degli anni Settanta, dallo storico della scienza Paolo Galluzzi (cfr. GALLUZZI 1979).
- 17** Sulla lingua del trattato vitruviano si vedano almeno: DEVOTO 1940, pp. 244-245; CALLEBAT 1982; ROMANO 1997; BOSAZZI 2000. Per l'interpretazione semantica del multidisciplinare vocabolario tecnico di Vitruvio, si rimanda all'opera lessicografica di CALLEBAT-FLEURY 1995.
- 18** Cfr. MACCAGNI 1996, p. 280.

i grecismi ereditati da Vitruvio¹⁹ – l'estensore della traduzione ottoniana, servendosi anche del processo di risemantizzazione, connota tecnicamente voci della lingua comune, le quali si funzionalizzano come termini per denotare e designare²⁰ ora strutture architettoniche o specifici elementi edilizi, ora dispositivi meccanici o semplici parti di essi.

Per ciascuna voce di seguito analizzata, fornisco alcune informazioni che ricordano i campi fondamentali di una scheda lessicografica. Preciso che per tutti i tecnicismi è valida l'etichetta di prima e unica attestazione, in quanto essi sono considerabili come *hapax* della letteratura scientifica artistico-architettonica e meccanica. Tale etichetta, però, non ha un valore assoluto, ma è condizionata dall'insieme dei riscontri che ho avuto modo di reperire nel *corpus* dei testi di riferimento spogliato in fase di redazione del *Glossario* per la compilazione del campo delle corrispondenze linguistiche²¹.

1. Lemma. Ciascuna entrata lessicale, posizionata al centro della pagina, è trascritta in carattere maiuscoletto secondo la grafia moderna²². Se necessario, chiarisco in corsivo le eventuali varianti di natura grafica, risultanti dalla ricognizione totale delle occorrenze della voce in tutto il testo.
2. Definizione. Riporto il significato di ogni termine attraverso una definizione ricostruita *ex novo*, tenendo conto del contesto in cui la

¹⁹ Su questo procedimento, iniziato da Francesco di Giorgio Martini e relativo alla formazione di un lessico nazionale dell'architettura attraverso la giustapposizione di lessico vitruviano con quello volgare e della tradizione mutuato dalle botteghe medioevali toscano-senesi, cfr. BIFFI 2001; ID. 2006, pp. 87-88.

²⁰ A tal proposito si vedano le considerazioni di PORRO (1973, p. 187): «I linguaggi della scienza e della tecnica, nella misura in cui rideterminano analiticamente il legame tra significante e significato, valorizzano al più alto livello quella che si suole definire funzione *denotativa* o *designativa*».

²¹ Per i riferimenti bibliografici relativi al *corpus* di testi spogliato, si veda MAZZONE 2018-2021, vol. II, pp. XIV-XXVIII.

²² Racchiudo tra parentesi quadre le forme al singolare non attestate nella traduzione.

- forma è collocata. Ciascun significato, segnalato tra virgolette alte scempie, è preceduto dall'indicazione della categoria grammaticale alla quale il tecnicismo appartiene. Quando vi sono più accezioni per uno stesso lemma, ho numerato le definizioni in ordine di importanza (per es.: **1.** '...'; **2.** '...'; ecc.). Per specificare se si tratti di un sostantivo maschile o femminile, utilizzo le abbreviazioni *s.m.* o *s.f.*
3. Esempi. Riproduco un brano significativo tratto dalla traduzione ottoboniana²³ che attesti l'utilizzo e il significato del tecnicismo. In ogni esempio sottolineo la forma analizzata; con un asterisco iniziale evidenzio gli esempi che accolgono una riflessione di carattere metalinguistico (come una glossa esplicativa che abbia intento definitorio). Segue la sigla *Ottob.* (Ottoboniano) e i rinvii numerici ai fogli del manoscritto (f./ff.), con l'indicazione abbreviata del *recto* o del *verso*. Nel caso in cui l'esempio sia stato tratto dalla fine di un foglio e dall'inizio di quello successivo di differente numerazione, il rinvio è indicato, per esempio, come: ff. 63v-64r. Se è stato estratto dalla fine del *recto* e dall'inizio del *verso* dello stesso foglio, scrivo, ad esempio: ff. 94r-v.
 4. Frequenza totale e *Index locorum*. Indico in ordine alfabetico le occorrenze totali del tecnicismo e per ciascuna di esse riproduco l'*Index locorum*, cioè l'insieme delle singole forme con il rinvio a tutte i fogli del manoscritto. Se una forma è presente più volte all'interno del medesimo *recto* o *verso*, ne do segnalazione con un numero arabo posto tra parentesi tonde, come nell'esempio: f. 29v (2).

23 I brani citati sono stati riprodotti secondo i criteri di trascrizione basati sulle indicazioni fornite da Arrigo Castellani (cfr. CASTELLANI 1982, pp. XVI-XIX), poi adattati da Marco Biffi per i testi quattro-cinquecenteschi e, in particolare, per l'edizione della *Traduzione* vitruviana di Francesco di Giorgio Martini (cfr. DI GIORGIO MARTINI/ BIFFI 2002, pp. CXIX-CXXIV). Ne ricordo solo alcuni: ogni brano, proposto con punteggiatura e accentazione moderna, è riprodotto fedelmente secondo l'originale per quanto riguarda i grafemi; tuttavia, si differenziano i due valori fonetici del grafema unico *u/v* e si trascrive *j* con *i*. Moderni sono sia l'inserimento delle maiuscole che la divisione logica delle parole. Tra parentesi tonde si sciolgono le abbreviazioni *ch(e)*, *p(er)*, *p(ar)*, ecc.; il *titulus* per le nasali è sciolto con (*n*) o (*m*), mentre la nota tachigrafica è resa con (*et*).

5. Lemmi correlati. Se presenti, fornisco quei lemmi inclusi nel *Glossario* che hanno la stessa derivazione etimologica o che sono connessi dal punto di vista semantico.

2.1. I termini artistico-architettonici

Riporto solo quattro casi – tratti dal *Glossario* dei termini artistico-architettonici, meccanici e idraulici della traduzione ottoboniana²⁴ – di voci appartenenti alla lingua comune (*cerchio*, *picciolo*, *quadro* e *ristretto*), le quali hanno subito, durante la fase di traduzione, commento e interpretazione del testo vitruviano, un processo di risemantizzazione in chiave specialistica da parte dell'anonimo estensore del volgarizzamento vaticano.

[PICCIOLO]

s.m. 'elemento che costituisce la struttura di una scala', 'scalino, gradino'

– *Ottob. f. 45v*: «Le gradatio(n)e delle scale, cioè e piccioli o scaglioni, si debbono dirizzare (et) fare i(n)tra a' cunei, cioè i(n)tra queglii ordini di sedie, e i(n)tra le sedie diri(m)petto agli a(n)guli de' quadrati alla prima preci(n)tione, cioè al p(r)imo p(re)cinto del theatro; dalla preci(n)tione i(n)fra quello anchora quelle del mezzo sieno fatte a dirittura, e i(n) so(m)ma ogni volta ch(e) si preci(n)gono se(m)pre crescono altrettanta(n)to».

Frequenza totale: 1

piccioli freq. = 1; 45v.

Prima e unica attestazione

²⁴ Cfr. *supra*, nota 8.

[QUADRO]

s.m. 'settore di muro quadrangolare posto al di sopra dell'architrave e compreso fra i due triglifi', 'metopa'

– *Ottob. f. 33v*: «Sopra l'architrave si debbe mettere e triglifi colle sue metoffe, cioè co(n) quegli quadri ch(e) sono allato a' triglifi, cioè ch(e) sara(n)no alti un mezzo modulo (et) larghi i(n) fronte uno modulo...».

Frequenza totale: 1

quadri freq. = 1; 33v.

Lemmi correlati: *metoche*, [*metoppa*]/ *methopha*/ [*metoffa*], *piano* (4.).

Prima e unica attestazione

RISTRETTO

grafia *-ct-*

s.m. 'rastremazione, graduale riduzione del diametro superiore di una colonna', 'restringimento della parte superiore di un elemento architettonico'

– **Ottob. f. 26v*: «(Et) e ristretti, cioè e ristignime(n)ti delle colo(n)ne da so(m)mo dallo hipotrach(e)le, cioè dal pianuzzo della colo(n)na, si debbono fare così, ch(e) se la colo(n)na sia almeno circa q(ui)ndici piedi, la grossezza da basso si debbe dividere i(n) sei p(ar)te e di quelle ci(n)que p(ar)te si facci quella da so(m)mo».

Frequenza totale: 6

ristrecto freq. = 1; 28v.

ristretti freq. = 2; 26v (2).

ristretto freq. = 3; 26v (2), 28v.

Matteo Mazzone

Lemmi correlati: *co(n)tractura*, [*ristrettura*], [*ristrignime(n)to*]/ *restrignime(n)to*.

Prima e unica attestazione

TO(N)DO

s.m. ‘elemento piano di forma rotonda, posto all’estremità inferiore o superiore di una colonna’, ‘spira, toro’

– *Ottob. f. 28r*: «(Et) lo aggetto della spera, cioè del to(n)do, sarà l’octava p(ar)te e la sextadecima p(ar)te della grossezza della colo(n)na».

Frequenza totale: 3

to(n)di freq. = 2; 27v, 29v.

to(n)do freq. = 1; 28r.

Lemmi correlati: *base/ basa* (1.), *cerchiellino*, *cerchio* (1.), *collarino*, *colo(n)na²/ colu(n)na* (1.), *ruota* (4.), *spira/ spera*, *thoro*, *trochilon/ trochilo*.

Prima e unica attestazione

2.2 I termini della meccanica applicata

Anche per i tecnicismi della meccanica applicata ho estrapolato quattro lemmi dal *Glossario* del volgarizzamento ottoboniano²⁵. Si tratta di unità lessicali appartenenti alla lingua comune (*addentellato*, *bastoncello*, *campana* e *codetta*) le quali hanno subito un processo di specializzazione semantica per la denominazione di taluni ingranaggi, congegni o parti di essi.

²⁵ Cfr. *supra*, nota 8.

ADE(N)TELLATO

s.m. 'elemento metallico aguzzo di un ingranaggio'

– *Ottob.* f. 95v: «(Et) sopra al piano, colla medesima ragio(n)e, si debbe collocare (et) mettere uno ade(n)tellato ri(n)chiuso nell'altro loculame(n)to, co' de(n)ti co(n)venie(n)ti al denticulo, el quale sarà fitto nel lato del seco(n)do ty(m)pano (et) i(n) quel ty(m)pano a(n)chora sieno fatti fori qua(n)te migla possa caminare i(n) un dì col carro».

Frequenza totale: 1

ade(n)tellato freq. = 1; 95v.

Lemmi correlati: *de(n)te* (1.), *de(n)tello* (2.), *de(n)ticello*, *de(n)ticulo*.

Prima e unica attestazione

BASTO(N)CELLO

1. s.m., 'perno di estremità di un albero'

– **Ottob.* f. 90r: «Quattro scapi, cioè base di materia trie(n)tale cioè di metallo, dua ne piglò p(er) traverso i(n)terposti, qua(n)to era la lu(n)ghezza della basa, (et) co(m)mese gli i(n)sieme (et) i(m)pio(m)bò cioè gli confichò (et) fermogli col pio(m)bo codace di ferro, cioè basto(n)cegli o legature di ferro a similitudine di code...».

Frequenza totale: 1

basto(n)cegli freq. = 1; 90r.

Lemmi correlati: *codace*.

Prima e unica attestazione

Matteo Mazzone

2. s.m. 'pistone, stantuffo'

– *Ottob. f. 93v*: «I(n) modo ch(e) dalla ba(n)da di sopra ne' trogolini gli emboli maschuli puliti a torno – emboli è vocabolo greco el quale i(n) lingua n(ost)ra significa getto, ma qui vuol dire quel basto(n)cello o simile a quello che si caccia nello schizzatoio...».

Frequenza totale: 1

basto(n)cello freq. = 1; 93v.

Lemmi correlati: [*embolo*], *schizzatoio*.

Prima e unica attestazione

CA(M)PANA

s.f. 'dispositivo della macchina idraulica che consiste in un imbuto rovesciato simile a una campana'; 'infundibulo'

– *Ottob. ff. 94r-v*: «Et i(n)tra l'arca cioè quella piastra o vero piano dove l'aqua è sostenuta, è i(n) tal maniera come i(n) uno i(n)fu(n)dibulo – infu(n)dibulo è uno cop(er)chio – a modo ch(e) uno cop(er)chio della ca(m)pana da stillare, el quale dalla ba(n)da di sopra e dadi alti circa a 3 dita, esse(n)do messi (et) posti di sotto, sospe(n)do (et) te(n)gono alto lo spatium da basso (et) intra e labbri, cioè gli orli da basso, el fo(n)do del phigeo (et) dell'altare – phigeo p(ro)p(r)io significa poggio o ara, poi el pianuzzo, ch(e) suono luoghi nella machina hydraulica così detti p(er) similitudine».

Frequenza totale: 1

ca(m)pana freq. = 1; 94v.

Prima e unica attestazione

[CODETTA]

s.f. 'prolungamento metallico, dente marginale di un organo o della corona di un ingranaggio'

– *Ottob.* ff. 95r-v: «(Et) dalla ba(n)da di sopra apresso alla cassa del charro co(n)fichasi fermame(n)te el loculame(n)to, el quale è come una cassetta ch(e) ha più sta(n)zoline fatto a similitudine d'una scarsella ch(e) ha più borsellini, el quale loculame(n)to abbia el ty(m)pano versatile, cioè ch(e) si volta, ch(e) sia collocato (et) posto i(n) coltello (et) ri(n)chiuso nello assicello, nella fronte del quale tympano facci(n)si de(n)ticegli, cioè codette, (et) i(n)tachature divisi equalme(n)te ch(e) p(er) numero sien 400 (et) ch(e) sieno co(n)venie(n)ti, cioè ch(e) risco(n)trino (et) ragua(n)gli(n)si al de(n)ticulo del ty(m)pano i(n)feriore».

Frequenza totale: 1

codette freq. = 1; 95v.

Lemmi correlati: *de(n)tatura*, *de(n)te* (1.), *de(n)tello* (2.), *de(n)ticello*, *de(n)ticulo*.

Prima e unica attestazione

3. Conclusioni

Questo piccolo campionario di esempi lessicali può essere sufficiente per confermare come, tra i procedimenti utilizzati per la creazione di lessici specialistici, quello della risemantizzazione in funzione tecnica di voci appartenenti alla lingua comune sia uno dei più fecondi anche in diacronia. Tale processo fornisce la possibilità a chi assimila, traduce e commenta testi tecnici di poter veicolare, in base al proprio bagaglio esperienziale e di conoscenze, quel sapere pratico che talvolta si fa anche – e solo – *sapere individuale*. Ciò porta a trarre questa conclusione: se da un lato la terminologizzazione è un procedimento altamente produttivo nella competenza linguistica dell'estensore della tra-

duzione ottoboniana, dall'altro lato i tecnicismi che ne derivano sono da considerare né più né meno che occasionalismi terminologici, ossia voci che non hanno trovato diffusione nella lingua tecnico-scientifica dei testi antichi, né risultano essere vitali in quella contemporanea²⁶.

In ambito artistico-architettonico, nel ms. *Ottoboniano* il sostantivo *picciolo*²⁷ assume il significato di 'elemento che costituisce la struttura di una scala', 'scalino, gradino'; il *quadro*²⁸ determina un 'settore di

26 Per un confronto con la lingua attuale, è parso interessante consultare il più completo vocabolario descrittivo dell'italiano contemporaneo: il *Grande Dizionario Italiano Dell'uso* (GRADIT) di Tullio De Mauro. La scelta non è causale, in quanto i criteri metodologici di allestimento delle voci del GRADIT prevedono anche l'inserimento di una specifica marca d'uso, ossia di una sigla posta in prossimità di ogni entrata lessicale e dopo la qualifica grammaticale, che permette di classificare a quale preciso strato del lessico italiano appartengono ciascun lemma ed eventualmente le sue singole accezioni. Tra le differenti marche d'uso utilizzate (per le quali si rimanda a GRADIT, vol. I, *Introduzione*, pp. xx-xxi), compare anche la sigla "TS", da riferirsi a vocaboli «legati a un uso marcatamente o esclusivamente tecnico-specialistico; sono così marcati 107.194 vocaboli usati e noti in gran parte soprattutto in rapporto a particolari attività, tecnologie, scienze [...]» (cfr. ivi p. xx). Nelle note che seguiranno, ho quindi deciso di riportare – e solo per le monorematiche – le differenti accezioni tecnico-specialistiche dei lemmi trattati (termini artistico-architettonici: *picciolo*, *quadro*, *ristretto*, *tondo*; termini meccanici: *addentellato*, *bastoncello*, *campana*, *codetta*), a dimostrazione della produttività del processo di terminologizzazione che queste voci d'uso comune subiscono in vari ambiti tecnici e in relazione a particolari attività.

27 Cfr. GRADIT, s.v. *picciolo*: «TS bot. formazione allungata a sottile sezione cilindrica, che sostiene la foglia collegandola al ramo».

28 Cfr. ivi, s.v. *quadro*: «TS giochi al pl., uno dei due semi rossi delle carte da gioco francesi, a forma di rombo; TS teatr. ciascuna delle parti in cui si divide un atto, introdotte da brevissime pause nella rappresentazione a palcoscenico vuoto, che si differenzia dall'atto perché non comporta variazioni di luogo e di scenario; TS cinem. durata di una singola azione ripresa senza interruzioni; TS telev. insieme delle righe che costituiscono le immagini su uno schermo televisivo e che tecnicamente è suddiviso in due semiquadri; TS elettr. parte di un impianto elettrico che contiene le apparecchiature di controllo dei cavi, il quadro di distribuzione o quello di manovra; TS milit. al plur., i ruoli degli ufficiali e dei sottufficiali dell'esercito | gli ufficiali e sottufficiali stessi [...]; TS giorn. articolo impaginato su due colonne e non collocato in testa alla pagina; TS geom. piano su quale si effettua la rappresentazione piana delle figure spaziali; TS ing., miner. quadro d'armatura;

muro quadrangolare posto al di sopra dell'architrave e compreso fra i due triglifi', 'metopa'; il *ristretto*²⁹ indica la 'rastremazione, graduale riduzione del diametro superiore di una colonna', 'restringimento della parte superiore di un elemento architettonico'; infine, il *tondo*³⁰ si definisce come un 'elemento piano di forma circolare, posto all'estremità inferiore o superiore della colonna, con funzione di base o di toro'.

Lo stesso procedimento è utilizzato per denominare ingranaggi e dispositivi meccanici o particolari elementi che determinano la struttura di macchine idrauliche. *L'addentellato*³¹ diviene un 'elemento metallico aguzzo di un ingranaggio'; il *bastoncello*³² si identifica ora in un 'perno di estremità di un albero meccanico', ora in un 'pistone o stantuffo'; la *campana*³³ si definisce come un 'dispositivo della macchina idraulica che consiste in un imbuto rovesciato simile a una campana; infundibulo'; infine, la *codetta*³⁴ non è altro che un 'prolungamento metallico, dente marginale di un organo o della corona di un ingranaggio'.

TS zoot. in mitilicoltura, castello di fusoli sporgenti dall'acqua tra i quali vengono tesi i libani; TS mar. vela quadra; TS telecom. antenna a telaio; TS tess. liccio».

- 29** Cfr. *ivi*, s.v. *ristretto*: «CO TS gastr. di sugo, salsa o liquido reso concentrato da una cottura prolungata; TS fin. mercato ristretto».
- 30** Cfr. *ivi*, s.v. *tondo*: «TS arte; nelle arti figurative, dipinto o bassorilievo di forma rotonda, diffuso spec. nel Rinascimento; TS tipogr. carattere latino con occhio dritto, contrapposto al corsivo; TS tecn. barra tonda d'acciaio, prodotta mediante laminazione, trafilatura o fucinatura; TS merceol. ciascuno dei pezzi di legna da ardere di ca. 10 cm di diametro».
- 31** Cfr. *ivi*, s.v. *addentellato*: «TS edil. l'insieme delle sporgenze poste in un muro per collegarlo con un altro».
- 32** Cfr. *ivi*, s.v. *bastoncello*: «TS anat., fisiol. cellula della retina di forma cilindrica, responsabile insieme ai coni della visione colorata; TS tipogr. carattere simile al bastone, ma di dimensioni ridotte».
- 33** Cfr. *ivi*, s.v. *campana*: «TS mus. parte finale allargata di alcuni strumenti a fiato; TS arch. nel capitello corinzio, parte interna avvolta dalla foglia di acanto o da altri motivi ornamentali; TS mar. parte girevole dell'argano intorno alla quale si avvolge il cavo».
- 34** Cfr. *ivi*, s.v. *codetta*: «TS ling. piccolo segno grafico a forma di virgola rovesciata posto sotto la *e*, indicante, nei manoscritti e nelle vecchie edizioni, il dittongo latino | TS fon. segno grafico posto sotto le vocali, indicante, in alcuni sistemi di trascr-

Le voci e le loro accezioni qui presentate sono il frutto dell'impegno che l'anonimo traduttore ha dedicato alla creazione del lessico specialistico architettonico e meccanico. Ma resta, in fondo, un tentativo di arricchimento terminologico, destinato però a rimanere confinato tra le righe del volgarizzamento ottoboniano e a non sedimentarsi, nel tempo, nel serbatoio del lessico specialistico delle singole discipline.

Bibliografia

- ARMANDO ET AL. 2022 = *Tra le righe dell'architettura: lingua, stile, testo*, a cura di Alessandro Armando, Giovanni Durbiano, Costanza Lucarini, Raffaella Scarpa, Milano-Udine, Mimesis, 2022.
- BELLUCCI 1997 = PATRIZIA BELLUCCI, *Gli usi speciali della lingua. Il linguaggio contemporaneo dell'architettura, con particolare riferimento al lessico*, «Quaderni del Dipartimento di linguistica» (Università degli Studi di Firenze), 8, 1997, pp. 153-212.
- BIFFI 1999 = MARCO BIFFI, *Sul lessico architettonico: alcuni casi controversi dalle traduzioni vitruviane*, in «Studi di lessicografia italiana», XVI, 1999, pp. 31-161.
- BIFFI 2001 = MARCO BIFFI, *Sulla formazione del lessico architettonico italiano: la terminologia dell'ordine ionico nei testi di Francesco di Giorgio Martini*, in *Le parole della scienza. Scritture tecniche e scientifiche in volgare (secoli XIII-XV)*, Atti del

zione fonetica, la pronuncia aperta | TS fon. Nell'ortografia polacca e nelle translitterazioni dell'alfabeto cirillico, segno grafico posto sotto alcune vocali, indicante un suono nasale; TS buocr. nella corrispondenza d'ufficio indirizzo posto a sinistra in alto del foglio; TS calz. estremità inferiore della tomaia; TS med. abrasione lineare all'estremo di una ferita da taglio; TS zool. spec. al pl., malattia dei bachi da seta; TS arm. prolungamento della culatta mobile, spec. nei moderni fucili da guerra e nelle armi da fuoco automatiche portatili; TS mar. cavo utilizzato da una nave, già ancorata di prora, per l'ormeggio di poppa | cavo posto generalmente a poppa ed utilizzato dalle navi per rimorchio o per ormeggio; TS ornit. com. → ballerina | → ²magnanina | → culbianco; TS mus. spec. al pl., periodo o gruppo di periodi musicali che concludono la prima o la terza parte di una sonata classica; TS gastr. decorazione per dolci costituita da granelli multicolori di zucchero o pagliuzze di cioccolato; TS elettr. nelle lampade elettriche, la parte interna al bulbo a cui si collegano i sostegni del filamento; TS agr. qualità inferiore di grano dai chicci piccoli».

Esempi di interferenza tra lingua comune e lingua specialistica

- Convegno (Lecce, 16-18 aprile 1999), a cura di Riccardo Gualdo, Galatina, Congedo, 2001, pp. 253-290.
- BIFFI 2003 = MARCO BIFFI, *Aspetti del lessico architettonico italiano*, in *Italia linguistica anno Mille. Italia linguistica anno Duemila*, Atti del xxxiv Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI) – Firenze, 19-21 ottobre 2000, a cura di Nicoletta Maraschio e Teresa Poggi Salani, Roma, Bulzoni, 2003, pp. 303-316.
- BIFFI 2005 = MARCO BIFFI, *Dal latino all'italiano e ritorno: il "De verborum vitruvianorum significatione" e la formazione del lessico architettonico italiano*, in *Bernardino Baldi (1553-1617). Studioso rinascimentale: poesia, storia, linguistica, meccanica, architettura*, Atti del Convegno di Studi di Milano (19-21 novembre 2003), a cura di Elio Nenci, Milano, Franco Angeli, 2005, pp. 143-174.
- BIFFI 2006 = MARCO BIFFI, *Il lessico dell'architettura nella storia della lingua italiana*, in *Fare storia 3: Costruire il dispositivo storico. Tra fonti e strumenti*, a cura di Jasenka Gudelj e Paola Nicolin, Milano, Mondadori, 2006, pp. 75-132.
- BIFFI 2009 = MARCO BIFFI, *Il teatro di Vitruvio: alcune osservazioni lessicali in margine alle prime traduzioni in volgare*, in *Saggi di letteratura architettonica da Vitruvio a Winckelmann*, vol. II, a cura di Lucia Bertolini, Firenze, Olschki, 2009, pp. 57-85.
- BIFFI 2017 = MARCO BIFFI, *Ingegneria linguistica tra Francesco di Giorgio e Leonardo* (LIII Lettura vinciana, Vinci 13 aprile 2013), Firenze, Giunti, 2017.
- BIFFI 2019 = MARCO BIFFI, *All'alba di un lessico intellettuale europeo dell'architettura*, in *Lingua delle arti e lingua di artisti in Italia fra Medioevo e Rinascimento*, a cura di Alessandro Aresti, Firenze, Cesati, 2019, pp. 37-60.
- BIFFI 2022 = MARCO BIFFI, *La lingua di architettura: alcune riflessioni a margine di venticinque anni di studi*, in *ARMANDO ET AL. 2022*, pp. 13-40.
- BOSAZZI 2000 = EMILIO BOSAZZI, *Il De Architectura di Vitruvio. Studi sulla lingua*, Trieste, Editreg, 2000.
- CALLEBAT 1982 = LOUIS CALLEBAT, *La prose du De architectura de Vitruve*, in «Aufstieg und Niedergang der römischen Welt», II, 30, 1, 1982, pp. 696-722.
- CALLEBAT-FLEURY 1995 = LOUIS CALLEBAT, PHILIPPE FLEURY, *Dictionnaire des termes techniques du 'De architectura' de Vitruve*, Hildesheim-Zürich-New York, Olms-Weidmann, 1995.
- CASTELLANI 1982 = ARRIGO CASTELLANI, *La prosa italiana delle origini. I. Testi di carattere pratico*, Bologna, Pàtron, 1982.
- CAVAGNOLI 2007 = STEFANIA CAVAGNOLI, *La comunicazione specialistica*, Roma, Carocci, 2007.
- CORTELAZZO 1994 = MICHELE A. CORTELAZZO, *Lingue speciali – la dimensione verticale*, Padova, Unipress, 1994.

Matteo Mazzone

- DEVOTO 1939a = GIACOMO DEVOTO, *Le cronache del calcio*, in «Lingua Nostra», I, 1939, pp. 17-21.
- DEVOTO 1939b = GIACOMO DEVOTO, *Dalle cronache della finanza*, in «Lingua Nostra», I, 1939, pp. 114-21.
- DEVOTO 1940 = GIACOMO DEVOTO, *Storia della lingua di Roma*, Bologna, Cappelli, 1940.
- DI GIORGIO MARTINI/BIFFI 2002 = FRANCESCO DI GIORGIO MARTINI, *La traduzione del De architectura di Vitruvio (dal ms. II.I.141 della Biblioteca Nazionale Centrale di Firenze)*, a cura di Marco Biffi, Pisa, Scuola Normale Superiore di Pisa, 2002.
- GALLUZZI 1979 = PAOLO GALLUZZI, *Per un lessico della terminologia meccanica nel '500*, in *Convegno nazionale sui lessici tecnici delle arti e dei mestieri*, Cortona «Il Palazzone», 28-30 maggio 1979, Firenze, Eurografica, 1979, pp. 143-163.
- GOTTI 1991 = MAURIZIO GOTTI, *I linguaggi specialistici. Caratteristiche linguistiche e criteri pragmatici*, Firenze, La Nuova Italia, 1991.
- GRADIT = *Grande Dizionario Italiano dell'uso*, a cura di Tullio De Mauro, Torino, UTET 2007 [si è consultata la versione aggiornata del 2007 su supporto digitale].
- GUALDO 2021 = RICCARDO GUALDO, *Introduzione ai linguaggi specialistici*, Roma, Carocci, 2021.
- GUALDO-TELVE 2021 = RICCARDO GUADO, STEFANO TELVE, *I linguaggi specialistici dell'italiano*, Roma, Carocci, 2021.
- MACCAGNI 1996 = CARLO MACCAGNI, *Cultura e sapere dei tecnici nel Rinascimento, in Piero della Francesca tra arte e scienza*, Atti del Convegno internazionale di studi (Arezzo, 8-11 ottobre 1992), (Sansepolcro, 12 ottobre 1992), a cura di Marisa Dalai Emiliani e Valter Curzi, Venezia, Marsilio, 1996, pp. 279-292.
- MANNI 1980 = PAOLA MANNI, *La terminologia della meccanica applicata nel Cinquecento e nei primi decenni del Seicento (Origini del lessico volgare scientifico)*, in «Studi di lessicografia italiana», II, 1980, pp. 139-213.
- MANNI 1981 = PAOLA MANNI, *Per una storia del lessico della meccanica applicata (i secoli XVI-XVIII)*, in *Convegno Nazionale sui lessici tecnici del Sei e Settecento*, Pisa, Scuola Normale Superiore, 1-3 dicembre 1980, Firenze, Eurografica, 2 voll., 1981, vol. I, pp. 89-109.
- MAZZONE 2017-2018 = MATTEO MAZZONE, *Un volgarizzamento adespoto del "De architectura" di Vitruvio del primo Cinquecento: il codice "Ottonobiano Latino 1653" della Biblioteca Apostolica Vaticana. Analisi, edizione e studio linguistico*, relatore prof. Marco Biffi, Firenze, Università degli Studi, 2017-2018.
- MAZZONE 2018-2021 = MATTEO MAZZONE, *La copia della traduzione vitruviana contenuta nel ms. Ottoboniano latino 1653: il ms. Italien 472 della Bibliothèq-*

Esempi di interferenza tra lingua comune e lingua specialistica

- ue Nationale de France. *Analisi, edizione e studio linguistico. Con l'aggiunta di un Glossario di oltre 930 voci tecniche tratte dal ms. Ottoboniano latino 1653* [vol. I: *Analisi, edizione e studio linguistico*; vol. II: *Glossario. Lettere A-L*; vol. III, *Glossario. Lettere M-Z*]. Tutore prof. Marco Biffi e co-tutrice prof.ssa Emanuela Ferretti, Firenze, Università degli Studi, 2018-2021.
- MAZZONE 2021 = MATTEO MAZZONE, *Osservazioni sulla lingua di un volgarizzamento cinquecentesco del "De architectura" di Vitruvio: il codice Ottoboniano latino 1653 della Biblioteca Apostolica Vaticana*, in «Studi di grammatica italiana», XL, 2021, pp. 37-94.
- MAZZONE 2022a = MATTEO MAZZONE, *Una nuova traduzione vitruviana del primo Cinquecento: il ms. Ottoboniano latino 1653 della Biblioteca Apostolica Vaticana*, in ARMANDO ET AL. 2022, pp. 161-176.
- MAZZONE 2022b = MATTEO MAZZONE, *Tradurre e interpretare il De architectura di Vitruvio: alcune riflessioni sul testo di una versione fiorentina di inizio Cinquecento*, in *L'italiano e la scienza tra Medioevo e Rinascimento. Le vie della lingua, della letteratura e dell'arte*, a cura di Lorenzo Bacchini et alii, Firenze, Cesati, 2022, pp. 101-109.
- NENCIONI 1995 = GIOVANNI NENCIONI, *Sulla formazione di un lessico nazionale dell'architettura*, in «Bollettino d'Informazioni del Centro di Ricerche Informatiche per i Beni Culturali», Scuola Normale Superiore di Pisa, v.2, 1995, pp. 7-33.
- NENCIONI 2000 = GIOVANNI NENCIONI, *Sulla formazione di un lessico nazionale dell'architettura*, in ID., *Saggi e memorie*, Pisa, Scuola Normale Superiore, 2000, pp. 51-74.
- PORRO 1973 = MARZIO PORRO, *I linguaggi della scienza e della tecnica*, in *I linguaggi settoriali in Italia*, a cura di Gian Luigi Beccaria, Milano, Bompiani, 1973, pp. 181-206.
- ROMANO 1997 = ELISA ROMANO, *Fra astratto e concreto. La lingua di Vitruvio*, in VITRUVIO, *De architectura*, a cura di Pierre Gros, traduzione e commento di Antonio Corso ed Elisa Romano, Torino, Einaudi, 1997, 2 voll., vol. I, pp. LXXIX-XCV.
- SERIANNI 2005 = LUCA SERIANNI, *Un treno di sintomi. I medici e le parole: percorsi linguistici nel passato e nel presente*, Milano, Garzanti, 2005.
- SOBRERO-MIGLIETTA 2006 = ALBERTO A. SOBRERO, ANNARITA MIGLIETTA, *Introduzione alla linguistica italiana*, Roma-Bari, Laterza, 2006.
- VISCONTI-MANFREDINI-COVERI 2020 = *Linguaggi settoriali e specialistici: sincronia, diacronia, traduzione, variazione*, Atti del xv Congresso SILFI Società Internazionale di Linguistica e Filologia Italiana (Genova, 28-30 maggio 2018), a cura di Jacqueline Visconti, Manuela Manfredini, Lorenzo Coveri, Firenze, Cesati, 2020.

Riassunto L'articolo si concentra sull'aspetto relativo all'interferenza tra lingua comune e lingua specialistica, e in particolare modo intende analizzare uno dei processi più produttivi nella creazione di lessici specialistici, ossia la rideterminazione semantica in chiave tecnica di voci d'uso quotidiano. L'analisi di tale procedimento risulta interessante se indagata in testi tecnici e di carattere pratico risalenti al Quattrocento e al Cinquecento, ossia agli albori della formazione di una lingua tecnica di alcune discipline specialistiche, come l'architettura e la meccanica applicata. Il saggio contiene un'analisi semantica di alcuni lemmi tratti dallo spoglio lessicale effettuato su quella che possiamo oggi considerare come la seconda traduzione cinquecentesca del *De architectura* di Vitruvio. Essa è databile tra il 1504 e il 1511 ed è contenuta nel manoscritto *Ottoboniano latino 1653* della Biblioteca Apostolica Vaticana. All'interno delle numerose precisazioni semantiche, l'anonimo estensore della traduzione ottoboniana connota tecnicamente voci della lingua comune, funzionalizzandole come termini per denotare e individuare sia strutture architettoniche o specifici elementi edilizi, sia congegni meccanici o parti di essi. Sono stati così esaminati quattro termini architettonici, ossia *picciolo*, *quadro*, *ristretto*, *tondo* e quattro tecnicismi meccanici, ovvero *addentellato*, *bastoncello*, *campana*, *codetta*. I lemmi sono stati estrapolati da una lista lessicale che è il frutto di una preliminare realizzazione di una versione elettronica della traduzione ottoboniana, processata attraverso lo strumento di analisi linguistica *StanfordNLP*, con il quale è stata possibile la tokenizzazione del testo e la successiva estrapolazione di una tabella contenente l'elenco delle forme affiancate dalla loro singola frequenza. Per ciascuna delle otto voci, sono state fornite delle informazioni che ricordano i campi fondamentali di una scheda lessicografica, così ripartite: il lemma, la definizione, un solo esempio (tratto dal volgarizzamento, volto ad attestare l'utilizzo e il significato del tecnicismo), la frequenza totale corredata da *index locorum* e la presenza di eventuali lemmi correlati.

Abstract The article focuses on the interference between common language and specialized language. In particular, it intends to analyze one of the most productive processes in the creation of specialized lexicons, that is the semantic redefinition of daily use items in a technical key. The analysis of this procedure is interesting if investigated in technical and practical texts dating back to the fifteenth and sixteenth centuries, namely at the beginning of the development of a technical language of specialized disciplines, such as architecture and applied mechanics. The essay contains a semantic analysis of some lemmas taken from the lexical analysis carried out on what we can consider now as the second sixteenth-century translation of *De architectura* by Vitruvius. It is dated between 1504 and 1511 and is contained in the ms. *Ottoboniano latino 1653* of the Vatican Apostolic Library. Among the numerous semantic details, the anonymous writer of the Ottobonian translation technically connotes common language entries, functionalizing them as terms to denote and identify architectural structures or spe-

Esempi di interferenza tra lingua comune e lingua specialistica

cific building elements, or parts thereof. Four architectural terms have been examined: *picciolo*, *quadro*, *ristretto*, *tondo* and four mechanical technicalities, namely *addentellato*, *bastoncello*, *campana*, *codetta*. The entries have been extracted from a lexical list which is the result of a preliminary realization of an electronic version of the Ottobonian translation, processed through the *StanfordNLP* linguistic analysis tool, with which it was possible to tokenize the text and the subsequent extrapolation of a table containing the list of shapes flanked by their single frequency. For each of the eight terms, information recalling the fundamental fields of a lexicographic card has been provided as follows: (senza the) lemma, definition, a single example (taken from the vulgarization, aimed to attesting the use and the meaning of technicality), the total frequency accompanied by *index locorum* and the presence of any related lemmas.

Il “dato silente” in un corpus di parlato schizofrenico

Simona Trillocco

1. Introduzione

Le malattie della mente, a partire dall'indirizzo psichiatrico medico-biologico (MINKOWSKI 1998), sono analizzate a differenti livelli (fisico, familiare, comportamentale, ecc.), ma questo non basta per avere un'idea chiara e precisa della patologia. È il caso della schizofrenia, su cui, sin dalle prime definizioni come *dementia praecox* (MOREL 1860; KRAEPELIN 1904), ancora oggi non esiste un pieno accordo nosografico, nonostante i molteplici tentativi di incasellarne le caratteristiche (ANDREASEN 1986; LIDDLE ET AL. 2002). Attualmente, secondo le direttive del DSM5 2013, si segue una distinzione in sintomi *positivi*, legati alla comparsa di fenomeni non comuni nei soggetti non affetti da schizofrenia (deliri, allucinazioni, linguaggio disorganizzato con deragliamento o incoerenza, comportamento disorganizzato o catatonico), e *negativi*, che, al contrario, rilevano una diminuzione o un'assenza di certe caratteristiche (espressione emotiva ridotta o mancanza di volizione).

In quest'ottica, tra i sintomi negativi si può ascrivere una riduzione delle capacità sociali (LUCARINI ET AL. 2020), per cui emergono difficoltà nella comunicazione interpersonale (COUTURE ET AL. 2006) e a livello pragmatico (KUPERBERG 2010a, 2010b). In correlazione a disfunzioni dell'emisfero destro (JOANETTE ET AL. 2008) a loro volta associate alla schizofrenia (HA ET AL. 2004; RIMOL ET AL. 2010; STEGMAYER ET AL. 2014), è stato osservato un impoverimento delle abilità extra- e para-linguistiche, tra cui la comprensione e l'inferenza del significato, il

riconoscimento del senso dell'umor, la spiegazione scritta o tramite disegni della metafora, l'analisi del discorso e la prosodia (PAWELCZYK ET AL. 2018).

Sul versante opposto, un sintomo positivo di particolare interesse è l'aumento del silenzio.

A partire da HEIKE 1981, infatti, le disfluenze in generale, e le pause in particolare, si considerano, più che un errore di produzione del parlato, un fenomeno fondamentale per il successo dell'atto comunicativo. Si distinguono errori *forestalled*, cioè previsti, non errori, funzionali all'elaborazione del parlato, ed errori *committed*, come *stalls*, “blocchi”, e *repairs*, “riparazioni”, che aiutano il parlante nella modulazione del discorso e concorrono alla sua fluidità. Ma se non rispettano certi standard, questi silenzi possono costituire un problema per la comunicazione. In ANDREASEN 1986 (cf. LIDDLE ET AL. 2002) uno dei diciotto sintomi della schizofrenia è il *blocking*, ovvero l'interruzione del pensiero seguita da una fase di silenzio che può durare da pochi secondi a qualche minuto, dopo cui è difficile riprendere il filo del discorso. In GOLDMAN-EISLER 1961 e BANFI 1999 la lunghezza delle pause è un chiaro discriminante tra parlato patologico e non patologico e sposta l'attenzione sul dato quantitativo. In questa direzione vanno le classificazioni in termini di durata (SAVY 2007; GIANNINI 2008; DOVETTO-GEMELLI 2013; DOVETTO ET AL. 2021), che individuano una “soglia minima”, sotto la quale il silenzio non va considerato (in GIANNINI 2008 è fissata a 180 ms, calcolato come valore medio di durata delle consonanti occlusive), e successivamente una serie di soglie secondarie (in GOLDMAN-EISLER 1961 il 70% delle pause del parlato non patologico è compreso tra 250 e 1000 ms; in SAVY 2007 la distinzione in <sp>, *small pauses*, le “pause brevi”, e <lp>, *long pauses*, le “pause lunghe”, cui si aggiunge in DOVETTO ET AL. 2021 <mp>, *medium pauses*, le “pause medie”).

In senso qualitativo, diversamente, le classificazioni si sono basate sulla posizione dei silenzi (*gaps/lapses*, rispettivamente pause semplici e pause più lunghe tra turni, e *pauses*, pause all'interno dei turni in HELDNER-EDLUND 2010; FORS 2011; similmente *inter-tours* e *intra-tours pauses* in DODANE-HIRSCH 2018) oppure sul loro legame con i processi di respirazione (*respiratoires* e *non respiratoires pauses* in FAUTH-THOUVAIN 2018).

