

# RiCOGNIZIONI

---

RIVISTA DI LINGUE, LETTERATURE E  
CULTURE MODERNE

6 • 2016 (III)



UNIVERSITÀ  
DI TORINO

DIPARTIMENTO DI  
LINGUE E LETTERATURE STRANIERE E  
CULTURE MODERNE



# RiCOGNIZIONI

---

RIVISTA DI LINGUE, LETTERATURE  
E CULTURE MODERNE

6 • 2016 (2)



UNIVERSITÀ  
DI TORINO

DIPARTIMENTO DI  
LINGUE E LETTERATURE STRANIERE E  
CULTURE MODERNE

**In copertina:** webtreats, *Amazing Tileable Grungy Fractal Tree Textures*, 2009, <https://www.flickr.com/photos/webtreatsetc/4730243915/in/photostream/> [retrieved 29/12/2016].

## COMITATO DI DIREZIONE

### Direttore responsabile

Paolo BERTINETTI (Università di Torino)

### Direttore editoriale

Carla MARELLO (Università di Torino)

## COMITATO DI REDAZIONE

Pierangela ADINOLFI (Università di Torino), Alberto BARACCO (Università di Torino),  
Elisabetta BENIGNI (Università di Torino), María Felisa BERMEJO CALLEJA (Università di Torino),  
Silvano CALVETTO (Università di Torino), Gianluca COCI (Università di Torino),  
Elisa CORINO (Università di Torino), Peggy KATELHOEN (Università di Torino),  
Massimo MAURIZIO (Università di Torino), Patricia KOTTELAT (Università di Torino),  
Enrico LUSSO (Università di Torino), Roberto MERLO (Università di Torino),  
Alessandra MOLINO (Università di Torino), Daniela NELVA (Università di Torino),  
Matteo REI (Università di Torino)

## SEGRETERIA DI REDAZIONE

Alberto BARACCO (Università di Torino), Gaia BERTONERI (Università di Torino),  
Elisa CORINO (Università di Torino), Roberto MERLO (Università di Torino),  
Daniela NELVA (Università di Torino), Barbara PAVETTO (Università di Torino),  
Matteo REI (Università di Torino)

## COMITATO SCIENTIFICO

Ioana BOTH (Universitatea «Babeş-Bolyai», Cluj-Napoca), Suranjan DAS (Università di Calcutta),  
Salvador GUTIÉRREZ ORDÓÑEZ (Universidad de León), Andrea CAROSSO (Università di Torino),  
Emanuele CICCARELLA (Università di Torino),  
Thierry FONTENELLE (Translation Center for the Bodies of the European Union, Luxembourg),  
Natal'ja Ju. GRJAKALOVA («Puškinskij Dom», Accademia delle Scienze di San Pietroburgo),  
Philip HORNE (University College, London), Krystyna JAWORSKA (Università di Torino),  
Ada LONNI (Università di Torino), Maria Grazia MARGARITO (Università di Torino),  
Fernando J.B. MARTINHO (Università di Lisbona), Francine MAZIÈRE (Université Paris 13),  
Riccardo MORELLO (Università di Torino), Francesco PANERO (Università di Torino),  
Virginia PULCINI (Università di Torino), Giovanni RONCO (Università di Torino),  
Michael RUNDELL (Lexicography MasterClass), Elmar SCHAFROTH (Universität Düsseldorf),  
Mikołaj SOKOŁOWSKI (Instytut Badań Literackich Polskiej Akademii Nauk, Warszawa),  
Michelguglielmo TORRI (Università di Torino), Claudia Maria TRESSO (Università di Torino),  
Jorge URRUTIA (Universidad «Carlos III», Madrid), Inuhiko YOMOTA (Kyoto University of Art & Design),  
François ZABBAL (Institut du Monde Arabe, Paris)

## EDITORE

### Dipartimento di Lingue e Letterature Straniere e Culture Moderne

Palazzo delle Facoltà Umanistiche  
Via Sant'Ottavio, 20, Torino  
<http://www.dipartimentolingue.unito.it/>

## CONTATTI

SITO WEB: <http://www.ojs.unito.it/index.php/ricognizioni/index>  
E-MAIL: [ricognizioni.lingue@unito.it](mailto:ricognizioni.lingue@unito.it)  
ISSN: 2420-7969



# SOMMARIO

---

## **CrOCEVIA • Dati empirici e risorse lessicali**

Editors: Elisabetta JEZEK & Isabella CHIARI

- 9 Elisabetta JEZEK, Isabella CHIARI • *Dati empirici e risorse lessicali*
- 15 M. Silvia MICHELI • *Limiti e potenzialità dell'uso di dati empirici in lessicografia: il caso del plurale delle parole composte*
- 35 Luca CIGNETTI, Silvia DEMARTINI • *Dal dato allo strumento. Quando l'errore è un valore: questioni teoriche e pratiche nell'allestimento del LISSICS (Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico)*
- 51 Berta GONZÁLEZ SAAVEDRA, Marco PASSAROTTI • *Verso un lessico di valenza del latino empiricamente motivato*
- 69 Anna FELTRACCO • *T-PAS: costruire una risorsa per l'italiano basata sull'analisi di un corpus*

## **ItINERARI**

- 81 Francesca GALFIONE • *Rifrazioni di un mondo caleidoscopico. Per una lettura dell'opera di V. Majakovskij*
- 93 Pino MENZIO • *Lo specchio del realismo. Conoscenza e affettività del testo letterario*
- 105 Riccardo IMPERIALE, Elmar SCHAFROTH • *Fraseologia italiana per apprendenti L2 e specialisti. La pratica dell'approccio costruzionista*

## **PaSSAGGI**

- 135 Giorgia ESPOSITO • *La traducción del texto poético. Los perros románticos de Roberto Bolaño*

## **SeGNALI**

- 147 Virgilio BERNARDONI • [recensione di] Ilaria BONOMI, Vittorio COLETTI (a cura di), *L'italiano della musica nel mondo*, Firenze, Accademia della Crusca-goWare, 2015-2016

- 
- 153 Akira KAWAMURA, Silvana MOSCA • [report dell'incontro] *La formazione iniziale degli insegnanti*, visita di ricerca comparativa di esperti giapponesi all'Università di Torino, 21-25 novembre 2016
- 157 Daniela MORO • [report del] *XL CONVEGNO INTERNAZIONALE AISTUGIA* (Associazione Italiana STUdi GIApponesi), Torino, 22-24 settembre 2016
- 161 Roberta SALA • [report del] *VI CONGRESSO ITALIANO DI SLAVISTICA*, Torino, 28-30 settembre 2016



# CrOCEVIA

---

DATI EMPIRICI E RISORSE LESSICALI

Editors: Elisabetta JEZEK & Isabella CHIARI



# DATI EMPIRICI E RISORSE LESSICALI

---

Elisabetta JEZEK, Isabella CHIARI<sup>1</sup>

## 1. Corpora e Risorse Lessicali

La disponibilità di dati linguistici in formato digitale è cresciuta in modo esponenziale negli ultimi 20 anni, stimolando lo sviluppo di modelli per la loro annotazione e di tecniche per la loro analisi statistica, al fine di condurre ricerca linguistica quantitativa e qualitativa e potenziare applicazioni computazionali che prevedono fasi di *machine learning* a partire da dati.

Ciò nonostante, la discussione fondamentale riguardo all'utilizzo di dati linguistici nella costruzione di risorse linguistiche, nella pratica lessicografica (tradizionale e computazionale, cfr. Hanks 2013) e in particolare nella elaborazione di teorie del linguaggio resta aperta e controversa (De Marneffe and Potts 2016), così come spesso sono insufficientemente esplicitati i limiti dell'uso di specifici corpora per la ricerca linguistica, in particolare quella lessicale.

Il termine "risorse lessicali" comprende oggi una vastissima gamma di oggetti: tra questi, versioni online di dizionari cartacei (alcuni tra questi *corpus-based* o *corpus-driven*); *dizionari elettronici* direttamente creati per essere distribuiti esclusivamente online; *dizionari collaborativi* creati da utenti ordinari in progetti volontari; *aggregatori di fonti lessicografiche* (come dictionary.com e thefreedictionary.com); corpora annotati (di lingua scritta, parlata, mista, di dominio specifico, multilingui); lessici computazionali monolingui e multilingui, pensati come *database* lessicali o basi di conoscenze finalizzate non tanto alla consultazione da parte di utenti, ma all'uso e integrazione in applicazioni computazionali, anche dati terminologiche. Sono diffuse le iniziative di standardizzazione degli schemi di annotazione e dei metadati (dati di alto livello, categorie generali volte a favorire interoperabilità e riusabilità delle risorse), metodi innovativi per l'acquisizione di dati (*crowdsourcing*, *gamification*), e iniziative di valutazione e validazione di metodi e risorse.