Il “dato silente” in un corpus di parlato schizofrenico

Su tali presupposti, il presente contributo mira a un'analisi quantitativa e qualitativa delle pause nel parlato patologico schizofrenico a partire da una prospettiva prosodico-pragmatica. Cornice teorica è la Teoria della Lingua in Atto (L-Act, CRESTI 2000; CRESTI-MONEGLIA 2005; MONEGLIA-RASO 2014), che si fonda sull'intuizione austiniana (AUSTIN 1962) secondo cui ciascun atto del parlare, detto atto linguistico, viene considerato un'attività umana ed è simultaneamente atto locutivo (l'atto in sé di dire qualcosa), un atto illocutivo (ciò che si fa nel dire qualcosa) e atto perlocutivo (ciò che ci si aspetta dal dire qualcosa). Fondante è l'«ipotesi che sia possibile decidere l'equivalenza tra unità del dominio dell'agire umano (atti) e unità linguistiche (enunciati)» (CRESTI 2000, p. 42), per cui l'*enunciato* diviene l'unità di riferimento del parlato e in relazione ad esso vengono misurati i silenzi.

Nello specifico: nel paragrafo 2 saranno definiti il *corpus* patologico e il gruppo di controllo su cui sono condotte le analisi; nel paragrafo 3 sarà illustrato il metodo adottato, a partire dall'allineamento testo-suono del parlato per enunciati fino ai criteri di classificazione adottati; nel paragrafo 4 verranno mostrate le analisi effettuate sia su base quantitativa che qualitativa, da cui saranno esplicitati i risultati nel paragrafo 5.

2. Il *corpus* CIPPS e il Gruppo di controllo

Oggetto di osservazione è il *corpus* CIPPS, Corpus di Italiano Parlato Patologico/Schizofrenico (DOVETTO-GEMELLI 2013), che nasce nel 2005 da una collaborazione tra la Scuola Sperimentale per la Formazione alla Psicoterapia e alla Ricerca nel Campo delle Scienze Umane Applicate della ASL NA1, diretta dal dott. Pastore, e il CIRASS, il Centro Interdipartimentale di Ricerca per l'Analisi e la Sintesi dei Segnali dell'Università di Napoli Federico II, all'epoca sotto la supervisione del prof. Albano Leoni.

Il *corpus* è costituito da 17 ore¹ di registrazione audio delle sedute di psicoterapia di quattro parlanti indicati per privacy con le lettere A,

¹ Del totale, solo 10 ore sono state trascritte, secondo le norme ortografiche di CLIPS (SAVY 2007).

B, C e D, ai quali la malattia è stata diagnosticata con sintomatologie diverse tra un soggetto e l'altro. In tabella 1, le specifiche:

	durata registrazione	sedute	tokens	diagnosi
A	02:33	3	2563	condizione predelirante (<i>Wahnstimmung</i>), assenza di allucinazioni
B	04:13	4	30021	schizofrenia paranoidea con delirio non strutturato, con allentamento dei nessi associativi e fuga delle idee, interpretazioni deliranti guidate da idee di riferimento e senza allucinazioni
C	02:14	2	10409	schizofrenia paranoidea con delirio strutturato megalomane con note di persecutorietà e allucinazioni verbali con idee di riferimento
D	00:28	1	1277	schizofrenia paranoidea con delirio
tot.	09:28	10	44270	

Tabella 1 Presentazione CIPPS (CRESTI ET AL. in stampa)

I quattro pazienti registrati sono tutti maschi provenienti dall'area napoletana, con un'età che va dai 35 ai 45 anni. L'italiano di tutti i soggetti è connotato dal dialetto, in particolar modo quello di D.

In tabella 2, i dati quantitativi del *corpus*, che comprendono il valore del *T-sounding*, ovvero la somma di suono e silenzio all'interno del turno del paziente (escludendo cioè le pause tra turni, sia di attacco che di chiusura, cfr. CRESTI ET AL. in stampa), il numero dei *tokens* e quello degli enunciati, per avere un'idea del parlato prodotto.

	<i>T-sounding</i>	tokens	enunciati
A	01:03	2563	619
B	03:43	30021	4204
C	01:26	10409	1552
D	0:17	1277	332
tot	6:29	44270	6707

Tabella 2 Descrizione quantitativa di CIPPS

Il “dato silente” in un corpus di parlato schizofrenico

Quanto al Gruppo di controllo, dal *corpus* LABLITA è stata fatta una selezione che consentisse una comparabilità per genere, per sesso, e per età e anche per la situazione comunicativa dell'intervista, anche se non con l'identica finalità psichiatrica che, per soggetti non affetti da patologia, risulta difficilmente replicabile. Purtroppo non è stato possibile bilanciare per diastria e diatopia, ma in ogni caso non risultano evidenze scientifiche secondo cui tali variazioni possano incidere in maniera significativa sulla pausazione.

Il *corpus* di confronto così ottenuto consta di otto parlanti, tutti di sesso maschile, di età simile a quella dei soggetti con schizofrenia, di varia provenienza². In tabella 3 i dati quantitativi per il campionamento descritto.

	<i>T-sounding</i>	tokens	enunciati
*prvdlr06-mace	00:09:43	1171	116
*prvdlr13-stum	00:11:34	1922	243
*pubdlr10-sodi	00:10:04	1454	219
*prvmnl06-ital	00:07:54	1400	133
*prvmnl01-cami	00:05:27	1397	112
*prvdlr20-att3	00:10:15	817	59
*prvdlr19-att2	00:10:19	1099	89
*prvdlr14-pell	00:08:24	1478	194
tot.	01:13:20	34398	4016

Tabella 3 Descrizione quantitativa del Gruppo di controllo

3. Metodo

La metodologia adottata per indagare le pause nel parlato patologico schizofrenico si compone essenzialmente di tre fasi: i. Allineamento te-

² Degli 8 soggetti, 3 sono di Firenze, 1 di Arezzo, 1 di L'Aquila, 1 di Milano e 2 di origine sconosciuta.

sto-suono su WinPitch (MARTIN 2004) per enunciati; ii. Annotazione con TextGrid su Praat (BOERSMA-WEENINK 2021); iii. Revisione manuale.

3.1 Allineamento testo-suono per enunciati

Occorre specificare preliminarmente che il testo allineato (e utilizzato per le analisi) è un adattamento di quello in DOVETTO-GEMELLI 2013, che, per motivi di leggibilità, è stato epurato di tutta una serie di indicazioni. Nella versione “nuda” (CRESTI *ET AL.* in stampa) ottenuta, non sono segnalati fenomeni quali il dialetto (<dialect>), i rumori di sottofondo (<NOISE>) né gli allungamenti³ (consonantici <cc> o vocalici <vv>), mentre è diversa la resa di alcuni fenomeni, come le produzioni foniche incomprensibili (precedentemente <unclear>, ora sono xxx) e le vocalizzazioni (prima <vocal>, ora hhh). Inoltre, in questa fase viene espunta l'indicazione delle pause vuote in base alla durata non numerica, differenziate in <sp>, le pause brevi, <lp>, quelle lunghe, e <P>, quelle medio/lunghe con interruzione del discorso⁴. Il testo è poi adattato, per ragioni di comparabilità con il materiale trascritto per il Gruppo di controllo, secondo i criteri del formato CHAT (Codes for Human Analysis of Transcripts)⁵.

L'operazione di allineamento testo-suono è condotta su WinPitch⁶ (MARTIN 2004), che consente la visualizzazione simultanea dell'onda sonora prodotta, con l'indicazione della frequenza fondamentale, per ogni enunciato, allineato in basso su una o più righe chiamate “Layer”, che nel nostro caso sono rinominate con delle sigle che indicano i parlanti.

- 3 In realtà, il caso degli allungamenti costituisce un aspetto da approfondire, poiché la loro presenza è particolarmente legata ai break prosodici non terminali e alle pause.
- 4 In questa fase del lavoro sono stati inseriti provvisoriamente il segno # per le pause “percepite” e la sua duplicazione ## nel caso di pausa sia particolarmente lunga.
- 5 Ideato da Brian MacWhinney nei primi anni '80 (MACWHINNEY 1991), in seguito implementato e adottato dal laboratorio fiorentino (MONEGLIA-CRESTI 1997).
- 6 Un software di analisi prosodica, «a method for easy and precise selection of alignment units, ranging from syllable to whole sentences in a hierarchical storing system of aligned data» (MARTIN 2004).

Il “dato silente” in un corpus di parlato schizofrenico

Contestualmente, il flusso di parlato è segmentato per enunciati, attraverso l'individuazione di break prosodici terminali secondo il quadro teorico di L-AcT. Quindi, “percettivamente” sono individuate le unità terminate, di cui è osservato il profilo prosodico e verificata l'interpretabilità pragmatica: ciascun enunciato deve risultare concluso, autonomo e perfettamente interpretabile in isolamento. L'enunciato viene allora delimitato dal segno di interruzione prosodica terminale //; qualora sia percettivamente scandito in più parti, l'enunciato è internamente diviso in unità prosodiche più piccole, marcate dal segno /, che indica break non terminali (Cresti 2000).

3.2 Annotazione con TextGrid su Praat e Revisione manuale

Il file .wpc2 viene successivamente esportato su Praat (BOERSMA-WEENINK 2021), per una rapida individuazione delle pause.

In realtà, per rendere possibile questa individuazione, è necessario disporre di file audio di una certa qualità, in cui il rumore non superi una certa soglia e non impedisca all'algoritmo di distinguere i silenzi dalle parole. Da questo punto di vista, purtroppo, il *corpus* CIPPS presenta una serie di problemi, motivo per cui, preliminarmente, si è resa indispensabile una fase di “pulitura della qualità sonora”: con il filtro “Reduce noise” di Praat, mediante un adeguamento diverso per ciascuna registrazione in base alle caratteristiche acustiche e principalmente all'intensità, sono stati limitati i rumori di fondo, con la creazione di una nuova traccia audio. Su quest'ultima si è eseguito infine lo script automatico⁷, con la segmentazione del parlato in porzioni in cui c'è la voce, il *sounding*, e in altre in cui non sono prodotte parole, il *silent*. In linea con la letteratura più recente (cfr. *art* and *pau* in FAUTH-TROU-

7 Nello specifico, su Praat Object si seleziona il file .wav che si intende analizzare, si clicca prima su “Annotate” e poi su “To TextGrid (silences)”. Nella finestra che si apre occorre controllare il “silence threshold (db)” (in genere il valore di default “-25.0” va abbassato se il default considera come silenzi anche i rumori di sottofondo, mentre va alzato se sfuggono parole) e poi generare il TextGrid.

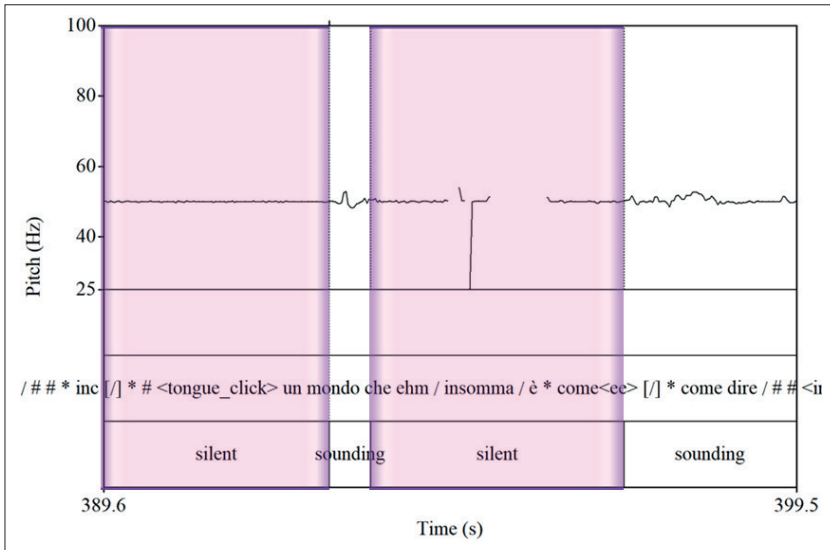


Figura 1 Esempio della segmentazione automatica dello script «sounding/silent»

VAIN 2018), il *sounding* comprende, oltre al parlato verbale lessicale, le disfluenze e i frammenti di parola, mentre nel *silent* ci sono anche rumori e fenomeni vocali non verbali, come le inspirazioni, i *tongue click* e le risate (ibidem per *pauses respiratoires* e *non respiratoires*).

Secondo questi criteri, è generato un file TextGrid in cui è presente un apposito livello per la pausazione, come si può vedere nella parte inferiore della figura 1.

In ultima istanza, è stato effettuato il controllo manuale.

Oltre agli errori di algoritmo, che possono sempre occorrere nel corso di questi procedimenti automatici, può capitare che lo script consideri silenzio anche la fase di preparazione delle occlusive, catturando solo il momento di esplosione della consonante. Per tale ragione, due revisori hanno controllato il TextGrid prodotto per il totale delle registrazioni, correggendo all'occorrenza i confini del *sounding/silent* (SACCONE-TRILLOCCO 2022). I *boundaries* così sistemati sono alli-

neati ai *pulses* della frequenza fondamentale, in modo da ridurre il più possibile il fattore percettivo e confermare quello acustico.

4. Analisi

Secondo il quadro teorico di L-AcT, le pause sono preliminarmente classificate qualitativamente, in base alla posizione occupata rispetto all'enunciato. Similmente all'approccio tipologico di ultima generazione (cfr. *inter-tours* e *intra-tours* in DODANE-HIRSCH 2018; *gaps/lapses* e *pauses* in HELDNER-EDLUND 2010; FORS 2011), ciascun *silent* viene rinominato in:

- T (<*turns*)⁸, quando la pausa occorre tra i turni di due parlanti differenti (in CIPPS il dottore e il paziente, nel Gruppo di controllo il parlante principale e l'intervistatore);
- UT (<*utterances*), quando la pausa occorre tra due enunciati dello stesso turno, dello stesso parlante;
- IU (<*informative units*), quando la pausa occorre tra due unità informative dello stesso enunciato⁹.

Per l'analisi quantitativa, si sceglie innanzitutto di suddividere i silenzi per soglie di durata, per individuare le quali, sulla base delle varie classificazioni presenti in letteratura (150-200 ms in GIANNINI 2006; DOVETTO-GEMELLI 2013; 201-250 ms in LEA-KLOKER 1975; DUEZ 1985;

⁸ Questo tipo di silenzio non è oggetto di discussione in questo contributo, poiché è fortemente dipendente dal contesto della comunicazione (che non è lo stesso per i due gruppi di parlato, dal momento che CIPPS è caratterizzato soltanto da interviste cliniche) e i risultati correlerebbero solo in parte con la presenza o meno della patologia.

⁹ È in corso una collaborazione con l'Universidade Federal de Minas Gerais (UFMG) di Belo Horizonte, dove il gruppo di ricerca Fale, guidato dal professor Tommaso Raso, ha accolto la classificazione delle pause per posizione aggiungendo una sottocategoria delle pause IU, ovvero le pause S (<*scansion*), che si collocano all'interno di una stessa unità informativa.

SAVY 2007; DOVETTO *ET AL.* 2021; 251-500 ms in MONEGLIA 2005; DOVETTO-GEMELLI 2013; DOVETTO *ET AL.* 2021; 501-1000 ms in SAVY 2007; DOVETTO-GEMELLI 2013; DOVETTO *ET AL.* 2021; 1001-5000 ms in SAVY 2007; DOVETTO-GEMELLI 2013), si divide il tempo in porzioni di 50 ms, al fine di ottenere una distribuzione più chiara e facilmente interpretabile.

La soglia minima considerata è di 150 ms (GIANNINI 2008), come è osservabile nei due grafici (figura 2 e 3), che mostrano la distribuzione in valori assoluti per il Gruppo di controllo (figura 2) e CIPPS (figura 3).

Sull'asse delle ascisse sono presenti le soglie di durata scansionate per 50 ms, mentre sull'asse delle ordinate è riportato il numero delle pause. Le colonne si riferiscono alla tipologia dei silenzi. Il grafico arriva fino a 2000 ms, che è mediamente la durata massima riscontrata nel Gruppo di controllo.

In via preliminare, si precisa che le pause di tipo IU sono chiaramente più frequenti di quelle UT e T perché dipendono dal numero di unità informative, che è necessariamente maggiore del totale delle unità terminate. Il dato interessante che si ricava è relativo piuttosto alla distribuzione: il numero delle pause tra unità informative è maggiore per le soglie comprese tra 150 e 550 ms, dopodiché si osserva un decrescere dell'occorrenza all'aumentare della durata, mentre le pause tra unità terminate mantengono grossomodo un andamento piatto, pur diminuendo gradualmente fino a scomparire.

Piuttosto diverso è il quadro per i silenzi del parlato patologico schizofrenico.

Due premesse fondamentali. La prima è che il grafico riportato è stato tagliato per motivi di leggibilità, dal momento che presentava una coda che arrivava fino a 60000 ms: essa si riferisce soprattutto al paziente A, che è molto restio a parlare, pur con le stimolazioni del dottore. La seconda premessa è, in questa prima fase di analisi, la non comparazione quantitativa dei due gruppi: l'asse delle ordinate in CIPPS arriva fino a 300 e nel Gruppo di controllo a 100, perché i due *corpora* raccolti non sono infatti bilanciati per tempo.

Ciò che invece è significativo è l'andamento delle pause. Mentre, infatti, per i silenzi di tipo IU si osserva un comportamento simile al non

Il “dato silente” in un corpus di parlato schizofrenico

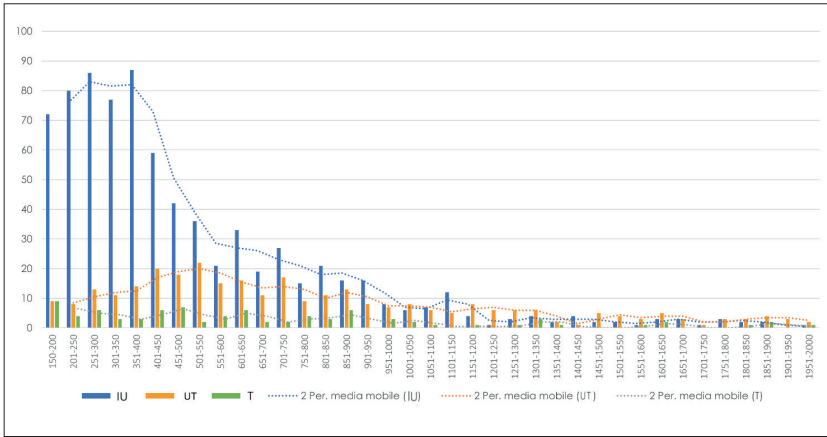


Figura 2 Distribuzione in valori assoluti delle pause del Gruppo di controllo

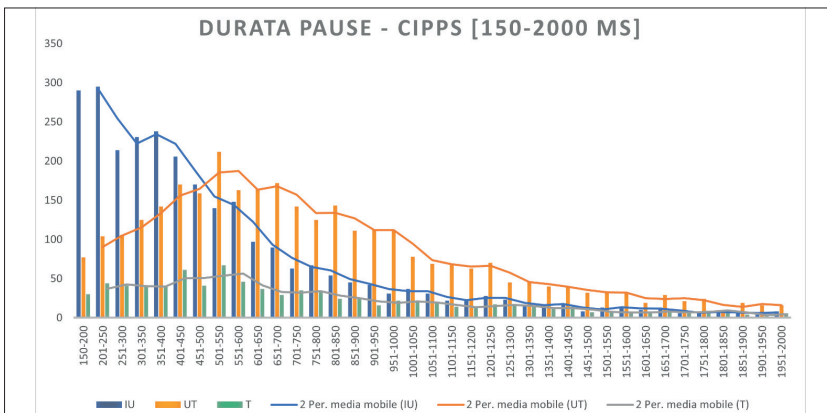


Figura 3 Distribuzione in valori assoluti delle pause di CIPPS

patologico, con una proporzionalità inversa tra numero di pause e loro durata, i silenzi di tipo UT rilevano un’atipia. Nella successione delle pause di tipo UT si osserva una proporzionalità diretta tra quantità e lunghezza delle pause, fino a un picco intorno ai 401 e gli 800 ms, dopodiché si registra una graduale decrescita.

Per meglio leggere e interpretare la distribuzione dei silenzi, e quindi poter comparare in termini statistici i risultati dei due *corpora*, i dati grezzi descritti sono stati convertiti in valori percentuali. Ciascun tipo di pausa è stato considerato in relazione al tipo di unità cui si riferisce (unità informativa, enunciato o turno). Nelle figure 4 e 5 gli istogrammi per i due gruppi.

Ancora per motivi di leggibilità, non sono state usate le soglie per 50 ms. Al loro posto sono stati individuati quattro range (in cui le barre orizzontali degli istogrammi sono state divise in 5 segmenti, che riportano nell'ordine: pause 150-250 ms, pause 251-500 ms, pause 501-1000 ms, pause >1000 ms e unità senza pause), sulla base dei risultati precedentemente ottenuti per i valori assoluti, e ad essi è aggiunto il numero di unità che non presenta pausa (i reticoli in bianco).

La prima osservazione è che la percentuale di turni e di *utterances* non accompagnati da pause è minore in CIPPS che nel Gruppo di controllo (26,2% vs 63,5% per T e 58,4% vs 68,8% per UT). Non esiste invece una differenza significativa per le pause IU (68,6% vs 67,7%), dove l'elemento che discrimina i due gruppi spetta invece alle pause di durata superiore ai 1000 ms (ovvero il quarto segmento orizzontale delle barre) che in CIPPS sono circa tre volte più frequenti (9,4% vs 3 %).

I segmenti indicanti le pause superiori a 1 sec sono ad ogni modo sempre più lunghi per il parlato schizofrenico, anche per le pause di tipo UT (18% vs 8,8%) e T (33,9% vs 8,1%), e non è molto dissimile il comportamento delle pause tra 500-1000 ms UT e T, che, ancora, in CIPPS sono più lunghe.

Le pause T evidenziano un *discrimen* fondamentale per la patologia, dal momento che i pazienti schizofrenici prima di un turno quasi sempre fanno una pausa (la somma delle percentuali dei diversi segmenti delle barre orizzontali restituisce circa il 74% del totale dei turni) e la fanno anche lunga (solo nel 4,80% producono pause tra i 150 e i 250 ms e la percentuale arriva al 12,96% per la soglia 251-500 ms, al 22,16% per quella 501-1000 ms e, addirittura, al 33,94% per le pause di durata superiore a 1s). Ma, come anticipato sopra, il contesto dell'intervista clinica condiziona in maniera decisiva le pause di tipo T (poiché la relazione medico-paziente incide inevitabilmente sui tempi di attacco

Il "dato silente" in un corpus di parlato schizofrenico

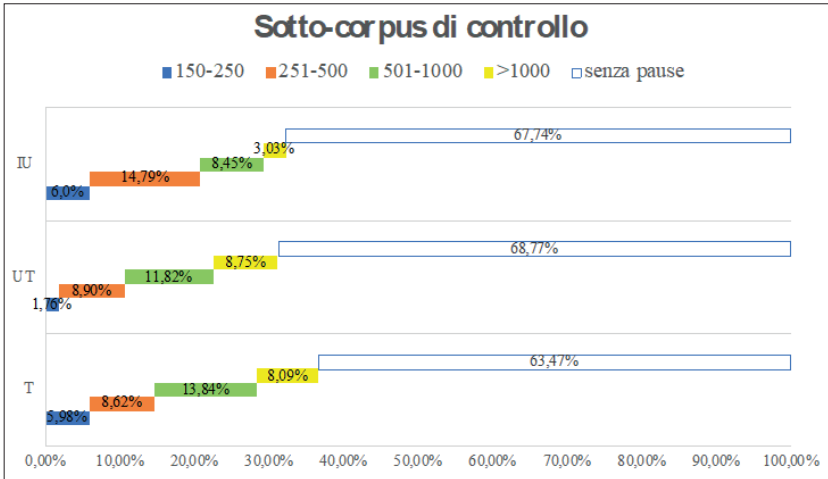


Figura 4 Frequenza delle pause nel Gruppo di controllo

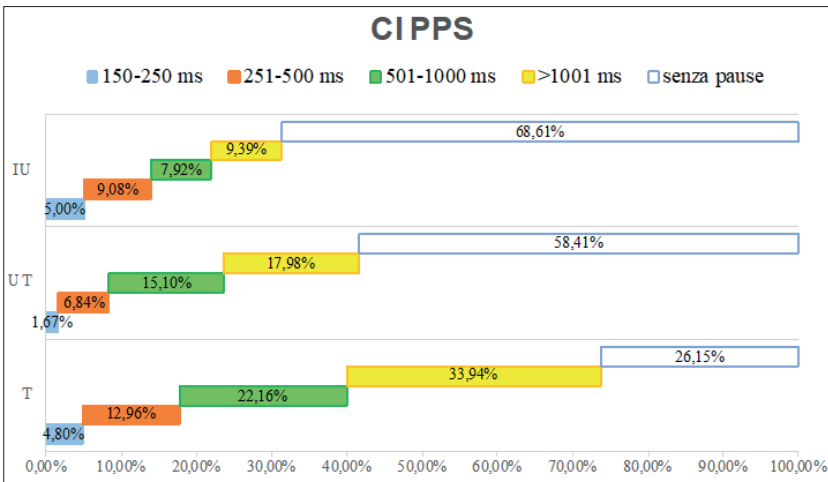


Figura 5 Frequenza delle pause in CIPPS

di turno del soggetto schizofrenico) che, pertanto, sono state escluse dalle analisi successive.

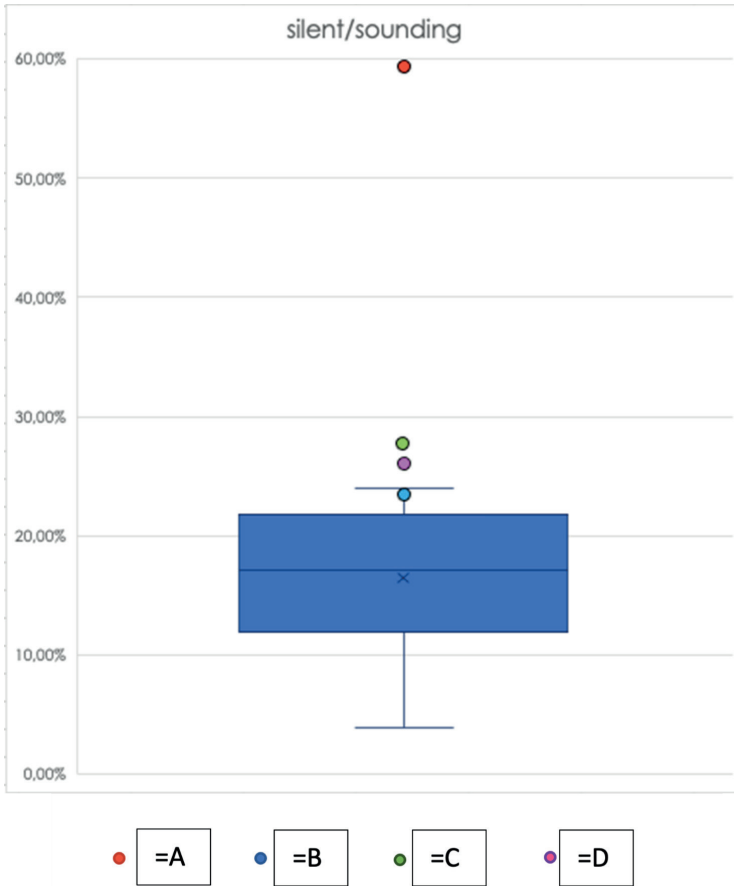


Figura 6 Silenzi IU e UT su T-sounding (Cresti et al. in stampa)

Di conseguenza, il *box plot* in Figura 6 riporta il valore percentuale delle sole pause IU e UT sul totale del *T-sounding*¹⁰ (che appunto esclude i silenzi tra turni) e offre un'immagine chiara e immediata della differente attitudine alla pausazione tra parlato patologico e non patologico.

¹⁰ Cfr. § 2.

Il “dato silente” in un corpus di parlato schizofrenico

Il rettangolo rappresenta la distribuzione maggioritaria delle pause IU e UT nel *T-sounding* del gruppo di controllo: come si legge in corrispondenza dell'asse delle ordinate, il valore oscilla tra il 12 e il 22%. I “baffi” (ovvero le linee inferiore e superiore) mostrano che nel parlato non patologico queste percentuali possono scendere fino a circa il 4% e salire fino al 24% del *T-sounding* totale.

I quattro pallini, invece, mostrano i quattro pazienti di CIPPS, tutti spostati nella parte alta del grafico. Anche se B (pallino più in basso, posto all'altezza del 24% circa di frequenza) è al limite del quartile superiore del Gruppo di controllo, infatti, il paziente mostra ugualmente di produrre una percentuale di silenzi maggiore rispetto alla maggior parte dei soggetti non patologici. C e D producono un parlato in cui si registra circa il 25/27% di pause e, dato più significativo di tutto il *corpus*, A arriva al 60% di silenzio sul totale del *T-sounding*.

Infine, per una validazione delle analisi effettuate, sono stati effettuati test statistici per valutare la variazione interna fra i pazienti CIPPS e quindi la loro maggiore o minore omogeneità. In primis, è stato applicato il test di Kruskal-Wallis, per verificare la similarità tra le mediane della durata delle pause nei 4 pazienti. Il test ha riportato sempre una differenza significativa ($p\text{-value} < 0,01$), evidenziando la non omogeneità del gruppo rispetto alla durata di ciascuno dei tre tipi di pause (UT, IU e T). Poi, è stato effettuato il test di Dunn, che per le pause UT e T ha mostrato la non omogeneità tra tutte le coppie ($p\text{-value} < 0,05$), mentre per le pause di tipo IU ha rilevato una differenza significativa del solo paziente B rispetto agli altri tre ($p\text{-value} < 0,01$ per le coppie B-A, B-C e B-D, diversamente $p\text{-value} > 0,4$ per A-D, A-C e C-D).

5. Conclusioni

I risultati generali confermano le aspettative basate sulla letteratura (GOLDMAN-EISLER 1961; BANFI 1999; HELDNER-EDLUND 2010; DOVETTO-GEMELLI 2013), nonostante la non omogeneità del *corpus*: le pause

del gruppo non patologico sono più corte e quelle del patologico più lunghe.

L'elemento di novità è costituito dalle analisi più fini effettuate sui due *corpora*. Già dalle prime osservazioni sugli andamenti, infatti, emerge una relazione significativa tra la durata e il tipo di pausa: mentre non sono state rilevate differenze per le pause tra unità informative, nel parlato schizofrenico è emerso un quantitativo maggiore di pause tra unità terminate superiori ai 500 ms. Più nel dettaglio, per il Gruppo di controllo, all'aumentare della durata, l'incidenza delle pause IU decresce significativamente (dal 70% nel range 250-500 ms al 36% nel range 1000-5000 ms), mentre quella delle pause UT resta grossomodo la stessa. Contrariamente, per il parlato patologico:

- le pause IU sono più numerose di quelle UT per il range di durata 150-500 ms (53% delle IU vs 35% delle UT);
- le pause IU diminuiscono significativamente nel range 500-1000 ms parallelamente a una preponderanza di pause UT (57% delle UT vs 30% delle IU).

Particolarmente interessanti sono i risultati sulla frequenza delle pause:

- la percentuale di turni e di *utterances* non accompagnati da silenzi è minore in CIPPS che nel Gruppo di controllo, in maniera evidente per le pause T e UT (26,15% vs 63,37% per T e 58,41% vs 68,77% per UT), in maniera trascurabile per le IU (68,61% vs 67,74%);
- nel parlato schizofrenico sono molto più frequenti pause di durata superiore a 1s in tutte e tre le tipologie (33,94% vs 8,09% per le pause T, 17,98% vs 8,75% per le UT, 9,39% vs 3,03% per le IU).

Bibliografia

- ANDREASEN 1986 = NANCY COOVER ANDREASEN, *Scale for the Assessment of Thought, Language, and Communication (TLC)*, in «Schizophrenia Bulletin», 12, 1986, pp. 473-482.
- AUSTIN 1962 = JOHN LANGSHAW AUSTIN, *How To Do Things With Words*, Oxford, The Clarendon Press, 1962, .
- BANFI 1999 = *Pause, interruzioni, silenzi. Un percorso interdisciplinare*, a cura di Emanuele Banfi, Trento, Dipartimento di Scienze Filologiche e Storiche, Labirinti 36, 1999.
- BOERSMA-WEENINK 2021 = PAUL BOERSMA, DAVID WEENINK, *Praat: doing phonetics by computer [Computer program]. Version 6.1.50*, retrieved 20 June 2021 from <http://www.praat.org/>
- COUTURE-PENN-ROBERTS 2006 = SHANNON M. COUTURE, DAVID L. PENN, DAVID L. ROBERTS, *The Functional Signification of Social Cognition in Schizophrenia: A Review*, in «Schizophrenia Bulletin», vol. 32, no. S1, 2006, pp. 44-63.
- COVINGTON ET AL. 2005 = MICHAEL A. COVINGTON, CONGZHOU HE, CATI BROWN, LORINA NAÇI, JONATHAN T. MCCLAIN, BESS SIRMON FJORDBAK, JAMES SEMPLE, JOHN BROWN, *Schizophrenia and the structure of language: the linguist's view*, in «Schizophr Res», 77, 2005, pp. 85-98.
- CRESTI 2000 = EMANUELA CRESTI, *Corpus di italiano parlato*, volume I, Firenze, Accademia della Crusca, 2000.
- CRESTI-MONEGLIA 2005 = C-ORAL-ROM. *Integrated reference corpora for spoken romance languages*, a cura di Emanuela Cresti e Massimo Moneglia, Amsterdam, John Benjamins, 2005.
- CRESTI-MONEGLIA 2017 = EMANUELA CRESTI, MASSIMO MONEGLIA, *Prosodic Monotony and Schizophrenia*, in «Lingua e patologia», Napoli, Aracne, 2017, pp. 147-197.
- CRESTI ET AL. in stampa = EMANUELA CRESTI, MASSIMO MONEGLIA, LORENZO GREGORI, VALENTINA SACCONI, SIMONA TRILLOCCO, *Segmentazione in enunciati del parlato schizofrenico e correlati della patologia nel parlato spontaneo. 4 casi di studio*, Atti del Convegno “Tra medici e linguisti 4: Parole dentro, parole fuori”, Napoli, 13-14 dicembre 2021, in stampa.
- DODANE- HIRSCH 2018 = CHRISTELLE DODANE, FABRICE HIRSCH, *L'organisation spatiale et temporelle de la pause en parole et en discours*, in «Langages», 211, 2018, pp. 5-12.

- DOVETTO-GEMELLI 2013 = *Il parlar matto. Schizofrenia tra fenomenologia e linguistica. Il corpus CIPPS*, a cura di Francesca Maria Dovetto e Monica Gemelli, Prefazione di Federico Albano Leoni, Napoli, Aracne, 2013.
- DOVETTO-CRESTI-ROCHA 2015 = FRANCESCA MARIA DOVETTO, EMANUELA CRESTI, BRUNO ROCHA, *Schizofrenia tra prosodia e lessico. Prime analisi*, in «Studi Italiani di Linguistica Teorica e Applicata», anno XLIV, numero 3, 2015.
- DOVETTO ET AL. 2021 = FRANCESCA MARIA DOVETTO, ALESSIA GUIDA, ANNA CHIARA PAGLIARO, RAFFAELE GUARASCI, LUCIA RAGGI, ASSUNTA SORRENTINO, SIMONA TRILLOCCO, *Corpora di italiano parlato patologico dell'età adulta e senile: CIPPS. CIPP-ma, CIPP-mci*, LIV Congresso Internazionale della Società di Linguistica Italiana, 8-10 settembre 2021.
- DSM 5 2013 = *The Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5) American Psychiatric Association, 2013.
- DUEZ 1985 = DANIELLE DUEZ, *Perception of silent pauses in continuous speech*, in «Language and Speech», vol. 28, part 4, Sage, New York, Sage, 1985, pp. 377-389.
- FAUTH-TRouvAIN 2018 = CAMILLE FAUTH, JÜRGEN TROUVAIN, *Détails phonétiques dans la réalisation des pauses en français: étude de parole lue en langue maternelle vs en langue étrangère*, in «Langages», 211, 2018, pp. 81-95.
- FORS 2011 = KRISTINA LUNDHOLM FORS, *Pause length variations within and between speakers over time*, in Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, Los Angeles, 2011, pp. 198-199.
- GOLDMAN-EISLER 1961 = FRIEDA GOLDMAN-EISLER, *The Rate of Changes in the Rate of Articulation*, in «Language and Speech», 4, 1961, pp. 171-174.
- GIANNINI 2008 = ANTONELLA GIANNINI, *I silenzi del telegiornale*, in *La comunicazione parlata (1)*, Atti del Congresso Internazionale (Napoli 23-25 febbraio 2006), Napoli, Liguori, a cura di Massimo Pettorino et al., 2008, pp. 97-108.
- HA ET AL. 2004 = TAK YOUN, KYOO SEOB HA, KYU SIK RHO, JONG MIN LEE, IN YOUNG KIM, SUN I KIM, JUN SOO KWON, *Gray matter abnormalities in paranoid schizophrenia and their clinical correlations*, in «Psychiatry Res.», 132 (3), 2004, pp. 251-260.
- HELDNER-EDLUND 2010 = MATTIAS HELDNER, JENS EDLUND, *Pauses, gaps and overlaps in conversations*, in «Phonetics», 38, 2010, pp. 555-568.
- HIEKE 1981 = ADOLF E. HIEKE, *A content processing view of hesitation phenomena*, in «Language and Speech», 24 (2), 1981, pp. 147-160.
- JOANETTE ET AL. 2008 = YVES JOANETTE, ANA INES ANSALDO, KARIMA KAHLA-OUI, HÉLÈNE CÔTÉ, VALERIA ABUSAMRA, ALDO FERRERES, ANDRÉ ROCH-LECOURS, *The impact of the right-hemisphere lesions on language abilities. Theoretic and clinical perspectives*, in «Neuropsychological Research. A Re-

Il "dato silente" in un corpus di parlato schizofrenico

- view», a cura di Peter Marien e Jubin Abutalebi, New York, Psychology Press, 2008, pp. 93-111.
- KRAEPELIN 2018 = EMILE KRAEPELIN, *Psychiatrie*, 1904, London, Forgotten Books, 2018.
- KUPERBERG 2010a = GINA R. KUPERBERG, *Language and schizophrenia. Part 1: an introduction*, in «Lang Linguist Compass», 2010, pp. 576-589.
- KUPERBERG 2010b = GINA R. KUPERBERG, *Language and schizophrenia. Part 2: what can psycholinguistic bring to the study of schizophrenia... and vice versa?*, in «Lang Linguist Compass», 2010, pp. 590-604.
- LABLITA = <http://corpus.lablita.it/>
- LEA-KLOKER 1975 = WAYNE A. LEA, DEAN R. KLOKER, *Prosodic Aids to Speech Recognition: VI. Timing Cues to Linguistic Structure and Improved Computer Programs for Prosodic Analysis*, 1975.
- LIDDLE ET AL. 2002 = PETER F. LIDDLE, ELTON T. C. NGAN, STEPHANIE L. CAISSIE, CAMERON M. ANDERSON, ALAN T. BATES, DIGBY J. QUESTED, RICHARD WHITE, ROWENA WEG, *Thought and Language Index: An Instrument for Assessing Thought and Language in Schizophrenia*, in «The British Journal of Psychiatry», 181, 2002, pp. 326-330.
- LUCARINI ET AL. 2020 = VALERIA LUCARINI, MARTINE GRICE, FRANCESCO CANGEMI, JULIANE T. ZIMMERMANN, CARLO MARCHESI, KAI VOGLEY, MATTEO TONNA, *Speech Prosody as a Bridge Between Psychopathology and Linguistics: The Case of the Schizophrenia Spectrum*, in «Frontiers in Psychiatry», 2020, pp. 1-8.
- MACWHINNEY 1991 = BRIAN MACWHINNEY, *The CHILDES project: tools for analyzing talk*, Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1991.
- MARTIN 2004 = PHILIPPE MARTIN, *WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora*, in «Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)», Lisboa, Portugal, European Language Resources Association (ELRA), 2004, pp. 537-540.
- MINKOWSKI 1998 = EUGÈNE MINKOWSKI, *La schizofrenia. Psicopatologia degli schizoidi e degli schizofrenici*, traduzione di Giuliana Ferri Terzian, Torino, Einaudi, 1998.
- MONEGLIA 2005 = MASSIMO MONEGLIA, *The C-Oral-Rom resource*, in C-ORAL-ROM. *Integrated reference corpora for spoken romance languages*, a cura di Emanuela Cresti e Massimo Moneglia, Amsterdam, John Benjamins, 2005, pp. 1-70.
- MONEGLIA-RASO 2014 = MASSIMO MONEGLIA, TOMMASO RASO, *Notes on Language into Act Theory (L-Act)*, in *Spoken corpora and linguistic studies*, a cura

- di Tommaso Raso e Heliana Mello, Amsterdam, John Benjamins, 2014, pp. 468-495.
- MOREL 1860 = BÉNÉDICT AUGUSTIN MOREL, *Traité des maladies mentales*, Paris, Masson, 1860.
- PAWEŁCZYK ET AL. 2018 = AGNIESZKA PAWEŁCZYK, MAGDALENA KOTLICKA-ANTCZAK, EMILA ŁOJEK, ANNA RUSZPEL, TOMASZ PAWEŁCZYK, *Schizophrenia patients have higher-order language and extralinguistic impairments*, in «Schizophrenia Research», 192, 2018, pp. 274-280.
- RIMOL ET AL. 2010 = LARS M RIMOL, CECILIE B. HARTBERG, RAGNAR NESVÅG, CHRISTINE FENNEMA-NOTESTINE, DONALD J. HAGLER JR., CHRIS J. PUNG, ROBIN G. JENNINGS, UNN K. HAUKVIK, ELISABETH LANGE, PER H. NAKSTAD, INGRID MELLE, OLE A. ANDREASSEN, ANDERS M. DALE, INGRID AGARTZ, *Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder*, in «Biol. Psychiatry», 68 (1), 2010, pp. 41-50.
- SAVY 2005 = RENATA SAVY, *Specifiche per la trascrizione ortografica annotata dei testi*, in *Italiano Parlato. Analisi di un dialogo*, a cura di Federico Albano Leoni e Rosa Giordano, Napoli, Liguori, 2005, pp. 1-37.
- STEGMAYER ET AL. 2014 = KATHARINA STEGMAYER, HELGE HORN, ANDREA FEDERSPIEL, NADJA RAZAVI, TOBIAS BRACHT, KARIN LAIMBÖCK, WERNER STRIK, THOMAS DIERKS, ROLAND WIEST, THOMAS J. MÜLLER, SEBASTIAN WALTHER, *Ventral striatum gray matter density reduction in patients with schizophrenia and psychotic emotional dysregulation*, in «Neuroimage Clin.», 4, 2014, pp. 232-239.

Riassunto Il contributo indaga il ruolo delle pause nel Corpus Italiano di Parlato Patologico Schizofrenico CIPPS (DOVETTO-GEMELLI 2013) nel quadro teorico della Teoria di Lingua in Atto (L-Act, CRESTI 2000; CRESTI-MONEGLIA 2010; MONEGLIA-RASO 2014). Due sono gli assunti teorici del lavoro: la nutrita letteratura sui silenzi a partire da HEIKE 1981, che considera fondamentale la pausa nel processo di programmazione del parlato, e le riflessioni di ANDREASEN 1986 e LIDDLE 2002 sul *blocking*, silenzio come indice di un'interruzione del pensiero nel parlato patologico. Su questi presupposti, il parlato schizofrenico è: i. sottoposto a uno script automatico su PRAAT (Boersma-Weenink 2021) per la detezione delle pause, che sono ricontrollate manualmente; ii. diviso in “*sounding*”, il parlato, e “*silent*”, che include pause vuote e piene (cfr. *art* and *pau* in FAUTH-THOUVAIN 2018); iii. segmentato in Enunciati e Unità Informative. Il “dato silente” è classificato per: i. durata, la cui soglia minima è di 150 ms (cfr. occlusive in GIANNINI 2008); ii. posizione, per cui le pause sono divise in inter-turni (T) e inter-enunciati (UT), secondo la letteratura

Il “dato silente” in un corpus di parlato schizofrenico

(cfr. *inter-tours* and *intra-tours* in DODANE-HIRSCH 2018, *gaps* and *pauses* in HELDNER-EDLUND 2010; FORS 2011) e anche in intra-enunciati (IU). I primi risultati, oltre a confermare pause complessivamente più lunghe per il parlato patologico (GOLDMAN-EISLER 1961; BANFI 1999; HELDNER-EDLUND 2010), rivelano dati molto significativi: mentre per il parlato normofasico all'aumentare della durata diminuisce l'occorrenza delle pause IU e si mantiene grossomodo costante quella delle UT, muta la situazione per CIPPS: se il numero delle pause IU è maggiore di quelle UT per le durate tra i 250 e i 500 ms, dopo tale soglia l'andamento s'inverte, con un picco significativo delle pause UT tra i 500 e i 550 ms.

Abstract The article investigates the role of pauses in the Italian Corpus of Pathological Schizophrenic Speech CIPPS (DOVETTO-GEMELLI 2013) within the theoretical framework of the Theory of Lingua in Atto (L-Act, CRESTI 2000; CRESTI-MONEGLIA 2010; MONEGLIA-RASO 2014). There are two theoretical assumptions of the work: the large literature on silences starting from Heike 1981, which considers the pause in the speech programming process as fundamental, and the reflections of ANDREASEN 1986 and LIDDLE 2002 on *blocking*, that is silence as indication of an interruption of thought in pathological speech. On these assumptions, schizophrenic speech is: i. subjected to an automatic script on PRAAT (BOERSMA-WEENINK 2021) for the detection of pauses, manually re-checked; ii. divided into “*sounding*”, speech, and “*silent*”, which includes empty and full pauses (cf. *art* and *pau* in FAUTH-TROUVAIN 2018); iii. segmented into *Utterances* and *Information Units*. The classification is made by: i. duration, whose minimum threshold is 150 ms (see stop-consonants in GIANNINI 2008); ii. position, whereby pauses are divided into inter-turns (T) and inter-utterances (UT), according to the literature (cf. *inter-tours* and *intra-tours* in DODANE-HIRSCH 2018, *gaps* and *pauses* in HELDNER-EDLUND 2010; FORS 2011) and also in intra-utterances (IU). The first results, in addition to confirming overall longer pauses for pathological speech (GOLDMAN-EISLER 1961; BANFI 1999; HELDNER-EDLUND 2010), reveal very significant data: while for not pathological speech, as the duration increases, the occurrence of IU pauses decreases and that of the UTs remains roughly constant, the situation changes for CIPPS: if the number of IU pauses is greater than the UT ones for durations between 250 and 500 ms, after this threshold the trend reverses, with a significant peak of UT pauses between 500 and 550 ms.

La raccolta e l'analisi dei dati linguistici in un'indagine su parlanti bilingui

Yasmina Moussaid

1. Introduzione

Negli studi sul bilinguismo e nella linguistica del contatto è stato fatto vasto utilizzo di tecniche statistiche e metodi quantitativi per rispondere ai vari interrogativi presi in esame. Tuttavia, poiché molte delle variabili studiate nelle ricerche condotte in contesti bilingui richiedono spesso anche un metodo qualitativo di analisi dei dati linguistici, molti degli studi recenti che si sono occupati del contatto linguistico, in circostanze di bi- e multilinguismo, hanno iniziato a combinare gli approcci quantitativi a nuovi approcci di tipo qualitativo. Alcuni tra i primi studiosi che hanno iniziato a servirsi del metodo qualitativo in questo ambito sono, tra gli altri, FISHMAN 1967 e GUMPERZ 1972, i quali intorno agli anni Settanta del secolo scorso hanno iniziato a interessarsi a nuovi aspetti legati al bilinguismo, come la diglossia, i registri stilistici, i dialetti e l'utilizzo delle lingue in ambienti multilingui dal punto di vista sociolinguistico. L'interesse verso queste nuove tematiche li ha portati a introdurre nuovi metodi qualitativi che permettessero l'osservazione dei comportamenti linguistici dei parlanti. Tra questi vi è l'osservazione diretta dei partecipanti, di cui si è servito FISHMAN 1967 in combinazione con la somministrazione di questionari per la raccolta di dati quantitativi, e la raccolta di registrazioni di interazioni verbali di parlanti bilingui di cui ha usufruito GUMPERZ 1972 per studiare come i parlanti creano, definiscono e riproducono il valore simbolico delle varietà linguistiche che padroneggiano. I loro

lavori hanno aperto la strada all'utilizzo di approcci multi-metodologici nelle ricerche sul bilinguismo, che per indagare gli usi linguistici dei parlanti bilingui hanno iniziato a servirsi sempre più di metodi come l'osservazione dei partecipanti, la registrazione di interazioni e le interviste.

Questi ultimi tre metodi appena citati, sono stati tutti utilizzati per realizzare il lavoro di ricerca che sarà illustrato nei paragrafi successivi. In una prima parte, descriverò brevemente lo studio condotto, i suoi obiettivi e le varie fasi di ricerca, mentre, nella seconda parte, presenterò alcune considerazioni sulla metodologia utilizzata, mettendo in evidenza, in particolare, i vantaggi e le difficoltà emerse durante una ricerca che si è avvalsa di una multi-metodologia per la raccolta di dati linguistici.

2. Metodologia e obiettivi della ricerca

La ricerca ha previsto due fasi di lavoro, entrambe volte a esaminare alcune caratteristiche linguistiche e sociolinguistiche di parlanti bilingui. In entrambe le fasi, gli informatori coinvolti sono stati parlanti bilingui italo-arabofoni, appartenenti a una fascia di età compresa tra i 17 e i 27 anni, e di seconda generazione, in quanto nati in Italia da genitori di madrelingua araba. Pertanto, essendo figli di parlanti madrelingua di arabo, hanno avuto modo di acquisire la lingua araba in ambiente familiare e apprendere, fin da bambini, la lingua italiana, grazie all'inserimento nei contesti scolastici e nei diversi contesti sociali. Si tratta dunque di un bilinguismo precoce che i soggetti considerati hanno mantenuto nel tempo, dal momento che, fino ad oggi, hanno tutti avuto modo di alternare costantemente le due lingue conosciute, con la possibilità di utilizzare la lingua araba sia in ambiente familiare sia mantenendo contatti con il Paese di origine e la lingua italiana nei restanti contesti sociali, tendenzialmente non familiari.

Il metodo scelto per condurre la prima indagine è stato quello delle interviste semi-strutturate, le quali, in parte, hanno previsto domande mirate per ottenere i dati linguistici ricercati e, in parte, erano costitu-

ite da domande aperte e discorsive per lasciare spazio anche alla raccolta di informazioni linguistiche non previste dallo studio (Appendice A). Le interviste hanno coinvolto, singolarmente, 11 parlanti italo-arabofoni e si sono svolte in luoghi pubblici. La lingua in cui sono state somministrate è quella italiana e la durata delle interviste, essendo costituite anche da domande aperte, è risultata piuttosto varia: la più breve ha avuto una durata di venti minuti, mentre l'intervista più lunga è durata circa un'ora. Le interviste sono state registrate e successivamente trascritte e grazie alle informazioni raccolte è stato possibile ricavare i dati linguistici di interesse per questa indagine.

In questa prima fase della ricerca, l'obiettivo, da una parte, è stato quello di indagare il legame tra la lingua e l'identità dei parlanti coinvolti e di identificare eventuali cambiamenti a livello della personalità e nel modo di pensare in relazione alle lingue padroneggiate; dall'altra, è stato quello di valutare la consapevolezza linguistica mostrata dai bilingui coinvolti, ovvero quanto questi fossero consapevoli del proprio bilinguismo, specialmente per quanto riguarda il passaggio da una lingua all'altra e le influenze linguistiche tra le due lingue.

Nella seconda fase di questo studio, sono partita dai risultati emersi dalla prima indagine e da quanto i bilingui avevano riferito durante le interviste della prima indagine riguardo agli usi che fanno delle lingue padroneggiate. Lo scopo è stato quello di osservare alcuni fenomeni linguistici che si possono manifestare nei discorsi bilingui, come l'alternanza linguistica, le scelte linguistiche e i vari fenomeni di contatto linguistico, oltre a riprendere anche l'aspetto della consapevolezza linguistica già osservato durante la prima indagine.

I bilingui che hanno preso parte alla seconda indagine della ricerca sono stati 5 e sono stati individuati tra gli 11 italo-arabofoni che hanno partecipato alla prima indagine. In questo caso, poiché l'indagine non si è limitata all'analisi delle informazioni direttamente riferite dai parlanti bilingui – come è accaduto nella prima indagine – ma ha voluto approfondire anche alcuni aspetti linguistici della lingua parlata dei partecipanti, la metodologia scelta ha richiesto ai partecipanti di realizzare delle registrazioni audio di discorsi bilingui per poterne ricavare dei *corpora* linguisticamente analizzabili. Ai bilingui è stato chiesto

di raccontare un episodio, aneddoto o aspetto inerente a determinati contesti proposti, immaginando di riferirsi a un interlocutore in grado di comprendere sia la lingua araba che la lingua italiana, in modo da poter alternare, il più possibile naturalmente, le due lingue. In totale sono stati raccolti un'ora e dieci minuti di discorsi bilingui, che sono stati trascritti manualmente, realizzando dei *corpora*, la cui analisi ha permesso di studiare – come dallo scopo della ricerca – le scelte linguistiche dei parlanti, le alternanze di codice e fenomeni di contatto linguistico.

In sintesi, gli obiettivi generali della ricerca, quindi di entrambe le indagini, sono stati i seguenti:

1. Riflettere sulla relazione tra lingue e identità nel caso di bilingui italo-arabofoni di seconda generazione, in base alle informazioni fornite dagli intervistati.
2. Valutare la consapevolezza linguistica dei partecipanti relativamente ai loro usi linguistici.
3. Osservare, in termini linguistici e attraverso l'analisi dei *corpora* raccolti, le scelte linguistiche, le alternanze tra le due lingue ed eventuali fenomeni di contatto.

3. Riflessioni sulla metodologia

In questo paragrafo sarà commentata la metodologia utilizzata, mettendo in evidenza anche alcuni dei dati linguistici più interessanti emersi nel corso del lavoro di ricerca grazie alla scelta di un approccio multi-metodologico. In particolare, poiché ogni tecnica di ricerca tende a presentare sia dei punti di forza sia delle limitazioni, e poiché lo stesso si può dire per quanto riguarda l'approccio combinato di cui si è servito questo studio, seguiranno osservazioni sia sui limiti dovuti alla scelta di questo tipo di metodologia, sia sui punti di interesse emersi proprio grazie all'utilizzo di un approccio simile.

3.1 Difficoltà e limiti

Come sopra accennato il metodo di ricerca scelto non ha portato soltanto a ottenere dei dati utili ai fini dello studio condotto, ma ha anche mostrato dei limiti che si sono manifestati sotto forma di difficoltà riscontrate durante il lavoro di ricerca.

A questo proposito, la difficoltà principale ha riguardato la gestione della *varietà* dei dati linguistici ricavati dalle risposte ricevute durante le interviste e dall'analisi dei discorsi bilingui. In entrambe le fasi della ricerca, infatti, la varietà dei dati linguistici raccolti è stata evidente. Per quanto riguarda i dati ricavati dalle interviste, questa varietà è stata frutto delle dichiarazioni soggettive dei bilingui coinvolti, specie quando le domande riguardavano l'aspetto identitario e della consapevolezza linguistica (Appendice A). Inoltre, anche nei dati ricavati dalla domanda relativa alla lingua utilizzata nei vari contesti proposti la varietà delle informazioni ricavate è stata evidente, in quanto ogni partecipante ha fornito una propria combinazione delle lingue che utilizza nei vari contesti che si differenziava da quella degli altri partecipanti. Allo stesso modo, durante l'analisi dei *corpora*, la disomogeneità ha riguardato le scelte linguistiche compiute dai parlanti, ma anche le tipologie di alternanze linguistiche e di code-switching che si sono manifestate. Pertanto, non è stato immediato, né lineare, categorizzare e interpretare i dati ottenuti attraverso la combinazione di due metodi qualitativi in una ricerca che si è occupata anche di aspetti legati alla sfera soggettiva dei parlanti bilingui. Di conseguenza, per categorizzare i dati e ottenere dei risultati – per quanto possibile – omogenei, è stato necessario condurre un'accurata analisi, non solo dei singoli dati, ma anche di una loro comparazione e dei contesti e degli atteggiamenti dei bilingui coinvolti nello studio.

3.2. Punti di interesse e strumenti della ricerca

Probabilmente, la difficoltà dovuta alla varietà dei dati ottenuti, descritta nel sottoparagrafo precedente, potrebbe essere stata meno evidente utilizzando un approccio non combinato o facendo uso di

metodi quantitativi in ricerche di questo tipo. Tuttavia, la scelta di un approccio di ricerca qualitativo, in uno studio sul bilinguismo interessato ad aspetti come la consapevolezza e l'identità linguistica dei parlanti, sembra essere stata la scelta più adeguata, in quanto ha permesso di osservare anche aspetti comportamentali e sociolinguistici dei bilingui presi in considerazione. Osservazioni di questo genere, infatti, non sarebbero state possibili scegliendo un approccio quantitativo. Inoltre, il metodo qualitativo ha permesso, allo stesso tempo, di raccogliere anche alcuni dati quantitativi utili, che generalmente sono raccolti attraverso i metodi quantitativi.

In particolare, nella prima fase della ricerca – ovvero quella delle interviste – chiedendo ai partecipanti di raccontare la loro biografia linguistica, sono state raccolte informazioni precise sui bilingui coinvolti, come l'età dei partecipanti, le lingue conosciute, l'età di esposizione alle due lingue e il paese di origine dei genitori in quanto parlanti di madrelingua araba. È stato, così, possibile realizzare dei grafici di carattere quantitativo come il seguente, ricavato dalla domanda che chiedeva ai partecipanti quale lingua utilizzassero in questi determinati contesti.

Come è possibile notare dal grafico (Figura 1) l'approccio metodologico scelto ha permesso di ricavare ugualmente dati quantitativi e percentuali precise riguardo all'uso che i bilingui fanno delle due lingue. In altri termini, la scelta del metodo delle interviste, oltre a dare la possibilità di ottenere dati quantitativi simili a quelli sopra rappresentati, che in genere si ottengono con la somministrazione di questionari, ha permesso di integrarli con ulteriori informazioni linguistiche, ricavate sia dal contenuto delle risposte date dagli intervistati, sia dall'osservazione diretta dei comportamenti dei partecipanti. Inoltre, l'osservazione diretta dei partecipanti è stata particolarmente utile per indagare l'aspetto della consapevolezza linguistica. Quest'ultima non è stata valutata solamente attraverso quanto i bilingui hanno riferito esplicitamente a parole, ma anche attraverso altre caratteristiche comportamentali, come le esitazioni, le ripetizioni, le pause e i silenzi, che si sono manifestate mentre i partecipanti rispondevano alle domande poste. Per esempio, il fatto che le risposte date alla domanda: "Utilizzi

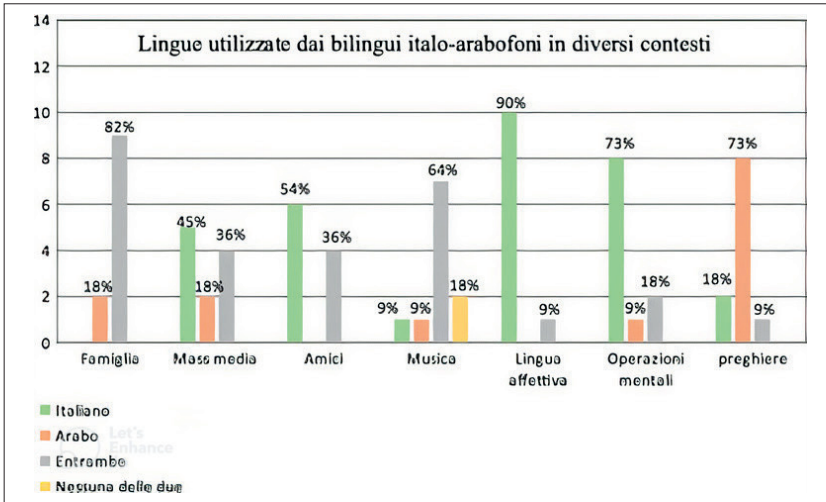


Figura 1 Grafico riportante in che misura i partecipanti italo-arabofoni coinvolti (asse delle ordinate=numero di partecipanti) utilizzano le lingue padroneggiate nei contesti proposti (asse delle ascisse)

maggiormente la lingua italiana o la lingua araba?” spesso non siano state immediate ma abbiano richiesto un tempo riflessione, o abbiano comportato delle esitazioni da parte degli informatori, è stato interpretato come segno di bassa consapevolezza linguistica.

Un ulteriore strumento interessante, legato all'utilizzo delle interviste, è quello delle domande aperte, le quali hanno permesso agli intervistati di rispondere liberamente e di raccontare talvolta anche aneddoti personali riguardo ai loro usi linguistici, fornendo quindi anche nuovi spunti utili per fare riflessioni ulteriori che inizialmente non erano previste dal lavoro di ricerca. Inoltre, è stato possibile spiegare le domande e il loro significato, chiarire eventuali fraintendimenti e richiedere esempi, se necessario. In particolare, nel caso di una delle domande riguardanti le alternanze linguistiche – ovvero quella che chiedeva ai partecipanti se alternassero le due lingue solo all'interno della stessa frase oppure anche all'interno della stessa parola – alcuni intervistati hanno chiesto chiarimenti su come interpretare la doman-

da posta. Questi ultimi, dopo aver inizialmente riferito di non alternare le due lingue all'interno di un unico termine, hanno poi risposto positivamente alla domanda e suggerito degli esempi su questa tipologia di contatto linguistico, di cui solitamente fanno uso. Infine, grazie a questo metodo, si è creata una relazione con i partecipanti che, in alcuni casi, li ha sollecitati a fornire informazioni e dati linguistici anche successivamente al momento dell'intervista. Per quanto riguarda l'aspetto della consapevolezza linguistica, per esempio, successivamente alle interviste, alcuni bilingui hanno specificato che pur riflettendo per giorni su alcune delle domande poste risultava comunque difficile stabilire in modo definitivo quando e quanto utilizzano una lingua anziché l'altra, determinare il momento del passaggio dall'una all'altra, nonché cercare a capire con quale lingua pensano.

In merito alla seconda parte della ricerca, invece, la metodologia adoperata – che, come già anticipatamente descritto nel § 2, è consistita nella realizzazione di *corpora* ricavati dalle registrazioni audio di discorsi bilingui – è stata proficua nei seguenti termini.

In primo luogo, ascoltando le registrazioni audio è stato possibile individuare le scelte linguistiche operate dai bilingui per parlare dei contesti proposti e conseguentemente realizzare, per ogni partecipante, una tabella come la seguente rappresentante le varie scelte linguistiche (tabella 1).

	Lingua di base	Presenza di elementi dell'altra lingua
Familiare/casalingo	Entrambe	
Universitario/lavorativo	Italiano	Sì
Mass media e social media	Italiano	Sì
Operazioni mentali	Italiano	Sì
Tempo libero	Italiano	Sì
Religioso/preghiere	Italiano	Sì
Amici	Entrambe	
Emozioni	Entrambe	

Tabella 1 Esempio di tabella riportante le scelte linguistiche adottate durante le registrazioni raccolte

Pertanto, anche in questo caso, il metodo scelto ha permesso di raccogliere dati di tipo quantitativo.

In secondo luogo, l'aver richiesto, in questa seconda fase, delle registrazioni di lingua parlata a ognuno dei partecipanti che avevano preso parte alla fase delle interviste – per poi ricavarne dei *corpora* linguisticamente analizzabili – ha permesso di individuare e analizzare le commutazioni di codice osservando direttamente i loro usi linguistici quotidiani e non soltanto attraverso i dati che avevano riferito durante le interviste. Quello che segue è uno degli esempi di alternanza linguistica estratti dai *corpus* realizzati.

Esempio 1

Ana ma 'raftš fin glsti nti, comunque la prima, quella dove c'è la lampada e dove ci sono *stilūāt* [...].

L'enunciato appena riportato rappresenta uno dei vari dati linguistici, ricavati dai *corpora*, che “completano” le informazioni fornite dagli intervistati. L'esempio, infatti, mette in evidenza – proprio come avevano riferito i partecipanti – come le commutazioni di cui essi fanno uso si possano manifestare sia a livello interfrasale, sia a livello intrafrasale. In altri termini, le alternanze linguistiche, nei dati raccolti, si sono presentate sia come singole parole pronunciate in una lingua all'interno di una proposizione pronunciata nell'altra lingua – come la parola *stilūāt* nell'es. 1 – sia come sequenze conversazionali più ampie come accade nella prima parte dell'enunciato dove l'intera proposizione iniziale è stata pronunciata in lingua araba. Un'ulteriore tipologia di *code-switching* che i bilingui hanno riferito di utilizzare, e che è stata successivamente riscontrata analizzando i *corpora*, è quella che consiste nella ripetizione dello stesso contenuto in entrambe le lingue utilizzate dai partecipanti, con lo scopo di risultare più chiari o di enfatizzare un determinato contenuto (cfr. MOUSSAID 2023). Infine, la scelta di questa metodologia, basata sull'analisi di *corpora* bilingui, è risultata efficace anche per proseguire con l'approfondimento della consapevolezza linguistica dei bilingui affrontato nella prima fase della ricerca attraverso le interviste e l'osservazione diretta degli atteggiamenti dei

bilingui. Più specificatamente, in questa seconda parte della ricerca, la consapevolezza linguistica è stata indagata attraverso la comparazione dei dati forniti dagli intervistati, relativamente alle lingue che hanno riferito di utilizzare nei vari contesti quotidiani (vedi grafico figura 1), con i dati ricavati dai *corpora* rispetto alle lingue che hanno effettivamente utilizzato nei loro discorsi per parlare di quegli stessi contesti.

Conclusioni

Qualsiasi indagine o ricerca quando giunge al termine lascia spazio a nuovi interrogativi, non solo in termini di nuove domande scientifiche a cui cercare di dare risposte, ma anche in termini metodologici per cercare di migliorare la qualità dei risultati delle future ricerche. Nel caso della ricerca illustrata finora, l'approccio combinato e i due metodi qualitativi scelti per condurre lo studio sembrano essere stati i più adatti per la raccolta e l'analisi dei dati in questo tipo di ricerca su parlanti bilingui. Infatti, hanno permesso di dare delle risposte alle domande su cui vertevano le due indagini, non solo analizzando singolarmente i dati ottenuti da ognuna delle metodologie utilizzate, ma anche confrontando le informazioni linguistiche ottenute dagli informatori con l'analisi linguistica dei dati ricavati dai *corpora* di lingua parlata realizzati. Tuttavia, come sottolineato nel § 3.1, l'utilizzo di un approccio combinato ha avuto anche dei limiti e presentato delle difficoltà, come quella della disomogeneità dei dati raccolti. Per quanto riguarda la metodologia utilizzata, infatti, uno dei principali interrogativi sorti durante il lavoro di ricerca riguarda proprio la gestione delle differenze che hanno caratterizzato i dati raccolti. Tali differenze corrispondono all'esigenza di trovare il metodo più adatto per analizzare e categorizzare dati linguistici come quelli qui esaminati, nonostante la loro varietà e nonostante il coinvolgimento della sfera soggettiva dei bilingui considerati, come è accaduto in questa ricerca che ha indagato aspetti come l'identità linguistica e la consapevolezza linguistica nell'ambito del bilinguismo.

Appendice A

Domande delle interviste

Parte i – Identità linguistica

1. Come ti chiami e come pronunci il tuo nome quando qualcuno ti chiede come ti chiami? In italiano o in arabo?
2. Raccontami la tua biografia linguistica.
3. Di quale cultura senti maggiormente di appartenere? (italiana o araba)
4. Ti senti una persona diversa quando utilizzi una lingua anziché l'altra? Perché?
5. Cambia il tuo modo di pensare in base alla lingua utilizzata?
6. Ci sono cambiamenti nel tuo modo di atteggiarti e comportarti quando parli in italiano rispetto all'arabo o viceversa?
7. Utilizzi atteggiamenti e comportamenti tipicamente italiani quando parli in arabo o viceversa?
8. Dimmi un vantaggio e una difficoltà nel crescere tra due mondi così differenti?
9. Pensa a un posto in cui ti senti a casa... Quale lingua si parla in questo posto?
10. "Parlare due lingue è come vivere due vite". Sei d'accordo? Per te è così?

Parte ii – Consapevolezza linguistica

11. Parli più arabo o più italiano?
12. Quando passi da una lingua all'altra, ci fai caso? Lo fai spesso?
13. Ti capita di mescolare le due lingue? Solo all'interno della stessa frase oppure anche all'interno della stessa parola? Mi sai fare qualche esempio?

Yasmina Moussaid

14. Ti faccio riferimento ad alcuni contesti e mi dici quale lingua utilizzi:
- Famiglia
 - Televisione e musica
 - Amici
 - Lingua affettiva (sentimenti): in quale lingua esprimi le tue emozioni e i tuoi sentimenti più profondi? Se sei arrabbiato quale lingua utilizzi?
 - Operazioni mentali e calcoli matematici
 - Preghiere
15. Pensi che parlare una seconda lingua sia un vantaggio oppure un inconveniente? Da quale punto di vista? Dimmi un'occasione in cui conoscere due lingue è un notevole beneficio. Perché?
16. In quale lingua pensi?
17. In quale lingua sogni?
18. Ti capita di parlare da solo/a? Parli più arabo o più italiano?

Bibliografia

- FISHMAN 1967 = JOSHUA A. FISHMAN, *Bilingualism with and without diglossia, diglossia with and without bilingualism*, In «The Journal of Social Issues», 23, 1967.
- GUMPERZ 1982 = JOHN GUMPERZ, *Discourse strategies*, Cambridge, Cambridge University Press, 1982.
- MOUSSAID 2023 = YASMINA MOUSSAID, *Il contatto linguistico tra identità e consapevolezza: uno studio su bilingui italo-arabofoni*, in «DILEF. Rivista digitale del Dipartimento di Lettere e Filosofia», 2, 2023, pp. 1-18.

Riassunto Indipendentemente dal tipo di osservazione che si vuole condurre, o dalla domanda scientifica a cui si cerca di dare risposta, le ricerche linguistiche comprendono

La raccolta e l'analisi dei dati linguistici

una fase di analisi dei dati, che ha inizio con la raccolta dei dati linguistici, prevede una loro elaborazione e termina con la presentazione dei risultati finali dell'analisi condotta. Ciò accade anche nell'ambito delle ricerche linguistiche condotte in contesti bilingui, o multilingui, come quella qui osservata, la quale è stata condotta con lo scopo di raccogliere e analizzare informazioni linguistiche legate al bilinguismo di parlanti italo-arabofoni. Le osservazioni riportate nei seguenti paragrafi vogliono riflettere sulle fasi di trattamento dei dati linguistici di tale ricerca, che ha combinato l'analisi delle informazioni ricavate da interviste semi-strutturate – rivolte a parlanti bilingui – con l'analisi di *corpora* bilingui, con l'intento di indagare alcune caratteristiche principali del bilinguismo, come le scelte linguistiche dei parlanti coinvolti, i fenomeni di contatto, e ulteriori questioni di tipo sociolinguistico e socioculturale, come l'identità e la consapevolezza linguistica.

Abstract Regardless of the type of observation a linguist wishes to conduct and regardless of which scientific question he seeks to answer, linguistic research includes different stages of data analysis, such as collecting data, processing them, and finally drawing conclusions based on the findings obtained. These phases are equally involved when the research studies are conducted in bilingual or multilingual contexts, as happened in the research study proposed in this paper, which was performed aiming to collect and analyse linguistic data related to the bilingualism of Italian-Arabic speakers. The following paragraphs discuss and reflect on the different stages needed to process the linguistic data of this study, which combined two different qualitative methods. In particular, the study integrated the analysis of linguistic information obtained from semi-structured interviews with the analysis of bilingual *corpora*. The purpose of the research was to investigate some features of bilingualism, such as the linguistic choices made by the speakers involved, language contact phenomena, and some additional sociolinguistic and sociocultural issues, as linguistic identity and language awareness.

Tra lingue e risorse linguistiche: metodi di analisi contrastiva della semantica di verbi di azione

Caterina Cacioli

1. Introduzione

La metodologia di ricerca negli studi contrastivi, in particolare dell'ambito semantico, si compone di diverse strategie di raccolta e analisi dei dati. Le sfide individuate nella ricerca contrastiva in semantica risiedono, in prima istanza, nella scelta delle risorse usate per le analisi, e in seconda istanza nei termini di paragone e nella comparabilità dei dati. Negli ultimi decenni la linguistica contrastiva si è affermata con una proliferazione di lavori grazie anche a risorse adesso a disposizione della ricerca (EBELING 2016; HASSELGÅRD 2020), che hanno promosso nuovi studi e la complementazione di studi precedenti. Si tratta, in particolare, di *corpora* multilingue e paralleli, che mirano a procurare dati altamente comparabili. Ulteriori tecnologie sono però a disposizione e forniscono possibilità per estendere la prospettiva delle ricerche contrastive: oltre ai *corpora*, che rimangono lo strumento più usato (JOHANSSON 2007; EGAN 2013; GRANGER E LEFER 2020), le analisi statistiche avanzate stanno diventando uno standard anche nella ricerca contrastiva (ENGHELS 2020, per esempi: GRIES 2013; LEVSHINA 2015; GRIES, JANSEGGERS E MIGLIO 2020; CACIOLI E VERNILLO 2023). Permane invece una scarsità di lavori che usano risorse multimodali, componente che nella ricerca in semantica risulta di fondamentale importanza per integrare le relazioni tra forme, significati e referenti nel mondo. Lo scopo di questo articolo è proporre l'uso combinato di due risorse linguistiche (una multimodale, *Imagact*, l'altra testuale) per la ricerca

contrastiva nella semantica dei verbi di azione, in relazione a questioni metodologiche, alla selezione dei termini per la comparazione e alla comparabilità dei dati. Da una prospettiva semantica, una questione che merita attenzione è l'individuazione dell'oggetto della ricerca da un punto di vista concettuale (= selezionare un dominio semantico) o linguistico (es., la categoria *verbi* o *aggettivi*), senza sovrapporre questi due livelli. In questo lavoro si affronterà questa questione all'interno degli studi contrastivi e attraverso l'illustrazione dei vantaggi della risorsa Imagact e della combinazione delle risorse, descrivendo il metodo proposto attraverso esempi di un caso di studio sui verbi d'azione.

2. Sfide metodologiche delle analisi contrastive

La linguistica contrastiva, una disciplina di ricerca inaugurata negli anni '50 (VINOGRADOV 2023; NORDRUM, EBELING E HASSELGÅRD 2016), si concentra sull'analisi dettagliata e sottile di somiglianze e differenze tra due lingue, o un piccolo gruppo di lingue, vicine dal punto di vista linguistico o socioculturale (GAST 2012). Le analisi contrastive sono indirizzate su un aspetto specifico, a qualsiasi livello di analisi, e la loro specificità fa emergere aspetti caratterizzanti le lingue in considerazione che solitamente sfuggono alle generalizzazioni tipologiche (KÖNIG 2012; ENGHELS, DEFRANCO E JANSEGGERS 2020) o all'osservazione di tali lingue in isolamento. Il vantaggio dell'uso di due (o poche più) lingue è anche di potersi concentrare su un'analisi qualitativa, a cui ricerche di larga scala non possono arrivare. Le due prospettive, contrastiva e tipologica, sono però collegate: se gli obiettivi delle analisi contrastive sono di scoprire cosa accomuna o differenzia due lingue su un aspetto specifico, al contrario di studi tipologici che ricercano schemi generali, è anche vero che gli studi contrastivi aiutano a indirizzare gli studi tipologici e a dare una prospettiva più specifica nella ricerca delle categorie universali (KÖNIG 2012). Gli studi contrastivi, infatti, non sono di per sé generalizzabili. Le somiglianze o differenze negli studi contrastivi sono definite internamente a un sistema, linguistico o concettuale, in riferimento alle lingue contrastate. L'aggiunta di

una lingua e la revisione dei parametri di paragone genererà risultati diversi oppure emergeranno nuovi aspetti da valutare. L'indagine è quindi *locale* rispetto alle lingue scelte, ma con un vantaggio di personalizzazione delle categorie da comparare: se in uno studio di larga scala si sceglie una categoria linguistica per la comparazione, che dovrà essere un'accomodazione generalizzabile a tutte le lingue (si veda HASPELMATH 2010 per una panoramica sulle categorie linguistiche e i concetti comparabili), negli studi contrastivi c'è un più alto margine di adattabilità al fenomeno che si sta studiando. Somiglianze e differenze sono quindi l'aspetto fondamentale su cui riflettere nella pianificazione e revisione delle ricerche contrastive, e ne rappresentano due grandi sfide legate tra loro: la prima, è relativa alla selezione dei termini di comparazione, la seconda all'effettiva comparabilità dei dati scelti.

2.1 *Tertium Comparationis* e semantica

Il termine di paragone comune alle lingue studiate, cosiddetto *tertium comparationis* (KRZESZOWSKI 1990) è la componente vitale di ogni ricerca contrastiva (CONNOR & MORENO 2005). Al fine di confrontare le strutture linguistiche, le forme o i significati tra le lingue, si deve trovare un riferimento comune, una categoria sovraordinata su cui confrontare il fenomeno. Infatti, a seconda del termine di paragone, due oggetti possono risultare simili o dissimili, e stabilire il punto comune nei termini di comparazione influenza quindi l'intera analisi di un fenomeno (JOHANSSON 2007; ENGHELS 2020). La ricerca di un *tertium comparationis* guida diversi aspetti della programmazione delle analisi contrastive: nel caso degli studi di semantica, un punto focale è la definizione dei concetti su cui concentrare la ricerca, unitamente all'osservazione delle strategie linguistiche usate per istanziare questi concetti e, di conseguenza, quali risorse sono utili per l'analisi e rendere operativa la domanda di ricerca (ENGHELS ET AL. 2020): per esempio, selezionare un concetto come "oggetto su cui sedersi" o delle realizzazioni linguistiche (sedia, poltrona, sgabello...) in più lingue, avranno come esito una diversa selezione di termini di comparazione e dei metodi

di raccolta dati (così come diversi risultati) considerato che le lingue hanno una varietà di categorizzazione linguistica nella denominazione degli artefatti (MALT 2010).

In semantica contrastiva, i termini di paragone sono ricercati in primis nelle risorse lessicografiche, nell'etimologia o negli approcci traduttologici: un'etimologia in comune (ad esempio, COUSSÉ, VAN DER AUWERA 2012), definizioni nei dizionari (bilingue o monolingue) o equivalenze di traduzione. Riguardo la selezione etimologica, studi hanno evidenziato quanto lingue geneticamente vicine possano differire nell'espressione dei domini semantici (MAJID, JORDAN E DUNN 2015): per esempio, in MAJID ET AL. 2007 si dimostra che verbi che in lingue germaniche hanno un'origine comune (inglese *break* e svedese *bräcka*) non sono equivalenti semantici. Già KRZESZOWSKI 1990 aveva mostrato come l'uso degli equivalenti di traduzione sia erroneo in principio per la ricerca in semantica: per esempio, l'uso di *corpora* paralleli per l'estrazione di traduttori ricorrenti non assicura equivalenza semantica, dal momento che una buona traduzione non richiede necessariamente equivalenza semantica.

Un approccio strutturato agli studi di semantica contrastiva emerge in GODDARD E WIERZBICKA 2009: qui si descrive una metodologia contrastiva sulla semantica dei verbi che parte dalla creazione di una lista più esaustiva possibile di qualità che definiscono la semantica di un verbo, analizzando gli usi del verbo nel loro contesto di occorrenza, da testare con giudizi di competenza di parlanti nativi. Il *tertium comparationis* è un dominio concettuale, quello delle *physical activities* in verbi come gli inglesi *cut* e *chop* (italiano *tagliare*), dove la selezione del dominio è operata tramite forme verbali, partendo dai verbi inglese e selezionando i traduttori in polacco e giapponese. Questo procedimento può essere considerato non obiettivo perché sovrappone il dominio concettuale a quello delle forme linguistiche, senza un chiaro termine di riferimento del dominio concettuale stesso. Un altro metodo che si riscontra nella letteratura è lo studio contrastivo di elicitazioni dirette sulla base di un supporto visivo (immagini o video) che limita il dominio concettuale selezionato (tra gli altri, MAJID, BOSTER E BOWERMAN 2008). In questo caso, ci sarà una forte comparabilità delle

risposte, ma l'elicitazione non prenderà in considerazione la varietà di uso in contesti di occorrenza naturale o l'uso in contesti non rappresentati nei supporti visivi.

2.2 Comparabilità dei dati

Una volta stabiliti i termini e le modalità di comparazione, è necessario valutare i dati presi in considerazione. La comparabilità dei *corpora* è sicuramente minore, rispetto a quella del dato elicitato, che può essere tenuto più sotto controllo dal ricercatore, ma i *corpora* rimangono necessari per raccogliere un'ampia prospettiva sui contesti di occorrenza e gli usi. Agli albori della disciplina, i dati più in uso derivavano da grammatiche o risorse lessicografiche con conseguente limitatezza dei dati più spontanei a disposizione e un ampio uso dell'introspezione come metodo di indagine (ENGHELS 2020). L'avvento dei grandi database testuali dagli anni '80 ha dato una svolta alla linguistica contrastiva e i *corpora* sono diventati lo strumento più in uso in questo campo (come testimoniato dai numerosi volumi dedicati, tra gli altri: GRANGER, LEROT, AND PETCH-TYSON 2003; JOHANSSON 2007), dal momento che coprono una grande varietà di generi testuali e di lingue. Per le ricerche contrastive si fa riferimento a *corpora* comparabili o *corpora* paralleli, sulla base degli obiettivi della propria ricerca, tenendo a mente che la piena comparabilità di testi raccolti in un *corpus* provenienti da lingue diverse non potrà essere piena. I *corpora* paralleli, in largo uso negli studi traduttologici (NORDRUM, EBELING E HASSELGÅRD 2016;) e nell'addestramento di risorse per la traduzione automatizzata, sono composti da testi tradotti in due o più lingue: dai *corpora* bilingue e bidirezionali (ovvero, con traduzioni e originali di entrambe le lingue) a *corpora* di lingue europee (EuroParl, con 11 lingue, KOHEN 2005). La comparabilità dei *corpora* paralleli è piena per il tipo di testi che si confrontano (stessi testi, con le loro traduzioni), ma riportano effetti di traduzione (del cosiddetto *traduttorese*, FUSCO 2006) che possono rendere falsificabile ed erronea la comparazione. VOLANSKY, ORDAN E WINTNER 2015 hanno infatti dimostrato che le caratteristiche del traduttorese sono

individuabili con metodi di apprendimento automatico addestrati sui *corpora*, che riescono a distinguere testi originali da testi tradotti. Questi fattori rendono i *corpora* paralleli una risorsa parzialmente affidabile per studi contrastivi di semantica, non applicati alla traduzione. Sicuramente, le preferenze o ricorrenze di traduzione offrono un dato semantico interessante, tuttavia si trova un'alternativa nei *corpora* comparabili. Questi, contengono testi originali, assimilabili per caratteristiche linguistiche o extralinguistiche: il genere testuale, la competenza linguistica dello scrivente (nel caso di *corpora* di apprendenti L2), il dominio di appartenenza del testo o il metodo di raccolta. L'uso di *corpora* paralleli e comparabili combinati è auspicabile (MAURANEN 2005) per complementare le informazioni apportate da ognuna di queste risorse.