## 2. Il dato linguistico

Il dato ha da sempre costituito un elemento centrale nel disegno di opere lessicografiche e nella ricerca linguistica. Come è noto, sono stati tuttavia di volta in volta privilegiati diversi tipi di dati. Per quanto riguarda la sola pratica lessicografica, si possono individuare ad es. (Chiari 2012: 97): a) l'introspezione: a.1) l'introspezione del lessicografo; a.2) l'introspezione dell'utente ordinario; b) l'uso del dizionario: b.1) l'analisi di questionari sull'uso dei dizionari; b.2) l'analisi del comportamento di utenti in situazioni sperimentali; b.3) l'analisi del comportamento di utenti in situazioni reali; c) i riferimenti: c.1) la descrizione proposta in altre opere lessicografiche (mono o plurilingui); c.2) la descrizione proposta in opere di riferimento come grammatiche, lavori didattici e pubblicazioni scientifiche descrittivi; d) le attestazioni: d.1) l'analisi di esempi d'uso selezionati o casuali tratti da selezioni di testi; d.2) l'analisi di esempi d'uso estratti da corpora di riferimento esistenti o costruiti *ad hoc* per fornire la base empirica dell'opera lessicografica.

---

<sup>1</sup> Il paragrafo 1.1 è stato scritto da E. Jezek, il paragrafo 1.2 è invece elaborato da I. Chiari, mentre il paragrafo 1.3 è stato scritto congiuntamente da I. Chiari ed E. Jezek.

---

Nella lessicografia contemporanea emergono alcuni nodi teorici e applicativi comuni quando si tenta di mettere in comunicazione dati estratti da corpora e *treebanks* e le risorse lessicali. Tra le questioni che si pongono più crucialmente al centro della discussione teorica e applicativa vi sono da una parte le diverse caratteristiche dei corpora disponibili che, per dimensione e composizione qualitativa, non sempre sono adeguati a rispondere ai singoli problemi di ricerca. Fenomeni relativamente rari (come la attestazione dei plurali dei composti italiani) necessitano di corpora molto ampi che tuttavia frequentemente non sono internamente organizzati in modo da permettere una osservazione delle diverse distribuzioni delle forme in competizione in tipologie testuali scritte e parlate diverse. L'uso del web per l'osservazione delle forme ha inoltre numerosi limiti di affidabilità, ripetibilità, verificabilità delle forme in contesto e di estrazione dei dati linguistici.

Il fatto linguistico e il dato empirico utilizzati per la costruzione di risorse lessicali e di *database* lessicografici hanno spesso uno statuto mediato da esigenze applicative e didattiche per le quali ci si muove in una linea sottile tra normatività, prescrittività, modelli su aspetti diversi dell'oggetto rappresentato (e delle sue dimensioni di variazione) e un approccio puramente descrittivo più vicino alla ricerca linguistica. Tale mediazione è spesso dovuta all'essenza applicativa e orientata sull'utente e ai diversi modelli linguistici che emergono nella vita sociale di una comunità linguistica. La costruzione di grandi *database* lessicali multilingui utilizzati in linguistica computazionale ha inoltre comportato la diffusione di risorse di dimensioni enormi ma con una verifica della qualità dei materiali pressoché assente. Si tratta infatti di risorse (esempi come BabelNet, gli allineamenti di Wordnet multilingui, ecc.) in cui l'aggregazione di dati linguistici provenienti da fonti diverse è operata in maniera quasi totalmente automatica e in cui la valutazione della consistenza teorica e linguistica dei dati prodotti è quasi inesistente. I vantaggi relativi dalla disponibilità di risorse di enormi dimensioni non è controbilanciata da una qualità che faccia procedere lo sviluppo della risorsa con gli avanzamenti della teoria e della descrizione linguistica.

Mentre infatti le diverse aree della ricerca linguistica sono state toccate e profondamente influenzate dal dibattito sulla natura del dato linguistico, sul suo posarsi su attestazione ed evidenza osservabile o su competenze e intuizioni del parlante (ideale o reale), lo sviluppo di risorse applicative non sempre ha percorso le stesse tappe.

La grande mole di materiale testuale a disposizione per l'elaborazione di risorse lessicali finisce per richiedere paradossalmente in maniera più evidente il ruolo di filtro, selezione, mediazione e astrazione del linguista. Tale mediazione è peraltro problematica, come emerge dalla complessità e non univocità delle operazioni di annotazione manuale, che richiedono addestramento, esplicitazione delle operazioni richieste e delle scelte – spesso anche delle convenzioni – operate. Diventa dunque cruciale il ruolo dell'annotatore e la valutazione della coerenza delle annotazioni operate da diversi utenti, siano essi utenti comuni della lingua o linguisti. Laddove invece le operazioni di annotazioni siano operate in maniera automatica il problema della coerenza e della valutazione si sposta a livello di verifica del tasso di errore. Il quadro che emerge conferisce centralità a una visione dei fenomeni lessicali e testuali come fenomeni con regolarità che si polarizzano secondo i tipi testuali e che necessitano di modelli flessibili che diano conto delle varietà attestate nello spazio linguistico delle lingue. Emerge dunque la complessità di dare conto di tale variabilità nella descrizione, spesso statica, fornita dalle risorse. La stessa adozione di terminologia come standard e non standard è un segnale di tale problematicità.

I contributi del numero mostrano in modo evidente l'attenzione ai problemi di trattamento (che sono appunto un prodotto dell'inadeguatezza degli strumenti a trattare il materiale autentico nella sua variabilità) e di annotazione, nonché al delicato problema del rendere conto in modo

sintetico e relativamente ‘statico’ di questioni dinamiche e variabili a livello di uso linguistico. Anche a livello di interrogazione le criticità che emergono dal trattamento del dato diventano evidenti in quanto, soprattutto quando si interrogano base dati di grandi dimensioni, l’annotazione con i suoi limiti costituisce l’unico filtro possibile per fruire del dato. E tutto ciò che non è restituito dalle possibilità di annotazione per via di incoerenze e inadeguatezze degli strumenti, risulta del tutto invisibile all’occhio di chi consulta le risorse o le integra in applicazioni più complesse.

A questi problemi si aggiunge inoltre un problema interno e costitutivo della lingua in uso stessa, ossia il veloce mutamento del lessico delle lingue (sia a livello di insieme di unità lessicali, ma soprattutto a livello di sviluppo e riorganizzazione dei sensi di lessemi esistenti) che rende le risorse relativamente obsolete in pochissimi anni, con la necessità di manutenzione di risorse e strumenti continua.

Rimane uno spazio da colmare che è quello determinato dai diversi obiettivi delle risorse che si mettono in relazione quando ad esempio si usano corpora per ‘informare’ risorse lessicali o quando si allineano risorse diverse. Questo spazio è il vero territorio di sfida per la costruzione delle risorse lessicali, poiché richiede una consapevolezza critica dei limiti dei diversi oggetti che si confrontano e del modo in cui intendono rappresentare i dati linguistici, che sola può garantire un esito qualitativamente accettabile e non occasionale nella produzione delle risorse stesse.

In questo quadro la costruzione di risorse lessicali si presenta come uno dei settori che necessitano di una profonda riflessione che metta in comunicazione aree molto diverse della linguistica contemporanea dalla teoria alla descrizione linguistica, dall’annotazione al trattamento, fino agli strumenti di interrogazione. Il rispetto della complessità del dato linguistico infatti richiede uno sforzo di rappresentazione, modellizzazione e descrizione che ancora non ha sempre esiti applicativi adeguati. I contributi del numero cercano di mettere in luce alcuni aspetti critici di questo processo.

### **3. I contributi del numero**

Sono di seguito raccolti quattro contributi che focalizzano l’attenzione sul rapporto tra dati empirici e risorse lessicali e che affrontano e problematizzano dal punto di vista metodologico il valore empirico e i limiti delle diverse fonti che possono essere considerate ‘dato’ in lessicologia, oltre al modo in cui queste fonti contribuiscono a definire la rappresentazione del lessico di una lingua. In che modo il dato linguistico, inteso in senso ampio, contribuisce a dar forma a diverse rappresentazioni e modelli del significato delle parole e delle loro relazioni semantiche e lessicali, della connessione tra dimensione semantica, comportamento sintattico e collocazionale e dimensione pragmatica (Jezek 2006)?

Dei quattro contributi, il primo focalizza l’attenzione sullo studio di un singolo fenomeno linguistico attraverso l’utilizzo di corpora, mentre i restanti tre descrivono nello specifico problematiche legate alla costruzione di lessici utilizzando dati empirici.

Il contributo di Silvia Micheli prende in esame il problema del modo in cui i dizionari rendono conto di fenomeni caratterizzati da ampia mutevolezza e in quale maniera le soluzioni proposte dai dizionari tengano in considerazione dati empirici estratti da corpora. Il problema viene affrontato mediante alcuni *case studies* relativi alla formazione del plurale dei composti italiani. Il saggio pone alcune questioni critiche rispetto all’uso dei corpora per rispondere a specifiche domande di ricerca a scopo lessicografico e prende in esame i limiti e i vantaggi dell’uso dei motori di ricerca sul web come strumento per la estrazione di dati linguistici e inoltre si prendono in esame due corpora che rispondono a domande di ricerca molto diverse: il corpus del Nuovo Vocabolario di Base e ItWac, mentre le risorse lessicografiche messe a confronto sono

---

il Gradit e il Devoto-Oli 2014. Si discute in particolare del modo in cui i diversi corpora (diversi per dimensione e per composizione) siano capaci di rendere conto di fenomeni linguistici rari e di come i dati da essi estratti siano generalizzabili per una presentazione lessicografica.