3. Combinazione di risorse

Come accennato nel § 2, la ricerca contrastiva si basa fortemente sui *corpora*. Tuttavia, con le nuove tecnologie, si incoraggia la combinazione di più risorse per ampliare le prospettive su un dato fenomeno e ottenere prove convergenti o divergenti dai risultati precedenti (MAURANEN 1999; GILQUIN 2000; ARPPE E JÄRVIKIVI 2007). Per questi motivi, si propone qui l'uso di un'ontologia multimodale e *corpora* comparabili per la ricerca contrastiva in semantica. Si prendono in considerazione esempi di verbi di azione, con particolare riferimento ad azioni su oggetti che provocano un cambiamento di stato irreversibile (ex. rompere un vaso), in tre lingue (italiano, inglese, svedese). L'applicazione della metodologia si inserisce nella letteratura di studi contrastivi su questo dominio concettuale (FILLMORE 1970; MAJID ET AL. 2007, VIBERG 2020) e sulla categoria linguistica dei verbi (VIBERG 2010, 2013; KOPECKA E NARASIMHAN 2012).

3.1 Imagact: scene visive come *tertium comparationis*

IMAGACT è un'ontologia visuale dell'azione, multimodale e multilingua, che integra informazioni visive e linguistiche (MONEGLIA ET

AL. 2012; MONEGLIA ET AL. 2014). Contiene 1010 azioni concrete rappresentate sottoforma di brevi videoclip, mappati su 16 lingue. Ogni video è annotato con i verbi che una lingua può usare per descriverlo e una frase descrittiva standardizzata (soggetto-verbo-oggetto). La risorsa è liberamente consultabile (www.imagact.it) e viene estesa a nuove lingue tramite una procedura di annotazione con parlanti nativi della lingua target. Ci sono tre funzioni di visualizzazione delle informazioni. Il primo è l'uso di una *galleria* di azioni: i video sono divisi in 9 meta-categorie e l'utente può navigare tra le categorie (per esempio, modifiche a oggetti o espressioni facciali) e cercare le azioni di interesse secondo ispezione puramente visiva, per poi accedere all'informazione linguistica associata. La funzione *dizionario* consente all'utente di inserire un verbo in una data lingua in una barra di ricerca e scegliere una lingua target. La risorsa restituisce i video con cui è taggato il verbo ricercato e i verbi associati a quei video nella lingua target. Nella funzione *comparativa*, l'utente può inserire nella barra di ricerca due verbi (in una lingua per il confronto intralinguistico o in due lingue per un confronto interlinguistico). La risorsa restituisce una schermata di confronto tra video con cui sono taggati i verbi oggetto della ricerca, divisi in 3 colonne. Nella colonna di sinistra si trovano i video taggati con il verbo A, nella colonna centrale i video che condividono la possibilità di riferimento con entrambi i verbi e nella colonna di destra i video taggati con il verbo B. Questa risorsa offre un ampio quadro della varietà di azioni che ricorrono nella nostra vita quotidiana con un potente strumento di contrasto. La funzione contrastiva è infatti particolarmente interessante perché offre una possibilità di contrasto dettagliato tra lingue.

L'input di partenza per la scelta dei termini di comparazione può avvenire su base visuale o linguistica. Su base visuale, si seleziona una caratteristica che accomuna delle azioni tramite la galleria e si procede alla selezione delle scene. Si prosegue integrando la componente linguistica, scoprendo i verbi che le descrivono e li si compara. L'altra modalità è su base verbale, per esempio, selezionando due traducanti – due verbi che nei dizionari bilingue sono dati come traduzione reciproca –. La scelta di uno dei due metodi è dettata dalle proprie necessi-

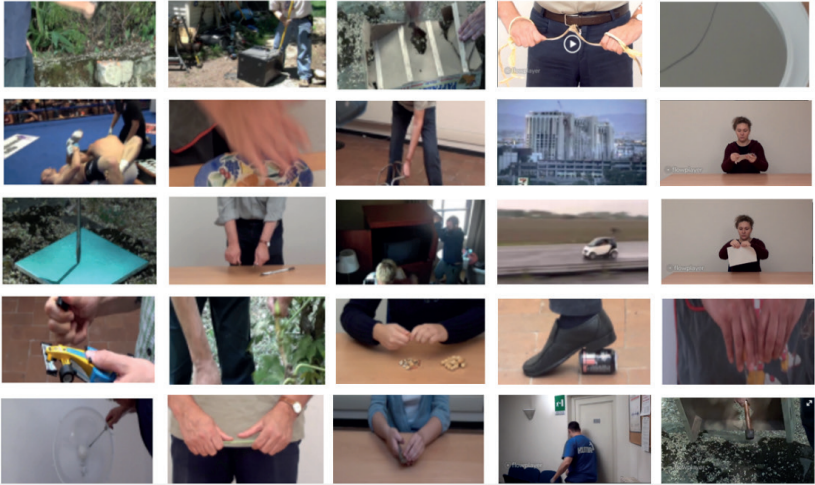


Figura 1 Fermoimmagine delle scene Imagact selezionate secondo l'ispezione visiva di azioni in cui un oggetto perde la sua integrità in maniera distruttiva. Le descrizioni delle scene sono riportate nell'appendice

tà e obiettivi di ricerca. Nel primo caso, si otterrà una selezione su base percettiva che porterà a una valutazione di un dominio semantico-cognitivo e a valutare il modo in cui si associano le forme linguistiche a tali azioni, in più lingue. Nel secondo caso, si otterrà una selezione linguistica che può essere ripetuta in più lingue.

Di seguito, illustriamo queste due possibilità di ricerca contrastiva con uno studio dell'area semantica dei cambi di stato di oggetti. Per l'area "modifica dell'oggetto" la galleria restituisce 313 video (rintracciabili al link: <http://www.imagact.it/imagact/query/gallery.seam>). Questa categoria è quindi piuttosto ampia e comprende azioni molto diverse tra loro: da lavare i piatti, a spostare oggetti, alla loro distruzione. Procediamo quindi selezionando un'area concettualmente coerente all'interno di questa categoria, per rendere possibile il lavoro di confronto ravvicinato: azioni in cui un oggetto perde la sua integrità in modo distruttivo. Così facendo, abbiamo una selezione di 25 scene (riportate in figura 1 e dettagliate nell'appendice).

In ognuna di queste scene, un oggetto perde la sua forma o parte della sua forma iniziale. Confrontando 3 lingue, vediamo la varietà di verbi a loro associati (tabella 1).

Italiano	Inglese	Svedese
rompere	break	göra/ta/ha sönder*
spaccare	smash	slå sönder
spezzare	snap	bryta, bryta av, bryta loss, bryta upp
spezzettare	open	Bräcka, bräcka upp
staccare	crack	falla sönder
sfasciare	crush	slita sönder
sfondare	damage	riva, riva sönder
schiacciare	bust	knäcka
aprire	break open	krossa
forzare	tear	spräcka
dividere		göra hål
sbattere		banka
danneggiare		dela
strappare		krocka
		förstöra
		öppna

* Queste tre varianti sono semanticamente equivalenti in svedese, con un'alternanza regionale (VIBERG 2020). La particella sönder si traduce letteralmente con "a pezzi" ed è unita a *ta* (lett. prendere), *ha* (lett. avere, rendere), *göra* (lett. fare).

Tabella 1 Verbi associati alle scene selezionate, nelle lingue italiano, inglese, svedese

Per prima cosa, possiamo valutare la selezione dei verbi. Dal punto di vista intra-linguistico, questo gruppo di verbi non sono necessariamente sinonimi o in relazioni lessicali (per esempio, di iperonimia). Essendo ogni video annotato con una varietà dei possibili verbi che possono riferirsi a quell'azione, la risultante annotazione restituisce un gruppo eterogeneo. Infatti, ogni azione può essere descritta facendo riferimento a diversi aspetti concettuali¹: possiamo citare il caso di

¹ Per un approfondimento sulla questione, si veda il concetto di equivalenze locali in MONEGLIA 1993; MONEGLIA 2000.

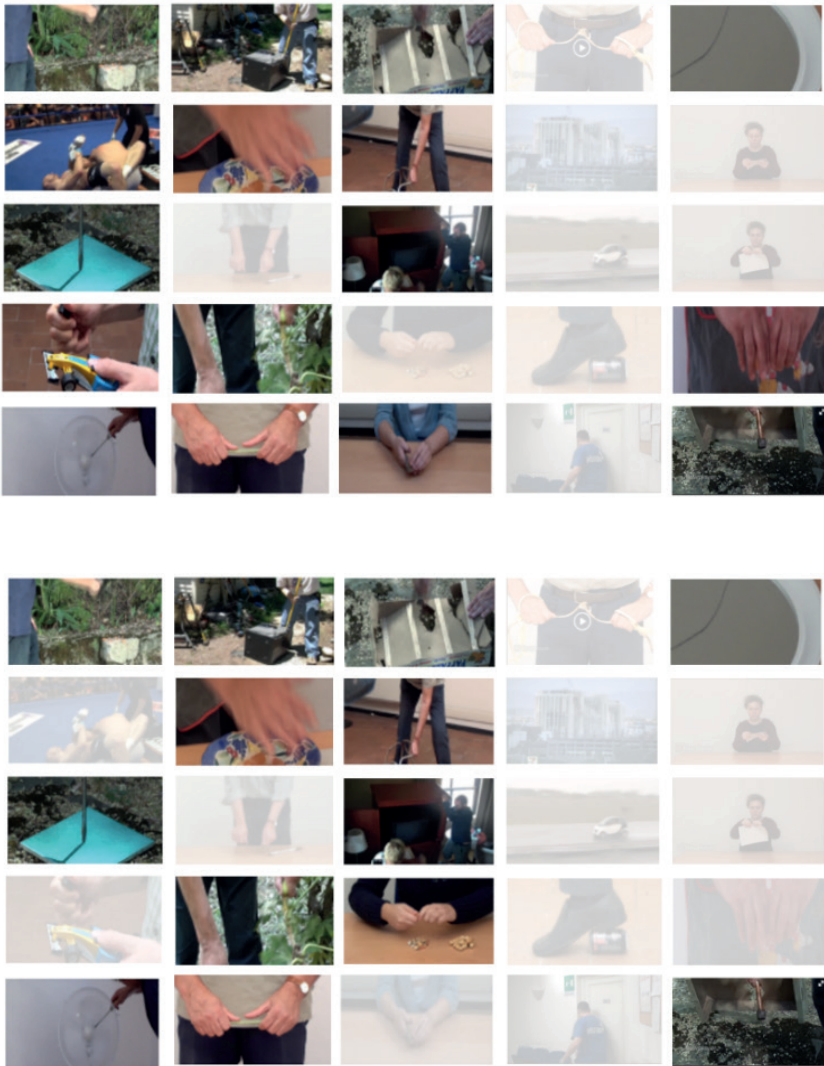


Figura 2 *Romper* e *break*, prima figura, sono interamente sovrapposti. Seconda figura, *ha sönder* – con una diversa selezione delle scene

rompere-aprire, nel contesto, molto frequente in cucina, della manipolazione di uovo al fine di accedere al suo contenuto (ex. Giulia rompe/apre un uovo) che selezionano un diverso aspetto dell'azione che si riferisce a una diversa area concettuale: da una parte l'accesso a un contenuto (*aprire*), dall'altra la manipolazione distruttiva dell'oggetto (*rompere*).

Da un punto di vista contrastivo, si nota una più ampia varietà di verbi svedesi rispetto all'italiano e all'inglese. Se consideriamo i verbi più generali che descrivono queste scene (e quindi, si applicano alla maggior parte delle scene), possiamo disegnare il nucleo linguistico delle scene: i verbi *rompere*, *break*, e *ha/ta/göra sönder* rispettivamente per l'italiano, l'inglese e lo svedese. Tutti e tre questi verbi delimitano eventi di rottura: ne deriva che un'ispezione visiva delle scene conduce a un raggruppamento piuttosto coerente di verbi e che i verbi nelle tre lingue si riferiscono a questa categoria concettuale con un certo grado di equivalenza.

A questo punto possiamo effettuare la ricerca su base linguistica. Per quanto riguarda questo dominio, la selezione visiva si sovrappone quasi completamente alla selezione linguistica effettuata ricercando i singoli verbi *rompere*, *break* e *ha sönder* – che abbiamo detto essere il nucleo centrale del dominio concettuale selezionato. Si aggiunge una sola scena: colpire lo schermo di un pc, che perde la schermata di funzionamento, riferibile a un evento di perdita di funzionalità ma non di integrità visibile esternamente (scena n. 26 nell'appendice). La situazione di sovrapposizione non si mantiene per altri verbi della selezione, che si estendono su altri domini concettuali. È il caso di verbi generali come *aprire/open/öppna* (rispettivamente, italiano-inglese-svedese) che solo in alcuni contesti di distruzione dell'oggetto al fine di accedere a un contenuto sono applicabili ai video selezionati, oppure del verbo *smash*, che ha un'estensione maggiore e condivisa in larga parte con *schacciare*. Lo studio contrastivo può passare alla valutazione delle associazioni dei verbi alle scene in queste lingue, per valutare la loro distribuzione. In svedese, le scene per *ta/ha/göra sönder* sono comprese nella stessa varietà di scene dell'italiano e dell'inglese, ma in minor numero. Si nota quindi che: a. c'è un'alta sovrapposizione dei verbi generali all'interno di un dominio selezionato visivamente, che indica che le tre lingue condividono una partizione concettuale delle azioni; b. *rom-*



Figura 3 Esempio dell'interfaccia utente di Imagact nell'uso della funzione compare con i verbi *spezzare* e *bryta*

pere e *break* hanno una varietà di applicazioni alle scene più ampia di *ha sönder*, che per generalità alterna con *bryta*.

Passando al confronto degli altri verbi, più specifici, che descrivono queste scene, otteniamo ulteriori informazioni di natura contrastiva relativa alla partizione semantica del dominio concettuale. Con la funzione *compare* di Imagact, possiamo valutare i verbi a coppie, inserendo un verbo in una lingua A e uno in una lingua B, fino a esaurire le possibili combinazioni. La figura 3 riporta un esempio dell'interfaccia di Imagact e la figura 4 riporta in forma riassunta l'esito di tale comparazione per tre lingue.

Emerge che *spezzare-snap-bryta* riportano scene in comune, venendosi così a costituire un altro cluster semantico, sottogruppo della selezione iniziale. *Spezzare* e *bryta* condividono due scene, ma non le stesse di *snap* e *bryta*, e ancora una volta non le stesse di *spezzare* e *snap*. Non sono totalmente equivalenti ma catturano una porzione di realtà in comune. Tramite il confronto delle scene denotate da questi verbi si trae informazione contrastiva riportata in figura 4.

Un altro sottogruppo è composto dai verbi *spaccare-smash-slå sönder*. Qui notiamo che *smash* si estende a scene che non erano presenti nella selezione iniziale – ovvero scene in cui non c'è una distruzione ma un cambio nella forma dell'oggetto (corrispondenti all'italiano *schiaccia-*

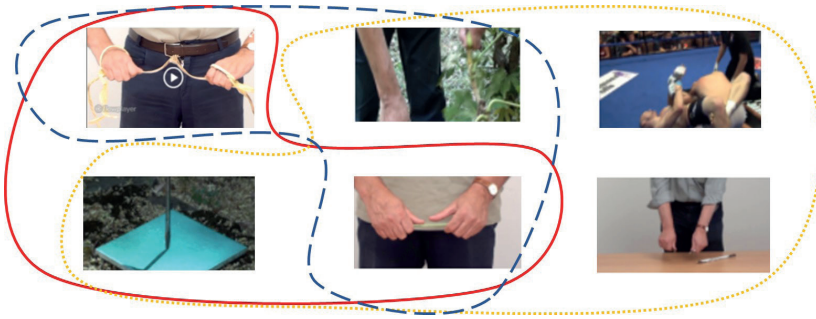


Figura 4 La linea continua delimita *spezzare*, la linea tratteggiata delimita *snap* e la linea puntata delimita *bryta*

re) –. Questo non accade con *spaccare* e *slå sönder*, che rimangono inclusi nella selezione di scene operata su base percettiva.

Alla luce dell'illustrazione della variazione dei verbi nel dominio semantico riportato sopra, vediamo che l'uso del mezzo visivo per disambiguare i concetti dalle forme linguistiche (e operare a un livello extralinguistico) è un grande vantaggio della risorsa, che consente di superare la questione relativa ai concetti e alle forme linguistiche nella ricerca in semantica. Le scene possono essere infatti usate in ottica contrastiva come *tertium comparationis*, il termine di riferimento comune su cui ricercare somiglianze e differenze che altrimenti non sarebbero esplicite. Infatti, questo tipo di informazione non è recuperabile nei dizionari o nei *corpora* (in cui mancando il mezzo visivo, l'azione non può essere giudicata). Esso fornisce una via d'uscita dalle implicazioni metalinguistiche che richiedono l'esplicitazione a mezzo linguistico del significato di un verbo e dal problema illustrato da GODDARD AND WIERZBICKA 2009 e all'alternativa delle descrizioni elicitate, nel § 2.

Sicuramente, l'estensione del *tertium comparationis* qui è limitata dalla natura stessa dei database che non possono contenere infinite scene visive. Catturare tutte le possibili azioni che possono essere compiute nella realtà è sicuramente impossibile. La risorsa è però

un'ampia e solida base di confronto tra verbi ai fini contrastivi, che può essere estesa ad altre lingue, mantenendo inalterato il riferimento di comparazione. Ciò permette inoltre di rendere meno *locali* le ricerche contrastive (come menzionato nel § 1).

Per uno studio contrastivo più completo sui verbi di azione, si deve però rendere conto di una dimensione importante: il contesto naturale di occorrenza e la varietà di contesto linguistico. Per queste ragioni, sarà importante combinare questa risorsa con i dati da *corpora*. Nel paragrafo successivo si suggerisce l'uso di *corpora* comparabili, presentando la famiglia di *corpora* TenTen, come strumento per esaminare la semantica del gruppo di verbi nei loro contesti naturali e privi di effetti di traduzione.

3.2 *Corpora* comparabili

I *corpora* qui proposti come valida risorsa per la linguistica contrastiva sono la famiglia di *corpora* TenTen (JAKUBÍČEK ET AL., 2013). Si tratta una collezione di *corpora* che conta ad oggi 43 lingue².

Lo scopo di questa famiglia di *corpora* è raccogliere materiale autentico dal web selezionando i testi seguendo uno stesso protocollo per tutte le lingue. Viene usato uno strumento di *crawling*, ossia uno strumento di pulizia di testi che elimina il contenuto linguistico indesiderato (generalmente, quello non testuale, come i link e le intestazioni) per poi procedere con tokenizzazione, lemmatizzazione ed etichettatura dei generi testuali. Campioni di testi sono controllati poi con procedure manuali e automatiche per raffinare la selezione ed eliminare testi doppi (per esempio, uno stesso articolo pubblicato due volte). Per molte di queste lingue, nuove versioni dei *corpora* vengono rilasciate a vari intervalli temporali. Per esempio, l'italiano è stato rilasciato nel 2010, 2016 e 2020. Un preciso protocollo di creazione del *corpus* assicura che ogni lingua sia comparabile su determinati parametri, anche se l'aspetto che non può essere controllato è l'esatta comparabilità del contenuto dei te-

² Consultabili sulla piattaforma Sketchengine al link <https://www.sketchengine.eu/documentation/tenten-corpora/>.

sti – la componente culturale può infatti alterare la varietà e qualità dei temi trattati –. Basta pensare al peso che hanno le news nei contenuti web e l’impatto locale che gli argomenti delle news hanno sui contenuti. Questo aspetto, che può porsi come un limite alla comparabilità, va però ad aumentare la validità ecologica dei *corpora* in questione e di un’analisi contrastiva, che non è presente in *corpora* che comparano, per esempio, traduzioni di romanzi in varie lingue. In quel caso, ci sarà piena comparabilità dei temi trattati, ma minore naturalezza di occorrenza.

Il caso di studio qui mostrato utilizza tre *corpora* della famiglia Ten-Ten: ItTenTen 2016 per l’italiano, EnTenTen2015 per l’inglese e SvTen-Ten2014 per lo svedese e un campione di 500 occorrenze per verbi. Nella restante parte di questo paragrafo si evidenzieranno due aspetti: la dimensione concreto-astratto della semantica dei verbi e i temi/pazienti che vengono modificati, per esemplificare l’analisi contrastiva da combinare con Imagact.

L’informazione di Imagact è legata alle azioni concrete. Considerata l’astrattezza dei contenuti linguistici, saranno i *corpora* il luogo in cui cercare una misura della divisione tra riferimenti a concetti concreti o astratti per ogni verbo. La tabella 2 mostra i risultati dell’annotazione manuale³ sulle occorrenze. I verbi riportati sono allineati nella tabella 2 secondo la loro vicinanza semantica, come discussa nel paragrafo precedente. Per lo svedese, si sono annotate due delle tre varianti regionali, *ha sönder* e *göra sönder*.

Italiano	Occorrenze concrete	Inglese	Occorrenze concrete	Svedese	Occorrenze concrete
Rompere	32%	Break	19.20%	Göra sönder Ha sönder	85.4% 89%
Spaccare	33.20%	Smash	60.20%	Slå sönder	60.60%
Spezzare	27.8%	Snap	16%	Bryta	14%

Tabella 2 Percentuali di occorrenze che descrivono un’azione concrete per i tre

3 L’annotazione manuale è stata condotta dall’autrice per le lingue inglese e svedese e da Paola Vernillo per la lingua italiana, con due annotatori esterni madrelingua, per ogni lingua, per la valutazione dell’accordo tra annotatori.

verbi generali delle tre lingue in osservazione in un campione di 500 occorrenze verbali produttive (escludi gli usi idiomatici) per verbo

Le occorrenze del campione mostrano una differenza netta nell'estensione di questi verbi. *Rompere* e *break* hanno un'alta componente metaforica, mentre il corrispettivo svedese si trova principalmente in contesti concreti (e simili percentuali di occorrenze concrete si ritrovano nelle due varianti svedesi *ha sönder* e *göra sönder*). L'italiano *spaccare* ha più usi metaforici rispetto ai vicini *smash* e *slå sönder*, mentre *spezzare*, mentre *snap* e *bryta* hanno più usi metaforici di *spezzare* (sebbene su percentuali inferiori).

Il ruolo tematico del paziente è significativo per l'indagine della semantica verbi di azione di modifica e manipolazione degli oggetti: le proprietà fisiche (come forma e dimensionalità) e funzionali (come l'oggetto viene usato e a quali fini) possono determinare le scelte verbali. Per il caso di studio qui discusso, l'annotazione di occorrenze mostra come i verbi svedesi *ha sönder* e *göra sönder* si estendono a oggetti di stoffa e vestiario in modo più sistematico rispetto all'italiano e all'inglese (1 occorrenza in *rompere*, nessuna in *break*, il 6% sia in *ha sönder* che *göra sönder*). Questo dato è rilevante dal momento che gli studi volti a indagare il concetto di *cambio di stato irreversibile in oggetti* rispettivamente ai verbi che lo realizzano (MAJID ET AL. 2007; MAJID, BOSTER E BOWERMAN 2008) hanno evidenziato, con task di elicitazione, come una dimensione principale della categorizzazione dei verbi d'azioni sia relativa ai materiali dell'artefatto e come la dimensione rigido-flessibile del materiale sia rilevante cross-linguisticamente nell'uso dei verbi, portando esempi di lingue in cui si fa una distinzione o meno (PYE 1996, lingua K'iche' Maya; BOWERMAN 2005, lingua inglese). Tramite *corpora*, si evidenziano sfumature nell'applicabilità di un verbo a categorie discrete (come rigido-flessibile) ed è possibile stimare il grado di estensione del fenomeno grazie alla varietà tematica ampia.

4. Conclusioni

Gli studi di linguistica contrastiva, e in particolare semantica contrastiva, affrontano numerose sfide metodologiche. Per questo motivo le riflessioni sui metodi e la valutazione di strumenti sono ricorrenti e, di fatto, auspicabili per raffinare le strategie della ricerca contrastiva e produrre risultati più robusti. In questo contributo si è fornita una revisione di alcuni di questi aspetti (la comparabilità dei dati e i termini di riferimento) e si è proposto l'uso combinato di due risorse, *Imagact* e *corpora* multilingue comparabili, per contrastare lingue diverse. In particolare, si è mostrato come *Imagact* offra una soluzione alla ricerca del *tertium comparationis* per un'analisi contrastiva sui verbi azionali, utile a superare questioni di selezione del dominio concettuale o linguistico nella ricerca in semantica. Inoltre, la combinazione con i dati testuali è utile a fornire informazioni mancanti (es., sulla componente metaforica dei verbi) o supplementari (es., le associazioni verbo-argomenti).

5. Bibliografia

- ARPPE-JÄRVIKIVI 2007 = ANTTI ARPPE, JUHANI JÄRVIKIVI, *Every Method Counts: Combining Corpus-Based and Experimental Evidence in the Study of Synonymy*, in «Corpus Linguistics and Linguistic Theory», 3 (2), 2007, pp. 131-59.
- BOWERMAN 2005 = MELISSA BOWERMAN, *Why Can't You 'Open' a Nut or 'Break' a Cooked Noodle? Learning Covert Object Categories in Action Word Meanings*, in *Building Object Categories in Developmental Time: 32nd Carnegie Symposium on Cognition*, a cura di L. Gershkoff-Stowe and D. Rakison, Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 33-62.
- CONNOR-MORENO 2005 = ULLA CONNOR, ANA MORENO, *Tertium Comparationis: A Vital Component in Contrastive Rhetoric Research*, in *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*, a cura di Paul Bruthiaux, Dwight Atkinson, William G. Eggington, William Grabe e Vai-dehi Ramanathan, 2005.
- COUSSÉ-VAN DER AUWERA 2012 = EVIE COUSSÉ, JOHAN VAN DER AUWERA, *Human Impersonal Pronouns in Swedish and Dutch: A Contrastive Study of Man and Men*, in «Languages in Contrast», 12 (2), 2012, pp. 121-38.

- DEFRANCQ 2015 = BART DEFRANCQ, *Contrasting Contrastive Approaches*, in «Languages in Contrast», 15 (1), 2015, pp. 1-3.
- EBELING 2016 = JARLE EBELING, *Contrastive Linguistics in a New Key*, in «NJES Nordic Journal of English Studies», 15 (3), 2016, pp. 7-14.
- EGAN 2013 = THOMAS EGAN, *Tertia Comparationis in Multilingual Corpora*, in *Advances in Corpus-Based Contrastive Linguistics: Studies in Honour of Stig Johansson*, a cura di Karin Aijmer and Bengt Altenberg, John Benjamins Publishing Company, 2013.
- ENGHELS-DEFRANCQ-JANSEGGERS 2020 = RENATA ENGHELS, BART DEFRANCQ, MARLIES JANSEGGERS, *Reflections on the Use of Data and Methods in Contrastive Linguistics*, in *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, a cura di Renata Enghels, Bart Defrancq e Marlies Jansegers, 2020, pp. 1-20.
- FILLMORE 1970 = CHARLES J. FILLMORE, *The Grammar of Hitting and Breaking*, in *Readings in English Transformational Grammar*, a cura di Roderick Jacobs e Peter Rosenbaum, Ginn and company, 1970, pp. 9-29.
- FUSCO 2006 = FABIANA FUSCO, *La Traduttologia: Concetti e Termini*, Udine, Forum, 2006.
- GAST 2012 = VOLKER GAST, *Contrastive Linguistics: Theories and Methods*, in *Dictionaries of Linguistics and Communication Science: Linguistic Theory and Methodology*, Berlin, Mouton de Gruyter, 2012.
- GILQUIN 2000 = GAËTANELLE GILQUIN, *The Integrated Contrastive Model: Spicing up Your Data*, in «Languages in Contrast», 3 (1), 2000, pp. 95-123.
- GODDARD-WIERZBICKA 2009 = CLIFF GODDARD, ANNA WIERZBICKA, *Contrastive Semantics of Physical Activity Verbs: 'Cutting' and 'Chopping' in English, Polish, and Japanese*, in «Language Sciences», 31 (1), 2009, pp. 60-96.
- GRANGER-LEROT-PETCH-TYSON 2003 = SYLVIANE GRANGER, JACQUE LEROT, STEPHANIE PETCH-TYSON, *Corpus-Based approaches to Contrastive Linguistics*, Leiden, The Netherlands, Brill, 2003.
- GRIES-JANSEGGERS-MIGLIO 2020 = STEPHAN THOMAS GRIES, TH MARLIES JANSEGGERS, VIOLA G. MIGLIO, *Quantitative methods for corpus-based contrastive linguistics*, in R. Enghels, B. Defrancq, & M. Jansegers (Eds.), *New approaches to contrastive linguistics: empirical and methodological challenges*, vol. 336, 2020, pp. 53-84.
- GRIES 2021 = STEFAN THOMAS GRIES, *Statistics for Linguistics with R: A Practical Introduction*, Berlin, Boston, De Gruyter Mouton, 2021.
- HASPELMATH 2010 = MARTIN HASPELMATH, *Comparative Concepts and Descriptive Categories in Crosslinguistic Studies*, in «Language», 86 (3), 2010, pp. 663-687.

- HASSELGÅRD 2020 = HILDE HASSELGÅRD, *Corpus-Based Contrastive Studies: Beginnings, Developments and Directions*, In *The Complementary Contribution of Comparable and Parallel Corpora to Crosslinguistic Studies*, a cura di Sylviane Granger and Marie-Aude Lefer, John Benjamins Publishing Company, 2020.
- JOHANSSON 2007 = STIG JOHANSSON, *Seeing through Multilingual Corpora*, in «Corpus Linguistics 25 Years on», a cura di Roberta Facchinetti, Leiden, The Netherlands, Brill, 2007, pp. 51-71.
- KOEHN 2005 = KOEHN PHILIPP, *Europarl: A parallel corpus for statistical machine translation*, in *Proceedings of Machine Translation Summit X: Papers*, 2005, pp. 79-86.
- KÖNIG 2012 = EKKEHARD KÖNIG, *Contrastive Linguistics and Language Comparison*, in «Languages in Contrast», 12 (1), 2012, pp. 3-26.
- KOPECKA-NARASIMHAN 2012 = ANETTA KOPECKA, BHUVANA NARASIMHAN, *Events of Putting and Taking: A Crosslinguistic Perspective*, Amsterdam, John Benjamins, 2012.
- KRZESZOWSKI 1990 = TOMASZ P. KRZESZOWSKI, *Contrasting Languages: The Scope of Contrastive Linguistics*, De Gruyter, 1990.
- LEVSHINA 2015 = NATALIA LEVSHINA, *How to Do Linguistics with R*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 2015.
- MAJID-JORDAN-DUNN 2015 = ASIFA MAJID, FIONA JORDAN, MICHAEL DUNN, *Semantic Systems in Closely Related Languages*, in «Language Sciences», 49, 2015, pp. 1-18.
- MAJID-BOSTER-BOWERMAN 2008 = ASIFA MAJID, JAMES S. BOSTER, MELISSA BOWERMAN, *The Cross-Linguistic Categorization of Everyday Events: A Study of Cutting and Breaking*, in «Cognition», 109 (2), 2008, pp. 235-250.
- MAJID ET AL. 2007 = ASIFA MAJID, MARIANNE GULLBERG, MIRIAM VAN STADEN, MELISSA BOWERMAN, *How Similar Are Semantic Categories in Closely Related Languages? A Comparison of Cutting and Breaking in Four Germanic Languages*, in «Cognitive Linguistics», 18 (2), 2007, pp. 179-194.
- MALT 2010 = BARBARA C. MALT, *Naming Artifacts: Patterns and Processes*, in «The Psychology of Learning and Motivation», 52, 2010, pp. 1-38.
- MAURANEN 1999 = ANNA MAURANEN, *Will 'translationese' Ruin a Contrastive Study?*, in «Languages in Contrast», 2 (2), 1999, pp. 161-185.
- MONEGLIA 1993 = MASSIMO MONEGLIA, *La Sottodeterminazione Del Significato Lessicale e l'equiestensionalità Locale Nel Paradigma Di «aprire»*, in «Studi di Grammatica Italiana», 15, 1993, pp. 297-324.

Caterina Cacioli

- MONEGLIA 2000 = MASSIMO MONEGLIA, *Cambiamenti Semantici Palesi e Nasosti Nel Lessico Verbale Italiano: Le Transizioni Di Possesso*, in «Lingua e Stile», XXXV, 4, 2000, pp. 629-664.
- MONEGLIA ET AL. 2012 = MASSIMO MONEGLIA, MONICA MONACHINI, OMAR CALABRESE, ALESSANDRO PANUNZI, FRANCESCA FRONTINI, GLORIA GAGLIARDI, IRENE RUSSO, *The IMAGACT Cross-Linguistic Ontology of Action. A New Infrastructure for Natural Language Disambiguation*, in *Proceedings of the Eight International Conference on Language Resources and Evaluation*, Paris, European Language Resources Association (ELRA), 2012, pp. 948-955.
- MONEGLIA ET AL. 2014 = MASSIMO MONEGLIA, SUSAN W. BROWN, FRONTINI FRANCESCA, GLORIA GAGLIARDI, KHAN FAHAD, MONICA MONACHINI, ALESSANDRO PANUNZI, *The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action*, in *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 3425-32, European Language Resources Association (ELRA), 2014.
- NORDRUM-EBELING-HASSELGÅRD 2016 = LENE NORDRUM, SIGNE OKSEFJELL EBELING, HILDE HASSELGÅRD, *Introduction-Languages in Contrast 20 Years On*, in «NJES Nordic Journal of English Studies», 15 (3), 2016, pp. 1-6.
- PYE 1996 = CLIFTON PYE, *K'iche' Maya Verbs of Breaking and Cutting*, in «Kansas Working Papers in Linguistics», 21, 1996, pp. 87-98.
- VIBERG 2013 = ÅKE VIBERG, *Posture Verbs: A Multilingual Contrastive Study*, in «Languages in Contrast», 13 (2), 2013, pp. 139-169.
- VIBERG 2020 = ÅKE VIBERG, *Contrasting Semantic Fields across Languages*, in *New Approaches to Contrastive Linguistics: Empirical and Methodological Challenges*, De Gruyter, 2020, pp. 265-312.
- VINOGRADOV 2023 = IGOR VINOGRADOV, *Pioneers of Contrastive Linguistics. Dominican Missionaries in Highland Guatemala*, in «Languages in Contrast»; 23 (1), 2023, pp. 34-59.

Sitografia

Imagact = www.imagact.it

Sketchengine = www.sketchengine.eu

Appendice

1. Rompere una bottiglia sbattendola su un muretto
2. Forzare l'apertura di una cassaforte
3. Colpire una scatola fino a distruggere interamente la sua forma originaria
4. Spezzare una corda tirando le estremità
5. Creare una frattura in uno specchio colpendo con uno strumento
6. Rompere un braccio
7. Schiacciare un uovo con una mano
8. Rompere una sedia separando schienale dalla base
9. Crollo di un palazzo
10. Strappare un foglio in due
11. Strappare un foglio in più parti
12. Dividere una mattonella in tre pezzi facendo pressione con un cacciavite
13. Spezzare un bastoncino in più parti
14. Distruggere i mobili di una stanza
15. Distruggere una macchina facendola scontrare con un muro
16. Staccare, rompendo, una ruota da una macchinina giocattolo
17. Staccare un ramo da una pianta
18. Sgusciare delle noccioline
19. Rompere un uovo
20. Spezzare una matita in due
21. Rompere delle noci
22. Dare una spallata a una porta chiusa a chiave e sondare la porta
23. Colpire una scatola di legno con un martello, creando un foro
24. Schiacciare una lattina con il piede
25. Rompere un ventilatore inserendo uno strumento nel meccanismo
26. Colpire un monitor

Caterina Cacioli

Riassunto La ricerca in semantica contrastiva si trova ad affrontare sfide metodologiche sulle risorse usate, la comparabilità dei dati interlinguistici e i criteri dei termini da contrastare. In questo lavoro si tratteranno le questioni menzionate per proporre l'uso combinato di una risorsa linguistica multimodale (Imagact) e di *corpora* per la ricerca in semantica contrastiva dei verbi d'azione. Nel caso di studio qui presentato si mostreranno i vantaggi della combinazione di risorse e di dati linguistici, il contributo della componente visuale della risorsa Imagact risulti nella disambiguazione dei livelli concettuali e linguistici del significato, e l'informazione suppletiva che forniscono i *corpora*.

Abstract Research in contrastive semantics have found methodological challenges along its way, related to the resources used, the comparability of inter-linguistic data and the criteria chosen for comparison. This article will tackle these issues and propose the combination of a multimodal, multilingual resource (Imagact) with *corpora* in the contrastive analysis of action verbs. In the case study presented here, it will be shown which are the advantages of combining resources and linguistic data, how the visual component of the Imagact resource helps disambiguating the conceptual and linguistic levels of meaning-making, and the complementing information provided by *corpora*.

Cov-I-Cor: un corpus di italiano istituzionale relativo alla gestione dell'emergenza sanitaria

Laura Occhipinti

1. Introduzione

La situazione di emergenza legata alla pandemia da Covid-19 ha messo in evidenza la necessità di un'indagine sul linguaggio istituzionale, varietà della lingua nazionale relativa alle comunicazioni ufficiali delle istituzioni.

Un'analisi del linguaggio istituzionale basata su *corpus* può essere interessante per evidenziare strategie e fenomeni linguistici problematici e/o complessi che potrebbero aver influito su una corretta comunicazione e dunque comprensione dei testi da parte dei cittadini. A questo proposito è stato costruito il *corpus* Cov-I-Cor, *corpus* di italiano istituzionale relativo alla gestione dell'emergenza sanitaria.

La comunicazione è sempre centrale in ogni tipologia di interazione che preveda uno scambio di informazioni efficace e lo è, ancor di più, in contesti di emergenza caratterizzati dall'incertezza. In queste situazioni, infatti, alla chiarezza espositiva, e dunque al messaggio, corrisponde una vera e propria risposta attiva da parte dei destinatari che può influire fortemente sull'andamento dell'emergenza stessa. Raffaella Bombi in *Comunicazione istituzionale e Covid-19 tra ricerca e formazione* sottolinea che «la pandemia, come numerosi altri eventi politici, economici, sociali e le trasformazioni scientifico-tecnologiche, [...] ha riportato alla luce il tema della *crisis* e *risk communication*, con particolare attenzione per la comunicazione tra lo Stato e i cittadini»¹. Il tema non

¹ BOMBI 2021, p. 5.

è nuovo dato che l'Organizzazione Mondiale della Sanità (Oms), già a partire dal 2017, ben prima quindi della diffusione del Covid-19, aveva evidenziato la necessità di una comunicazione del rischio chiara e non ambigua, nonostante la consapevolezza dei contesti incerti in cui avviene, attraverso la pubblicazione delle linee guida relative alla comunicazione in contesti di emergenza. Il documento, *Communicating risk in public health emergencies A WHO (World Health Organization) guideline for emergency risk communication (ERC) policy and practice*, facilmente reperibile sul sito dell'OMS², è stato elaborato per i responsabili politici e decisionali nelle situazioni di emergenza e pone al centro numerosi elementi e strategie che sono stati talvolta sottovalutati da parte delle istituzioni o non sempre messi in pratica nel giusto modo, in questa, come in altre occasioni di emergenza.

La pianificazione di un testo e la comprensione possono infatti influire fortemente sul successo (o insuccesso) delle indicazioni e informazioni fornite alla popolazione per rallentare e/o fermare la diffusione dei contagi. Le istituzioni, dunque, sono centrali in processi comunicativi che dovrebbero mirare al coinvolgimento e alla rassicurazione della popolazione colpita, che ha bisogno di sentirsi coinvolta e di fidarsi degli "addetti ai lavori", nazionali, ma anche locali, esistenti per la risposta all'emergenza.

Si è ritenuto dunque interessante costruire una risorsa che miri a raccogliere i testi istituzionali nazionali, scritti, con cui è stata gestita l'emergenza sanitaria. Alle scelte relative alla tipologia di italiano istituzionale preso in esame (§ 2), punto non poco problematico considerando l'ampia varietà del campo di analisi, seguirà una descrizione dettagliata del *corpus* (§ 3), con le diverse fasi di costruzione e raccolta dei testi (§ 4). Si concluderà con una panoramica delle possibilità di analisi da poter condurre su questo *corpus* e sugli obiettivi ultimi del mio lavoro di ricerca (conclusioni).

² Si rimanda al sito dell'Oms per la lettura integrale del documento <https://apps.who.int/iris/bitstream/handle/10665/259807/9789241550208-eng.pdf?sequence=2&isAllowed=y>.

2. Il linguaggio istituzionale

Per avere un campione rappresentativo e bilanciato della varietà di italiano istituzionale che si vuole analizzare, sono necessarie una serie di scelte, qualitative *in primis* e quantitative³ poi, abbastanza problematiche. In effetti, quello del linguaggio istituzionale è un campo di indagine molto ampio: più che di italiano istituzionale è possibile, se non necessario, parlare di linguaggi istituzionali, a seconda «degli scopi e delle sfere di attività delle specifiche organizzazioni del settore pubblico e privato nelle loro comunicazioni ufficiali»⁴. L'attenzione alla selezione del dato linguistico si rivela essenziale dal momento che quest'ultimo rappresenta «l'evidenza empirica su cui fondare lo sviluppo di modelli e teorie linguistiche»⁵.