Il secondo contributo riguarda questioni teoriche e applicative nella costruzione del Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico. In questo saggio Luca Cignetti e Silvia De Martini presentano un corpus di italiano scritto da bambini e ragazzi della Svizzera Italiana in contesto scolastico. Emergono dunque problemi di trattamento di fenomeni non standardizzati tipici della scrittura giovanile e spesso influenzati da forme di comunicazione mediata dalle tecnologie. Il saggio tratta sia di problemi applicativi e di trattamento sia di possibili approcci didattici all'uso dei materiali raccolti.

Gli ultimi due contributi rivolgono l'attenzione a risorse lessicali *corpus-driven* che uniscono scopi di ricerca linguistica con scopi computazionali. Il contributo di Berta González Saavedra e Marco Passarotti "Verso un lessico di valenza del latino empiricamente motivato" presenta un lessico in cui l'attenzione è volta alle strutture sintattiche associate alle parole dotate di proprietà argomentali. Dopo una rapida introduzione ai lessici di valenza e alle modalità di costruzione (distinte in *intuition-based* e *corpus-driven*) gli autori presentano Latin Vallex, un lessico di valenza per il latino realizzato in stretta connessione con l'annotazione semantico-pragmatica di due *treebank* latine comprensive di testi di epoche e generi diversi, descrivendone la struttura delle entrate lessicali. Il contributo si sofferma in particolare sulle modalità di codifica di tre costruzioni sintattiche: le proposizioni passive, le proposizioni infinitive, l'ablativo assoluto. Il contributo si conclude con una illustrazione delle modalità di interrogazione che mettono ancora una volta in luce lo stretto legame tra il lessico valenziale e la banca dati da cui esso è estratto: il modo cioè in cui una *frame entry* è connessa alle sue occorrenze testuali. Il contributo si conclude con una interessante considerazione relativa ai limiti dell'indagine *corpus-driven*.

Il contributo di Anna Feltracco "T-PAS: costruire una risorsa per l'italiano basata sull'analisi di un corpus" riporta la costruzione di una raccolta di strutture predicato-argomenti con informazione semantica sul tipo semantico delle posizioni argomentali, associata a un corpus di occorrenze e a un repertorio di tipi semantici. La risorsa T-PAS rappresenta un interessante modello complementare ai lessici di valenza sopra descritti, in cui l'attenzione è sulla semantica delle costruzioni anziché sulla sintassi. Nel contributo, l'autrice descrive la procedura di acquisizione delle T-PAS e lo stato dell'arte della risorsa alla quale ha contribuito per la parte lessicografica. Oltre alle specifiche dell'estrazione, che mettono in luce la metodologia utilizzata nella generalizzazione delle strutture a partire dalle occorrenze del corpus, sono riportati dati e caratteristiche salienti della risorsa, e i risultati dei primi esperimenti in cui la risorsa è stata utilizzata. Questi includono un esercizio di disambiguazione del significato del verbo in contesto. Sono infine riportati dati relativi alla valutazione di una porzione della risorsa e le problematiche legate alla annotazione delle relazioni tra T-PAS. Il contributo si chiude con osservazioni relative alle possibilità di espansione semi-automatica della risorsa.

Si ringraziano i membri del Comitato Scientifico del laboratorio "Dati Empirici e Risorse Lessicali" organizzato in occasione del XLIX Congresso internazionale di Studi della SLI (Università di Malta, 24-26 settembre 2015), dove i contributi sono stati presentati: Silvia Bernardini, Marco Biffi, Federica Casadei, Ulrich Heid, Alessandro Lenci, Francesco Urzi. Si ringrazia inoltre Elisa Corino per la cura redazionale e Carla Marengo per la proposta di accogliere i contributi nella rivista *RiCognizioni*.

**BIBLIOGRAFIA**

- De Marneffe, M-C., Potts, C. (2016), *Developing linguistic theories using annotated corpora*, in I. Nancy, J. Pustejovsky (eds.), *The Handbook of Linguistic Annotation*, Berlin, Springer, in corso di stampa.
- Chiari, I. (2012), *Il dato empirico in lessicografia: dizionari tradizionali e collaborativi a confronto*, in Bollettino di Italianistica. Per Tullio De Mauro, II: 94-125.
- Hanks, P. (2013), *Lexical Analysis: Norms and Exploitations*, Cambridge - MA, The MIT Press.
- Jezek, E. (2006), *Argument Structure, Verb Patterns and Dictionaries*, in: C. Marelllo, E. Corino, C. Onesti (a c. di), *Euralex XII Proceedings*, Torino, Edizioni dell'Orso: 1169-1180.

**ELISABETTA JEZEK** • E. Jezek is Associate Professor in Linguistics at Università di Pavia where she has taught syntax and semantics and applied linguistics since 2001. Her research interests include lexical semantics, verb classification, theory of argument structure, event structure in syntax and semantics, lexicon/ontology interplay, word class systems, and computational lexicography. She has edited a number of major works in lexicography and published contributions focusing on the interplay between corpus analysis, research methodology, and linguistic theory.

**E-MAIL** • jezek@unipv.it

**ISABELLA CHIARI** • is Researcher (Assistant professor) in "La Sapienza" University of Rome, where she is holding courses in General linguistics and Computational linguistics for Bachelor, Master and PhD degrees since 2000. She is also Associate at the Institute of Cognitive Sciences and Technologies (ISTC-CNR) and President of Amal for Education, a non profit organization focused on education for refugees. She wrote several essays and articles, directed and participated in numerous research projects. Her research interest focus on statistical and corpus linguistics, Italian lexicography, methodology of linguistic research.

**E-MAIL** • isabella.chiari@uniroma1.it





# LIMITI E POTENZIALITÀ DELL'USO DI DATI EMPIRICI IN LESSICOGRAFIA

Il caso del plurale delle parole composte

---

*M. Silvia* MICHELI

**ABSTRACT** • The goal of this paper is twofold: on the one hand, it deals with the main methodological problems related to the study of Italian compound words using linguistic corpora; on the other hand, it aims at evaluating to what extent data extracted from two corpora of contemporary Italian, which are very different in size and content, can be used to improve Italian dictionaries in regards to the inflection of Italian compounds, about which speakers often have linguistic doubts.

**KEYWORDS** • Compounds, Corpus Linguistics, Inflection, Lexicography

## 1. Introduzione

Il contributo che il dato empirico può offrire a una risorsa lessicografica riguarda tipicamente la costruzione del lemmario, la composizione delle glosse e la loro articolazione in accezioni: su questi elementi si è concentrata la riflessione sull'uso di dati estratti da corpora in lessicografia. Poca attenzione è stata finora dedicata al contributo che i dati posso fornire per migliorare le informazioni di natura morfologica che accompagnano i lemmi di un dizionario, su cui non di rado si concentrano i dubbi dei parlanti, rispetto ai quali l'intuizione non è efficace: come osservato da Iacobini, Giuliani (2001: 332), infatti, in morfologia il giudizio del parlante non è dirimente nello stabilire se una parola è possibile o attestata. In particolare, la formazione del plurale di determinate parole – che per diverse ragioni possono costituire delle eccezioni alle tendenze generali<sup>1</sup> – rappresenta un motivo molto frequente per cui un parlante interroga un dizionario. Un interessante, quanto complesso, caso di questo tipo è costituito dalle parole composte, una categoria del lessico italiano che negli ultimi decenni è stata oggetto di numerosi contributi teorici<sup>2</sup>, orientati prevalentemente alla definizione della categoria e alla classificazione dei suoi membri. Nella riflessione sulla natura dei composti, l'osservazione della flessione

---

<sup>1</sup> Si pensi ad esempio al plurale delle parole in *-cia* e *-gia* o, come si vedrà in questo contributo, a quello delle parole composte.

<sup>2</sup> Rendere conto della vastissima bibliografia sulle parole composte in italiano non rientra negli obiettivi di questo contributo, che si concentra piuttosto sulle metodologie con cui questo tipo di parole può essere studiato; oltre ai lavori su specifiche tipologie di composto a cui si farà riferimento nel corso dell'articolo, si rimanda a Masini, Scalise (2012) e Iacobini, Thornton (2016: 209-210) per un quadro generale dello stato dell'arte sulla composizione italiana.

costituisce un aspetto particolarmente rilevante, spesso sottovalutato, perché può aiutare a comprendere come i parlanti concepiscono questo tipo di parole<sup>3</sup> – come entità lessicalizzate e immagazzinate nel lessico mentale o come entità trasparenti e scomponibili nelle proprie parti – e quale grado di analizzabilità conservano nella coscienza linguistica.

Requisito fondamentale per l'analisi di questo fenomeno è l'osservazione di dati empirici, di natura sia qualitativa sia quantitativa, la cui raccolta pone tuttavia numerosi problemi metodologici legati a specifiche caratteristiche formali e semantiche proprie delle parole composte. È su questo aspetto che il presente contributo intende soffermarsi, con l'obiettivo di valutare in che modo, ed entro quali limiti, i risultati ottenuti da un'indagine empirica possono essere considerati realmente *dati*, su cui basarsi per integrare o mettere in discussione quanto riportato dalle risorse lessicografiche: allo stato attuale, infatti, i dizionari non sempre si dimostrano strumenti efficaci per i lettori, in alcuni casi perché riportano informazioni discordanti, in altri perché registrano più forme plurali, senza indicare quale sia effettivamente la variante più frequente nell'uso.

Attraverso l'analisi di alcuni *case studies*, si intende quindi proporre una riflessione sul rapporto tra dati empirici e risorse lessicografiche<sup>4</sup>; in particolare, dopo aver discusso quali risorse permettono, o non permettono, di raccogliere dati affidabili e rappresentativi, si metterà a confronto quanto riportato da due dizionari – il Devoto Oli e il GradiT – rispetto alla formazione del plurale di alcune tipologie di composto, e si analizzerà in che misura le informazioni ricavate da due corpora di italiano contemporaneo – itWaC e il corpus del Nuovo Vocabolario di Base – possono integrarlo o metterlo in discussione.

## 2. Questioni preliminari

Un primo elemento che caratterizza le parole composte e ne rende più complesso il loro studio attraverso i corpora riguarda la bassa frequenza con cui queste ricorrono nell'uso reale dei parlanti: esse rappresentano un insieme di entità lessicali mediamente rare per ragioni di tipo semantico e pragmatico, in quanto veicolano un insieme di significati a cui i parlanti ricorrono solo in determinati contesti o domini testuali. Considerando ad esempio i composti Verbo+Nome – tipologia particolarmente produttiva in italiano – si osserva che essi vengono generalmente utilizzati in funzione agentiva o strumentale: spesso individuano strumenti o figure professionali difficili da incontrare al di fuori di particolari ambiti specialistici o contesti (si pensi ad esempio a *provavalvole* o *narrastorie*). Questo vale in massima parte anche per le altre tipologie di composto: le forme del tipo Aggettivo+Aggettivo (Grossmann, Rainer 2009; D'Achille, Grossmann 2009), ad esempio, si formano quasi esclusivamente per indicare i giocatori di una squadra di calcio (ad es. *bianconero*), l'appartenenza a uno schieramento politico (ad es. *nazional socialista*) o la provenienza geografica (ad es. *serbocroato*). La bassa frequenza con cui le parole composte compaiono nei testi è quindi dovuta in primo luogo a ragioni pragmatiche: nella maggior parte dei casi, esse sono create dai parlanti per nominare concetti o oggetti che non

<sup>3</sup> Su come i composti vengono processati dai parlanti sono disponibili numerosi studi di psicolinguistica: per un quadro generale rimando al volume curato da Libber, Jarema (2007).

<sup>4</sup> I casi particolari che si discuteranno nei paragrafi successivi non sono ovviamente da considerarsi rappresentativi e/o esplicativi in modo esaustivo del fenomeno, quanto piuttosto funzionali a mettere in luce i problemi metodologici legati allo studio dei composti su basi empiriche: l'obiettivo centrale del contributo non è quindi descrivere la formazione del plurale nelle parole composte in italiano, ma discuterne le possibili metodologie di indagine ed evidenziare limiti e potenzialità del contributo che due corpora di italiano contemporaneo possono fornire ai dizionari.

si trovano nell'orizzonte quotidiano degli individui. A tale questione se ne aggiunge un'altra legata più in particolare allo studio del plurale di un sostantivo, non necessariamente composto: nell'uso reale dei parlanti la distribuzione tra singolare e plurale è determinata da fattori difficilmente prevedibili perché strettamente legati al contesto in cui viene prodotto l'enunciato, e molto spesso è sbilanciata a favore del singolare. Questo rende ancora più arduo riuscire a estrarre dai corpora dati quantitativamente significativi relativi alle forme flesse.

### **3. I dati empirici: fonti, problemi, metodi**

La difficoltà con cui i composti vengono intercettati dai corpora potrebbe indurre a ritenere efficace l'utilizzo del web come corpus da cui estrarre dati empirici, in virtù delle enormi dimensioni e del contenuto eterogeneo. L'uso dei motori di ricerca, come Google o Yahoo, per interrogare il web nell'ambito di analisi linguistiche è stato – e, anche se in misura minore, è ancora – un tema a lungo dibattuto in letteratura (cfr. Kilgarriff, Grafenstette 2003; Crystal 2006; Lüdeling *et al.* 2007; Kilgarriff 2007): la maggior parte degli studiosi intervenuti nel dibattito si è detta contraria all'utilizzo di Google in linguistica, evidenziandone i numerosi limiti e mettendone in discussione la scientificità<sup>5</sup>. Le argomentazioni con cui i detrattori dell'uso di Google hanno sostenuto la loro posizione discendono tutte da una caratteristica intrinseca e ineliminabile del web: il suo contenuto cambia continuamente ed è perciò di fatto inconoscibile per l'utente; ne consegue che le interrogazioni nei motori di ricerca non sono replicabili e le informazioni che se ne possono estrarre non sono confrontabili e quindi, semplicemente, non possono essere definiti *dati*<sup>6</sup>, sulla base dei quali elaborare/verificare teorie linguistiche o migliorare il contenuto di grammatiche e dizionari. Le grandi potenzialità della rete possono piuttosto essere sfruttate per creare corpora di grandi dimensioni, come testimonia il progetto WaCky, nell'ambito del quale è nato il webcorpus di italiano contemporaneo itWaC. Nonostante i numerosi problemi metodologici, l'uso di Google in linguistica gode comunque di una certa vitalità: nell'ambito degli studi di morfologia, la sua validità come strumento di ricerca è stata sostenuta da Hathout *et al.* (2008) e Montermini (2015)<sup>7</sup>. In particolare, nei due contributi gli autori sostengono che, con le dovute cautele, il web possa essere utilizzato proficuamente per estrarre dati quantitativi<sup>8</sup> al fine di valutare la produttività di meccanismi di formazione delle parole o la portata di fenomeni morfologici anche rari e quindi difficilmente osservabili attraverso i corpora tradizionali. Costituirebbe inoltre uno degli aspetti più interessanti del web la possibilità di analizzare la creatività lessicale in contesti informali e spontanei quali forum, blog, etc. (ivi: 72): tuttavia, questo non sembra un valido argomento per sostenerne la maggiore efficacia rispetto ai corpora tradizionali, dal momento che esistono corpora costituiti anche da pagine di forum o blog, come il già citato itWaC, o pensati appositamente per lo studio della Comunicazione Mediata dal Computer (CMC), come uno dei sottocorpora che costituiscono il Nuovo Vocabolario di Base<sup>9</sup>, o Web2Corpus\_IT (Beißwenger *et al.* 2016; Chiari 2016).

<sup>5</sup> Per una più dettagliata disamina dei limiti dei motori di ricerca si rimanda a Kilgarriff (2007).

<sup>6</sup> Per una delimitazione del concetto di dato linguistico si vedano Iannàccaro (2000) e Lehmann (2004).

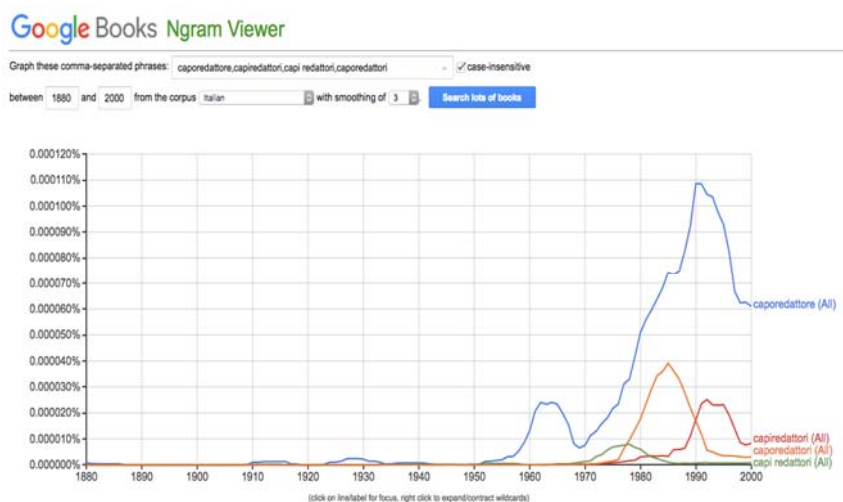
<sup>7</sup> Nei due contributi l'uso del web come fonte di dati è applicato allo studio di alcuni suffissi del francese e dell'italiano.

<sup>8</sup> Dati di cui, pur riconoscendone la diversità rispetto a quelli estratti dai corpora tradizionali, si sostiene la affidabilità nell'ambito di un'indagine scientifica: «In no case should this 'new' kind of data be taken as a weakness for the theories deduced from it» (Hathout *et al.* 2008: 82).

<sup>9</sup> Per la descrizione del corpus si veda il paragrafo 5.