Ogni linguaggio istituzionale è caratterizzato da una serie di scelte linguistiche e stilistiche legate al contesto extralinguistico, ossia dipendenti dai fattori reali in cui viene utilizzato, che ne determinano forma linguistica e finalità. In primis, è importante ricordare che i destinatari a cui la comunicazione pubblica si rivolge sono molteplici e vanno dal singolo cittadino alla collettività. Con il termine *collettività* si intende «una pluralità di persone considerate nel loro insieme»⁶, come un tutt'uno e dunque si fa riferimento sia all'insieme dei cittadini che alle altre istituzioni, enti o gruppi di persone autonomamente individuabili. Sebbene dunque «tutti i linguaggi istituzionali [siano] mezzi espressivi funzionali a far conoscere i contenuti informativi degli atti giuridici e amministrativi al fine di far attuare le loro disposizioni ai vari livelli di governo»⁷, non è possibile pensarli come un agglomerato unico.

3 LENCI 2005, scheda informativa 1.3 (*Costruire un corpus*).

4 VELLUTINO 2018, p. 84.

5 LENCI 2005, p. 23.

6 Si veda <https://www.treccani.it/vocabolario/collettivita/>.

7 VELLUTINO 2018, p. 85.

La prima macro-distinzione di cui bisogna tener conto nel quadro teorico necessario alla definizione dei criteri di selezione che precede la raccolta vera e propria dei testi è quella tra linguaggi istituzionali speciali e medial⁸. Con i primi ci si riferisce ai linguaggi normativo e amministrativo che, per definizione, hanno delle caratteristiche più rigide legate al codice del diritto che li regola. I testi di queste varietà sono classificati come «vincolanti»⁹ da Sabatini, in relazione al rapporto tra autore e destinatario: «l'autore determina un vincolo più o meno forte nell'interpretazione del testo da parte del destinatario»¹⁰.

Con l'espressione linguaggi istituzionali medial, invece, si intendono i linguaggi più propriamente legati alla comunicazione e informazione sociale: i linguaggi che si propongono di dare informazioni di pubblica utilità, promuovendo la cittadinanza attiva; il linguaggio giornalistico, finalizzato a informare gli organi di stampa; il linguaggio pubblicitario, finalizzato a comunicare l'identità istituzionale e utilizzato dalle campagne di comunicazione pubblica per far conoscere diritti, servizi, opportunità e sensibilizzare ai temi di interesse generale¹¹.

Si è deciso di soffermarsi sui linguaggi istituzionali vincolanti, pensando alla necessità di un intervento che miri ad affrontare il problema “alla base”, a partire dai testi sorgente che regolano la collettività. Ad avvalorare questa scelta risuonano le parole presenti nella Direttiva sulla semplificazione dei testi amministrativi, redatta dal Dipartimento della funzione pubblica nel 2005:

I numerosi atti prodotti dalle pubbliche amministrazioni, sia interni (circolari, ordini di servizio, bilanci) sia esterni, devono prevedere l'utilizzo di un linguaggio comprensibile, evitando espressioni burocratiche e termini tecnici. Anche gli atti amministrativi in senso stretto, che producono effetti giuridici diretti e immediati per i destinatari, devono essere progettati e scritti

8 *Ibidem.*

9 SABATINI 1990, p. 97.

10 RASO 2005, p. 57.

11 Per una disamina completa si rimanda a VELLUTINO 2018, pp. 95-111.

pensando a chi li legge. [...] Devono, perciò, essere sia legittimi ed efficaci dal punto di vista giuridico, sia comprensibili, cioè di fatto efficaci, dal punto di vista comunicativo¹².

Per il concetto stesso di vincolante «il destinatario tendenzialmente non può dare contributi interpretativi al testo e quindi esso deve fornire al lettore tutte le istruzioni per una decodifica precisa»¹³: il linguaggio deve essere chiaro, lineare e comprensibile.

Per bilanciare la raccolta dei testi ai fini della rappresentatività del campione, si è provato a distinguere l'ambito amministrativo, inteso come linguaggio della Pubblica Amministrazione, e quello giuridico, più genericamente inteso come linguaggio che attiene alla sfera giuridico-normativa, cercando di individuare le tipologie di testi connessi a questi due campi. Ad oggi, però, manca una classificazione generale delle tipologie testuali che rientrano nei linguaggi istituzionali vincolanti¹⁴ e spesso c'è confusione o sovrapposizione tra gli ambiti: «i testi del diritto e dell'amministrazione costituiscono un oggetto “dai confini più permeabili e sfrangiati”, non sempre facilmente circoscrivibili nelle griglie ristrette di precise tipologie testuali»¹⁵. In effetti, il confine tra linguaggio giuridico e amministrativo è molto difficile da delimitare, dal momento che molti testi si collocano a metà, basti pensare alle ordinanze, che presentano caratteristiche normative e applicative. Dal punto di vista linguistico, inoltre, il confine sembra ancora meno netto: il linguaggio amministrativo è molto influenzato dal linguaggio

12 Dipartimento della funzione pubblica, *Direttiva sulla semplificazione dei testi amministrativi*, p. 2.

13 RASO 2005, p. 57.

14 Per la classificazione più esaustiva presente in letteratura si rimanda a VIALE 2008. La classificazione proposta risulta abbastanza soddisfacente per quanto riguarda i testi amministrativi, ma alcune tipologie testuali possono rientrare in entrambi gli ambiti. Il criterio utilizzato per distinguerli, che si basa sulla tipologia di destinatario a cui si rivolgono, non è sufficiente: ad entrambi gli ambiti corrispondono sia testi “esterni” che testi “interni”.

15 LUBELLO 2017, p. 11.

del diritto e questo implica che molte forme linguistiche ormai risiedano in entrambe le varietà.

Per i fini più ampi della mia ricerca, che mira ad automatizzare le procedure di semplificazione dei linguaggi istituzionali vincolanti, non è dunque necessaria una separazione tra le due varietà¹⁶. D'altronde, come sostenuto da Raso, i testi amministrativi, anche definiti burocratici, «risentono delle caratteristiche dei testi che ne costituiscono la fonte, cioè i testi legislativi e l'ampia produzione di norme che a essi si collega»¹⁷.

Diversi sono i libri e gli articoli scientifici che trattano il linguaggio giuridico e quello amministrativo come un unico oggetto, basti citare a titolo d'esempio un saggio di Piero Fiorelli del 1994¹⁸ o il più recente lavoro di Lubello del 2017:

si tratta di settori strettamente contigui, [il linguaggio amministrativo], peraltro, è storicamente una variante particolarmente estesa del linguaggio giuridico, con cui intrattiene un legame strutturale, dal momento che quest'ultimo rappresenta la fonte primaria della normativa burocratica [...]. A differenza del linguaggio giuridico, quello amministrativo [...] non è tecnico-specialistico stricto sensu e conosce un variegato spettro di realizzazioni testuali, di contesti d'uso e di destinazioni, applicandosi a un ambito molto ampio di comunicazione: ad accomunare testi molto eterogenei non sono né l'emittente né il destinatario, ma un insieme di scelte linguistiche che delineano un codice scritto formale, tendenzialmente conservativo nelle sue strutture.

16 Dal momento che l'obiettivo più ampio della ricerca è quello di costruire una risorsa parallela semplificata che possa contribuire all'avanzamento dell'automatizzazione dei processi di semplificazione, a interessarci sono principalmente i fenomeni linguistici presenti, come ribadito nel corpo del testo, in entrambi gli ambiti linguistici. Inoltre, è bene ricordare che i dati necessari per le procedure di semplificazione automatica sono, ad oggi molto pochi, in particolare per la lingua italiana. Questo rappresenta sicuramente uno dei problemi centrali nell'avanzamento e nello sviluppo di questo *task* linguistico.

17 RASO 2005, p. 29.

18 FIORELLI 1994.

Si è deciso dunque di bypassare questa divisione e di procedere alla raccolta dei testi su base tematica, in modo da rendere il *corpus* rappresentativo e bilanciato per un'analisi linguistica della gestione pandemica da parte delle istituzioni nazionali. L'interesse di base, infatti, è quello di evidenziare i punti di oscurità e complessità linguistica presenti nelle decisioni che sono state prese e messe in atto, sia in ambito normativo che amministrativo, con lo scopo di proporre una semplificazione, necessaria per rendere i testi più accessibili e fruibili per la popolazione. Questo obiettivo, già teorizzato a partire dagli anni '90¹⁹, si rivela cruciale in questa fase storica, dal momento che la comprensione delle misure adottate è alla base della democraticità di un Paese ed è sinonimo di sostenibilità sociale.

3. Corpus

Il linguaggio istituzionale che si è scelto di analizzare è dunque quello relativo alle istituzioni nazionali. Questa decisione si è rivelata essenziale per ridurre il campo di indagine e condurre un'analisi più focalizzata. Si è deciso di escludere i testi prodotti da istituzioni regionali, provinciali e comunali, per non incorrere in fenomeni ipotetici legati alla variazione diatopica del testo, nonostante ci sia la consapevolezza che questa varietà ha una possibilità minore, seppur non nulla, di incorrere in problematiche di questo tipo²⁰. Inoltre, in un'ottica di sem-

19 Il tema della semplificazione normativa e amministrativa è al centro di un dibattito molto ampio, non solo contemporaneo e relativo a un'unica area: interessa tutte le amministrazioni occidentali. In Italia, il tema viene portato avanti e dibattuto principalmente da professori universitari che, a partire dagli anni '90, con la figura di Sabino Cassese, hanno proposto una serie di iniziative e linee guida per rendere i testi più fruibili per i cittadini. Le varie iniziative «non si sono mai trasformate, però, in una vera e propria campagna, caratterizzata da sistematicità e continuità» (CORTELAZZO 2021, p. 64). Per una rassegna sulle iniziative passate si veda CORTELAZZO 2021 pp. 63-74 e LUBELLO 2017, pp. 98-110.

20 Diversi sono gli studi diacronici e sincronici che sono stati condotti sui testi istituzionali di una regione o città specifica. Si cita a titolo di esempio il saggio di NOBILI 2021.

plificazione che si rivolga potenzialmente²¹ all'intera popolazione, si è deciso di concentrarsi sui testi a cui tutti i cittadini italiani, o residenti in Italia, sono stati esposti. Sono state inoltre escluse le audizioni o le conferenze-stampa, sebbene siano state frequenti durante questo “biennio pandemico”, dal momento che fanno riferimento a un canale di trasmissione più vicino al campo dell'oralità, evitando così variazioni problematiche legate all'asse diamesico.

I testi raccolti, dunque, sono tutti i testi nazionali ufficiali scritti, prodotti per fronteggiare l'emergenza sanitaria, in un arco di tempo che va dal 21 gennaio 2020, data delle prime informazioni relative al Covid provenienti dalla Cina, ancor prima che arrivasse in Italia²², al 31 marzo 2022, che segna la fine dello stato di emergenza nel nostro Paese.

Nell'arco di questi due anni, o poco più, sono stati prodotti – esclusivamente a livello nazionale – e raccolti 904 testi, atti prodotti da diverse istituzioni, tra cui Governo, presidente del Consiglio, presidente della Repubblica, Protezione Civile, Aifa (Agenzia italiana del farmaco) e i diversi ministeri.

Si è deciso di selezionare questi testi perché riconosciuti come fonti ufficiali delle comunicazioni a tema Covid. La gran parte di questi atti, in effetti, è stata pubblicata sulla Gazzetta Ufficiale della Repubblica Italiana. Quest'ultima può essere considerata la fonte ufficiale di conoscenza delle norme in vigore in Italia: è uno «strumento di diffusione, informazione e ufficializzazione di testi legislativi, atti pubblici e privati che devono giungere con certezza a conoscenza dell'intera comunità»²³, come si legge nella schermata ufficiale della Gazzetta. Affinché

21 È necessaria l'aggiunta dell'avverbio *potenzialmente* dal momento che si è consapevoli che parlare di semplificazione in senso assoluto potrebbe essere limitante. Consapevoli che la semplificazione dovrebbe essere fatta *ad personam*, o almeno prestando attenzione alle necessità di gruppi più o meno ampi di individui, sulla base del loro *background* culturale, si è optato per una semplificazione generale guidata dalle linee guida prodotte durante questi anni dalle istituzioni e dai progetti universitari portati avanti.

22 Il primo caso ufficialmente certificato sul suolo italiano risale al 29 gennaio 2020, con relativo annuncio di Giuseppe Conte.

23 Si rimanda al sito ufficiale <https://www.gazzettaufficiale.it/>.

una legge ufficiale o un provvedimento entrino in vigore è necessaria la pubblicazione in Gazzetta, motivo per cui i testi che vengono pubblicati dovrebbero essere accessibili all'intera popolazione.

Secondo l'art. 54 della Costituzione italiana, in effetti, «tutti i cittadini hanno il dovere di essere fedeli alla Repubblica»²⁴, ma, per esserne fedeli, è necessario che comprendano tutti i principi e le leggi che regolamentano la vita della e nella Repubblica. È utile ancora ricordare l'art. 5 del nostro Codice penale: «Nessuno può invocare a propria scusa l'ignoranza della legge»²⁵. Questa norma riprende il principio *ignorantia iuris (legis) non excusat* e sembra non ammettere eccezioni. A questo proposito è intervenuta la Corte costituzionale che ha ammesso la scusabilità dell'ignoranza della legge²⁶ nei casi in cui questa ignoranza sia inevitabile, citando a esempio casi di assoluta oscurità legislativa, nonché linguistica. Il confine tra ciò che è oscuro o chiaro non è così nettamente delineabile e non può prescindere dal contesto globale in cui si inserisce: una disposizione che potrebbe essere chiara per qualcuno, con un certo grado di istruzione e un dato contesto socioeconomico, potrebbe essere del tutto oscura per qualcun altro con caratteristiche totalmente opposte.

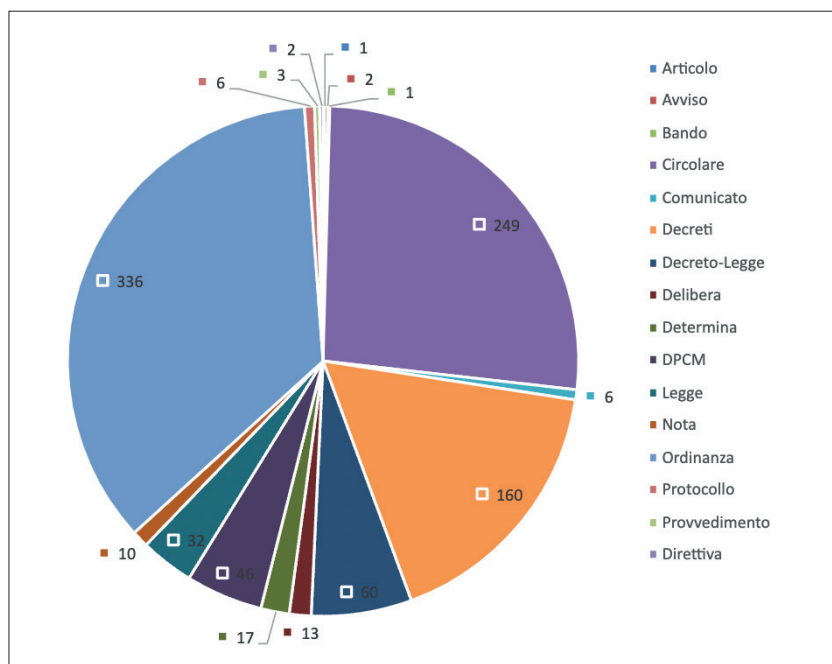
Cov-I-Cor, dunque, può essere descritto come un *corpus* monolingue (italiano), scritto, specialistico²⁷, sincronico. Contiene al suo interno 904 testi, per un numero complessivo di parole pari a circa 4.000 *token*.

24 Costituzione italiana, art. 54. Si rimanda a <https://www.senato.it/istituzione/la-costituzione>.

25 Codice penale, art. 5. Si rimanda a https://www.gazzettaufficiale.it/atto/serie_generale/caricaArticolo?art.versione=2&art.idGruppo=1&art.flagTipoArticolo=1&art.codiceRedazionale=030U1398&art.idArticolo=5&art.idSottoArticolo=1&art.idSottoArticolo1=10&art.dataPubblicazioneGazzetta=1930-10-26&art.progressivo=0.

26 Si veda la sentenza n. 364 del 1988 prodotta dalla Corte costituzionale.

27 Si parla di *corpus* specialistico perché è rappresentativo di una data varietà dell'italiano, ma si vuole chiarire che il linguaggio di cui ci stiamo occupando è sicuramente un linguaggio settoriale ma non del tutto speciale, dal momento che contiene al suo interno diversi linguaggi speciali, afferenti ai diversi campi di cui si occupa.



4. Raccolta testi

Dopo aver fissato i criteri di progettazione del *corpus*, è stato necessario procedere all'acquisizione dei testi: la raccolta è avvenuta con diverse modalità. È importante sottolineare che tutti i testi presi in considerazione sono di natura pubblica, motivo per cui non è stato necessario richiedere permessi particolari legati alla questione, problematica in questo ambito, del *copyright* e, inoltre, a partire dalla legge n. 69 del 18 giugno 2009²⁸ e dal decreto legislativo del 7 marzo 2005 n.

²⁸ La legge n. 69/2009 e, in particolare, l'art. 32, ha stabilito che dal 1° gennaio 2010 «gli obblighi di pubblicazione di atti e provvedimenti amministrativi aventi effetto di pubblicità legale si intendono assolti con la pubblicazione, da parte delle am-

82²⁹, tutti i documenti di interesse per questa ricerca sono pubblici e consultabili online. Tutto ciò ha sicuramente facilitato la raccolta dei testi che è avvenuta, *in primis*, attraverso l'utilizzo di BootCat (Bootstrapping Corpora and Terms)³⁰, la cui funzione principale è quella di creare in maniera semi-automatica *corpora* specialistici di dimensioni medio-piccole estraendo i testi dal web sulla base di *query* fornite dal linguista. Questo strumento è stato utilizzato per provare ad automatizzare il processo di ricerca dei testi dal web: sono state inserite delle *tuple*³¹ di parole chiave da utilizzare come *query* che successivamente sono state trovate e aperte nel browser. Attraverso questa ricerca sono stati evidenziati rapidamente i siti attinenti alle parole fornite in input ed è iniziata l'analisi puntuale di ogni singolo sito web; sono stati così progressivamente esaminati tutti i siti web delle Istituzioni a livello nazionale. Il primo sito emerso dalla ricerca con BootCat è stato sicuramente quello del Governo, o meglio della Presidenza del Consiglio dei ministri, a cui sono seguiti quello del Parlamento, della Presidenza della Repubblica, e i siti dei singoli ministeri, a partire dal Ministero della Salute, punto di riferimento centrale in questa emergenza pandemica. Hanno affiancato il lavoro del Ministero della Salute sicuramente la Protezione Civile, l'Istituto superiore di sanità e l'Aifa, motivo per cui si è ricorso alla consultazione anche dei siti di queste istituzioni. In effetti, numerosi sono stati anche i testi promulgati dai non pochi commissari nominati durante questa fase di emergenza.

Tutti i testi rilevati erano già in formato elettronico; si è proceduto a scaricarli e a categorizzarli per data, tipologia e istituzione di riferi-

ministrazioni e degli enti pubblici obbligati, nei propri siti informatici, o nei siti informatici di altre amministrazioni ed enti pubblici obbligati, ovvero di loro associazioni».

29 Decreto legislativo n. 82 del 7 marzo 2005; si rimanda alla fonte ufficiale per un approfondimento, <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07;82>.

30 BARONI-BERNARDINI, 2004.

31 Si rimanda alla pagina ufficiale di BootCat relativa a "Corpus creation mode": https://docs.sslmit.unibo.it/doku.php?id=bootcat:help:corpus_creation_mode#custom_tuples_advanced.

mento in un documento Excel; questo ha permesso di avere una panoramica chiara del materiale su cui lavorare.

Dopo aver raccolto i materiali digitali, è stato necessario trasformarli nel formato più adeguato al trattamento computazionale, provvedendo alla loro codifica e annotazione. In effetti i testi scaricati e/o salvati non erano in formato txt, ma tutti in formato pdf. A volte i testi disponibili erano già in formato pdf, altre volte è stato necessario salvare le pagine html. Al formato pdf, però, corrispondevano sia documenti in formato testuale, sicuramente meno problematici per la conversione e normalizzazione testuale, sia in formato OCR: ci si è trovati dinanzi a immagini scannerizzate di documenti, la cui conversione è più problematica e presenta documenti che successivamente necessitano di una “pulizia” maggiore.

È stato necessario trasformare tutti i testi in formato txt, un documento di testo standard contenente testo non formattato³². Questa operazione è essenziale sia ai fini dell’etichettatura del testo sia per rendere consultabile i testi senza incorrere in problemi di interscambiabilità e riusabilità: «presenta l’innegabile vantaggio di poter essere gestito da programmi diversi indipendentemente dal sistema operativo»³³.

La conversione dei testi è avvenuta in diversi modi. La maggior parte dei documenti è stata trasformata mediante l’utilizzo di una libreria Python chiamata PyPDF2³⁴ che ha permesso la conversione automatica dei file dal formato pdf a txt. Per alcuni pdf, che erano stati generati da Mac, non è stato possibile convertire i file in questo modo, a causa di problemi di incompatibilità. Pertanto, si è ricorsi all’utilizzo di BootCat che permette di trasformare facilmente i testi in questo

³² Si ricorda che «i formati come doc (il formato dei file di Microsoft Word) o pdf (il formato dei file di Acrobat Adobe) [...] presentano la peculiarità di strutturare un testo digitale in maniera estremamente fruibile per il lettore umano, con tutte le informazioni di formattazione utili alla sua composizione editoriale e alla sua visualizzazione» (LENCI 2005, p. 67).

³³ LENC I 2005, p. 67.

³⁴ <https://pypi.org/project/PyPDF2/>.

formato. Per i file, invece, contenenti immagini di testo scannerizzate, è stato necessario ricorrere a sistemi di riconoscimento OCR. Essendo pochi i documenti in questo formato, è stato utilizzato Adobe Pro³⁵ per riconoscere i caratteri e per salvarli in un nuovo file con la stessa estensione.

È stato necessario poi ripulire i testi, intervenendo su tutti quegli elementi che potrebbero interferire con la consultabilità e linearità del testo:

- sono state eliminate tutte le tabelle presenti, i grafici, le figure, con le relative *caption*, gli indici, le appendici e i numeri di pagina: tutti quegli elementi di difficile leggibilità per un computer e che rappresentano dunque rumore, soprattutto nel cambio codifica;
- tutte le note sono state posizionate alla fine del documento, in modo che non interrompessero il corpo centrale del testo;
- le parole che sono state separate erroneamente nella trasformazione dei documenti, o che si interrompevano nell'andata a capo tra una riga e l'altra, sono state ripristinate;
- sono stati cancellati tutti i simboli relativi a codifiche particolari e dunque non appartenenti al formato Unicode UTF-8;
- sono state sostituite tutte le vocali apostrofate³⁶ con vocali accentate e tutti i *bullet* particolari nei punti elenco con il carattere “-”;
- sono stati cancellati tutti i tab e gli spazi “superflui” presenti nel corpo del testo.

Questa fase di normalizzazione del testo è avvenuta in gran parte manualmente, ma alcune delle operazioni sono state effettuate attraverso il riconoscimento di pattern ricorrenti nei documenti mediante la libreria *re*³⁷ di Python, che consente, attraverso le espressioni regolari di eliminare o trasformare i pattern individuati nel *corpus*.

35 <https://www.adobe.com/it/acrobat/how-to/ocr-software-convert-pdf-to-text.html>.

36 Formattazione utilizzata dalla Gazzetta Ufficiale.

37 <https://docs.python.org/3/library/re.html>.

Una volta ripuliti, i testi sono stati annotati in formato CONLL-U: schema di annotazione afferente alle Universal Dependencies³⁸. I testi sono stati automaticamente annotati mediante l'utilizzo della catena di analisi Stanza³⁹ e della libreria os⁴⁰, necessaria per iterare sui file presenti nelle diverse cartelle. Ogni file, dunque, è stato letto e poi annotato su un nuovo *file script* e presenta il testo segmentato in frasi e per ogni frase c'è un'annotazione in colonne con dati relativi a ID, FORM, LEMMA, UPOS, XPOS, FEATS, HEAD, DEPREL DEPS, MISC.

Si è optato per questo schema di annotazione per due motivi principali: il primo fa riferimento al *task* che ci proponiamo, relativo alla semplificazione linguistica automatica; il secondo fa riferimento alla rappresentatività interlinguistica delle UD.

Diversi studi in letteratura, infatti, mostrano che la semplificazione automatica basata su dipendenze (*dependency-based*) ottiene risultati migliori rispetto alle semplificazioni *constituency-based*⁴¹. Inoltre, la costruzione del *corpus* parallelo che ne seguirà potrebbe essere di aiuto per lo sviluppo di sistemi di semplificazione automatica anche per lingue diverse dall'italiano, per l'interscambiabilità dei dati.

5. Conclusioni

In questo lavoro è stato presentato e descritto il *corpus* Cov-I-Cor. Dopo aver motivato la necessità di un *corpus* specialistico di questo tipo, sono stati illustrati nel dettaglio i criteri teorici che hanno portato alla deli-

³⁸ Le *Universal Dependencies* costituiscono un tentativo di unificazione e conformità rappresentativa interlinguistica: si mira a un'annotazione coerente che riesca a fornire un inventario di categorie e linee guida che facilitino lo sviluppo di un parser multilingue. Per un approfondimento si consulti <https://universaldependencies.org/>.

³⁹ <https://stanfordnlp.github.io/stanza/index.html>.

⁴⁰ <https://docs.python.org/3/library/os.html#os-file-dir>.

⁴¹ Per un approfondimento si veda SIDDHARTHAN 2010 e SIDDHARTHAN 2011.

neazione del campione in esame e la metodologia con cui questi testi sono stati raccolti, ripuliti e annotati.

La costruzione di questa risorsa rappresenta solo il primo step di un lavoro più ampio che mira alla costruzione di un *corpus* parallelo, a partire da una sottosezione di Cov-I-Cor, che presenterà testi originali e semplificati allineati. Il proposito è quello di automatizzare alcuni dei meccanismi di semplificazione proposti e di fornire uno strumento che possa aiutare il cittadino – nella comprensione dei testi ufficiali – e chi scrive ad avere un modello concreto di testo semplificato. Questo strumento può essere molto utile sia per i propositi di semplificazione in generale, sia in questa fase transitoria in cui le numerose linee guida fornite ai “comunicatori ufficiali” non sono state realmente messe in atto nel processo di ideazione e scrittura dei testi.

Bibliografia

- ALEBACHEW ET AL. 2017 = World Health Organization, *Communicating risk in public health emergencies. A WHO guideline for emergency risk communication (ERC) policy and practice*, Geneva, Schweiz, 2017.
- ATKINS-CLEAR-OSTLER 1992 = SUE ATKINS, JEREMY CLEAR, NICHOLAS OSTLER, *Corpus Design Criteria*, in «Literary and Linguistic Computing», volume 7, 1992, pp. 1-16.
- BARONI-BERNARDINI 2004 = MARCO BARONI, SILVIA BERNARDINI, *BootCaT: Bootstrapping Corpora and Terms from the Web*, in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, 2004.
- BARONI-UEYAMA 2006 = MARCO BARONI, *Building general- and special-purpose corpora by Web crawling*, in *Proc. 13th NIJL International Symposium, Language Corpora: Their Compilation and Application*, Tokyo, 2006, pp. 31-40.
- BIBER 1993 = DOUGLAS BIBER, *Representativeness in corpus design*, in «Journal of Literary and Linguistic Computing», 8 (4), 1993.
- BOMBI 2021 = *La comunicazione istituzionale ai tempi della pandemia. Da sfida a opportunità*, a cura di Raffaella Bombi, *Lingue, culture e testi*, Roma, Il Calamo, 2021.
- BRUNATO-DELL'ORLETTA-VENTURI 2022 = DOMINIQUE BRUNATO, FELICE DELL'ORLETTA, GIULIA VENTURI, *Linguistically-Based Comparison of Different*

- Approaches to Building Corpora for Text Simplification: A Case Study on Italian*, «Frontiers in Psychology», vol. 13, Switzerland, Frontiers Research Foundation, 2022.
- BRUNATO ET AL. 2019 = DOMINIQUE BRUNATO, ANDREA CIMINO, FELICE DELL'ORLETTA, SIMONETTA MONTEMAGNI, GIULIA VENTURI, *Trattamento Automatico della Lingua per la comunicazione della Pubblica Amministrazione*, in Proceedings of Ital-IA 2019, 18-19 March, Rome, Italy, 2019.
- CORTELAZZO 2021 = MICHELE A. CORTELAZZO, *Il linguaggio amministrativo*, Roma, Carocci editore, 2021.
- DE MARNEFFE ET AL. 2021= MARIE-CATHERIN DE MARNEFFE, CHRISTOPHER D. MANNING, JOAKIM NIVRE, DANIEL ZEMAN, *Universal Dependencies*, in «Computational Linguistics», vol. 47, no. 2, 2021, pp. 255-308.
- DIPARTIMENTO DELLA FUNZIONE PUBBLICA 2005 = Dipartimento della funzione pubblica, *Direttiva sulla semplificazione del linguaggio delle Pubbliche amministrazioni*, Roma, 2005.
- FIORITTO 1997 = ALFREDO FIORITTO, *Manuale di stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche*, Bologna, il Mulino, 1997.
- LENCI-MONTEMAGNI-PIRELLI 2016 = ALESSANDRO LENCÌ, SIMONETTA MONTEMAGNI, VITO PIRELLI, *Testo e computer. Elementi di linguistica computazionale* (IV edizione), Roma, Carocci editore, 2016.
- LUBELLO 2015 = SERGIO LUBELLO, *Il linguaggio burocratico*, Roma, Carocci editore, 2015.
- NOBILI 2021 = CLAUDIO NOBILI, *Per lo studio dell'italiano burocratico in area campana: ancora sul progetto CUR e presentazione di CorTIBuS*, in BOMBI 2021.
- RASO 2005 = TOMMASO RASO, *La scrittura burocratica*, Roma, Carocci, 2005.
- SABATINI 1999 = FRANCESCO SABATINI, *Rigidità-esplicitzza vs elasticità-implicitzza: possibili parametri massimi per una tipologia dei testi*, in *Linguistica testuale comparativa*, a cura di Francesco Sabatini, Gunver Skytte, Copenhagen, Museum Tusulanum Press, 1999.
- SIDDHARTHAN 2010 = ADVAITH SIDDHARTHAN, *Complex lexico-syntactic reformulation of sentences using typed dependency representations*, in Proceedings of the 6th International Natural Language Generation Conference (INLG 2010), Dublin, Ireland, 2010.
- SIDDHARTHAN 2011 = ADVAITH SIDDHARTHAN, *Text Simplification using Typed Dependencies: A Comparison of the Robustness of Different Generation Strategies*, in Proceedings of the 13th European Workshop on Natural Language Generation, Nancy, France, 2011.
- VELLUTINO 2018 = DANIELA VELLUTINO, *L'italiano istituzionale per la comunicazione pubblica*, Bologna, il Mulino, 2018.

Cov-I-Cor: un corpus di italiano istituzionale

VIALE 2008 = MATTEO VIALE, *Studi e ricerche sul linguaggio amministrativo*, Padova, Cleup, 2008.

Riassunto In questo contributo vengono presentate e discusse le decisioni teoriche e metodologiche che sono state prese per la costruzione del *corpus* Cov-I-Cor, una risorsa linguistica che raccoglie i testi istituzionali utilizzati per gestire l'emergenza sanitaria da Covid-19.

Abstract This paper presents and discusses the theoretical and methodological decisions that were made in the construction of the Cov-I-Cor *corpus*. Cov-I-Cor is a linguistic resource that collects institutional texts used to manage Covid-19 health emergency.

CORMIP

(Corpus Multimodale dell'Italiano Parlato): questioni intorno al trattamento del dato linguistico multimodale

Luca Lo Re

1. Introduzione

Il lavoro presentato in queste pagine propone un metodo per la costruzione di un corpus multimodale dell'italiano parlato, inserendosi nella tradizione dei *corpora* di italiano parlato compilati dal gruppo di ricerca LABLITA dell'Università di Firenze. Il lavoro di ricerca ha dato vita a un piccolo corpus pilota con l'intento di sperimentare il metodo che verrà illustrato e per porre così le basi per ricerche future. Questa ricerca nasce dall'esigenza di approcciarsi alla lingua come sistema multimodale che costruisce ed esprime i significati attraverso diversi indici e canali, in particolare gesto e parlato. Gli studi sul gesto (*gesture studies*) hanno sviluppato approcci e metodi diversi così che ad oggi manca uno standard rispetto alla raccolta e al trattamento dei dati. Inoltre, la complessità dell'oggetto di studio richiede un grande sforzo nella compilazione dei *corpora* portando a una quantità di dati ridotta (dati elicitati in laboratorio o raccolte annotate per specifici obiettivi di ricerca). Il lavoro di ricerca ha l'intento di proporre una metodologia per la compilazione dei *corpora* multimodali, basata su un metodo pragmatico e percettivo.

2. La nozione di multimodalità

Il termine 'multimodalità', coniato a metà degli anni '90, è ampiamente usato da diversi studiosi che a oggi non sono riusciti a costruire una

definizione condivisa e unica dell'oggetto di studio della multimodalità. In modo generico è possibile affermare che l'oggetto di interesse è l'uso di diverse modalità utilizzate per la creazione del significato (GODWIN 2000; KRESS-VAN LEEUWEN 2001).

L'interesse della presente ricerca si rivolge all'italiano parlato spontaneo, concependo la lingua come un'azione che si realizza attraverso diversi mezzi: prosodia, parola, gesto, espressioni facciali e postura; e si basa sui risultati e i riferimenti teorici degli studi sul gesto di Kendon, McNeill, Enfield. Kendon definisce il gesto come «a name for visible action when it is used as an utterance or as a part of an utterance» (KENDON 2004) e vede l'enunciato come «any unit of activity that is treated by those co-present as a communicative 'move', 'turn' or contribution. Such units of activity may be constructed from speech or from visible bodily action or from combinations of these two modalities» (*ibidem*). Invece Enfield parla di un *composite utterance* definendolo «as a communicative move that incorporates multiple signs of multiple types» (ENFIELD 2013). Così l'idea di un enunciato multimodale sembra essere un concetto teorico, basato su prove empiriche, ma che non può essere considerato come unità di riferimento per l'analisi linguistica. Infatti, non esiste alcuna definizione basata su caratteristiche pratiche oltre che sull'enunciato parlato.

3. Il dibattito intorno alla natura multimodale del linguaggio

La nozione di linguaggio come sistema multimodale ci porta a riconoscere che la natura del gesto è verbale (inteso come elemento della comunicazione orale) e che è un canale che esprime diversi valori linguistici, ponendosi in contrapposizione con la nozione di lingua tradizionale caratterizzata dall'arbitrarietà (HOCKETT 1960) e dalla doppia articolazione (MARTINET 1960).

Il punto di vista multimodale sul linguaggio implica che il sistema linguistico sia dinamico e semiologicamente eterogeneo. Infatti, la significatività di un atto linguistico multimodale nasce da una continua dialettica con il contesto in cui si realizza e la concomitanza di diverse

modalità di significazione che non hanno lo stesso grado di convenzionalizzazione. Quindi il sistema lingua non può essere definito su parametri come la doppia articolazione e l'arbitrarietà, poiché farebbero escludere i gesti da questo sistema riducendoli a elementi accessori del linguaggio. La lingua appare come un sistema composto da diversi tipi di modalità semiologiche che interagiscono tra loro per realizzare un'unità globale.

La nozione di linguaggio elaborata da DE MAURO 2000 rappresenta un importante contributo al dibattito. Egli ha sostenuto che non è il carattere dell'audioralità a definire le lingue umane, piuttosto esse si distinguono dagli altri codici comunicativi per la capacità di produrre nuovi significati e nuove parole. Questa possibilità è data dalla capacità, del sistema lingua, di riarticolare i propri segni sia a livello semantico che a livello del significante. L'indeterminatezza del segno permette la riformulazione della relazione tra significante e significato per mezzo del contesto d'uso; infatti, una delle caratteristiche della lingua elencate da DE MAURO 2000 è il carattere locale del funzionamento dei segni e la necessità della realizzazione di intese tra i parlanti. Questa flessibilità, quindi, si realizza nell'uso del linguaggio e rende possibile la ridertiminatezza dei segni linguistici così da estendere il loro significato fino al raggiungimento dell'autoreferenzialità; ed è attraverso l'uso metalinguistico che i parlanti controllano e gestiscono la duttilità del segno nella negoziazione sociale. In questo modo il linguaggio si caratterizza per la sua indeterminatezza data dall'arbitrarietà e dall'uso sociale. Così, De Mauro ha sostenuto che il linguaggio è uno strumento sociale che riafferma continuamente la relazione tra significante e significato nell'uso attraverso l'arbitrarietà e la negoziazione sociale. Ogni sistema linguistico utilizza la sua modalità, la lingua dei segni con il segno e la lingua orale con le parole, ma come abbiamo visto sopra ogni sistema linguistico è multimodale.

In questo modo viene offerta una nozione dinamica di linguaggio in cui il gesto può vedere riconosciuto il suo ruolo e le sue funzioni. In particolare, lo studioso ne individua quattro: 1) appoggio extrafunzionale (inteso come ruolo di scansione e sottolineatura, «un ruolo di appoggio alla scansione sintattica e alla determinazione del senso complessivo di un enunciato», *ibidem*); 2) integrazione alla semantica

di singoli lessemi o gruppi sintagmatici di lessemi; 3) sostituzione semioticamente equivalente; 4) sostituzione semioticamente equivalente (linguaggi speciali degli operatori di borsa, di pescatori, degli addetti aeroportuali ecc.).

Su una prospettiva simile si collocano gli studi di ENFIELD 2009 che cerca di superare la staticità della nozione tradizionale di lingua attraverso un approccio socio-interazionale arrivando a concettualizzare il *composite utterance*. L'enunciato è visto come una unità del comportamento sociale che ha una ben definita relazione causa-effetto e ogni mossa comunicativa scaturisce da determinati presupposizioni e dai *commitments* che vengono riportati nello scambio dei turni dialogici e a cui l'interlocutore è tenuto a rispondere. È proprio il riconoscimento dell'anatomia casuale/condizionale e normativa delle sequenze di interazione, in cui ogni mossa porta a un nuovo obiettivo con conseguenze per i parlanti coinvolti, che permette a Enfield il superamento della staticità del segno linguistico sassuriano. La dinamicità semantica del *composite utterance* è determinata anche dalla diversa natura dei segni che lo compongono: segni convenzionali, segni non convenzionali e segni simbolici indessicali (ENFIELD 2009).

La questione della multimodalità del linguaggio ha portato a rivedere la definizione di lingua da un punto di vista semiotico-linguistico, inglobando l'aspetto sociale e interazionale. A ciò è necessario aggiungere anche l'aspetto cognitivo che influenza in toto l'espressione linguistica, rispetto alla concettualizzazione e all'uso di diverse strategie espressive.

4. Il corpus CORMIP

4.1 La raccolta dei dati: metodi e strumenti

Una prima questione da affrontare per la compilazione del *corpus* CORMIP è stata la definizione di spontaneità dei dati dell'italiano parlato e la definizione di dati multimodali. Di fronte alla complessità delle questioni da affrontare, si è ritenuto opportuno iniziare a elaborare una proposta metodologica attraverso l'uso di un *corpus* pilota che includa

dati sulla gestualità, oltre che sul parlato inteso come canale, lasciando in secondo piano la rappresentatività del *corpus*.

Abbiamo progettato la raccolta dati cercando di garantire la diversificazione delle tipologie interazionali, un livello minimo di variabilità diatopica e la spontaneità degli eventi linguistici.

Per la tipologia interazionale abbiamo seguito il parametro del numero dei locutori «che determina l'ossatura dell'evento comunicativo e permette di distinguere i testi orali in monologhi, dialoghi e conversazioni» (CRESTI 2000) includendo nei dati tre tipologie interazionali:

- monologico, in cui ritroviamo solamente un parlante in un contesto interattivo dove l'interlocutore non può togliere il turno in modo (es. una lezione universitaria);
- dialogico, in cui ritroviamo due parlanti che possono interagire fra loro prendendo il turno liberamente;
- conversazionale, dove troviamo più di tre parlanti che interagiscono potendo prendere il turno liberamente.

La quasi totalità delle registrazioni è ascrivibile a un contesto sociale privato tra persone con un grado di conoscenza reciproca alto, fatta eccezione di un solo brano che riporta una lezione universitaria e dunque è possibile parlare di contesto comunicativo pubblico. Tutte le interazioni sono spontanee e raccolte in un contesto naturale.

Rispetto alla variabilità diatopica, sono stati inclusi solamente due punti di raccolta dati riferibili a due città, Firenze e Catania. Per ciascuna delle due città abbiamo raccolto un brano per ogni tipologia interazionale. Questa nostra scelta intende includere un livello minimo di variazione diatopica.

Raccogliere dati spontanei porta a tener conto del paradosso dell'osservatore (LABOV 1972). Per motivi etici e legali, non è stato possibile realizzare la registrazione senza l'accordo dei partecipanti portando a cercare di attenuarne gli affetti. Così, i parlanti sono stati informati (attraverso l'informativa sulla privacy) sulla finalità delle registrazioni senza però specificare il campo di studi e gli elementi di interesse. È stata utilizzata strumentazione non invasiva, nello specifico: una tele-

camera GoPro Hero 6 e un audio registratore Zoom H6 con un microfono panoramico (120°). Il setting garantiva ai parlanti libertà di movimento, non sono state date loro regole o accorgimenti da seguire e gli argomenti trattati sono totalmente spontanei.

In totale sono state registrate sei diverse situazioni comunicative, tre generi conversazionali per due diverse città. I partecipanti hanno un range di età che va dai vent'anni ai sessanta e il livello di istruzione più basso è il diploma di scuola media superiore. Nella tabella 1 viene illustrato il *dataset* del *corpus*.

Genere conversazionale	Evento comunicativo	Città	Durata
Monologo	Lezione letteratura	Firenze	5'.20"
Dialogo	Dialogo Scout	Firenze	5'.46"
Conversazione	Arbitri pallamano	Firenze	5'.42"
Monologo	Racconto di vita	Catania	5'.27"
Dialogo	Conversazione studenti	Catania	5'.52"
Conversazione	Conversazione viaggio	Catania	7'.12"

Tabella 1 Data-set del corpus CORMIP

4.2 Trascrizione e annotazione dei dati

I sistemi di annotazione del gesto sono diversi e calibrati su metodi e obiettivi propri: NEUROGES (LAUSBERG 2013) (LAUSBERG-SLOETJES 2016), CoGesT (GIBBON ET AL. 2003) (TRIPPEL ET AL. 2004), e LASG (BRESSEM-LADEWIG-MÜLLER 2013).

Questa ricerca ha l'obiettivo di creare un sistema di trascrizione e annotazione che possa identificare le unità strutturali su base percettiva dando centralità alla diversità delle modalità da analizzare. Infatti, i diversi metodi di significazioni di ciascuna modalità (gesto e parlato) dipendono fortemente dalla loro fisicità, pur scaturendo entrambe da un medesimo processo cognitivo.

L'annotazione del parlato fa riferimento alla Teoria della Lingua in Atto (CRESTI 2000), che identifica gli enunciati e le unità di intona-

zione della modalità parlata sulla percezione uditiva; per la trascrizione e l'annotazione gestuale il riferimento teorico è rappresentato dagli studi di KENDON 1972 e MCNEILL 1992. I due livelli di annotazioni sono mantenuti separati in modo da lasciare le informazioni linguistiche dei due canali indipendenti l'uno dall'altro, per poter così ridurre le possibili influenze e per poter indagare in modo dettagliato come le due modalità si correlano. La multimodalità dell'azione linguistica emerge dall'annotazione dell'illocuzione, che rappresenta l'elemento linguistico che caratterizza a nostro avviso l'uso di elementi semantici, intonativi e gestuali.

4.3 La trascrizione e l'annotazione del parlato

La Teoria della Lingua in Atto (CRESTI 2000) si basa sulla teoria degli atti linguistici di AUSTIN 1962. La proposta si poggia su due tipi di unità di riferimento individuate prosodicamente: l'*utterance* e la *stanza*. L'*utterance* è l'unità linguistica minima e principale caratterizzata da un confine prosodico terminato e compie un unico atto linguistico; la *stanza* è formata da una sequenza di *comment* deboli che non corrispondono a una sequenza di enunciati. Le unità di riferimento del discorso sono entità linguistiche basate su caratteristiche semantiche, pragmatiche e prosodiche e la loro identificazione avviene prosodicamente attraverso il riconoscimento percettivo dei confini tonali da parte dell'annotatore. La struttura informativa del parlato è costruita intorno all'unità necessaria e sufficiente chiamata *Comment* e che potrebbe essere accompagnata da altre unità opzionali con le quali forma lo schema informativo. Le unità aggiuntive assumono diverse funzioni: *Topic*, *Parenthesis*, *Appendix*, *Locutive Introducer* e *Discourse Markers*.

La centralità della prosodia all'interno della Teoria della Lingua in Atto si appoggia al modello prosodico elaborato dai lavori di IPO (t HART-COLLIER-COHEN 1990) dimostrando che tra la struttura informativa e struttura prosodica esiste una corrispondenza (MONEGLIA-RASO 2014).

Il quadro teorico della Teoria della Lingua in Atto, oltre a restituirci un metodo di analisi che non può prescindere dalle caratteristiche fi-

siche del parlato, ci fornisce gli strumenti utili per poter segmentare il flusso del parlato in unità prosodiche su base percettiva.

Il formato di trascrizione utilizzato è CHAT-LABLITA ed è stato creato in conformità con l'approccio teorico che implementa il formato CHAT, creato nell'ambito del progetto CHILDES, includendo l'intonazione e la sua funzione di demarcazione delle unità di enunciazione e di informazione (CRESTI 2000; CRESTI-MONEGLIA 2005). Il flusso del parlato è segmentato percettivamente in unità tonali segnate da pause prosodiche che possono essere terminate o non terminate. L'unità prosodica terminata determina i confini dell'enunciato ed è rappresentata con due barre //; mentre l'unità prosodicamente non terminata identifica le altre unità prosodiche all'interno dell'enunciato ed è rappresentata con una sola barra /. Per la trascrizione di altri fenomeni il formato fornisce un repertorio completo come è illustrato nella tabella 2.

Simbolo	Valore
//	Break prosodico terminale
?	Break prosodico terminale con intonazione interrogativa
...	Break prosodico terminale con intonazione sospensiva
+	Break prosodico terminale per sequenza interrotta
/	Break prosodico non terminale
[/]	Falsa partenza con ripetizione
&	Vocalizzazione o frammento di parola
hhh	Fenomeno paralinguistico o non linguistico come tosse o risata
xxx	Parola non comprensibile

Tabella 2 Simboli del sistema di trascrizione CHAT-LABLITA

La natura dialogica dell'evento è stata riportata nella costruzione del *template* di annotazione attraverso l'uso del software ELAN.

Per il modello di Cresti la struttura informativa è pienamente corrispondente alla struttura prosodica, quindi le unità prosodiche, delimitate dai break prosodici, esprimono un valore informativo.

Nell'ottica della Teoria della Lingua in Atto è possibile rintracciare una corrispondenza tra le unità prosodiche di tipo *root* ('t HART-COL-

LIER-COHEN 1990), che sono necessarie e sufficienti per la realizzazione di un pattern tonale, alle unità informative di tipo *comment*. Mentre alle unità prosodiche di tipo *prefix* corrispondono le unità informative di *topic* e a quelle prosodiche di tipo *suffix* le unità informative di *appendice* (FIRENZUOLI 2003; MONEGLIA-RASO 2014). La corrispondenza tra *pattern tonale* e *pattern informativo* risulta rappresentate in modo efficace da una tabella presente in MONEGLIA-RASO 2014 e che riportiamo di seguito:

Prosodic pattern		Information pattern	
Root		Comment	
(Prefix)	(Suffix)	(Topic)	(Appendix)
(Introducer)		(Locutive Introducer)	
(Parenthetical)			
(Incipit)	(Phatic)	(Incipit)	(Phatic)

Tabella 3 Corrispondenza struttura prosodica e struttura intonativa (adattata da Moneglia-Raso 2014)

Il tags-set utilizzato per l'annotazione del parlato tiene conto, pertanto, delle unità prosodiche con l'aggiunta di ulteriori tag per il discorso riportato, contrassegnando le diverse etichette con “_r”. Abbiamo inoltre dedicato un'etichetta per le unità interrotte e una per le unità prosodiche non decifrabili. Nella tabella 4 riportiamo il tag-set per il parlato.

TAG	DEFINIZIONE
ROOT	unità prosodia prominente, necessaria e sufficiente per la realizzazione dell'enunciato
PREFIX	unità prosodica opzionale e subordinata, che occupa una posizione temporalmente antecedente a una Root, una Suffix o un'altra Prefix
SUFFIX	unità prosodica opzionale e subordinata, temporalmente segue le unità di Root o di Prefix

TAG	DEFINIZIONE
INCIPIIT	unità prosodica intonativamente opzionale e subordinata, occorre a inizio turno o enunciato ed è lessicalmente caratterizzate
PHATIC	unità prosodiche subordinata e opzionale, occorre in qualsiasi posizione all'interno dell'enunciato e svolge la funzione comunicativa per il mantenimento dell'apertura del canale
PARENTHETICAL	unità intonative con profilo prosodico basso e realizzata con velocità maggiore rispetto al resto dell'enunciato di cui rappresenta un'inserzione contenutistica
ROOT_r	unità ROOT di parlato riportato
PREFIX_r	unità PREFIX di parlato riportato
SUFFIX_r	unità SUFFIX di parlato riportato
INCIPIIT_r	unità INCIPIIT di parlato riportato
PHATIC_r	unità PHATIC di parlato riportato
PARENTHETICAL_r	unità PARENTHETICAL di parlato riportato
INTERRUPTED	unità prosodica interrotta
EMPTY	unità prosodica non interpretabile

Tabella 4 Tag-set della struttura intonativa di CORMIP

Il tag-set appena riportato ci ha permesso di annotare il flusso del parlato, segmentato percettivamente, restituendo un elenco di unità di base la cui concatenazione, insieme alla sincronizzazione con gli altri canali di espressione linguistica, permette la codifica di valori informativi, illocutivi e semantici. Di seguito vedremo su quali basi abbiamo costruito la trascrizione e l'annotazione del gesto, cercando di mantenere fede all'approccio percettivo e dando dunque risalto al movimento e al canale visivo. L'annotazione sul software ELAN apparirà come nella figura 1.

LUI-UTTERANCE <small>[79]</small> LUI-Intonation Units <small>[119]</small>	va be' / novantotto / arrivi //		
	PREFIX	ROOT	SUFFIX

Figura 1 Esempio di annotazione del parlato di CORMIP su ELAN

4.4 Trascrizione e annotazione del gesto

In un quadro linguisticamente complesso è necessario affermare che il carattere multimodale della lingua – e dunque la sua natura poli-semiotica – debba necessariamente indurci a indagare come i diversi sistemi semiotici riescano a formare un sistema unico e coerente. Alla luce di quanto affermato da DE MAURO 2000, i gesti esprimono diverse funzioni in relazione al loro grado di convenzionalizzazione a cui si associano un crescente grado di iconicità e articolazione. Così, al Kendon's continuum (MCNEILL 1992) si potrebbe associare un continuum di funzioni che vede la corrispondenza tra un gesto iconico e una funzione lessicale e a gesti meno iconici, in cui è difficile distinguere elementi di articolazione, funzioni soprasegmentali.

Nell'attuare il principio di trascrizione e annotazione su base percettiva abbiamo scelto di utilizzare le unità di analisi di KENDON (1972, 1980, 2004). Se nel parlato il flusso fonico è percepito uditivamente, e quindi trascritto e annotato sulla base di ciò che l'annotatore percepisce¹, per il gesto entrano in campo il movimento, l'iconicità e di conseguenza il canale visivo.

La segmentazione è avvenuta su diversi livelli, secondo l'architettura del gesto illustrata da KENDON 2004. Nel livello più alto abbiamo segmentato le *Gesture Unit* (G-UNIT), l'unità gerarchicamente maggiore che ingloba l'intera escursione del movimento, ed è visivamente riconoscibile perché definita dall'inizio del movimento delle mani fino al loro ritorno in posizione di riposo. Il secondo livello di segmentazione è rappresentato dalle *Gesture Phrase* (G-PHR) ed è l'unità che suddivide la *gesture unit*. Può essere definita come l'unità a cui corrisponde un significato comunicativo e si identifica per un particolare movimento nello spazio o per una particolare configurazione delle mani. Il livello più basso è rappresentato dalle *Gesture Phases*, cioè le unità che compongono una *Gesture Phrase* e si distinguono: la *preparation* che è la fase di preparazione del movimento, l'unità di *stroke* che è la fase culminante del gesto ed è l'unità necessaria e suffi-

¹ Nei passaggi più complessi abbiamo risolto i dubbi attraverso l'uso di Praat.

ciente e la *retraction*, che è la fase in cui le mani o le braccia tornano in una posizione di riposo. Inoltre, riconoscendo l'importanza assunta dall'espressione facciale nell'esprimere significati o nel modificarli attraverso l'espressione di un'attitudine o di un'azione linguistica, abbiamo ritenuto necessario aggiungere al primo livello di segmentazione un'etichetta generica che segni la presenza di un'espressione facciale.

La fase di segmentazione e trascrizione del gesto, che è stata fatta separatamente per la mano destra e per la mano sinistra, corrisponde a un'unica fase che si declina su tre livelli.

Nella tabella 5 riportiamo in sintesi i livelli di trascrizione e i simboli utilizzati.

LIVELLO	SIMBOLO	DEFINIZIONE
I	G	unità gerarchicamente maggiore e ingloba l'intera escursione del movimento
I	FACIAL EXPRESSION	espressione facciale
II	PHR	l'unità a cui corrisponde un significato comunicativo e si identifica per un particolare movimento nello spazio o per una particolare configurazione delle mani
III	STROKE	fase culminante del gesto ed è l'unità necessaria e sufficiente
III	PREPARATION	fase di preparazione del movimento
III	RETRACTION	la fase in cui le mani o le braccia tornano in una posizione di riposo

Tabella 5 Tag-set per la trascrizione gestuale di CORMIP

Mentre nella figura 2 mostriamo un estratto di trascrizione del gesto estratta dal software, in cui è possibile notare l'organizzazione dei tiers che riporta la natura gerarchica dell'architettura del gesto.

LUI-G-UNITS [16]		G		
LUI-G-Phrases [29]		PHR		
LUI-G-Phases-RG [67]		STROKE	STROKE	RETRACTION
LUI-G-Phases-LF [64]				
LUI-G-Type [29]		Non-Pictorial		

Figura 2 Esempio di annotazione del gesto in CORMIP su ELAN

Nel *corpus* non è stata aggiunta nessuna annotazione gestuale che si basi su categorizzazioni o classificazioni di funzioni di relazioni con il parlato e/o di gradi convenzionalità del gesto. Perciò sono state utilizzate etichette di natura generale per restituire minime informazioni di base utili ai ricercatori. Queste categorie si basano sull'importanza dell'iconicità per il gesto che si esprime in modo graduale anche in relazione alle funzioni espresse dal gesto².

Così i gesti sono stati divisi in *Pictorial*, *Non-Pictorial* e *Conventional*. Nello specifico, con l'etichetta *Pictorial* abbiamo cercato di raggruppare tutti i gesti che assumono visivamente una forma riconducibile a un'immagine, a un contorno di un oggetto o a un'azione con riferimento a oggetti e cose del mondo reale. Sotto questa etichetta rientrano tutti i gesti che possono essere classificabili come iconici, pittografici, ideografici o rappresentativi, sulla base delle strategie di rappresentazione individuate da MÜLLER 2013. Mentre con l'etichetta *Non-Pictorial* sono annotati tutti quei gesti che assumono movimenti ritmici (come i batonici) o forme non assimilabili a oggetti nel mondo (come forme geometriche). Mentre con l'etichetta *Conventional* sono etichettati quei gesti che hanno raggiunto un grado di convenzionalità tale che, in un

² Tra gli altri (EFRON 1972; EKMAN-FRIESEN 1969) anche KENDON 2004 ha stilato le diverse funzioni espresse ed esprimibili dai gesti distinguendo le funzioni pragmatiche (quelle di parsing, quelle modali che esprimono la modalità di interpretazione di un enunciato e quelle performative che indicano il tipo di atto linguistico), funzioni interattive o interpersonali relativi alle funzioni dialogiche interazionali e le funzioni referenziali.

determinato sistema linguistico, a quel gesto è possibile associare un valore semantico. Nella tabella 6 riportiamo il tagset del gesto.

TAG	DEFINIZIONE
PICTORIAL	gesti che riproducono una forma riconducibile a un'immagine o a un'azione
NON-PICTORIAL	gesti che riproducono movimenti ritmici o forme non assimilabili a oggetti nel mondo
CONVENTIONAL	gesti convenzionali a cui è possibile associare un valore semantico riconosciuto dalla comunità dei parlanti

Tabella 6 Tag-set per l'annotazione gestuale in CORMIP

È necessario che l'indagine sulla lingua da un punto di vista multimodale si traduca nello studio del comportamento di unità di analisi dei diversi canali per evitare categorizzazioni che rischiano di ridurre il gesto a mero ausilio del canale verbale.

4.5 Dall'azione linguistica all'analisi multimodale

La Teoria della Lingua in Atto (CRESTI 2000), sulla base della simultaneità degli atti linguistici teorizzata da AUSTIN 1962, ha derivato la possibilità di porre in relazione l'atto linguistico con l'enunciato (CRESTI 2005). In questa relazione la prosodia è considerata l'interfaccia tra l'atto illocutivo e quello locutivo e rappresenta il mezzo necessario per trasdurre la concezione pragmatica in entità concreta e udibile, che è l'*utterance* (CRESTI 2020). Nella struttura dell'enunciato, la forza illocutiva è espressa dall'unità informativa di *Comment* e che prosodicamente corrisponde all'unità di *Root*.

Nel corso degli studi su *corpora* portati avanti dal gruppo di ricerca LABLITA, è stato costituito un repertorio di tipi illocutivi distribuiti su cinque classi generali identificate in:

CORMIP (Corpus Multimodale dell'Italiano Parlato)

- *Rifiuto*: un atteggiamento di libertà e separazione dal parlante dall'interlocutore, che permette uno scontro con quest'ultimo, una richiesta di sua trasformazione;
- *Asserzione*: un atteggiamento di certezza del parlante nei confronti dell'interlocutore, sicurezza che consente di proporre giudizi, scoperte, valutazioni, rappresentazioni, come oggetti nuovi al mondo;
- *Espressione*: un atteggiamento di manifestazione "estetica" di stati d'animo, emozioni e credenze;
- *Rito*: un atteggiamento esterno di assolvimento di compiti linguistici che hanno effetti legali e sociali e che possono essere compiuti con la minima partecipazione affettiva.

Queste cinque classi generali sono state usate per l'annotazione dell'illocuzione, il cui valore è stato determinato dall'interpretazione dell'azione multimodale, il cui nucleo espressivo di riferimento rimane l'unità tonale di tipo Root. Nella tabella 7 riportiamo il tag-set per l'annotazione delle classi illocutive.

TAG	DEFINIZIONE
ASSERTION	Atto linguistico assertivo
DIRECTION	Atto linguistico direttivo
EXPRESSION	Atto linguistico espressivo
RITE	Atto linguistico rituale
REFUSAL	Atto linguistico di rifiuto

Tabella 7 Tag-set dei valori illocutivi in CORMIP

Nello specifico la nostra annotazione dell'illocuzione ha seguito tre principi:

- se il parlante esprime l'azione linguistica attraverso l'uso esclusivo del canale verbale, viene annotato il valore illocutivo espresso dall'unità prosodica di tipo Root;
- se il parlante esprime l'azione linguistica attraverso l'uso esclusivo del canale gestuale (compreso l'espressione faccia, o la co-occorren-

- za tre gesto manuale ed espressione facciale), viene annotato il valore illocutivo espresso dall'unità gestuale corrispondente;
- se il parlante esprime l'azione linguistica attraverso l'uso concomitante del canale verbale e gestuale compreso l'espressione faccia, o la co-occorrenza tre gesto manuale ed espressione facciale, viene annotato il valore illocutivo valutando in modo complessivo espresso dalla co-occorrenza delle relative unità gestuale e verbali.

I principi appena elencati si basano su due presupposti. Il primo è che il valore illocutivo – che esprime l'intenzione comunicativa del parlante di natura affettiva – si trasforma fisicamente in azione linguistica attraverso il potenziale uso delle diverse modalità (CRESTI 2020) e, aggiungiamo noi, anche attraverso l'uso dei gesti e dell'espressione facciale. Il secondo presupposto è il giudizio dell'annotatore. Infatti, se da un lato la teoria della lingua in atto ha sistematizzato secondo quali parametri pragmatici, semiologici e cognitivi si definiscono i diversi valori illocutivi (CRESTI 2005; 2020) dall'altro mancano dei parametri corrispettivi per giudicare il valore illocutivo espresso dalla gestualità o dalla co-occorrenza di gesto e parlato. Pertanto, in questi casi, abbiamo fatto fede al giudizio analitico dell'annotatore in quanto, come già detto, ci avviciniamo all'annotazione dell'illocuzione nel nostro *corpus* sperimentalmente.

La natura sperimentale dell'operazione mira a utilizzare l'annotazione dell'azione linguistica come mezzo per indagare l'esistenza e le caratteristiche dell'unità multimodale. L'atto linguistico non rappresenta un'unità di analisi, ma un approccio metodologico di tipo pragmatico all'intenzione comunicativa espressa dai parlanti attraverso l'uso di diversi canali e/o della loro coordinazione. Al momento non siamo in grado di poter esprimere definizioni inerenti alle nozioni di atto linguistico multimodale, ma senza dubbio il *corpus* rappresenta uno strumento utile a indagarlo.

5. Conclusioni

In queste pagine è stato riportato e discusso il metodo usato per realizzare un *corpus* multimodale pilota dell'italiano parlato in contesto spontaneo (CORMIP). L'obiettivo è stato quello di creare uno strumento per indagare la lingua come sistema multimodale. Un ruolo importante nell'annotazione è svolto dall'azione linguistica che, all'interno del template annotativo, svolge un ruolo chiave di interpretazione del carattere multimodale dell'enunciato. In conclusione, possiamo affermare che questo nostro approccio ci ha permesso di costruire un *corpus* pilota – che raccoglie sei brani di interazioni spontanee in contesti diversi e in diverse tipologie di interazioni – e di indagare le interazioni comunicative da un punto di vista multimodale senza però creare uno strumento che possa apparire centrato maggiormente su una delle due modalità. Questo è stato possibile grazie all'approccio percettivo e pragmatico al fenomeno lingua. È possibile, inoltre, sostenere la necessità di *corpora* multimodali sempre più grandi contenenti dati spontanei e realizzati attraverso l'uso di standard largamente condivisi.

Bibliografia

- AUSTIN 1962 = JOHN LANGSHAW AUSTIN, *How to do things with words*, Oxford, Oxford University Press, 1962.
- BRESSEM-LADEWIG-MÜLLER 2013 = JANA BRESSEM, SILVA H. LADEWIG E CORNELIA MÜLLER, *Linguistic Annotation System for Gestures*, in *Body - Language - Communication. An International Handbook on Multimodality in Human Interaction*, edited by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Tessendorf, «Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk)», 38/1, Berlin-Boston, De Gruyter Mouton, 2013, pp. 1098-1124.
- CRESTI 2005 = EMANUELA CRESTI, *Per una nuova classificazione dell'illocuzione*, in *Tradizione e innovazione. Atti del VI Convegno Internazionale della SILFI – Gerhard-Mercator Universit (Duisburg, 28 giugno-2 luglio 2000)*, a cura di Elisabetta Burr, Firenze, Franco Cesati Editore, pp. 233-246.

- CRESTI-MONEGLIA 2005 = C-ORAL-ROM. *Integrated reference corpora for spoken romance languages*, eds. by Emanuela Cresti, Massimo Moneglia, Amsterdam, John Benjamins, 2005.
- CRESTI 2000 = EMANUELA CRESTI, *Corpus di italiano parlato*, Firenze, Accademia della Crusca, 2000.
- CRESTI 2020 = EMANUELA CRESTI, *The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory*, in *In search of basic units of spoken language: A corpus-driven approach*, eds. by Shlomo Izre'el, Heliana Mello, Alessandro Panunzi, Tommaso Raso, Amsterdam, John Benjamins, 2020, pp. 181-219.
- DE MAURO 2000 = TULLIO DE MAURO, *Vocalità, gestualità, lingue segnate e non segnate*, in *Viaggio nella città invisibile*, a cura di C. Bagnara, G. Chiappini, M.P. Conte, M. Ottolini, Pisa, Edizioni del Cerro, 2000, pp. 17-45.
- EFRON 1972 [1941] = DAVID EFRON, *Gesture, race and culture*, The Hague, Mouton, 1972.
- EKMAN-FRIESEN 1969 = PAUL EKMAN E WALLACE V. FRIESEN, *The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding*, in «Semiótica», 1, 1, 1969, pp. 49-98.
- ENFIELD 2009 = NICK J. ENFIELD, *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*, Language Culture and Cognition, Cambridge, Cambridge University Press, 2009.
- ENFIELD 2013 = NICK J. ENFIELD, *Composite Utterances approach to meaning, in Body - Language - Communication. An International Handbook on Multimodality in Human Interaction*, edited by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Tessendorf, «Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk)», 38/1, Berlin-Boston, De Gruyter Mouton, 2013.
- ENFIELD 2009 = NICK J. ENFIELD, *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*, Language Culture and Cognition, Cambridge, Cambridge University Press.
- FIRENZUOLI 2003 = VALENTINA FIRENZUOLI, *Le Forme Intonative di Valore Illocutivo dell'Italiano Parlato: Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA)*, Tesi di Dottorato, Università degli Studi di Firenze.
- GIBBON ET AL. 2003 = DAFYDD GIBBON, ULRIKE GUT, BENJAMIN HELL, KARIN LOOKS, ALEXANDRA THIES, THORSTEN TRIPPEL, *A Computational Model of Arm Gestures in Conversation*, in Eighth European Conference on Speech Communication and Technology, Geneva, Switzerland, 2003.
- GOODWIN 2000 = CHARLES GOODWIN, *Action and embodiment within situated human interaction*, «Journal of Pragmatics», 32, 2000, pp. 1489-1522.

- 't HART-COLLIER-COHEN (1990) = JOHAN 't HART, RENE COLLIER, ANTOINE COHEN, *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody*, Cambridge University Press, Cambridge, 1990.
- KRESS-VAN LEEUWEN (2001) = GUNTHER KRESS, THEO VAN LEEUWEN, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, London, New York, Edward, 2001.
- HOCKETT 1960 = CHARLES F. HOCKETT (1960), *The origin of speech*, in «Scientific American», 203, 1960, pp. 88-96.
- KENDON 1972 = ADAM KENDON, *Some relationships between body motion and speech: an analysis of an example*, in *Studies in Dyadic Communication*, edited by Aron Wolfe Siegman e Benjamin Pope, Elmsford, New York, 1972, pp. 69-89.
- KENDON 1980 = ADAM KENDON, *Gesticulation and speech: Two aspects of the process of utterance*, in *The relationship of verbal and nonverbal communication*, edited by Mary R. Key, Berlin-New York, De Gruyter Mouton, 1980, pp. 207-228.
- KENDON 2004 = ADAM KENDON, *Gesture: Visible Action as Utterance*, Cambridge, Cambridge University Press, 2004.
- LABOV 1972 = WILLIAM LABOV, *Sociolinguistic Patterns*, Philadelphia, PA, University of Pennsylvania Press, 1972.
- LAUSBERG 2013 = HEDDA LAUSBERG, *NEUROGES. A Coding System for the Empirical Analysis of Hand Movement Behaviour as a Reflection of Cognitive, Emotional, and Interactive Processes*, in *Body - Language - Communication. An International Handbook on Multimodality in Human Interaction*, edited by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Tessendorf, «Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk)», 38/1, Berlin-Boston, De Gruyter Mouton, 2013, pp. 1022-1037.
- LAUSBERG-SLOETJES 2016 = HEDDA LAUSBERG, HAN SLOETJES, *The revised NEUROGES-ELAN: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture*, «Behavior Research Methods», 48, 2016, pp. 973-993.
- MARTINET 1960 = ANDRÉ MARTINET, *Éléments de linguistique Générale*, Paris, Armand Colin, 1960.
- MCNEILL 1992 = DAVID MCNEILL, *Hand and Mind: What Gestures Reveal about Thought*, Chicago, University of Chicago Press, 1992.
- MONEGLIA-RASO 2014 = MASSIMO MONEGLIA, TOMMASO RASO, *Notes on Language into Act Theory (L-Act)*, in *Spoken corpora and linguistic studies*, edited by Tommaso Raso, Heliana Mello, Amsterdam, John Benjamins, 2014, pp. 468-495.

MÜLLER 2013 = CORNELIA MÜLLER, *Gestural Modes of representation as techniques of depiction*, in *Body - Language - Communication. An International Handbook on Multimodality in Human Interaction*, edited by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Tessendorf, «Handbücher Zur Sprach-Und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (Hsk)», 38/1, Berlin-Boston, De Gruyter Mouton, 2013, pp. 1687-1702.

TRIPPEL ET AL. 2004 = THORSTEN TRIPPEL, DAFYDD GIBBON, ALEXANDRA THIES, JAN-TORSTEN MILDE, KARIN LOOKS, BENJAMIN HELL, AND ULRIKE GUT, *CoGesT: A Formal Transcription System for Conversational Gesture*, in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (Lrec'04)*, Lisbon, Portugal, 2004.

Riassunto Questa ricerca affronta la questione metodologica per la compilazione del corpus multimodale di lingua parlata CORMIP, proponendo una possibile soluzione alle questioni legate alla gestione dei dati linguistici multimodali. I problemi riguardano tutte le fasi della compilazione, dalla raccolta dei dati fino alla loro trascrizione e annotazione.

Abstract This research addresses the methodological issue for the compilation of the multimodal spoken language corpus CORMIP. It proposes a possible solution to the issues related to the management of multimodal language data. The problems concern all stages of compilation, from data collection to data transcription and annotation.

Il dato del Paesaggio Linguistico: una proposta di classificazione con alcuni esempi dal mercato di S. Lorenzo (FI)

Lorenzo Cambi

1. Cos'è il Paesaggio Linguistico?

Gli studi sul *Linguistic Landscape*, in italiano paesaggio (o panorama) linguistico, nascono come branca degli studi sociolinguistici a partire dagli anni Settanta, tuttavia la prima definizione si ha nel 1997 con il lavoro di Landry e Bourhis sul Canada:

The language of public road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings combines to form the linguistic landscape of a given territory, region, or urban agglomeration. (LANDRY-BOURHIS 1997, p. 25)

Sostanzialmente, si tratta dello studio di tutto ciò che di scritto (solo scritto?) compare in uno spazio pubblico¹. Questa prima concettualizzazione ha dato vita a un produttivo filone di ricerca che si è arricchito di anno in anno. I primi casi di studio hanno riguardato nazioni, regioni o città, caratterizzate da uno statuto quantomeno bilingue (per es. Canada, Israele, Belgio e poi Paesi Baschi, Frisia, ecc.²); l'obiettivo di questo tipo di analisi, ancora esclusivamente quantita-

- ¹ Alcuni studiosi sono andati a indagare anche spazi privati come le abitazioni di persone immigrate o spazi semi-pubblici come gli interni di ristoranti, musei, ecc.
- ² Per una panoramica invece sugli studi sul paesaggio linguistico italiano si veda BELLINZONA 2021.

tivo, era quello di verificare o attestare lo spazio rappresentativo che veniva dedicato alle varie lingue, soprattutto a quelle di minoranza, sia da parte delle istituzioni, e quindi nei cosiddetti segni *top-down*, sia da parte dei privati, e quindi nei segni *bottom-up*³. Gli studi sul paesaggio linguistico (d'ora in poi PL) di questo genere comparavano i dati socio-demografici e le politiche linguistiche delle zone di inchiesta con le percentuali di rappresentazione di una determinata lingua nel PL.

Questo approccio si è conservato fino alla fine del primo decennio degli anni Duemila, quando si è ritenuto che un'analisi di questo tipo non fosse sufficiente per caratterizzare un'area contemporanea, soprattutto urbana. Fondamentale per questo cambio di prospettiva è stato il concetto di *Superdiversity* introdotto dall'antropologo Steven Vertovec (cfr. VERTOVEC 2007 e 2023); con questo, l'autore vuole sottolineare la complessità nella definizione di categorie etniche e sociali causata dall'accelerazione nei processi di globalizzazione e dai continui flussi migratori, sempre più diversi dal punto di vista socio-economico. Considerato ciò, e la conseguente difficoltà nello stabilire dei confini netti tra le varie comunità, soprattutto per quanto riguarda quelle di "minoranza" e quelle "immigrate", ci si è resi conto di come un mero dato statistico ricavato dallo studio del PL non potesse essere sufficiente per rilevare le "condizioni di salute" di una lingua in un determinato contesto; sicuramente questo dato quantitativo costituisce un aspetto rilevante, tuttavia rimane generico senza un approfondimento di diversa natura, cioè etnografico.

Un autore fondamentale per un nuovo tipo di approccio, stavolta qualitativo, è stato Jan Blommaert (BLOMMAERT 2010 e 2013). Questi, specialmente in *Chronicles of Complexity* (2013), ha sottolineato l'importanza di un metodo etnografico per l'analisi del PL: solamente con uno studio approfondito del contesto e tramite il coinvolgimento di chi il paesaggio linguistico lo crea (istituzioni e privati cittadini), e di chi lo "vive", si può arrivare alla comprensione di un ambiente sociale superdiverso quali sono gli spazi urbani, ma non solo, al giorno d'oggi. Infatti, in contesti dinamici come lo sono quelli contemporanei, anche

3 Per la prima definizione di segni *top-down* e *bottom-up*, cfr. BEN-RAFAEL ET AL. 2006.

un dato statisticamente irrilevante può rappresentare una tendenza futura o un cambiamento sociale in atto.

2. La necessità di un approccio misto

Fino a pochi anni fa il metodo quantitativo e quello qualitativo sono stati tenuti prevalentemente separati e i vari studiosi sceglievano quello che più si addiceva al loro oggetto di ricerca⁴; negli ultimi anni, invece, si è cercato di sollecitare un approccio di tipo misto (si veda, per es., GORTER-CENOZ-VAN DER WORP 2022), poiché i primi due metodi, presi singolarmente, sembravano presentare dei limiti importanti.

Oltre a non tenere conto della progressiva *superdiversity* del contesto sociale, e fermo restando la difficoltà a considerare le statistiche sul PL come rappresentative di una realtà urbana nella sua interezza, l'approccio quantitativo mostra il proprio limite più rilevante nel mancato coinvolgimento di attori e spettatori del PL: trattandosi di segni che vanno a costruire un spazio sociale, è necessario interpellare coloro i quali questo spazio lo vivono, sia concependolo che percependolo, poiché sono questi, e le loro reazioni, a caratterizzarli⁵. Il PL influenza gli atteggiamenti delle persone che a loro volta influenzano la futura costituzione del PL (cfr. GORTER 2021).

Altro aspetto problematico per una metodologia quantitativa è stabilire l'oggetto di ricerca; si consideri, per esempio, come GORTER 2013 abbia aggiornato la definizione di PL che avevano dato Landry e Bourhis:

[...] recent technological developments have added many new types of signs: electronic flat-panel displays, LED neon lights, foam boards, electronic mes-

4 Virtù, o vizio, del PL è il suo carattere marcatamente interdisciplinare, il quale ha portato studiosi delle aree di ricerca più diverse a confrontarsi col tema e ad adottare sistemi specifici per i loro interessi.

5 Cfr. LEFEBVRE 1991 per i concetti di spazio concepito, percepito e vissuto.

sage centers, interactive touch screens, inflatable signage, and scrolling banners. (GORTER 2013, p. 191)

Tralascieremo, in questa sede, il filone “semiotico” degli studi sul PL⁶. Rimanendo nell’ambito dei segni strettamente linguistici, i problemi di definizione, soprattutto in prospettiva analitica, sono numerosi: una volta individuato il segno da catalogare (e non è facile), l’unità di analisi è il segno stesso, preso singolarmente, o il contesto in cui compare, per es. un negozio, una strada? Nel senso, è opportuno costruire delle statistiche analizzando segno per segno, oppure ha più senso condurre delle analisi su macrounità⁷, per es. i segni di quel negozio/di quella strada, nel loro insieme? Si ritiene che questa seconda strada sia preferibile, poiché esprime più chiaramente le scelte linguistiche dei vari attori del PL; il rischio, altrimenti, è quello di “drogare” i risultati, sovraestendendo le scelte di alcuni a discapito di quelle altri. Tuttavia, queste sono questioni che tutt’ora non hanno trovato una soluzione e come possiamo intuire sono aspetti molto importanti.

D’altra parte, anche il metodo qualitativo presenta dei limiti, su tutti quello del presentare indagini che risultano significative solo per lo specifico contesto in cui vengono condotte, e quindi molto difficilmente “riproducibili”. Esemplicativo di questo approccio è il raffinato lavoro di Blommaert su un quartiere di Anversa, che, a partire dall’analisi di un numero molto limitato di segni, consente allo studioso di giungere a considerazioni di ordine generale (cfr. BLOMMAERT 2013). Questo genere di analisi, tuttavia, è solo minimamente ripetibile in un contesto diverso e ciò si scontra con uno degli obiettivi, se non l’obiettivo principale, della ricerca sul PL negli ultimi anni: trovare un metodo “standard” il più riproducibile possibile. Inoltre, sempre in merito alla selezione dei segni sui quali costruire l’analisi etnografica, la mancata contestualizzazione numerica di questi (rispetto al totale di quelli

6 Si segnalano qui i lavori di SCOLLON-SCOLLON 2003 (*geosemiotics*), JAWORKI-THURLOW 2010 (*semiotic landscape*) e PENNYCOOK-OTSUJI 2015 (*smellscape*).

7 Tra i vari punti di vista, si rimanda qui al concetto di *unità* dato prima da CEÑOZ-GORTER 2006 e poi ripreso da CALVI 2018.

presenti in una determinata area) rischia di far apparire ordinari dei fenomeni in realtà molto particolari, o viceversa, e il tutto è subordinato alla soggettività degli studiosi; anche capire che impatto statistico ha un segno in un determinato contesto, insomma, risulta un aspetto importante per comprenderne il reale valore.

Altra questione non di poco conto è che questo tipo di approccio tende a un'analisi più generalmente semiotica del segno, arrivando talvolta a “dimenticarsi” degli aspetti linguistici⁸. In studi di questo tipo, infatti, il contesto diventa spesso prevalente sui segni, i quali fungono “semplicemente” da punto di partenza per riflessioni più generali di carattere prevalentemente sociale.

Ancora oggi, le due metodologie sembrano muoversi in parallelo, senza produrre una sintesi che invece sarebbe auspicabile⁹.

Per procedere in questa direzione, occorrerebbe prima di tutto individuare un campione definito (per es. una via, una piazza, un quartiere, ma anche un singolo negozio) del quale registrare tutte le occorrenze, organizzando i dati secondo un sistema di catalogazione con criteri condivisi; successivamente si potrebbe procedere a una contestualizzazione più approfondita attraverso un'indagine etnografica. Tutto ciò, tenendo conto dell'esistenza, nei luoghi indagati, di eventuali politiche linguistiche in grado di interferire e di orientare la percezione di chi quei paesaggi linguistici li vive¹⁰.

- 8 Ci si riferisce qui, per es., all'analisi morfo-sintattica e/o lessicale dei segni, agli eventuali fenomeni di *translanguaging* (cfr. GARCIA-WEI 2014), ai rapporti di traduzione nei segni multilingue, ecc..
- 9 In merito agli ultimi sviluppi dei due approcci si veda: AMOS-SOUKUP 2020 e LYONS 2020 per quello quantitativo e BLOMMAERT-MALY 2014 (e poi BLOMMAERT-MALY 2019a) per quello qualitativo.
- 10 Si veda per es. il caso di Prato in BARNI-BAGNA 2010: la città ha vietato insegne monolingui non-italiane (cinesi?), costringendo a inserire una traduzione in italiano, e ciò indubbiamente ha avuto un riflesso sul PL urbano (e sulla sua percezione).

3. E il paesaggio linguistico virtuale?

A complicare il quadro si è aggiunto il filone di ricerca sul *Virtual* (o *Online*) *Linguistic Landscape*, teorizzato per la prima volta da IVKOVIC-LOTHERINGTON 2009:

the virtual LL serves to delineate the linguistic community and to mark language status in expressed power relations among the coexisting linguistic choices in the cyberspace community. [...]. In this way, the VLL functions as an identity marker, providing choice in textual access and expression. (IVKOVIC-LOTHERINGTON 2009, p. 19)

Non discuteremo in questa sede sugli elementi costitutivi del PL virtuale, anche perché il dibattito è ancora aperto e sembra lontano dall'arrivare a una conclusione¹¹, tuttavia è importante sottolineare come lo spazio virtuale sia al giorno d'oggi un aspetto non trascurabile, o meglio, da non trascurare, poiché rappresenta una costante nella realtà sociale odierna.

Lo spazio virtuale nella realtà contemporanea, soprattutto urbana¹², non è da considerarsi uno spazio alternativo a quello fisico: attraverso lo spazio virtuale soddisfiamo alcuni dei nostri bisogni, anche fisici, e parte di quelli che vengono soddisfatti nello spazio fisico sono comunque aiutati da infrastrutture e canali digitali (KELLERMAN 2014). In una società *superdiversa*, insomma, è imprescindibile tenere in considerazione anche ciò che avviene nello spazio virtuale.

Blommaert e Maly parlano di *online-offline nexus* (BLOMMAERT-MALY 2019b) nel PL, tema che approfondiremo in seguito, riferendosi ai collegamenti "diretti" tra paesaggio linguistico in senso stretto, fisico, e paesaggio linguistico virtuale: per es. rimandi a siti internet, a pagine sui social network, ecc. Questo è senza dubbio un aspetto da tenere in assoluta considerazione per una concettualizzazione del PL virtuale, tuttavia

¹¹ In merito, tra i vari, si vedano HIIPPALA ET AL. 2019, KALLEN-DHONNACHA-WADE 2020 e YAO 2021.

¹² Si veda KELLERMAN 2019.

si ritiene che da solo non sia sufficiente: non sono anche i post geotagati in un determinato luogo a creare il panorama virtuale dello stesso? E gli hashtag? La questione, come detto, è molto complessa e dibattuta e non sarà questa la sede per trarre delle conclusioni che, proprio per la mancanza di accordo nella comunità scientifica su ciò che costituisce il PL virtuale, richiederebbero di essere ampiamente argomentate.

4. Quali criteri di catalogazione?

Stabilire dei criteri per la catalogazione appare un passaggio essenziale per un'indagine quantitativa sul PL e altrettanto importante è dotarsi di strumenti i più accessibili possibile; infatti, se l'obiettivo è rendere la ricerca "replicabile", è necessario che lo strumento con cui si realizza sia facilmente disponibile.

In merito alle categorie da adottare per catalogare i tipi di unità che abbiamo selezionato (a riguardo, si veda la definizione di segno data da Backhaus: «A sign was considered to be any piece of written text within a spatially definable frame», BACKHAUS 2006, p. 55), due ottimi punti di partenza teorici ci sembrano essere il lavoro di Amos sulla "Chinatown" di Liverpool (AMOS 2016) e il già citato contributo di AMOS-SOUKUP 2020.