Si fondano su informazioni ricavate da Google le considerazioni di Montermini (2008) sulla formazione del plurale dei composti Nome+Nome e Verbo+Nome: l'analisi si basa su un campione di composti di dimensioni molto ristrette del quale viene osservata la flessione interrogando il motore di ricerca, facendo precedere ciascuna forma dal proprio articolo determinativo (quindi *i/gli* per i composti maschili, *le* per i femminili)<sup>10</sup>, al fine di osservare quali siano i pattern più diffusi e se fattori come il genere e la vocale finale dei costituenti abbiano un ruolo nella formazione del plurale. Oltre alla ristrettezza del campione esaminato, un limite fortissimo al lavoro nel suo insieme è costituito da due aspetti metodologici: *in primis*, va considerato che antepoendo l'articolo determinativo flessso a tali forme di fatto vengono esclusi dai risultati i casi in cui i composti sono preceduti da un articolo indeterminativo o da nessun articolo; inoltre, i valori numerici forniti (ivi: 174) si riferiscono al numero delle pagine web indicizzate da Google nel momento in cui è stata effettuata la ricerca<sup>11</sup>, non al numero delle occorrenze.

I vantaggi offerti dalla rete – dimensioni, eterogeneità di contenuto, facilità di interrogazione – si rivelano quindi solo apparenti se commisurati ai numerosi problemi che conseguono dal suo utilizzo nell'ambito di indagini linguistiche di natura empirica, per le quali, è bene ribadirlo, gli strumenti scientificamente più validi e affidabili sono i corpora. Come osservato da Kilgarriff (2007), il contributo che il web può dare agli studi linguistici è quindi limitato alla costruzione di corpora costituiti da testi provenienti dalla rete, come quelli del progetto WaCky (Baroni, Kilgarriff 2006), o il corpus di Google Ngram Viewer (Michael *et al.* 2012)<sup>12</sup>. In particolare, nell'ambito di uno studio sulla formazione del plurale delle parole composte, tale risorsa può essere utilizzata per verificare, in diacronia, la presenza di forme plurali concorrenti nell'uso dei parlanti, in modo da integrare i dati estratti dai corpora, che fotografano il fenomeno solo sul



<sup>10</sup> Montermini (2008:166; nota 8).

<sup>11</sup> Questo ovviamente implica che le ricerche non sono né confrontabili né replicabili.

<sup>12</sup> Pur nei suoi limiti, Google Ngram Viewer (d'ora in poi GNV) costituisce uno strumento utilizzabile nell'ambito di un'indagine linguistica perché permette di attingere a un vastissimo corpus annotato e chiuso, costituito da una parte consistente dei testi digitalizzati da Google Libri, e da cui possono essere estratti dati replicabili e confrontabili. È importante sottolineare che il contenuto di GNV non coincide precisamente con quello di Google Libri (sui cui limiti si veda Gomez Gane 2008): mentre infatti il secondo viene continuamente ampliato tramite l'aggiunta di nuovi testi digitalizzati, il primo è stato finora aggiornato soltanto una volta, nel 2012, e ha quindi contenuto costante.

piano sincronico. In Figura 1 si può osservare, a titolo di esemplificazione, il grafico relativo al composto *caporedattore*.

Figura 1. Attestazioni delle forme del lemma *caporedattore* nel corpus di Google Ngram Viewer

Diversamente da quanto riportato dal Gradit e dal Devoto Oli, il composto risulta attestato prima del 1962; in particolare, la più antica occorrenza di *caporedattore* risalirebbe al 1883, anno in cui la forma è attestata nel seguente brano tratto dalla rivista *La civiltà cattolica*:

E tanto più realmente invalida, quanto più apparentemente validissima agli occhi degli ignari, è in primo luogo quella difesa che pel primo scoperse L. Wogue gran rabbino di Parigi e Caporedattore dell'*Univers israelite* [...]<sup>13</sup>.

Un interessante contributo che GNV può dare alla lessicografia riguarda infatti la possibilità di retrodatare la prima attestazione di una parola con una certa facilità, rendendo accessibile un vastissimo repertorio di testi digitalizzati in pochi secondi. Rispetto alla formazione del plurale, il grafico permette di attestare la presenza di due forme plurali in concorrenza (*caporedattori* e *capiredattori*), che, con alterne fortune, hanno convissuto nel corso della seconda metà del Novecento<sup>14</sup>.

Nonostante contribuiscano a definire un quadro più chiaro del fenomeno, le informazioni ricavate da GNV possono soltanto integrare, ma non sostituire, il contributo di natura quantitativa e qualitativa dei corpora tradizionali, a cui è necessario rivolgersi per integrare o rivedere il contenuto di risorse lessicografiche o di grammatiche.

Nella scelta del corpus più adatto da cui attingere i dati, due aspetti sembrano particolarmente importanti da valutare: le dimensioni e il contenuto. Quanto debba essere grande un corpus per poter essere considerato rappresentativo e fonte di dati attendibili costituisce una questione tutt'altro che risolta nella linguistica dei corpora. Marc Brysbaert e Boris New (2009) hanno osservato come le dimensioni ottimali di un corpus dipendano strettamente dalla frequenza delle parole oggetto dell'indagine: più esse sono rare, maggiori devono essere le dimensioni del corpus<sup>15</sup>. D'altra parte, gli stessi studiosi sottolineano che l'utilizzo di risorse troppo estese, il cui contenuto potrebbe non essere rappresentativo della lingua o di un suo particolare dominio, è indubbiamente rischioso e non sembra garantire risultati migliori<sup>16</sup>.

Nello studio dei composti non va inoltre sottovalutata la possibilità di analizzare tutti i contesti d'uso attraverso le concordanze<sup>17</sup>, la cui osservazione è imprescindibile per poter andare oltre il dato quantitativo e valutare il comportamento delle forme nell'uso dei parlanti. Nello studio del plurale delle parole composte questo elemento è particolarmente importante anche

<sup>13</sup> La rivista è stata fondata da un gruppo di gesuiti a Napoli nel 1850 ed è ancora attiva. La citazione è tratta dalla sezione "Cronaca contemporanea" (*La civiltà cattolica*, 1883, vol. 12, p. 606).

<sup>14</sup> Come si vedrà nel paragrafo 7.1, tale dato sarà confermato da due corpora di italiano contemporaneo.

<sup>15</sup> Per i dati su cui si basano le riflessioni dei due studiosi rimando al loro contributo: Brysbaert, New (2009: 977-90).

<sup>16</sup> In particolare, gli autori del contributo ritengono che «For most practical purposes, a corpus of 16-30 million words suffices for reliable word frequency norms. In particular, there is no evidence that a corpus of 3 billion words is much better than a corpus of 30 million words»; d'altra parte, un corpus con meno di 16 milioni di occorrenze non fornisce dati significativi per parole che presentano una frequenza inferiore a 100 occorrenze. Occorre sottolineare che quanto sostenuto dai due studiosi si riferisce in primis all'inglese (Ivi: 980).

<sup>17</sup> In quest'ottica, un corpus come il CORIS (Favretti *et al.* 2002), che per ragioni legate ai diritti d'autore permette di visualizzare solo una parte delle concordanze, non costituisce una risorsa efficace nello studio della flessione dei composti.

perché permette di individuare quando un composto viene usato come invariabile, caso frequente nelle forme Verbo+Nome.

#### 4. Il campione esaminato: consistenza e metodologia di raccolta

I dati relativi ai *case studies* che si analizzeranno sono stati raccolti nell'ambito di un'indagine volta a descrivere la formazione del plurale di più tipologie di composti (Micheli 2016); di seguito si riporta una sintetica descrizione della metodologia adottata per creare il campione di parole composte di cui si è analizzata la flessione.

In una prima fase sono stati estratti dal Devoto Oli 2014 tutti i lemmi classificati come 'composti' e appartenenti alle tipologie Nome+Nome, Aggettivo+Nome, Nome+Aggettivo, Aggettivo+Aggettivo, Verbo+Nome.

| Composti selezionati dal Devoto Oli 2014 |                |                |                     |            |        |
|--|----------------|----------------|---------------------|------------|--------|
| Nome+Nome                                | Aggettivo+Nome | Nome+Aggettivo | Aggettivo+Aggettivo | Verbo+Nome | Totale |
| 516                                      | 177            | 105            | 55                  | 997        | 1860   |

Tabella 1. Lemmi selezionati dal Devoto Oli 2014: distribuzione quantitativa rispetto alla tipologia di composto

Com'è noto, per il plurale delle prime quattro tipologie l'italiano ammette tre tipi di flessione: una interna, in cui solo il primo costituente viene flesso; una esterna, in cui la marca di plurale è posta sul margine destro del lessema; una doppia, in cui entrambi i costituenti vengono flessi. Per i composti Verbo+Nome, in cui il primo costituente rimane sempre invariabile, si può invece distinguere tra composti che rimangono invariabili e composti che presentano due forme distinte, una per il singolare e una per il plurale.