Nell'articolo del 2016 Amos distingue le seguenti categorie (ove possibile, si riporta la traduzione in italiano):

- lingua/e: le eventuali lingue contenute nell'unità;
- *multilingualism*: il rapporto tra le lingue presenti nel testo (se sono traduzioni esatte, se il contenuto presentato in una lingua è diverso da quello presentato in un'altra, ecc.);
- funzione comunicativa: il ruolo pragmatico del testo;
- luogo: il luogo esatto in cui compare l'unità;
- materialità: come è costituita materialmente l'unità;
- autore: la classe di appartenenza dell'autore/responsabile del testo;
- cornice: tipo di luogo in cui l'unità compare (ristorante, negozio, edificio residenziale, ecc.);
- "campo" (*field*): il tipo di discorso associato al testo.

Questa categorizzazione ha funzionato da modello per quella di AMOS-SOUKUP 2020, che, invitando ad aderire ai principi della socio-linguistica laboviana (individuare un'unità di analisi stabile; raccogliere tutti i dati che si trovano sul campo, non trascurando quelli "non-eccezionali"; selezionare una lista di variabili indipendenti), suddivide le varie categorie in due gruppi: quelle "fisiche" e quelle relative al discorso. Per quanto riguarda le prime, le fisiche, gli autori distinguono:

- la collocazione fisica del segno;
- la grandezza;
- la materialità;
- il formato di applicazione del testo (se scritto a mano, stampato, ecc.).

Per quanto riguarda invece le variabili relative al discorso:

- l'autorialità (ufficiale, privata o non autorizzata, con relative sottocategorizzazioni);
- contesto (*contextual setting*);
- il tipo di discorso (artistico, infrastrutturale, regolatorio, trasgressivo, ecc.).

Questo modello di classificazione, pur dettagliato, manca degli aspetti relativi alle lingue contenute nei segni, tuttavia rimane il tentativo più accurato di definire un modello standard di classificazione delle unità di PL.

A nostro avviso, una proposta molto interessante arriva dagli sviluppatori dell'applicazione per telefoni mobili *Lingscape* (<https://lingscape.uni.lu/>), la categorizzazione dei quali risulta essere rinforzata proprio dalla presenza di uno strumento gratuito e facilmente accessibile per chiunque abbia a disposizione una connessione a Internet e un dispositivo mobile. *Lingscape* permette inoltre di confrontare i propri dati con quelli del database generale, nel quale confluiscono tutte le raccolte dati (la visibilità di queste dipende dalla volontà di chi gestisce i progetti, i quali possono essere mantenuti privati).

È tramite questa applicazione che chi scrive si propone di catalogare i dati del PL del mercato di S. Lorenzo, a Firenze. È bene specificare

che le varianti erano così al giorno 26/01/2023; nei mesi successivi queste sono state aggiornate e parzialmente cambiate. Adesso riporteremo i vari criteri di catalogazione presenti sul sito di *Lingscape* (ove possibile traducendoli), spiegandoli e talvolta corredandoli di foto scattate nel luogo di inchiesta.

4.1 Directedness¹³

Con questa voce si intendono distinguere i diversi tipi di autore, se istituzionali/pubblici o privati, quindi se i segni sono *top-down* o *bottom-up* (cfr. *supra*).



Top-down



Bottom-up

Va precisato che con questo tipo di distinzione la categoria *bottom-up* rimane piuttosto generica, comprendendo dai privati cittadini, alle attività commerciali, agli anonimi autori di graffiti, ecc.

¹³ Per la lista delle categorie con tutte le variabili: <https://lingscape.uni.lu/taxonomies/>.

Lorenzo Cambi

4.2 Discorso

Ci si riferisce al dominio socio-pragmatico dei testi presenti nel PL; si possono avere testi commemorativi, informativi, commerciali (nel mercato sono questi a prevalere), politici, ecc.



Commerciale



Commemorativo

4.3 Dominanza

Con questa voce si intendono le diverse gerarchie possibili all'interno di un segno: se a prevalere è il colore, lo sfondo, il carattere tipografico, ecc.

4.4 Dynamics

Ci si riferisce con questa voce alle dinamiche proprie del segno; queste possono riguardare sia la sua collocazione (se fissa, come un'insegna, o mobile/temporanea, come può essere un cartello di un prodotto in vendita o la locandina di un evento), che il suo contenuto, il quale può essere interattivo, statico o dinamico.

4.5 Form (“forma del segno”)

Con questa categoria gli sviluppatori dell'app intendono riferirsi al tipo di segno, se si tratta quindi di un post-it, di un'insegna a neon, di un display, di un cartello stradale, di un adesivo, ecc.



Insegna “stradale” all’interno del mercato



Insegne di una macelleria



Adesivo di una delle attività del mercato

4.6 Stratificazione

Questa variabile si riferisce agli eventuali livelli di stratificazione che possono essere riportati dal segno; gli interventi possono essere materiali (per esempio si aggiunge un segno sopra un altro), culturali, storici, linguistici, ecc.



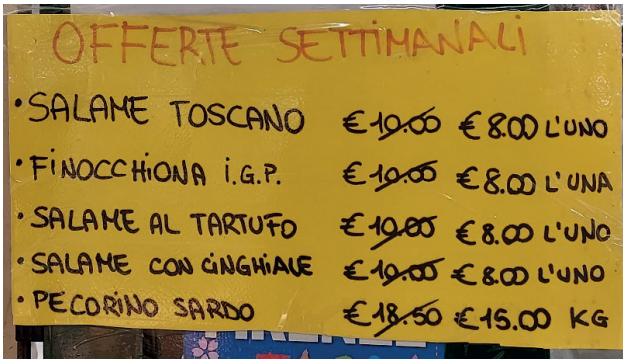
Stratificazione materiale e culturale (il nuovo proprietario, che è musulmano, ha oscurato la precedente scritta “salumi”)

La foto appena presentata è molto interessante perché ci dimostra come anche l’“assenza” all’interno del PL sia parlante: la cancellazione della scritta “salumi”, a parte la persona proprietaria del banco che lavora lì, è l’unica testimone dell’identità culturale del nuovo rivenditore, il quale si è guardato dal sostituire l’insegna dell’attività precedente (molto tradizionale, anche per la presenza del dialettismo *civaie* ‘legumi secchi e cereali’)¹⁴.

14 Le questioni di tradizionalità e autenticità sono aspetti fondamentali nel discorso commerciale, soprattutto per ciò che riguarda l’enogastronomia, e il dialetto risulta un codice perfetto per veicolare questo tipo di valori simbolici.



Parziale cancellazione del testo in giapponese



Qui abbiamo una stratificazione che potremmo definire "commerciale"

Lorenzo Cambi

4.7 Linguality (“lingue”)

Questo criterio di catalogazione evidenzia il numero di lingue presenti in un segno, il quale può essere monolingue, bilingue, trilingue, multi-lingue, translingue (*code/tag switching*), ecc.



Monolingue



Bilingue



Multilingue

4.8 Materiale

Con questa voce ci si riferisce ai materiali che costituiscono il segno; se quindi questo è di carta, vetro, legno, cartone, ferro, ecc.

4.9 Modo

Con questa variabile si intendono i vari modi in cui il segno è stato prodotto, cioè come è stato “scritto”, quindi se è scritto a mano¹⁵, se è stampato, se è scolpito, ecc.

¹⁵ In merito risulta molto interessante l'articolo di Wei-Hua su un mercato londinese (WEI-HUA 2021).

Lorenzo Cambi



Scritti a mano



Stampati

4.10 Modificazione

Questa categoria è molto affine a quella della “stratificazione”, perché vuole distinguere i diversi modi di alterazione dei segni; rispetto al criterio precedente, questo sembra interessarsi principalmente al contenuto semantico dell’intervento; le varianti infatti sono: chiarimento, commento, correzione, estensione (del messaggio), ecc.

4.11 Script (“alfabeto/sistema di scrittura”)

Ci si riferisce qui al sistema grafico presente nel segno, quindi alfabeto latino, arabo, cirillico, ebraico, caratteri cinesi, giapponesi, coreani, ecc.



Alfabeto latino, caratteri coreani e giapponesi



Alfabeto arabo

Lorenzo Cambi

4.12 Dimensioni

Si distingue la grandezza dei segni in base al formato di carta secondo lo standard internazionale (AO e derivati), con l'aggiunta delle seguenti categorie: $1\text{m}^2\text{-}4\text{m}^2$; $4\text{m}^2\text{-}10\text{m}^2$; $> 10\text{m}^2$.

4.13 Stato

Questa categoria descrivere l'operatività del segno, se è cioè ancora valido, operativo, oppure se è "scaduto" (come può essere una locandina di un evento passato o un'insegna di un'attività dismessa), ecc.



Questa insegna, presente in uno dei banchi del mercato, è il rimasuglio di una vecchia attività ormai sostituita

4.14 Status

Questa variabile si riferisce al diritto di un segno di stare dove si trova, se è cioè un segno autorizzato, non-autorizzato o "riconosciuto" (*recognized*; cioè se da non-autorizzato è stato poi accettato e conservato, come per es. può essere un'opera di *street art*).



Non-autorizzato



“Riconosciuto”

Lorenzo Cambi

4.15 Supplementi

L'app permette anche di caricare materiale fotografico aggiuntivo come per es. foto di archivio, cartoline, ecc.



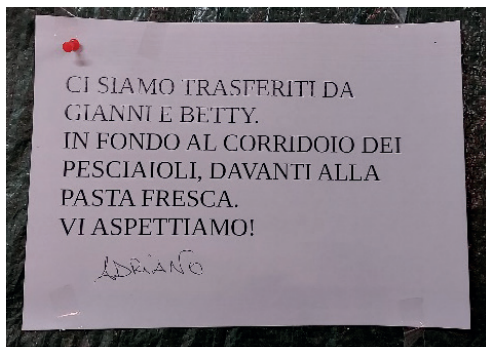
“Biglietto da visita” della macelleria islamica

4.16 Temporality

Ci si riferisce con questa voce strettamente agli aspetti temporali del segno, cioè se si tratta di un segno permanente (un'insegna), temporaneo (un cartello “torno subito” scritto su di un foglietto di carta), *event-related*, ecc.



Event-related



Temporaneo

4.17 Traduzione

Questa voce si riferisce specificamente ai segni multilingue e indaga i rapporti di traduzione tra le varie lingue presenti nel segno. In questo caso si ritiene importante riportare tutte le varianti associate a questa variabile, pertanto si hanno “traduzioni”:

- complementari: i messaggi nelle varie lingue costituiscono un unico significato, di cui ogni lingua costituisce una parte;
- duplicanti: il messaggio è ripetuto in modo (più o meno) identico nelle varie lingue;
- frammentarie: il significato completo è dato in una sola lingua, delle altre si hanno traduzioni parziali;
- sovrapposte: alcune parti del messaggio sono ripetute nelle varie lingue, altre sono espresse in una lingua sola.

Lorenzo Cambi



Traduzione complementare

4.18 Varietà

Con quest'ultima categoria si intendo le varietà linguistiche relative alle lingue presenti nel segno, quindi se si tratta di una varietà standard, non-standard, locale (con varie distinzioni), storica, ecc.



Si ha un tratto locale (anche se non percepito come tale) quale *nostrale* 'nostrano'

Queste sono le varie categorie di classificazione che l'applicazione *Lingscape* ci fornisce; la versione base, utilizzabile da chiunque, consente di registrare esclusivamente quali lingue sono presenti nel segno, ma con la funzione "progetti"¹⁶, concessa gratuitamente dagli sviluppatori a chi la richiede, si possono attivare (o non attivare) queste diciotto variabili.

Come si può vedere, alcune si riferiscono più strettamente ad aspetti linguistici (*linguality*, *script*, traduzione, varietà), altre all'analisi del discorso, altre ancora agli aspetti materiali del segno e quindi al complesso semiotico dello stesso; in base alle necessità degli studiosi e alle domande di ricerca si possono selezionare solamente le variabili che si intendono investigare.

Tra le varie proposte per un'analisi quantitativa del PL, questa pare la più solida, anche, o forse soprattutto, per lo strumento che mette a disposizione della ricerca: il database generale è infatti consultabile da chiunque e i dati dei vari *progetti* sono resi pubblici a seconda della volontà dei vari ricercatori. Inoltre, aspetto da non sottovalutare, questa applicazione permette di far registrare i dati a "non addetti ai lavori", consentendo così quello che Purschke chiama *crowdsourcing* (PURSCHKE 2017 e 2021).

Nonostante ciò, per quanto riguarda la ricerca che si sta conducendo sul mercato centrale di Firenze, vi è un aspetto che rimane non trattato, e che invece sarebbe auspicabile considerare, cioè il punto di incontro tra il PL fisico e quello virtuale: quello che Blommaert e Maly hanno definito come *online-offline nexus*¹⁷.

16 Per caricare poi le foto in un progetto, e non direttamente nel database generale, come avviene quando si usa la versione base, è necessario inserire una password decisa dall'amministratore del progetto; tutte le informazioni a riguardo si trovano alla pagina web dell'applicazione.

17 Chi scrive si è messo in contatto con gli sviluppatori dell'applicazione per implementare le categorie di classificazione anche con la variabile *online-offline nexus*.

5. Cos'è l'*online-offline nexus*? E come si manifesta nel PL?

Blommaert e Maly parlano così del nesso tra spazio fisico (e quindi spazio sociale fisico) e spazio virtuale (quindi spazio sociale virtuale):

We live our lives in an online-offline nexus. This simple observation renders us aware of the fact that social actions can be organized, set up, “staffed” and distributed in online as well as offline spaces; and it helps us realize that much of what we observe in the way of social action in superdiverse (offline, geographical) areas has, at least, been *conditioned* and perhaps even *made possible* by online infrastructures, in terms both of *actors* and of *topography*. (BLOMMAERT-MALY 2019b, p. 3; corsivi nell'originale)

I due studiosi parlano di questo nesso come di una costante nelle nostre vite contemporanee, soprattutto in ambienti “superdiversi” come le aree urbane e soprattutto quando ci riferiamo a quelli che definiamo “quartieri etnici”. Al giorno d'oggi è veramente difficile ipotizzare una nostra azione che non coinvolga, almeno in parte, lo spazio virtuale (come abbiamo già visto in precedenza riferendoci ai lavori di Kellerman, cfr. *supra*). Per questo motivo ci chiediamo: è possibile analizzare uno spazio sociale senza considerare, quantomeno parzialmente (un'analisi totale è praticamente impossibile vista la velocità dei cambiamenti in rete e la mole di dati ivi presente), lo spazio virtuale? Che conclusioni si possono trarre ignorandolo?

Aspetto importante, che suggerisce una risposta negativa alla domanda appena posta, è che nel PL fisico sempre più spesso compaiono rimandi allo spazio virtuale, che siano questi a pagine sui social network o a siti internet.

Come si può osservare, già in un solo segno si ha il rinvio alla pagina Facebook, l'informazione sulla possibilità di effettuare gli acquisti online (il che lascia presupporre l'esistenza di un sito, che effettivamente esiste) e, in alto, il sito del consorzio che raggruppa una parte dei rivenditori (<http://www.storicomercatocentrale.it/>) con accanto i loghi di Instagram e Facebook, a testimoniare la presenza anche su queste piattaforme; inoltre si ha il QR code che rimanda alla pagina Facebook.

Il dato del Paesaggio Linguistico



Questa è la storia di un'azienda di fruttigie a di una famiglia. Anzi di due.

Perché la Tripperia Bambi nasce dalla esperienza di due nipoti Bambi, nella tradizione della famiglia trippera, intraprendi Trippa da solo e cento anni dopo c'è la nuova generazione che con entusiasmo e passione scrive pagine di una storia ancora sempre in equilibrio tra l'antico e il nuovo, la voglia di scoprire il mondo, la qualità artigianale della lavorazione del prodotto e l'uso originale dei mezzi di comunicazione. La mostra illustra il processo di come preparare e servire uno dei piatti della tradizione trippera: la trippa a siron solo.

Una storia che nasce al cruce di San Frediano (Bambi), produttori di Trippa e Lampredotto e 5 generazioni.

BAMBI
Trippa e Lampredotto
FIRENZE
1910

Trippa si mangia!

Dal primo di Lampredotto a Trippa, passate per il proprio de bene e bene a siron, con tutti gli ingredienti.

This is a story about an offal company and a family. Actually - two families.

Tripperia Bambi was born from the efforts of grandmothers Bambi and the vision of the Trippera family, both butchers for over a century.

Today, the new generation is passionately engaged to live a new story, one that balances tradition and modernity, the love of one's homeland with the desire to discover the world, the artisan quality of the product and its original means of communication.

This is a story that explains how to prepare and serve one of the most traditional Florentine dishes.

This is a story that begins in San Frediano. Where Bambi producing trippa and Lampredotto for 5 generations under the name of Bambi.

www.bambitrippaelampredotto.it



Come si può intuire, i casi non sono isolati e la tendenza è verosimilmente in aumento, anche considerando il fenomeno di gentrificazione/turisticizzazione che sta investendo il centro di Firenze¹⁸ e quindi anche la zona del mercato di S. Lorenzo: diventa essenziale, al fine di attirare nuova clientela composta soprattutto da turisti, avere delle pagine sui social network attive e molto curate.

Pertanto, anche senza discutere qui la questione teorica del PL virtuale, appare fondamentale tenere in qualche modo in considerazione anche ciò che avviene in rete, poiché è anche lì che si vanno a costruire gli spazi sociali in cui viviamo, e di conseguenza anche i paesaggi linguistici.

18 Il centro città, infatti, oltre a un progressivo fenomeno di spopolamento e di aumento del costo degli immobili, sta attraversando un periodo di crescita esponenziale di attività commerciali dedicate alla gastronomia (cfr. LODA-BONATI-PUTTILLI 2020; PUTTILLI-BONATI-PORTINARO 2020 e il n. CXIX, fasc. 4, della RGI, 2022) e di strutture ricettive per turisti. In merito alla possibile correlazione tra gentrificazione e cibo si veda SBICCA 2018.

6. Conclusioni

In conclusione, si è cercato di fornire una breve panoramica sulle questioni metodologiche connesse allo studio del paesaggio linguistico e sui criteri secondo i quali catalogarne i dati. Come si è potuto vedere il dibattito è ancora aperto e una soluzione non è ancora stata trovata; è importante infatti preservare gli aspetti positivi connessi ai due approcci tradizionali, quantitativo e qualitativo, sviluppando una metodologia mista che permetta sia di rendere lo studio “ripetibile”, sia di fare analisi più approfondite dei luoghi, tramite il coinvolgimento di attori e spettatori del PL.

Con questi propositi, si è cercato di suggerire un modello per la catalogazione dei dati del PL, cioè quello dell’applicazione per dispositivi mobili *Lingscape*, riportato qui con alcuni esempi fotografici tratti dal mercato centrale di Firenze; in aggiunta a questo si è proposta la variabile *online-offline nexus*, ritenendo fondamentale considerare anche ciò che avviene nello spazio virtuale, ormai non più spazio alternativo a quello fisico, ma coesistente.

Altro tema importante qui accennato è appunto quello del PL virtuale; si è ritenuto importante sottolineare come una metodologia nuova e aggiornata per lo studio e la catalogazione dei dati debba necessariamente considerare, almeno in parte, lo spazio virtuale, poiché questo, in modo più o meno diretto, influenza la nostra percezione dello spazio, e quindi anche del PL.

Bibliografia

- AMOS 2016 = WILLIAM AMOS, *Chinatown by numbers: defining an ethnic space by empirical linguistic landscape*, in «Linguistic Landscape», 2 (2), 2016, pp. 127-156.
- AMOS-SOUKUP 2020 = WILLIAM AMOS, BARBARA SOUKUP, *Quantitative 2.0: Towards Variationist Linguistic Landscape Study (VaLLS) and a standard canon of LL variables*, in *Reterritorialising linguistic landscapes: Questioning boundaries and opening spaces*, edited by David Malinowski, Stefania Tufi, London, Bloomsbury, 2020, pp. 56-76.

- BACKHAUS 2006 = PETER BACKHAUS, *Multilingualism in Tokyo: A Look into the Linguistic Landscape*, in *Linguistic Landscape. A New Approach to Multilingualism*, edited by Durk Gorter, Bristol, Multilingual Matters, 2006, pp. 52-66.
- BARNI-BAGNA 2010 = MONICA BARNI, CARLA BAGNA, *Linguistic Landscape and Language Vitality*, in *Linguistic Landscape in the City*, edited by Elana Shohamy, Eliezer Ben-Rafael, Monica Barni, Bristol, Blue Ridge Summit: Multilingual Matters, 2010, pp. 3-18.
- BELLINZONA 2021 = MARTINA BELLINZONA, *Linguistic Landscape. Panorami urbani e scolastici nel XXI secolo*, Milano, Franco Angeli, 2021.
- BEN-RAFAEL ET AL. 2006 = ELIEZER BEN-RAFAEL, ELANA SHOHAMY, MUHAMMAD HASAN AMARA, NIRA TRUMPER-HECHT, *Linguistic Landscape as Symbolic Construction of the Public Space: The Case of Israel*, in «International Journal of Multilingualism», 3 (1), 2006, pp. 7-30.
- BLOMMAERT 2010 = JAN BLOMMAERT, *The Sociolinguistics of Globalization*, Cambridge, Cambridge University Press, 2010.
- BLOMMAERT 2013 = JAN BLOMMAERT, *Ethnography, Superdiversity and Linguistic Landscapes: Chronicles of Complexity*, Bristol/Buffalo/Toronto, Multilingual Matters, 2013.
- BLOMMAERT-MALY 2014 = JAN BLOMMAERT, ICO MALY, *Ethnographic Linguistic Landscape Analysis and social change: A case study*, «Tilburg Papers in Culture Studies», 100, 2014.
- BLOMMAERT-MALY 2019a = JAN BLOMMAERT, ICO MALY, *Digital Ethnographic Linguistic Landscape Analysis (ELLA 2.0)*, «Tilburg Papers in Culture Studies», 233, 2019.
- BLOMMAERT-MALY 2019b = JAN BLOMMAERT, ICO MALY, *Invisible Lines in the Online-Offline Linguistic Landscape*, «Tilburg Papers in Culture Studies», 223, 2019.
- CALVI 2018 = MARIA VITTORIA CALVI, *Espanol e Italiano en el paisaje linguistico de Milan ¿Traducción, mediación o translanguaging?*, in «Lingue e Linguaggi», 25, 2018, pp. 145-172.
- CENOS-GORTER 2006 = JASONE CENOS, DURK GORTER, *Linguistic Landscape and Minority Languages*, in *Linguistic Landscape. A New Approach to Multilingualism*, edited by Durk Gorter, Bristol, Multilingual Matters, 2006, pp. 67-80.
- GARCIA-WEI 2014 = OFELIA GARCIA, LI WEI, *Translanguaging. Language, Bilingualism and Education*, London, Palgrave Pivot, 2014.
- GORTER 2013 = DURK GORTER, *Linguistic Landscapes in a Multilingual World*, in «Annual Review of Applied Linguistics», 33, 2013, pp. 190-212.
- GORTER 2021 = DURK GORTER, *Multilingual inequality in public spaces: towards an inclusive model of linguistic landscapes*, in *Multilingualism in the public spa-*

- ces: *Empowering and transforming communities*, edited by Robert Blackwood, Deirdre A. Dunlevy, London, Bloomsbury, 2021, pp. 1-22.
- GORTER-CENOZ-VAN DER WORP 2022 = DURK GORTER, JASONE CENOZ, KARIN VAN DER WORP, *Global and Local Forces in Multilingual Landscapes: A Study of a Local Market*, in *Spaces of Multilingualism*, edited by Robert Blackwood, Unn Røyneland, New York, Routledge, 2022, pp. 188-211.
- HIIPPALA ET AL. 2019 = TUOMO HIIPPALA, ANNA HAUSMANN, HENRIKKI TENKANEN, TUULI TOIVONEN, *Exploring the linguistic landscape of geotagged social media content in urban environments*, in «Digital Scholarship in the Humanities», 34 (2), 2019, pp. 290-309.
- IVKOVIC-LOTHERINGTON 2009 = DEJAN IVKOVIC, HEATHER LOTHERINGTON, *Multilingualism in cyberspace: Conceptualising the virtual linguistic landscape*, in «International Journal of Multilingualism», 6, 2009, pp. 17-36.
- JAWORSKI-THURLOW 2010 = ADAM JAWORSKI, CRISPIN THURLOW, *Semiotic Landscapes. Language, Image, Space*, New York, Continuum, 2010.
- KALLEN-DHONNACHA-WADE 2020 = JEFFREY L. KALLEN, ESTHER NÍ DHONNACHA, KAREN WADE, *Online Linguistic Landscapes: Discourse, Globalization, and Enregisterment*, in *Reterritorialising linguistic landscapes: Questioning boundaries and opening spaces*, edited by David Malinowski, Stefania Tufi, London, Bloomsbury, 2020, pp. 96-116.
- KELLERMAN 2014 = AHARON KELLERMAN, *The Internet as a Second Action Space*, Londra/New York, Routledge, 2014.
- KELLERMAN 2019 = AHARON KELLERMAN, *The Internet City. People, Companies, Systems and Vehicles*, Cheltenham/Northampton, Edward Elgar, 2019.
- LANDRY-BOURHIS 1997 = RODRIGUE LANDRY, RICHARD Y. BOURHIS, *Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study*, in «Journal of Language and Social Psychology», 16 (1), 1997, pp. 23-49.
- LEFEBVRE 1991 = HENRI LEFEBVRE, *The Production of Space*, Oxford, Blackwell, 1991.
- LODA-BONATI-PUTTILLI 2020 = MIRELLA LODA, SARA BONATI, MATTEO PUTTILLI, *History to eat. The foodification of the historic centre of Florence*, in «Cities», 103, 2020, pp. 1-11.
- LYONS 2020 = KATE LYONS, *The Quality of Quantity*, in *Reterritorialising linguistic landscapes: Questioning boundaries and opening spaces*, edited by David Malinowski, Stefania Tufi, London, Bloomsbury, 2020, pp. 31-55.
- PENNYCOOK & OTSUJI 2015 = ALASTAIR PENNYCOOK, EMI OTSUJI, *Making scents of the landscape*, in «Linguistic Landscape», 1, 2015, pp. 191-212.
- PURSCHKE 2017 = CHRISTOPH PURSCHKE, *Crowdsourcing the linguistic landscape of a multilingual country. Introducing Lingscape in Luxembourg*, in «Linguistik online», 85, 2017, pp. 181-202.

Lorenzo Cambi

- PURSCHE 2021 = CHRISTOPH PURSCHE, *Crowdscapes. Participatory research and the collaborative (re)construction of linguistic landscapes with Lingscape*, in «Linguistics Vanguard», 7, 2021, pp. 1-12.
- PUTTILI-BONATI-PORTINARO 2020 = MATTEO PUTTILI, SARA BONATI, LAURA PORTINARO, *Urban meatification. Esplorazioni visuali tra tutela e consumo del patrimonio culturale nel centro storico di Firenze*, in «Geotema», 62, 2020, pp. 61-70.
- RGI = *Rivista geografica italiana*, Milano, Franco Angeli.
- SBICCA 2018 = JOSHUA SBICCA, *Food, Gentrification, and the Changing City*, in «Boletín ecos», 43, 2018.
- SCOLLON-SCOLLON 2003 = RON SCOLLON, SUZIE WONG SCOLLON, *Discourses in place. Language in the Material World*, London, Routledge, 2003.
- VERTOVEC 2007= STEVEN VERTOVEC, *Super-diversity and its implications*, in «Ethnic and Racial Studies», 30 (6), 2007, pp. 1024-1054.
- VERTOVEC 2023 = STEVEN VERTOVEC, *Superdiversity. Migration and Social Complexity*, New York, Routledge, 2023.
- WEI-HUA 2021 = LI WEI, ZHU HUA, *Making sense of handwritten signs in public spaces*, in «Social Semiotics», 31 (1), 2021, pp. 61-87.
- YAO 2021 = XIAOFANG YAO, *Metrolingualism in online linguistic landscapes*, in «International Journal of Multilingualism», 2021, DOI: 10.1080/14790718.2021.1887197.

Sitografia

Lingscape = <https://lingscape.uni.lu/>

Riassunto In questo contributo si affronta il tema del *Linguistic Landscape*, soffermandosi sulle questioni metodologiche che riguardano in particolare la raccolta e la classificazione dei dati. Si è posta particolare attenzione agli aspetti positivi, ma soprattutto ai limiti, degli approcci tradizionali, quantitativo e qualitativo, suggerendo lo sviluppo di una metodologia mista che preveda sia una parte quantitativa, la più oggettiva e ripetibile possibile, sia una qualitativa, etnografica, coinvolgendo attori e spettatori del LL. Si è poi accennato alla questione del PL virtuale, sottolineando l'importanza di tenere in considerazione anche lo spazio online per la caratterizzazione di spazi superdiversi quali sono quelli contemporanei, soprattutto in contesto urbano. La parte centrale del

lavoro riporta tre proposte di classificazione dei dati del paesaggio linguistico: quella di AMOS 2016, quella di AMOS-Soukup 2020 e quella dell'applicazione per dispositivi mobili *Lingscape*. In riferimento a quest'ultima, è stata data una descrizione più puntuale delle variabili, le quali sono state presentate congiuntamente a del materiale fotografico raccolto durante i sopralluoghi al mercato centrale di S. Lorenzo. Il contributo si conclude con un'ultima parte che invita a considerare anche la variabile dell'*online-offline nexus*, non contemplata da nessun altro modello, come punto di contatto tra il paesaggio linguistico in senso stretto, fisico, e quello virtuale. Come accennato in precedenza, lo spazio virtuale è fondamentale per una più profonda comprensione delle dinamiche presenti negli spazi sociali contemporanei.

Abstract This contribution deals with the topic of the Linguistic Landscape, dwelling on the methodological issues that particularly concern data collection and classification. Particular attention was paid to the positive aspects, but above all to the limitations, of the traditional approaches, quantitative and qualitative, suggesting the development of a mixed methodology involving both a quantitative part, as objective and repeatable as possible, and a qualitative, ethnographic one, involving actors and viewers of the LL. The virtual PL was then touched upon, emphasising the importance of taking online space into account for the characterisation of super-diverse spaces such as contemporary ones, especially in an urban context. The central part of the paper reports three proposals for the classification of linguistic landscape data: that of AMOS 2016, that of AMOS-Soukup 2020 and that of the mobile application *Lingscape*. Concerning the latter, a more detailed description of the variables was given, which was presented in conjunction with photographic material collected during the visits to the central market of St. Lorenzo. The contribution concludes with a final part that invites us to also consider the variable of the online-offline nexus, not contemplated by any other model, as a point of contact between the linguistic landscape in the strict, physical sense and the virtual one. As mentioned earlier, the virtual space is crucial for a deeper understanding of the dynamics present in contemporary social spaces.

Codifica XML-TEI: proposta di mark-up per i dizionari metodici

Barbara Patella

1. Vocabolari metodici: da prodotti “statici” a prodotti “dinamici”

Poiché il fulcro della ricerca di dottorato ricade sulla marcatura dei principali dizionari metodici italiani dell'Ottocento¹ – quindi su un'elaborazione informatica finalizzata a rendere interrogabili testi retrodigitalizzati² che, raccolti in una banca dati, saranno consultabili su una

- 1** Per un focus sulla tipologia della lessicografia metodica (o sistematica) si vedano in particolare MARELLO 1980 – riferimento fondamentale per questo filone di ricerca –, SESSA 1979 e APRILE 2023.
- 2** I dizionari inclusi nella banca dati sono i seguenti: 1) FRANCESCO TARANTO E CARLO GUACCI, *Vocabolario domestico italiano ad uso de' giovani, ordinato per categorie*, Napoli, Dalla Stamperia del Vaglio, 1849 (2^a ed., 1851); 2) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti, ad alcuni mestieri, a cose domestiche, e altre di uso comune, per saggio di un vocabolario metodico della lingua italiana. Parte prima. Vocabolario domestico*, Torino, Stamperia Reale, 1846 (2^a ed., 1851); 3) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti [...]. Parte seconda. Vocabolario metodico d'arti e mestieri*, Torino, Stamperia Reale, 1853; 4) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti [...]. Parte terza postuma contenente il vocabolario dei veicoli su terra, e dei veicoli su acqua, e frammenti relativi ai vocaboli mercantili, alla zecca, ed al cavalcare*, Torino, Stamperia Reale, 1860; 5) STEFANO PALMA, *Vocabolario metodico-italiano. Parte che si riferisce all'agricoltura e pastorizia, arti ed industrie che ne dipendono*, I-II, Milano, Libreria editrice di Educazione e d'Istruzione di Paolo Carrara, 1870; 6) STEFANO PALMA, *Prontuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico*, Milano, Paolo Carrara Editore-Libraio, 1877; 7) ANGIOLINA BULGARINI, *Prontuario di voci concernenti i lavori donneschi*, Torino, Stamperia Reale di G.B. Paravia e comp., 1878; 8) AURELIO GOTTI, *Vocabolario metodico della lingua italiana. Casa*, Torino, Ditta

piattaforma online –, questo contributo vuole offrire una riflessione di metodo e, al contempo, una serie di applicazioni dimostrative riguardo a un sistema di mark-up in parte sperimentale. Parliamo di un sistema sperimentale perché il linguaggio che adoperiamo, l'XML-TEI³, è architettato e collaudato per funzionare prevalentemente su repertori semasiologici (in cui la ricerca procede da un significante noto al relativo significato ignoto) e non su repertori come i nostri che, oltre a fondarsi sul criterio opposto, quello onomasiologico (nei quali si parte da significati noti per giungere a significanti ignoti), presentano un ordinamento non alfabetico: le entrate, infatti, sono disposte all'interno di categorie tematiche (come “piante”, “abiti e accessori”, “cibi e bevande”, “casa e sue parti” ecc.) secondo associazioni concettuali e relazioni semantiche (soprattutto in base a rapporti di iponimia e meronimia).

Si è pensato, quindi, di informatizzare i dizionari metodici poiché rappresentano – per struttura, intenti e contenuti – fonti preziose per più rami della linguistica italiana (lessicografia, storia della lessicografia, storia della lingua, dialettologia), soprattutto per il tentativo rinnovatore dei compilatori, che furono, da un lato, attenti al recupero di quei settori lessicali trascurati dalla lessicografia tradizionale (dunque solleciti alla raccolta di voci tecniche di arti e mestieri, voci scientifiche, voci della conversazione) e, dall'altro, sensibili a forme ed espressioni dell'uso parlato (in controtendenza rispetto alla prassi lessicografica, fino ad allora fortemente letterariocentrica)⁴ – non a caso, furono promotori di inchieste sul campo *ante litteram* (condotte principalmente in Toscana).

G.B. Paravia e comp., 1883; 9) FRANCESCO CORAZZINI, *La città e lo stato. La casa e la famiglia. Dizionario metodico*, Torino, Ermanno Loescher, 1885. Per una descrizione del progetto si rimanda a PATELLA 2023.

- 3 Per la marcatura dei dizionari metodici siamo ricorsi allo standard di codifica internazionale XML-TEI e alle linee guida distribuite dal consorzio TEI – con particolare riguardo al nono capitolo, riservato proprio all'elaborazione dei repertori lessicali (cfr. TEI P5 2021, pp. 291-329) –, giacché forniscono un assetto di istruzioni capace di regolamentare il mark-up e, al tempo stesso, di personalizzarlo, favorendo inoltre una sostenibilità del progetto e un'interoperabilità dei dati marcati.
- 4 Tradizionalmente, «nei dizionari della lingua italiana si era dato largo spazio [...] al lessico “alto” della prosa e della poesia, attraverso gli esempi e le citazioni tratte

Addentrandoci, allora, nei processi di trasformazione che da dati “statici” (vocabolari metodici cartacei) conducono a prodotti interrogabili e dinamici (vocabolari metodici digitalizzati e marcati), mostriamo alcuni dei criteri e degli espedienti che abbiamo adottato con lo scopo di potenziare la consultabilità e la consultazione di tali risorse. Considerate le particolari caratteristiche (a cominciare dall’ordinamento per temi e dall’eterogeneità della microstruttura), è stato necessario – in funzione del lavoro informatico – uno studio preliminare sia dei repertori stessi sia delle regole dell’XML-TEI, dal momento che una certa disomogeneità si riscontra non solo fra un dizionario metodico e l’altro (variabilità esterna), ma anche all’interno di uno stesso vocabolario (variabilità interna).

2. Mark-up a livello di macrostruttura

Relativamente agli apparati, per i nostri dizionari abbiamo optato per una marcatura essenzialmente “leggera”, che ha interessato l’annotazione di tali sezioni di corredo mediante l’impiego dei tag <front>, <back>, <div>, <head>, <pb>, <p>: ci siamo serviti dell’elemento <div> (*text division*) per delimitare qualunque blocco di testo, ossia ciascun esotesto⁵ (o apparato) che può trovarsi in apertura (nel *front matter*: <front>⁶) o in chiusura (nel *back matter*: <back>⁷) di un dizionario⁸;

dagli scrittori, ma non c’era posto per le parole “basse” e comuni della vita quotidiana, che non erano nominate, ovviamente, nelle opere letterarie. Proprio per colmare questa lacuna, e per andare incontro alle richieste di un pubblico che sempre più spesso cercava nei dizionari il termine “nazionale” con cui designare gli oggetti che avevano nomi diversi da regione a regione, nell’Ottocento [...] esplose la moda dei dizionari metodici o “sistematici”» (DELLA VALLE-PATOTA 2006, pp. 138-139).

⁵ Cfr. RICCIO 2016, p. 64.

⁶ Cfr. TEI P5 2021, pp. 193-196.

⁷ Cfr. *ivi*, pp. 198-200.

⁸ Oltre a ciò, il tag <div> (previsto non solo negli apparati ma anche all’interno del <body>) serve inoltre per etichettare le suddivisioni contenute nel corpo del dizionario, dal momento che i dizionari metodici, organizzando le entrate in sezioni

a esso è subordinato il tag <head>, usato per racchiudere i titoli dei blocchi testuali; <pb> è invece il marcatore dei numeri di pagina (*page beginning*), mentre <p> (*paragraph*) circoscrive ogni porzione di testo contenuta all'interno di <div>. In base a ciò, mediante il lavoro di mark-up è possibile isolare e rendere interrogabili (nell'ottica di una piattaforma online) le seguenti sezioni:

- prefazione/premessa/introduzione
- avvertimenti
- dedica
- postfazione
- note a piè di pagina
- altra sezione (senza specificazione)

A sostegno di quanto detto, forniamo due esempi tratti dal *Pronuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico* di Stefano Palma (1877).

Nella prima delle due immagini (figura 1) riportiamo la tipica modalità di marcatura delle prefazioni: contenuta negli apparati iniziali (quindi inclusa nella sezione <front>⁹), la prefazione è marcata con <div> (che ne indica l'inizio e, mediante attributo, il relativo valore: *type="preface"*), sotto cui si trovano annidati il titolo (delimitato da <head>) e il testo della prefazione (contenuto in <p>); infine, il tag di chiusura col quale si segnala la fine della sezione (</div>). Per giunta, se lo si ritiene necessario, il testo può essere suddiviso in più parti (cioè si può inserire un elemento <p> per ogni porzione testuale che si vuole etichettare), secondo lo schema seguente:

tematiche, presentano il lemmario suddiviso e articolato in parti, capitoli e/o paragrafi.

- ⁹ Grazie alla modalità di visualizzazione dell'*outline* (osservabile nella colonna di sinistra sia della figura 1 sia della figura 2), si può notare l'impianto strutturale del *Pronuario* di Palma: vediamo, ad esempio, che il *front matter* si compone di due blocchi testuali (annotati con <div>) oppure che il corpo del dizionario (*body*) è articolato in più capitoli.

```

<div type = "preface">
  <head> TITOLO DELLA PREFAZIONE </head>
  <p> testo </p>
  <p> testo </p>
  [...] [...] [...]
  <p> testo </p>
</div>

```

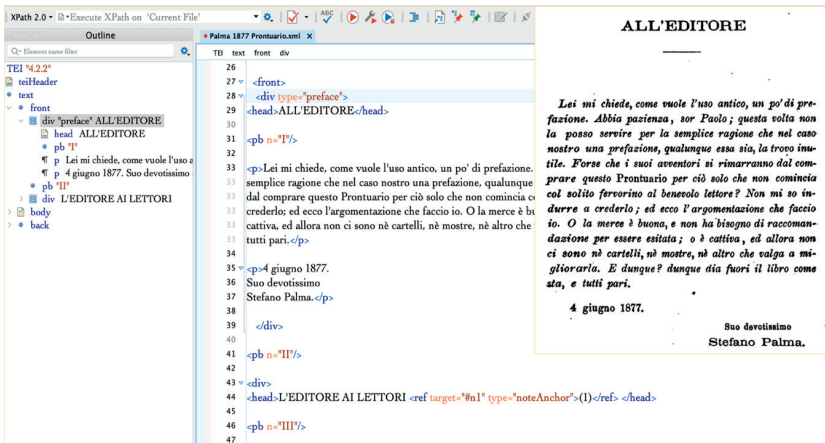


Figura 1 Schermata tratta da *Oxygen XML Editor* (a sinistra) con mark-up applicato alla prefazione contenuta in PALMA 1877 (riprodotta a destra, p. 1)

Nella seconda immagine (figura 2) osserviamo un altro apparato iniziale (che segue la prefazione appena vista), sempre delimitato da <div>, ma questa volta senza alcun attributo (@type) né valore di specificazione (come “preface”, “dedication” o simili). Qui, inoltre, notiamo che l’esotesto contiene un ulteriore elemento, cioè una nota a piè di pagina, che andrà annotata mediante l’uso di due stringhe:

```

<ref target="#n1" type="noteAnchor"> (1) </ref>
<note type="footnote" xml:id="n1"> testo della nota </note>

```

La prima stringa, a cui sono associati il riferimento al testo della nota (“#n1”) e il valore “noteAnchor”, è impiegata per marcare il numero – o, in altri casi, il simbolo – collegato alla nota; la seconda, contrassegnata dal valore “footnote” e da un identificatore univoco (“n1”), serve a marcare il testo della nota.