Nella seconda fase di raccolta sono state raccolte le forme plurali; al fine di intercettare nei corpora anche le forme non previste dalle grammatiche o dai dizionari<sup>18</sup>, si è scelto di includere nel campione tutti i possibili tipi di plurale (con flessione interna, esterna e doppia). Ogni composto è stato quindi flesso al plurale in tre modi differenti: per *capostazione*, ad esempio, sono state create le forme *capistazione*, *capostazioni*, *capistazioni*. Nel caso di composti del tipo Verbo+Nome è stato flesso solo il secondo costituente, essendo il primo un elemento verbale invariabile. Il vantaggio di lavorare con un formario esploso è che permette di intercettare tutti i possibili tipi di plurale formulabili sfruttando le possibilità morfologiche dell'italiano, a

<sup>18</sup> Le indicazioni contenute nei dizionari e nelle grammatiche, insieme alla tradizione scolastica, costituiscono per Serianni le «classiche fonti della norma linguistica» (Serianni 2014, : 239). Il concetto di norma è stato, ed è tuttora, ampiamente discusso in letteratura da vari studiosi: in particolare, due posizioni diverse sono state assunte da Serianni (2004; 2014) e Sgroi (2010; 2016). Legato al concetto di norma è quello di errore, a cui è dedicato un volume recentemente curato da Grandi (2015), in cui sono raccolti numerosi contributi che, da punti di vista differenti, definiscono e interpretano il rapporto tra grammatica ed errore, regole ed eccezioni, nelle lingue naturali. Nel caso del plurale dei composti, come messo in luce in Micheli (2016: 229-33) e come si vedrà nei paragrafi successivi, le grammatiche e i dizionari non sono concordi nel definire una norma chiara rispetto alla formazione del plurale di queste parole; in assenza di una norma, parlare di 'errori' in riferimento agli usi incerti dei parlanti non è del tutto appropriato, essendo l'errore un concetto che si definisce sempre in relazione a una norma.

prescindere dalla loro attestazione nei dizionari o nelle grammatiche, ed evitare i numerosi errori che il POS tag compie con questo tipo di forme<sup>19</sup>.

Un ulteriore elemento di cui si è tenuto conto è stato la forma grafica con cui le forme compaiono nei testi: soprattutto nel caso delle neoformazioni sono infatti molto frequenti nell'uso oscillazioni grafiche tra la forma unverbata, quella con il trattino e quella in cui i due costituenti sono separati dallo spazio. Al fine di intercettare tutti i composti, a prescindere dalla grafia presentata, ciascuna forma del campione, singolare e plurale, è stata trascritta nelle tre varianti: il lemma “capostazione”, ad esempio, compare nel campione nelle forme *capostazione*, *capostazione*, *capo stazione*, *capostazioni*, *capo stazioni*, *capo-stazioni*, *capistazioni*, *capi stazioni*, *capi-stazioni*, *capistazione*, *capi stazione*, *capi-stazione*.

Il campione raccolto attraverso le due fasi risulta costituito da 27.609 forme, di cui nella Tabella 2 si riporta la distribuzione quantitativa rispetto alla tipologia di composto.

| Tipologia di composto | AA  | AN    | NA    | NN     | VN    | Totale |
|-----------------------|-----|-------|-------|--------|-------|--------|
| Forme dei lemmi       | 828 | 2.411 | 4.550 | 12.886 | 6.934 | 27.609 |

Tabella 2. Forme dei lemmi raccolte per l'indagine: distribuzione quantitativa rispetto alla tipologia di composto

### **5. Le risorse lessicografiche: il Gradit e il Devoto Oli 2014**

Come risorse lessicografiche di riferimento si è scelto di adottare due dizionari molto diversi per architettura e postulati teorici – il Grande Dizionario Italiano dell'Uso<sup>20</sup> e il Devoto Oli 2014<sup>21</sup> – in modo da osservare il trattamento riservato a tale fenomeno tanto da parte di un dizionario che dichiaratamente intende porsi il più vicino possibile all'uso reale dei parlanti quanto da parte di uno di impostazione più tradizionale. I due dizionari strutturano le informazioni riguardanti il plurale dei lemmi in modo diverso: la versione elettronica del Gradit fornisce sempre la forma plurale di ciascun lemma, sia essa regolare o atipica, indicando in nota eventuali forme concorrenti o meno comuni; il DO riporta il plurale soltanto nei casi in cui si abbia una flessione interna o doppia, mentre laddove non si forniscono indicazioni si intende implicitamente che il plurale presenti una regolare flessione esterna, come se si trattasse di un lessema semplice.

### **6. Il Corpus del Nuovo Vocabolario di Base e itWaC: limiti e potenzialità**

Nella scelta della fonte di dati empirici per la presente indagine sono stati presi in considerazione tre fattori: le dimensioni, la possibilità di consultare le concordanze, il contenuto. Tra i corpora di riferimento attualmente disponibili per l'italiano, la risorsa che è parsa più vicina

<sup>19</sup> I motivi per cui si è scelto di lavorare con un formario esplosivo sono essenzialmente due: 1) nei corpora annotati attualmente disponibili per lo studio dell'italiano non è previsto un tag specifico per le parole composte che permetta di estrarle automaticamente; 2) le frequenti oscillazioni riguardanti la grafia con cui i composti ricorrono nei testi determina numerosi errori nel POS tag, limitandone l'efficacia nell'estrazione di dati relativi a questo tipo di parole.

<sup>20</sup> In particolare, del Grande Dizionario Italiano dell'Uso si è consultata la versione elettronica, del Devoto Oli quella cartacea.

<sup>21</sup> D'ora in poi Gradit e DO.

a soddisfare tali requisiti è il Corpus del Nuovo Vocabolario di Base, curato da Isabella Chiari e Tullio De Mauro (cfr. Chiari, De Mauro 2012); tale corpus, costituito da 18 milioni di occorrenze, permette di consultare le concordanze ed è bilanciato in sei sottocorpora a seconda della tipologia testuale (Stampa, Saggistica, Spettacolo, Comunicazione Mediata dal Computer, Letteratura, Parlato). Un limite di questa risorsa è legato alle sue dimensioni: essa ha infatti una estensione commisurata all'obiettivo per cui è stata creata, ossia lo studio del vocabolario di base, ma insufficiente per lo studio di entità lessicali più rare come le parole composte. L'utilizzo di questo solo corpus limiterebbe quindi l'estrazione di dati quantitativi in grado di descrivere il fenomeno: per questo motivo si è scelto di affiancargli un altro corpus di dimensioni maggiori come secondo riferimento. In virtù della sua estensione si è scelto di utilizzare itWaC (cfr. Baroni *et al.* 2009): il corpus è costituito da circa due miliardi di occorrenze e rappresenta lo strumento più esteso per lo studio dell'italiano contemporaneo; le significative dimensioni ne fanno quindi una risorsa in grado di intercettare anche entità lessicali a bassa frequenza. L'utilizzo di itWaC nell'ambito di indagini morfologiche non è raro<sup>22</sup>, ma presenta almeno due limiti che è necessario considerare: l'assenza di un bilanciamento per generi testuali e l'assenza di una documentazione sul contenuto. Entrambi gli aspetti dipendono dalla procedura di acquisizione automatica dei testi, che di fatto rende opaco il contenuto effettivo del corpus e non permette di strutturarlo secondo criteri di bilanciamento.

|               | Corpus del NVDB  | itWaC                    |
|---------------|--|--------------------------|
| Estensione    | 18 milioni di occorrenze   | 2 miliardi di occorrenze |
| Tipologia     | Corpus di riferimento  | Web corpus               |
| Bilanciamento | Bilanciato secondo sei tipologie testuali (Letteratura, Stampa, Saggistica, Comunicazione Mediata dal Computer, Spettacolo, Parlato) | Nessun bilanciamento     |

Tabella 3. Caratteristiche strutturali dei due corpora

Si tratta di strumenti molto diversi tra loro quanto a struttura, contenuto e finalità, che possono quindi fornire informazioni diverse nel corso dell'indagine: il corpus per il NVdB permette di ottenere i dati sulla frequenza delle forme attestate nei vari sottocorpora e di osservare le concordanze, fondamentali per chiarire il comportamento delle forme, soprattutto di quelle invariabili; itWaC è in grado – per la sua estensione – di fornire dati più consistenti sulla frequenza delle forme che andranno confrontati con quelli del corpus del NVdB. Non va tuttavia sottovalutato il fatto che la decisione di utilizzare due corpora molto diversi pone ovviamente il problema di dover interpretare dati solo in parte confrontabili, perché estratti da risorse diverse. Nei paragrafi successivi si osserverà in che misura i due corpora possono fornire dati significativi sulla flessione di alcune particolari tipologie di parole composte.

<sup>22</sup> Si veda ad esempio Baroni *et al.* 2007.