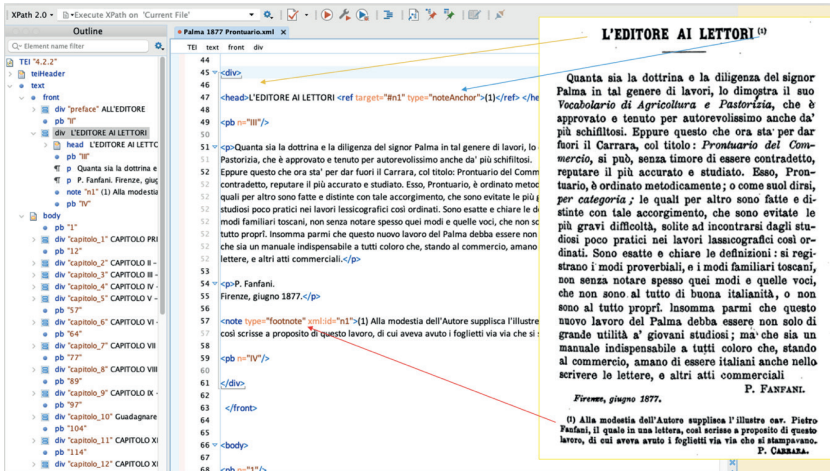


Figura 2 Schermata tratta da Oxygen XML Editor (a sinistra) con mark-up applicato a uno degli apparati iniziali contenuti in PALMA 1877 (riprodotto a destra, p. III)

2.1 Il tag <div> nel body

Considerato che tutti i vocabolari sistematici inclusi nella nostra banca dati prevedono – in base ai temi trattati – una suddivisione interna in più parti e/o in più capitoli (e talvolta persino in paragrafi) entro cui sono distribuite le entrate, ricorreremo all’elemento <div> anche nel corpo del dizionario (<body>), con lo scopo di isolare ogni suddivisione presente ed eventualmente specificarne il valore:

- <div> [...] </div> → sezione senza specificazione
- <div type = “part”> [...] </div> → parte (non numerata)

```

<div type = "part" n= "4"> [...] </div>    → parte 4
<div type = "chapter" n= "3"> [...] </div> → capitolo 3

<body>
<pb n="1"/>
<div type="chapter" n="1" xml:iid="capitolo_1">
<head> CAPO I. Dei lavori di maglia.</head>
<entry>
<form type="lemma" subtype="phrase"><orth>Lavori di maglia</orth></form> <sense>sono quelli che si fanno intrecciando il filo con ferri o
con altri arnesi per modo che formino tanti anelli l'uno attaccato all'altro; e sono: la <re
type="hyponym"></form><orth>Maglia</orth></form></re> propriamente detta, il <re
type="hyponym"></form><orth>Modano</orth></form></re>, l'<re type="hyponym"></form><orth>Uncinetto</orth></form></re>, il <re
type="hyponym"></form><orth>Chiacchierino</orth></form></re>, la <re type="hyponym"></form type="phrase"><orth>Trina a
tombolo</orth></form></re>.</sense>
</entry>
<entry>
<form type="lemma"><orth>Maglia</orth>
<form type="not-recommended"><orth> Punto </orth></form>
</form> <sense>Così chiamasi ciascun anello dei suddetti lavori.</sense>
</entry>

[...]   [...]   [...]

```

Figura 3 Schermata tratta da *Oxygen* con esempio di mark-up applicato al corpo del dizionario di BULGARINI 1878 (p. 1)

3. Mark-up a livello di microstruttura

Quanto alla microstruttura dei vocabolari metodici, alquanto composita, la vera sfida del progetto consiste nel valorizzare e nel far emergere, da un lato, i numerosissimi sottoelementi che sono presenti all'interno degli articoli lessicografici – quali parole, locuzioni e proverbi non registrati come entrate autonome oppure i commenti dei compilatori – e, dall'altro, le relazioni semantiche sottese sia allo sviluppo verticale (costituito dal lemmario) sia a quello orizzontale (rappresentato dal materiale lessicale annidato sotto le voci a espone), vale a dire i rapporti di iponimia-iperonimia, olonimia-meronimia e sinonimia-antonimia, sui quali la lessicografia sistematica fonda la concatenazione di entrate e sottoentrate. Per tali ragioni, accanto agli elementi – per così dire – canonici (quali lemmi, campo

del significato e rinvio ad altre voci)¹⁰ ci siamo prefissati di restituire informaticamente anche una serie di informazioni che differenziano la lessicografia metodica: i commenti, la giustapposizione di voci a esponente, la tipologia delle entrate registrate (forme flesse, locuzioni, proverbi), le voci correlate (al lemma o a voce interna) annidate all'interno degli articoli.

Alla luce di ciò, secondo lo schema configurato per la microstruttura, prevediamo la marcatura dei seguenti elementi:

lemmi	
Proposta di mark-up a livello di microstruttura :	campi correlati (collocati, sinonimi, meronimi, polirematiche, proverbi ecc.)
	commenti dei lessicografi
	indicatori d'uso / di registro
	rinvio ad altre voci

Nelle tabelle che presentiamo di seguito illustriamo, più in dettaglio, la serie degli elementi lessicali, semantici o testuali che sono stati oggetto di mark-up. Nella colonna di sinistra sono indicate le stringhe XML-TEI (con tag, attributi e valori), mentre a destra i corrispondenti elementi marcati a livello di microstruttura e di mediostruttura¹¹:

¹⁰ All'opposto, nel nostro modello di marcatura non abbiamo contemplato elementi considerati abituali negli articoli lessicografici, quali categorie grammaticali, esempi, etimologie, in quanto informazioni fornite sporadicamente o in modo irregolare e incoerente nei dizionari metodici, nei quali la redazione delle voci è spesso eterogenea e non predicibile.

¹¹ Cfr. RICCIO 2016, pp. 65-68.

Serie di tag, attributi e valori	Elementi marcati
<code><form type="lemma"><orth></orth></form></code>	<i>Lemma</i> (unità lessicale semplice)
<code><form type="lemma" subtype="inflected"><orth></orth></form></code>	<i>Lemma</i> (unità lessicale in forma flessa)
<code><form type="lemma" subtype="phrase"><orth></orth></form></code>	<i>Lemma</i> (unità lessicale di due o più parole)
<code><form type="variant"><orth></orth></form></code>	<i>Variante del lemma</i>
<code><form type="obsolete"><orth></orth></form></code>	<i>Forma obsoleta</i>
<code><form type="not-recommended"><orth></orth></form></code>	<i>Parola/ Locuzione sconsigliata dal lessicografo</i>
<code><re><form><orth></orth></form></re></code>	<i>Voce correlata</i> al lemma o a <code><re></code>
<code><re type="colloc"><form><orth></orth></form></re></code>	<i>Parola co-occorrente</i> con il lemma o con <code><re></code>
<code><re type="hyperonym"><form><orth></orth></form></re></code>	<i>Iperonimo</i> (del lemma o di <code><re></code>)
<code><re type="hyponym"><form><orth></orth></form></re></code>	<i>Iponimo</i> (del lemma o di <code><re></code>)
<code><re type="meronym"><form><orth></orth></form></re></code>	<i>Meronimo</i> (del lemma o di <code><re></code>)
<code><re type="derivation"><form><orth></orth></form></re></code>	<i>Base/Derivato</i> del lemma o di <code><re></code>
<code><re type="antonym"><form><orth></orth></form></re></code>	<i>Antonimo</i>
<code><re type="synonym"><form><orth></orth></form></re></code>	<i>Sinonimo</i>
<code><re type="geo-synonym"><form><orth></orth></form></re></code>	<i>Geosinonimo</i>
<code><re type="foreign"><form><orth></orth></form></re></code>	<i>Forestierismo</i>

Figura 4 Tabella con tipologie di lemmi e di voci correlate

Serie di tag, attributi, valori	Elementi marcati
<code><xr type="hyperonym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>iperonimo</i> (lemmatizzato)
<code><xr type="hyponym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>iponimo</i> (lemmatizzato)
<code><xr type="meronym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>meronimo</i> (lemmatizzato)
<code><xr type="holonym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>olonimo</i> (lemmatizzato)
<code><xr type="derivation"><ref target="#lemma"> </ref></xr></code>	Rinvio al <i>derivato</i> o alla <i>base</i> del lemma
<code><xr type="antonym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>antonimo</i> (lemmatizzato)
<code><xr type="synonym"><ref target="#lemma"> </ref></xr></code>	Rinvio a un <i>sinonimo</i> (lemmatizzato)
<code><entry type="holonym"></code>	<i>Olonimo</i> (di altre <i>entries</i> " <i>meronym</i> ")
<code><entry type="meronym"></code>	<i>Meronimo</i> (di una <i>entry</i> " <i>holonym</i> ")
<code><entry type="hyperonym"></code>	<i>Iperonimo</i> (di altre <i>entries</i> " <i>hyponym</i> " annidate)
<code><entry type="hyponym"></code>	<i>Iponimo</i> (di una <i>entry</i> " <i>hyponym</i> ")
<code><entry type="xref"></code>	Entrata con rinvio a un'altra voce
<code><sense> </sense></code>	Campo del significato

Serie di tag, attributi, valori	Elementi marcati
<sense value="comment" n="35"> </sense>	Note di commento del lessicografo numerate
<sense value="comment"> </sense>	Note di commento del lessicografo non numerate
<usg> </usg>	Indicazioni d'uso, di registro o simili
<superEntry></superEntry>	Gruppo di voci annidate
 	Porzione di testo esclusa dalla marcatura

Figura 5 Tabella con tipologie di rinvii, entrate e altri elementi

3.1 Voci sconsigliate dai lessicografi: il caso del *Prontuario* di Bulgarini

Un'ulteriore caratteristica che contraddistingue i vocabolaristi meto-
dici – in linea col fatto di concepire i propri dizionari come “manuali
di guida” – riguarda la registrazione di parole e locuzioni vitande per
mettere in guardia il lettore. Per questo genere di voci, un caso partico-
lare è rappresentato dal *Prontuario di voci concernenti i lavori donneschi*, in
cui Angiolina Bulgarini allestisce un apparato a piè di pagina destinato
a un uso particolare, come si legge in apertura del volume:

Ho poi notato a piè di pagina i vocaboli stranieri o inesatti che un uso,
secondo me men buono, o certi scrittori di guide di lavori donneschi ci fanno
adoperare invece di quelli più italiani e più propri che pur il buon uso vivente
ci porge, e del quale appunto sono le voci da me segnate¹².

In tal caso è stato necessario progettare una soluzione alternativa,
rispetto all'originale cartaceo, per trattare informaticamente queste
voci, così da recuperarle e restituirle in funzione delle ricerche sulla
piattaforma: nel mark-up abbiamo perciò rinunciato a etichettarle
come note a piè di pagina, perché ciò che ci interessava non era il luogo
occupato nella pagina, bensì il valore di questo materiale lessicale; da
qui si è deciso di farle “risalire” in fase di marcatura accanto ai lemmi,
etichettandole come “voci non raccomandate” tramite la stringa

¹² BULGARINI 1878, *Avvertenza*, p. IV.

<form type="not-recommended"> <orth> voce scongiata </orth> </form>

Mostriamo allora un'immagine del *Prontuario* (BULGARINI 1878, p. 7) con la corrispondente resa nell'editor XML (*Oxygen*), in cui è possibile osservare due casi di voci non raccomandate: il primo riguarda l'entrata *maglie sode*, a cui si accompagna una numerazione (posta fra parentesi tonde) che rimanda a due voci scongiate (*briglie*, *barrette*) – e per questo collocate a piè di pagina dalla lessicografa –; il secondo caso concerne il lemma *pippiolini*, al quale è associato – mediante la stessa modalità di *maglie sode* – il termine *picot* (francesismo da evitare). In entrambi i casi, in fase di mark-up, abbiamo fatto risalire le voci non raccomandate, affiancandole al lemma a cui sono connesse:

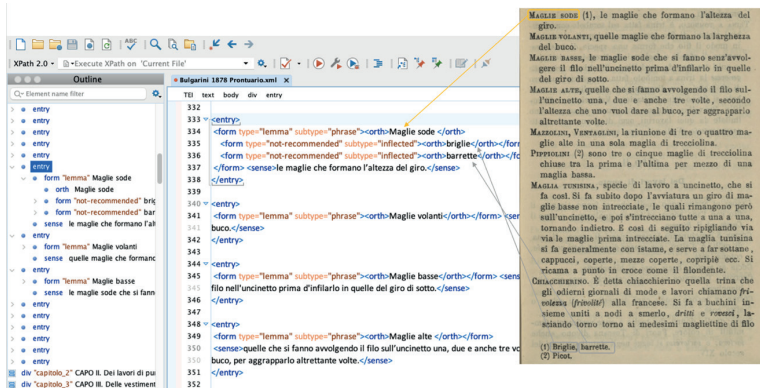


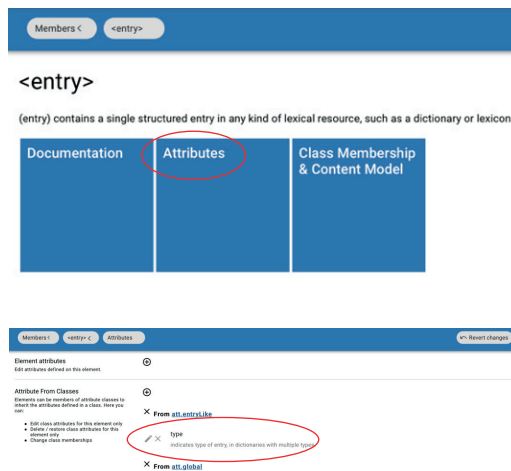
Figura 6 Schermata tratta da *Oxygen* (a sinistra) con mark-up applicato a BULGARINI 1878 (riprodotto a destra, p. 7)

4. Personalizzazione dello schema con Roma web application

Come abbiamo accennato a più riprese, poiché i vocabolari metodici presentano una larga eterogeneità strutturale, è stato fondamentale poter personalizzare il modello di mark-up (lo schema XML-TEI) e adattarlo alle nostre esigenze al fine di estrapolare informazioni ed elementi linguisticamente interessanti: per questo ci siamo serviti di

un'applicazione web fornita dal consorzio TEI, ossia *Roma web application*¹³, che consente di aggiungere, rimuovere o modificare elementi, attributi e valori prestabiliti rispetto allo standard di riferimento; questo strumento permette di diversificare il mark-up in base ai nostri scopi e potenziare la consultabilità dei vocabolari metodici (a partire dalla marcatura delle relazioni semantiche), rimanendo tuttavia aderenti alle norme imposte dalle linee guida (*P5 Guidelines*)¹⁴.

Presentiamo allora alcuni passaggi dell'intero procedimento con cui è possibile personalizzare lo schema¹⁵, prendendo come esempio le modifiche relative a <entry>, tag che delimita le entrate. Dopo aver selezionato l'elemento di nostro interesse (<entry>), tramite l'opzione "Attributes", possiamo scegliere relativi attributi e valori da modificare; per questa simulazione scegliamo l'attributo @type:



13 Specificiamo che abbiamo testato la nuova versione di questo tool (raggiungibile all'indirizzo <https://romabeta.tei-c.org>), che, pur essendo ancora in versione beta, funziona correttamente.

14 Cfr. TEI P5 2021.

15 Di seguito riportiamo alcune schermate esemplificative tratte dall'applicazione *Roma*.

Decidiamo, quindi, se aggiungere o eliminare i valori associati a *entry* (mediante *@type*) in base al disegno di tagset prestabilito e alle funzionalità di ricerca che vorremo rendere disponibili. Mostriamo il caso in cui aggiungiamo dei valori non contemplati dall'XML-TEI, fra cui ad esempio “meronym” e “hyperonym” (per i quali forniamo anche una descrizione della funzione che assumono nel sistema di marcatura):

Values
Set values for this attribute.

hom (homograph) groups information relating to one homograph within an entry.

Description
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2008-02-01" xml:lang="en"
  >groups information relating to one
  homograph within an entry.</desc>
```

xref (cross reference) a reduced entry whose only function is to point to another main entry (e.g. for forms of an irregular verb or for variant spellings: was pointing to be, or esthete to aesthete).

meronym

Description
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2023-02-23" xml:lang="en">
  an entry that is a meronym of another
  entry </desc>
```

hyperonym

Description
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2023-02-23" xml:lang="en">
  an entry that is a hyperonym for one or
  more entries </desc>
```

Una volta completato il procedimento, l'applicazione permette di generare l'XML *schema* personalizzato e salvare il documento in più formati (*RelaxNG schema*, *W3C schema*, *DTD*, ecc.); a questo punto lo schema può essere associato a un determinato testo all'interno di un editor XML – nel nostro caso specifico in *Oxygen*, dove la validità è confermata dai valori che (correttamente inseriti con *Roma* e riconosciuti dall'editor) appaiono nel menù a tendina accanto all'attributo *@type* (collegato a *<entry>*). Come si vede dalla schermata, vengono suggeriti sia i valori preimpostati dalla TEI (“abbr”, “affix”, “foreign”, “hom”, “main”, “supplemental”, “xref”) sia quelli aggiunti da noi (“hyperonym”, “derivation”, “holonym”, “hyponym”, “meronym”):

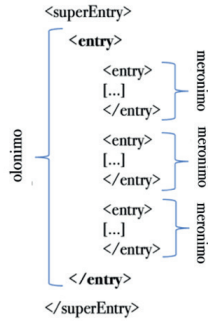


Figura 7 Schermata di Oxygen contenente il testo del *Vocabolario domestico di Carena* (1851)

4.1 Il tag <superEntry>

Restando nell'ambito della personalizzazione dello schema di marcatura, segnaliamo che in qualche caso siamo andati a reinventare o a ridefinire la funzione dei tag, come ad esempio per <superEntry>. Secondo le norme, <superEntry> può essere impiegato per raggruppare «a sequence of entries within any kind of lexical resource, such as a dictionary or lexicon with function as a single unit, for example a set of homographs»¹⁶; noi ne abbiamo rideterminato il ruolo – estendendolo – in conformità all'ordine ideologico dei vocabolari metodici, impiegandolo come tag “contenitore” per raggruppare non solo eventuali omografi, ma anche entrate legate da rapporti semantici (es. olonimo/meronimi o iperonimo/iponimi) o lessicali (es. corradicali; base/derivato), con lo scopo di poterle isolare e individuare nelle opzioni di ricerca del sito. Pertanto, per un olonimo e i relativi meronimi (come pure per iperonimi e iponimi) il criterio gerarchico da rispettare sarà il seguente:

¹⁶ TEI P5 2021, p. 292.



In conclusione mostriamo la modalità d'impiego di <superEntry>, applicandola a uno dei repertori della banca dati; nella fattispecie, proponiamo un caso di meronimia presente nel *Vocabolario domestico* di Carena (1851). Nel primo capitolo, dedicato al *Vestire e alle sue accompagnature*, il lessicografo fa seguire all'olonimo *anello* i meronimi che ne dipendono (*gambo, castone, fondo del castone, fascia del castone, foglia*); grazie al mark-up, siamo in grado di restituire la relazione implicata e raggruppare le entrate semanticamente connesse all'interno di <superEntry>, in base al seguente schema di annidamento:

```

<entry>
  <entry>
    [...] [...] [...]
  </entry>
</entry>

```

Osserviamo allora da vicino, relativamente agli elementi in questione, la marcatura che abbiamo realizzato all'interno dell'editor:

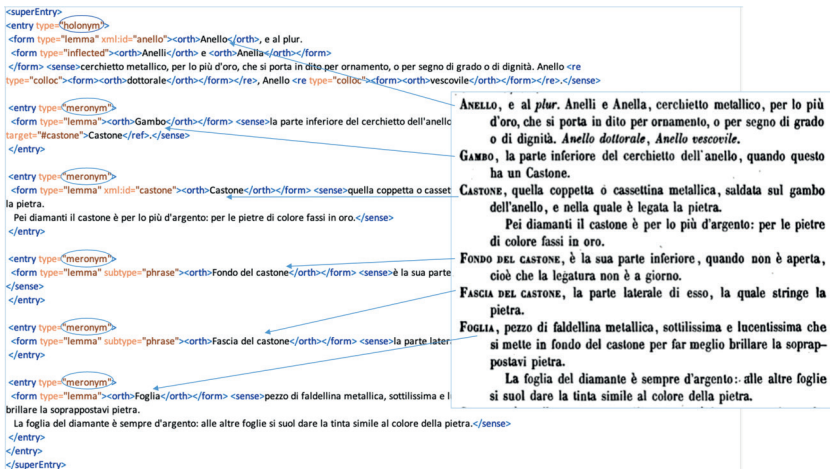


Figura 8 Schermata di *Oxygen* (a sinistra) con mark-applicato a CARENA 1851 (riprodotto a destra, p. 32)

Conclusioni

In queste pagine, che si inseriscono nel solco della lessicografia digitale, abbiamo voluto offrire un piccolo saggio dei criteri e dei procedimenti di mark-up adottati per i dizionari metodici, una particolare tipologia lessicografica che, per il criterio onomasiologico e per l'ordinamento non alfabetico, ha rappresentato un banco di prova per il linguaggio XML-TEI. Attraverso esempi e applicazioni di marcatura, abbiamo dato dimostrazione delle operazioni con cui è possibile trasformare dizionari metodici cartacei in prodotti digitali interrogabili, mostrando peraltro alcune delle soluzioni necessariamente escogitate qualora lo standard TEI non contemplasse condizioni analoghe a quelle dei nostri repertori.

Pertanto, con l'informatizzazione dei vocabolari metodici si prospettano notevoli vantaggi – in termini sia quantitativi sia qualitativi – in previsione delle ricerche che verranno predisposte sulla piattafor-

ma online e, di conseguenza, dei risultati che il supporto informatico consentirà di estrapolare ed elaborare.

Bibliografia

- APRILE 2023 = MARCELLO APRILE, *I dizionari metodici nell'Ottocento*, in Emiliano Picchiorri e Maria Silvia Rati (a cura di), *La lessicografia italiana dell'Ottocento. Bilanci e prospettive di studio*, Atti del Convegno, Università "G. D'Annunzio" di Chieti-Pescara (Chieti, 24-25 maggio 2022), Firenze, Cesati, 2023, pp. 101-123.
- BULGARINI 1878 = ANGIOLINA BULGARINI, *Prontuario di voci concernenti i lavori donneschi*, Torino, Stamperia Reale di G.B. Paravia e comp., 1878.
- CARENA 1851 = GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti, ad alcuni mestieri, a cose domestiche, e altre di uso comune, per saggio di un vocabolario metodico della lingua italiana. Parte prima. Vocabolario domestico*, Torino, Stamperia Reale, 1846 (2^a ed., 1851).
- DELLA VALLE-PATOTA 2006 = VALERIA DELLA VALLE E GIUSEPPE PATOTA, *L'italiano. Biografia di una lingua*, Milano, Sperling & Kupfer, 2006.
- MARELLO 1980 = CARLA MARELLO, *Lessico ed educazione popolare. Dizionari metodici dell'800*, con introduzione di Giovanni Nencioni, Roma, Armando, 1980.
- PALMA 1877 = STEFANO PALMA, *Prontuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico secondo il buon uso toscano, in servizio delle scuole tecniche e commerciali. Con parole preliminari di P. Fanfani*, Milano, Paolo Carrara Editore-Libraio, 1877.
- PATELLA 2023 = BARBARA PATELLA, *Dizionari metodici dell'Ottocento: verso una piattaforma interrogabile*, in Emiliano Picchiorri e Francesco Montuori (a cura di), *In fieri, 4. Ricerche di linguistica italiana*, Atti della IV Giornata dell'ASLI per i dottorandi (Firenze, Accademia della Crusca, 2-4 dicembre 2021), Firenze, Cesati, 2023, pp. 241-247.
- RICCIO 2016 = ANNA RICCIO, *Gli strumenti per la ricerca linguistica. Corpora, dizionari e database*, Roma, Carocci, 2016.
- SESSA 1979 = MIRELLA SESSA, *Arti e mestieri nei vocabolari metodici*, in Atti del "Convegno Nazionale sui lessici tecnici delle arti e dei mestieri" (Cortona «Il Palazzone», 28-30 maggio 1979), 1979, pp. 37-78.

Barbara Patella

TEI P5 2021 = *Guidelines for Electronic Text Encoding and Interchange* (April, 2022), Text Encoding Initiative Consortium (consultabile all'indirizzo <https://tei-c.org/release/doc/tei-p5-doc/it/html/index.html>).

Riassunto Il contributo, incentrato sull'informatizzazione dei principali dizionari metodici italiani dell'Ottocento, si focalizza in particolar modo sulle scelte di mark-up (a livello di macrostruttura e di microstruttura) che, basate sull'XML-TEI, sono state adottate – e in parte personalizzate – per rendere interrogabili questa tipologia di repertori onomasiologici non ordinati alfabeticamente, così da potenziarne la consultazione e garantirne modalità di ricerca avanzate (nell'ottica di una piattaforma che, realizzata con la collaborazione degli informatici, sarà fruibile in rete); inoltre, ai criteri teorici si accompagnano alcune applicazioni dimostrative che sono alla base della costruzione della banca dati dedicata alla lessicografia metodica, in modo tale da offrire un piccolo saggio sia dell'attività di marcatura sia dei procedimenti che dal dato cartaceo conducono al dato digitale.

Abstract This contribution, centred on the computerisation of the main Italian non-alphabetic dictionaries of the 19th century, focuses in particular on the mark-up choices (at the level of macrostructure and microstructure) which, based on XML-TEI, have been adopted – and partly customised – to make this type of onomasiological repertories searchable, so as to enhance its consultation and guarantee advanced research options (with a view to an online platform that will be created with the collaboration of computer scientists). Moreover, the theoretical criteria are accompanied by some demonstrative applications which are at the basis of the construction of the database dedicated to non-alphabetic lexicography, in order to offer a small sample of the mark-up activity as well as of the procedures that lead from the paper data to the digital data.

Indice dei nomi

- Adler, Alfred 155n
Aissen, Judith 4
Albano Leoni, Federico 195
Alcuino di York 99n
Alfonzetti, Giovanna 52, 53, 55n, 59n
Amos, William 293n, 295, 296
Angelucci, Arnaldo 158
Aprile, Marcello 321n
Ascoli, Graziadio Isaia 22n
Assagioli, Roberto 165
Auer, Peter 53, 53n, 55, 55n, 59n, 62, 63
- Bacchelli, Riccardo 158, 161
Backhaus, Peter 295
Bagna, Carla 293n
Ballarè, Silvia 46, 48n
Bárány, Andras 3
Barni, Monica 293n
Baroni, Marco 261n
Beccaria, Gian Luigi 152n, 171n
Bellinzona, Martina 289
Bellucci, Patrizia 174n, 175n
Benincà, Paola 3
Ben-Rafael, Eliezer 290n
Bernardini, Silvia 261n
- Berni, Francesco 73
Berretta, Monica 3, 135, 136, 144
Berruto, Gaetano 46, 51n, 52, 59, 172, 172n
Biffi, Marco 173n, 174n, 176n, 177n
Birago, Francesco 101n
Biscioni, Antonio Maria 80
Bleuler, Eugen 155n
Blommaert, Jan 290, 292, 293n, 294, 311, 312
Boerio, Giuseppe 104n
Bohorič, Adam 116, 116n
Bombi, Raffaella 251, 251n
Bonaparte, Marie 156
Bonati, Sara 314n
Bonghi, Ruggero 93n
Bosazzi, Emilio 175n
Bossong, Georg 3
Bottari, Bartolomeo 74n
Bourhis, Richard 289, 291
Bowerman, Melissa 232, 244
Bressema, Jana 274
Brezina Vaclav 49
Bricchi, Mariarosa 93n, 97n
Brusa, Emilio 76

I dati linguistici. Metodologie e strumenti della ricerca

- Bulgarini, Angiolina 321n, 327, 330, 330n, 331
Buonarroti il Giovane, Michelangelo 73, 80, 100n
Cacioli, Caterina 229
Callebat, Louis 175n
Calvi, Maria Vittoria 292n
Capponi, Gino 94n, 96
Cardinaletti, Anna 22n, 23n
Carena, Giacinto 78n, 321n, 334, 335, 336
Caro, Annibale 73, 80
Carotenuto, Aldo 153n, 155n
Castellani, Arrigo 177n
Cavagnoli, Stefania 171n
Cecchi, Giovan Maria 73, 80, 92, 96, 100, 100n, 103, 104, 105
Cellini, Benvenuto 80
Cenoz, Jasone 291, 292n
Cerruti, Massimo 45n, 46, 61
Cesari, Antonio 74, 88, 89, 91, 94, 95, 95n, 104n
Cesarotti, Melchiorre 75, 92
Cherubini, Francesco 72, 73, 74, 77n, 80, 94, 96n, 97, 97n, 99n, 100n, 103, 103n, 104
Ciccolone, Simone 52n
Cimino, Guido 155n
Cioni, Gaetano 95
Cohen, Antoine 275, 277
Collier, Rene 275, 276-277
Compagni, Dino 80
Comrie, Bernard 4
Corazzini, Francesco 322n
Corsa, Rita 164n
Cortelazzo, Michele 152, 152n, 171n, 172, 172n, 257n
Corti, Maria 76n, 77, 77n, 79n, 86n
Coveri, Lorenzo 171n
Cresti, Emanuela 46, 195, 196, 198, 198n, 199, 206, 212, 213, 273, 274, 275, 276, 282, 284
Croft, William 4
d'Alberti di Villanova, Francesco 104n
Dalmatin, Jurij 116
d'Ambra, Francesco 80, 100n
Dal Negro, Silvia 52n
Dardi, Andrea 74n, 75n, 76n, 78n
David, Michel 151, 152, 154, 155n, 160
Dazzi, Nino 155n
De Angelis, Alessandro 3, 10, 21, 23n
Defrancq, Bart 230
Della Valle, Valeria 323n
De Mauro, Tullio 184n, 271, 279
De Sanctis, Sante 159, 159n, 163
De Sarlo, Andrea 164
Devoto, Giacomo 171n, 175n
Dhonnacha, Esther Nì 294n
Di Caro, Vincenzo Nicolò 22
Di Giorgio Martini, Francesco 173n, 176n, 177n
Ebeling, Jarle 229, 233
Egan, Thomas 229
Enfield, Nick J. 270, 272
Enghels, Renata 229, 230, 231, 233
Faggiuoli, Giovanni Battista 80, 103, 103n
Fanciullo, Franco 21n
Fauriel, Claude 93n
Feroci Luti, Giovanna 73, 96, 100, 102, 102n
Ferrari, Jacopo 77n, 97n
Fiacchi, Luigi 92
Fillmore, Charles J. 234

- Fiorelli, Piero 256
 Firenzuola, Agnolo 73, 80, 100, 100n
 Firenzuoli, Valentina 277
 Fishman, Joshua A. 215
 Fleury, Philippe 175n
 Fliess, Wilhelm 155n
 Forcellini, Egidio 103, 103n
 Foscolo, Ugo 75
 Fresu, Rita 133, 134, 135, 136, 137, 142, 144
 Freud, Anna 156
 Freud, Sigmund 151, 153, 154, 154n, 155, 155n, 156, 160
 Fusco, Fabiana 143, 243
 Fusi Daniele 138

 Gaj, Ljudevit 116n
 Galilei, Galileo 80, 91n
 Galluzzi, Paolo 175n
 Garcia, Ofelia 293n
 Gast, Volker 230
 Gelli, Giovan Battista 80
 Gherardini, Giovanni 97
 Ghirardi, Sabina 94n, 100n, 104n
 Giacalone Ramat, Anna 135
 Gilquin, Gaëtanelle 234
 Giusti, Giuliana 22n, 23n
 Giusti, Giuseppe 93
 Goddard, Cliff 232, 241
 Goodwin, Charles 270
 Goria, Eugenio 46
 Gorter, Durk 291, 292, 292n
 Gotti, Aurelio 321n
 Gotti, Maurizio 171n
 Grazzini, Anton Francesco (il Lasca) 80, 100, 100n, 105, 105n
 Gries, Stefan Thomas 229
 Grossi, Tommaso 71n, 95
 Guacci, Carlo 321n

 Gualdo, Riccardo 152n, 153n, 171n, 172n
 Guardiano, Cristina 3
 Gumperz, John 52, 53n, 58, 59, 62, 63n, 215

 Hart, Johan 275, 276
 Haspelmath, Martin 231
 Hasselgård, Hilde 229, 230, 233
 Hiippala, Tuomo 294n
 Hockett, Charles 270
 Hua, Zhu 303n

 Iemmolo, Giorgio 3
 Isella, Dante 71n, 77n, 78, 78n, 79n, 80n, 81n, 82n, 83n, 84n, 86n, 89
 Ivkovic, Dejan 294

 Jaworski, Adam 292n
 Jefferson, Gail 48
 Johansson, Stig 229, 231, 233
 Jung, Carl Gustav 155n

 Kallen, Jeffrey 294n
 Kellerman, Aharon 294, 294n
 Kendon, Adam 270, 275, 279, 281n
 König, Ekkehard 230
 Kress, Gunther 270

 Labov, William 62, 135, 273
 Ladewig, Silva H. 274
 Lakoff, Robin 141
 La Mésangère, Pierre de 72, 94, 94n, 97n, 104n
 Landry, Rodrigue 289
 Lausberg, Hedda 274
 Lausberg, Heinrich 5
 Ledgeway, Adam 3, 4, 6, 9, 17, 21, 21n, 23n

I dati linguistici. Metodologie e strumenti della ricerca

- Lefebvre, Henri 291n
Lenci, Alessandro 142, 253n, 262n
Leopardi, Giacomo 75
Levi Bianchini, Marco 156, 162
Levshina, Natalia 229
Libri, Guglielmo 95, 95n
Lippi, Lorenzo 73, 80, 95
Loda, Mirella 314n
Lombardi, Girolamo 74
Lombardi Vallauri, Edoardo 3
Lorusso, Paolo 22n, 23
Lotherington, Heather 294
Lubello, Sergio 255n, 256, 257n
Luti, Emilia 96, 96n
Lyons, Kate 293n
- Maccagni, Carlo 175, 175n
Machiavelli, Niccolò 80
Magalotti, Lorenzo 73, 80
Majid, Asifa 232, 234, 244,
Malt, Barbara C. 232
Maly, Ico 293n, 294, 312
Manfredini, Manuela 171n
Manni, Paola 175n
Manzini, Maria Rita 22n, 23, 24
Manzoni, Alessandro 71, 72, 73, 74,
75, 77, 78, 79, 81, 81n, 82, 84, 85, 88,
89, 91, 91n, 92, 92n, 93, 93n, 94n, 95,
95n, 96n, 97n, 98, 100, 100n, 101n,
102, 103, 103n, 104n, 105, 106, 112
Marello, Carla 321n
Marhaba, Sadi 155n
Martinet, André 270
Martino, Paolo 6
Matteucci, Luigi 97, 97n, 105n
Mauri, Caterina 46
Mazzone, Matteo 173n, 174n, 176n
McEnergy Tony 49
McNeill, David 275, 279
- Miani, Stefano 151n
Miglietta, Annarita 171n
Milroy, Lesley 135
Minucci, Paolo 80, 80n, 95
Moneglia, Massimo 46, 195, 198n,
202, 234, 235, 237n, 275, 276, 277
Monosini, Agnolo 103
Montemagni, Simonetta 142
Monti, Vincenzo 74, 75, 76
Morselli, Enrico 160, 163
Moussaid, Yasmina 223
Müller, Cornelia 274, 281
Musatti, Cesare 155n
- Nencioni, Giovanni 75n, 76n, 77, 92n,
174n
Niccolini, Giovanni Battista 95
Nobili, Claudio 257n
- Palma, Stefano 321n, 324, 324n, 325, 326
Papini, Giovanni 158
Patella, Barbara 322n
Patota, Giuseppe 323n
Pauli, Sebastiano 94n
Pederzani, Giuseppe 74n
Pelillo, Giulia 152n
Pirrelli, Vito 142
Pitrè, Giuseppe 23
Pitteri, Francesco 74
Plauto, Tito Maccio 94
Plebani, Tiziana 132
Ponza, Michele 95
Porro, Marzio 176n
Portinaro, Laura 314n
Pozzato, Maria Pia 135
Puttilli, Matteo 314n
Pye, Clifton 244
- Ranchetti, Michele 153

- Rank, Otto 156
 Raso, Tommaso 195, 201n, 254n, 255n, 256, 256n, 275, 277
 Redi, Francesco 91n
 Regis, Riccardo 45n, 59
 Riccio, Anna 323n, 328n
 Rohlfs, Gerhard 3, 4, 10, 21n, 22n, 23, 23n
 Romano, Elisa 175n
- Saba, Umberto 161, 164
 Sabatini, Alma 143
 Sabatini, Francesco 254, 254n
 Sacchetti, Franco 73, 80
 Salviati, Leonardo 80, 100n
 Salvini, Anton Maria 73, 80, 84, 91n, 96, 96n, 100, 100n
 Savoia, Leonardo Maria 3, 17, 22n, 23, 24
 Scollon, Ron 292n
 Scollon, Suzie Wong 292
 Serdonati, Francesco 104n
 Serianni, Luca 142, 173n
 Sessa, Mirella 321n
 Siddharthan, Advaith 264n
 Silverstein, Michael 4
 Sloetjes, Han 274
 Sobrero, Alberto A. 171n
 Soldati, Mario 152
 Sorrento, Luigi 22n
 Soukup, Barbara 293n, 295, 296
 Suzuki, Shingo 3
- Taranto, Francesco 321n
- Telve, Stefano 152n, 171n
 Terenzio Afro, Publio 94, 95n
 Thurlow, Crispin 292n
 Tissi, Silvio 158
 Tommaseo, Niccolò 75, 92n, 93n
 Treccani degli Alfieri, Giovanni 76
 Trillocco, Simona 200
 Trippel, Thorsten 274
 Trubar, Primož 115
- Ungaretti, Giuseppe 158
- Van der Worp, Karin 291
 Van Leeuwen, Theo 270
 Vannetti, Clementino 74
 Varchi, Benedetto 80, 82
 Vellutino, Daniela 253n, 254n
 Vertovec, Steven 290
 Viale, Matteo 255n
 Villani, Giovanni 80
 Villani, Paola 133
 Visconti, Jacqueline 171n
- Wade, Karen 294n
 Wei, Li 293n, 303n
 Weill-Tessier, Pierre 49
 Weiss, Edoardo 152, 153, 155, 156, 157, 158, 160, 160n, 161, 164, 164n, 165
 Wierzbicka, Anna 232, 241
- Zanotti, Paolo 74
 Zatta, Antonio 10, 11
 Zweig, Stefan 161

Studi e ricerche del Dipartimento di Lettere e Filosofia

Antichità e Filologia

1. FRANCESCO CANNIZZARO, *Sulle orme dell'Iliade. Riflessi dell'eroismo omerico nell'epica d'età flavia*, 2023.
2. *Noster delectat error. L'errore tra filologia e letteratura*, a cura di Elisa Migliore, Matilde Oliva, Claudio Vergara, 2024.

Letteratura italiana e Romanistica

1. *L'illustre volgare. Riletture, riscritture e traduzioni dantesche nelle lingue romanze*, a cura di Michela Graziani, Michela Landi e Salomé Vuelta García, 2023.
2. «*La sintassi del mondo*». *La mappa e il testo*, a cura di Laura Bardelli, Elisa Caporiccio, Ugo Conti, Antonio D'Ambrosio, Carlo Facchin, Martina Romanelli, 2023.
3. *La violenza nella letteratura italiana. Forme, linguaggi e rappresentazioni*, a cura di Rebecca Bardi, Camilla Bencini, Chiara Canali, Andrea Carnevali, Alice Petrocchi, Alessandro Privitera, Andrea Talarico, 2023.

Linguistica

1. «*La sua chiarezza séguita l'ardore*». *Studi di linguistica e filologia offerti a Paola Manni*, a cura di Barbara Fanini, 2023.
2. *I dati linguistici. Metodologie e strumenti della ricerca*, a cura di Caterina Cacioli, Serena Carlamaria Crespi, Stefano Miani, Barbara Patella, Ersilia Russo, Carmelina Toscano, 2024.

Finito
di stampare
nel mese di luglio 2024 da Rotomail Italia S.p.A.

Volume stampato con tecnologia print on demand

Come raccogliere, interpretare, trattare e condividere i dati linguistici: queste le principali linee di indagine del volume, che ospita i contributi di dottorandi e studiosi di Linguistica e Umanistica digitale. Dal mutamento alla variazione, dalla fraseologia al lessico specialistico, dalla lessicografia alla semantica, fino all'utilizzo di linguaggi di mark-up e di *corpora*, l'attenzione si sofferma su approcci e metodologie di lavoro differenti, con lo scopo di condividere pratiche di studio e di ricerca – sia in campo teorico sia in campo applicativo – sul dato linguistico, oggetto esplorabile secondo molteplici sfaccettature e sistemi d'indagine.

CATERINA CACIOLI è dottoranda in Umanistica Digitale presso l'Università degli Studi di Firenze. Si occupa di semantica e metodi sperimentali per lo studio del significato in linguistica cognitiva.

SERENA CARLAMARIA CRESPI si è addottorata presso l'Università degli Studi di Firenze. È ingénieur d'étude nell'ambito dell'umanistica digitale e della filologia moderna presso l'Université Grenoble Alpes.

STEFANO MIANI è assegnista di ricerca presso l'Università degli Studi di Firenze. Si occupa di linguaggi e terminologie specialistiche e di lingua e discriminazione.

BARBARA PATELLA è assegnista di ricerca e docente a contratto presso l'Università degli Studi di Firenze. Si occupa di lessicografia informatica e di vocabolari metodici.

ERSILIA RUSSO è assegnista di ricerca presso l'Università degli Studi di Firenze. Si occupa della fraseologia dei *Promessi sposi* e del *Vocabolario del fiorentino contemporaneo*.

CARMELINA TOSCANO si è addottorata presso l'Università degli Studi di Firenze. Si occupa di contatto linguistico tra greco e romanzo e di dialettologia storica.