## 7. Risultati dell'indagine: un quadro generale

Un primo dato che va osservato per valutare le potenzialità e i limiti dei due corpora riguarda il numero dei lemmi del campione che vi risultano attestati (Tabella 4).

|  | Corpus del NVdB | itWaC     |
|--|-----------------|-----------|
| Lemmi del campione                         | 1886            |           |
| Lemmi attestati nel corpus                 | 789             | 1468      |
| Forme dei lemmi attestate nel corpus       | 1232            | 2869      |
| Occorrenze delle forme dei lemmi attestate | 19.066          | 1.648.885 |

Tabella 4. Composti attestati nei due corpora

Come prevedibile date le dimensioni significative del corpus, in itWaC è attestata una parte considerevole del campione di forme (circa il 70%); diversamente, nel corpus del NVdB ne è attestato circa il 40%. Tali dati offrono tuttavia soltanto una panoramica generale rispetto alle potenzialità dei due corpora: per valutare entro quali limiti essi possono fornire dei dati significativi sulla flessione delle parole composte occorre infatti osservare quanti e quali lemmi sono attestati con un numero di occorrenze sufficiente per poter comprendere e descrivere fondatamente il comportamento delle forme. La frequenza dei lemmi costituisce infatti un aspetto fondamentale per valutare l'attendibilità del contributo di un corpus; nondimeno, osservare quali sono i composti che ricorrono più frequentemente nei due corpora permette di ricavare informazioni anche sul loro contenuto.

| Corpus del Nuovo Vocabolario di Base |              |           | ITWAC |                |           |
|--------------------------------------|--------------|-----------|-------|----------------|-----------|
| Rango                                | Lemma        | Frequenza | Rango | Lemma          | Frequenza |
| 1                                    | Mezzogiorno  | 408       | 1     | Centrosinistra | 81.381    |
| 2                                    | Mezzanotte   | 388       | 2     | Piattaforma    | 54.421    |
| 3                                    | Capolavoro   | 385       | 3     | Mezzogiorno    | 52.566    |
| 4                                    | Pomodoro     | 356       | 4     | Salvaguardia   | 51.348    |
| 5                                    | Portafoglio  | 352       | 5     | Capolavoro     | 43.279    |
| 6                                    | Marcia piede | 259       | 6     | Centrodestra   | 39.904    |
| 7                                    | Palcoscenico | 246       | 7     | Capogruppo     | 37.044    |
| 8                                    | Piattaforma  | 239       | 8     | Portavoce      | 31.915    |
| 9                                    | Asciugamano  | 233       | 9     | Capoluogo      | 31.834    |
| 10                                   | Pianoforte   | 228       | 10    | Resoconto      | 25.343    |
| 11                                   | Portavoce    | 200       | 11    | Pomodoro       | 25.322    |
| 12                                   | Centrodestra | 195       | 12    | Portafoglio    | 19.602    |
| 13                                   | Reggisenò    | 185       | 13    | Passaporto     | 19.141    |
| 14                                   | Passaporto   | 181       | 14    | Fine Settimana | 18.595    |

| Corpus del Nuovo Vocabolario di Base |                |     | ITWAC |                  |        |
|--------------------------------------|----------------|-----|-------|------------------|--------|
| 15                                   | Fine Settimana | 164 | 15    | Capoverso        | 18.591 |
| 16                                   | Capogruppo     | 163 | 16    | Buonsenso        | 17.914 |
| 17                                   | Banconota      | 159 | 17    | Pianoforte       | 17.821 |
| 18                                   | Buonsenso      | 152 | 18    | Mezzanotte       | 16.874 |
| 19                                   | Centrosinistra | 149 | 19    | Palcoscenico     | 15.834 |
| 20                                   | Anno Luce      | 136 | 20    | Marciapiede      | 14.547 |
| 21                                   | Gentiluomo     | 127 | 21    | Arcobaleno       | 12.224 |
| 22                                   | Capoluogo      | 122 | 22    | Buonafede        | 12.120 |
| 23                                   | Buonafede      | 119 | 23    | Cassintegrazione | 10.521 |
| 24                                   | Salvaguardia   | 116 | 24    | Cortometraggio   | 9.861  |
| 25                                   | Altopiano      | 108 | 25    | Altopiano        | 9.232  |
| 26                                   | Passatempo     | 106 | 26    | Pallavolo        | 8.433  |
| 27                                   | Pianoterra     | 105 | 27    | Banconota        | 8.383  |
| 28                                   | Grattacielo    | 104 | 28    | Pianoterra       | 8.354  |
| 29                                   | Arcobaleno     | 97  | 29    | Lungometraggio   | 7.659  |
| 30                                   | Caporedattore  | 94  | 30    | Bianconero       | 7.240  |

Tabella 5. Lista di frequenza dei composti nei due corpora (rango 1-30)

Confrontando le prime trenta posizioni delle liste di frequenza lemmatizzate – relative alle parole composte – dei due corpora (Tabella 5) emergono elementi interessanti circa il loro contenuto e l’influenza che esso determina sui risultati dell’indagine. Osservando le prime posizioni della lista (ranghi 1-5) emergono subito alcune differenze: *in primis*, si nota che nel corpus del NVdB le prime due posizioni sono occupate dalle indicazioni temporali *mezzogiorno* e *mezzanotte*, appartenenti alla fascia d’uso FO del Gradit, laddove in itWaC ai medesimi ranghi si trovano i composti *centrosinistra* e *piattaforma*, parole tipiche del linguaggio della politica, nel primo caso, e del web, nel secondo. In itWaC, in generale, si nota una significativa presenza, ai ranghi più alti, di termini tipici del linguaggio della politica e dei giornali: *centrosinistra* (rango 1), *centrodestra* (rango 6), *capogruppo* (rango 7), *portavoce* (rango 8). Tali forme risultano attestate anche nella lista di frequenza del corpus del NVdB ma a ranghi più bassi: *portavoce* (rango 11), *centrodestra* (rango 12), *capogruppo* (rango 16), *centrosinistra* (rango 19). I ranghi più alti nel corpus del NVdB sono invece occupati da parole di stretto uso quotidiano, appartenenti alla fascia d’uso FO del Gradit, alcuni dei quali assenti dalla lista di itWaC<sup>23</sup>: *pomodoro* (rango 4), *portafoglio* (rango 5), *marciapiede* (rango 6), *asciugamano* (rango 9). Dal confronto tra le due liste di frequenza si può quindi notare come in itWaC risultino sovrastimati alcuni termini particolarmente frequenti nel linguaggio della politica e del giornalismo, e sottostimate le parole di significato più concreto, appartenenti all’orizzonte quotidiano dei parlanti. Tali discrasie sono strettamente collegate al contenuto non bilanciato del corpus, costituito da testi scaricati

<sup>23</sup> In particolare, dei lemmi della lista di frequenza del corpus del NVdB, sono assenti da quella di itWaC le seguenti forme: *asciugamano*, *reggiseno*, *anno luce*, *gentiluomo*, *passatempo*, *grattacielo*, *caporedattore*.

automaticamente dalla rete, tra i quali è molto probabile si trovi un significativo numero di articoli giornalistici o pagine web dedicate all'analisi politica.

Per quanto riguarda il numero delle occorrenze con cui ricorrono i composti, da itWaC si possono estrarre dei dati quantitativi notevolmente più consistenti: l'alto numero di occorrenze registrato, se da un lato costituisce un aspetto positivo, perché fornisce un solido fondamento empirico all'analisi, dall'altro rappresenta un limite non indifferente, perché di fatto impedisce, o quantomeno limita, l'analisi delle concordanze e dei contesti d'uso. Non va inoltre sottovalutato il fatto che, trattandosi di un corpus dal contenuto parzialmente sconosciuto, non si è spesso in grado di verificare l'attendibilità della fonte.

Il corpus del NVdB, pur fornendo dati significativi per un insieme più limitato di forme, garantisce la possibilità di osservarne i contesti d'uso e verificare la distribuzione quantitativa del composto rispetto alla tipologia testuale.

|    | Lemma        | Freq. Tot. | Stampa | Narrativa | Saggistica | Spettacolo | CMC   | Parlato |
|----|--------------|------------|--------|-----------|------------|------------|-------|---------|
| 1  | Mezzogiorno  | 408        | 78.7   | 87.8      | 31.6       | 37.6       | 18    | 160.5   |
| 2  | Mezzanotte   | 388        | 53.1   | 97.1      | 36.7       | 64.1       | 42.9  | 118.7   |
| 3  | Capolavoro   | 385        | 95.1   | 33.6      | 97.4       | 44.6       | 192.6 | 7.6     |
| 4  | Pomodoro     | 356        | 63.4   | 90.6      | 32.9       | 34.9       | 115   | 59.8    |
| 5  | Portafoglio  | 352        | 86.9   | 75.7      | 10.1       | 76.7       | 42.9  | 87.4    |
| 6  | Marciapiede  | 259        | 40.9   | 117.7     | 16.4       | 43.2       | 11.1  | 38      |
| 7  | Palcoscenico | 246        | 69.5   | 17.7      | 65.8       | 112.9      | 12.5  | 16.1    |
| 8  | Piattaforma  | 239        | 111.4  | 15.9      | 79.7       | 7          | 44.3  | 11.4    |
| 9  | Asciugamano  | 233        | 9.2    | 72.9      | 6.3        | 50.2       | 119.1 | 16.1    |
| 10 | Pianoforte   | 228        | 30.7   | 86.9      | 22.8       | 19.5       | 20.8  | 55.1    |

Tabella 6. Primi dieci composti in ordine di frequenza attestati nel corpus del NVdB: distribuzione quantitativa rispetto ai sottocorpora

Dai dati normalizzati dei primi dieci composti più frequenti nel corpus del NVdB (Tabella 6) si può osservare il modo in cui essi si distribuiscono tra le diverse tipologie testuali. Come prevedibile, i primi due composti – che designano indicazioni temporali – si concentrano soprattutto nel parlato e in letteratura. Registrano alte frequenze nel sottocorpus di Narrativa anche altri composti i cui *designata* ricorrono frequentemente nell'orizzonte quotidiano dei parlanti – come *pomodoro*, *marciapiede*, *asciugamano*, *pianoforte* – e che in itWaC risultano sottostimati a causa del numero presumibilmente limitato di testi letterari contenuti nel corpus.

### 7.1 I composti *capo* + Nome

I composti con *capo*- fanno parte della vasta ed eterogenea categoria dei composti Nome Nome e costituiscono una serie molto numerosa in italiano: il Devoto Oli registra 129 forme; un indizio della vitalità del pattern *capo* + nome in italiano contemporaneo ci è dato dalla presenza di cinque neoformazioni di questo tipo nei repertori di Adamo e Della Valle<sup>24</sup>. Questo insieme di

<sup>24</sup> In particolare, si tratta delle forme *capoazienda*, *capodelegazione*, *caposcafista*, *capo-staff* (Adamo, Della Valle 2003b e Adamo, Della Valle 2006). Va ricordato che la rappresentatività dei due repertori è molto limitata, dal momento che essi sono stati raccolti dai due studiosi attraverso lo spoglio manuale di un corpus di quotidiani nazionali di piccole dimensioni (cfr. Adamo, Della Valle 2003a).

composti, apparentemente omogeneo dal punto di vista formale, si compone in realtà di forme costituite da due sostantivi tra i quali intercorrono relazioni diverse. Riprendendo la classificazione proposta da Serianni (1989: 154) nella sua grammatica, questo insieme di forme può essere classificato in tre tipologie a seconda che *capo* individui: 1) «colui che è a capo di qualcosa ('x è a capo y')»; 2) «colui che è a capo di qualcuno ('x è capo tra  $x_1, x_2, x_3...$ ')»; 3) «ciò che si segnala tra altri oggetti omogenei come 'preminente', 'eccellente' (= 'un capo-x')». Adottando la classificazione proposta da Scalise, Bisetto (2009), i composti del primo tipo si possono inserire nel gruppo dei subordinativi (es. *capostazione, capoclasse, etc.*); quelli del secondo e del terzo tipo appartengono al tipo appositivo, con testa a sinistra nel primo caso (es. *caporedattore, capocuoco, etc.*), a destra nel secondo (es. *capoluogo*).

Nella Tabella 7 si riportano i dati relativi alla distribuzione quantitativa delle forme plurali e delle occorrenze rispetto alle tre tipologie individuate attraverso il criterio semantico.

| Tipologia di composto              | Tipo flessione    | Corpus del NVDB |     |            |      | itWaC        |      |            |      |
|------------------------------------|-------------------|-----------------|-----|------------|------|--------------|------|------------|------|
|                                    |                   | Numero forme    | %   | Occorrenze | %    | Numero forme | %    | Occorrenze | %    |
| 'x è il capo di y'                 | Flessione interna | 28              | 90  | 89         | 89.9 | 83           | 79   | 13.700     | 88.7 |
|                                    | Flessione esterna | 2               | 6.5 | 9          | 9.1  | 14           | 13.3 | 1.644      | 10.6 |
|                                    | Doppia flessione  | 1               | 3.5 | 1          | 1    | 8            | 7.6  | 93         | 0.7  |
| 'x è capo tra $x_1, x_2, x_3...$ ' | Flessione interna | 0               | 0   | 0          | 0    | 1            | 23.8 | 11         | 1.9  |
|                                    | Flessione esterna | 3               | 50  | 4          | 57.1 | 10           | 47.6 | 275        | 47.9 |
|                                    | Doppia flessione  | 3               | 50  | 3          | 42.9 | 10           | 47.6 | 288        | 50.2 |
| 'un capo-x'                        | Flessione interna | 0               | 0   | 0          | 0    | 5            | 41.6 | 41         | 0.2  |
|                                    | Flessione esterna | 2               | 100 | 130        | 100  | 5            | 41.6 | 19.009     | 99.7 |
|                                    | Doppia flessione  | 0               | 0   | 0          | 0    | 2            | 16.8 | 13         | 0.07 |

Tabella 7. Composti con *capo*:- distribuzione quantitativa delle occorrenze rispetto alla classificazione di Serianni (1989)

Nel gruppo dei composti determinativi si può individuare una tendenza molto netta a formare il plurale modificando soltanto il primo costituente, che rappresenta l'elemento testa del composto. La significativa frequenza con cui queste forme ricorrono nell'uso non sembra aver determinato una diminuzione nel grado di trasparenza del composto né lo spostamento della flessione sul margine destro del lessema. Anche per i composti in cui *capo* designa un elemento che si segnala come preminente rispetto ad altri non si registrano oscillazioni significative nella formazione del plurale: nella quasi totalità delle forme la marca di plurale è posta sul costituente testa, posto questa volta a destra.

Le occorrenze dei composti in cui *capo* designa 'colui che è a capo di qualcuno', tipologia meno produttiva delle altre, si distribuiscono quasi equamente tra forme plurali che modificano

soltanto il secondo costituente e forme plurali con entrambi i costituenti modificati: i due tipi di flessione sembrano quindi in concorrenza nell'uso dei parlanti<sup>25</sup>. Del composto più frequente appartenente a tale tipologia, *caporedattore*, itWaC attesta ben quattro forme plurali: *capiredattori* (65 occ.)/*capi redattori* (57 occ.), *caporedattori* (64 occ.) e *capiredattore* (3 occ.).

L'osservazione del contesto in cui le singole forme vengono utilizzate completa il quadro su questo tipo di composti: in relazione a questo aspetto, il corpus del NVdB fornisce un contributo fondamentale perché permette di osservare le concordanze; le enormi dimensioni di itWaC rendono invece di fatto impossibile il suo utilizzo in questo senso. Rispetto alla tipologia dei composti con *capo-*, si può notare che solamente quattro composti, *capofamiglia*, *capolinea*, *capobranco*, *capofila*, in un numero di casi molto limitato (2 occorrenze), vengono utilizzati come invariabili.

Delle tre tipologie, quella in cui si rilevano più discrasie tra i due dizionari, rispetto alla flessione al plurale, è la seconda: nella seguente tabella (Tabella 8) se ne riportano alcuni esempi che permettono di osservare tre casi in cui i dati empirici possono essere utilizzati per migliorare le informazioni lessicografiche. Come si può notare, i dati estratti dal corpus del NVdB per queste forme non sono sufficienti a osservare il fenomeno; in questo caso, quindi, si farà riferimento a quelli estratti da itWaC.

| Lemma          | Forme plurali attestate | itWaC           |             | Corpus del NVdB |             | Plurale Devoto Oli                   | Plurale Gradit         |
|----------------|-------------------------|-----------------|-------------|-----------------|-------------|--------------------------------------|------------------------|
|                |                         | Freq. plurale   | Freq. sing. | Freq. plurale   | Freq. sing. |                                      |                        |
| capocannoniere | capocannonieri          | 36              | 1.105       | 0               | 12          | capicannonieri                       | capocannonieri         |
|                | capicannonieri          | 1               |             | 0               |             |                                      |                        |
| capocomico     | capocomici              | 31              | 754         | 1               | 4           | capocomici, capicomici (meno comune) | capocomici, capicomici |
|                | capicomici              | 3               |             | 0               |             |                                      |                        |
| capocronista   | capocronisti            | 3               | 99          | 0               | 0           | capicronisti                         | capicronisti           |
|                | capicronisti            | 6               |             | 0               |             |                                      |                        |
| capocuoco      | capocuochi              | 3               | 117         | 0               | 18          | capocuochi, capicuochi               | capocuochi, capicuochi |
|                | capicuochi              | 0               |             | 1               |             |                                      |                        |
| capomastro     | capomastri              | 53              | 288         | 2               | 12          | capomastri, capimastri               | capomastri, capimastri |
|                | capimastri              | 49              |             | 0               |             |                                      |                        |
| capo operaio   | capo-operai             | 1               | 114         | 0               | 0           | capi operai                          | capooperai             |
|                | capi operai             | 37              |             | 0               |             |                                      |                        |
| caporedattore  | caporedattori           | 64              | 2.244       | 1               | 91          | capiredattori                        | caporedattori          |
|                | capiredattori           | 122             |             | 2               |             |                                      |                        |
|                | capiredattore           | 3 <sup>26</sup> |             | 0               |             |                                      |                        |

<sup>25</sup> In questi casi si potrebbe parlare di «sovrabbondanza», riprendendo il termine proposto da Thornton (2012: 183) in riferimento alla polimorfia nei paradigmi flessivi verbali dell'italiano, definito come il fenomeno per cui «two or more forms are available to realize the same cell in an inflectional paradigm».

<sup>26</sup> Effettuando su itWaC una query relativa alla forma *capiredattore*, si trovano inizialmente 11 risultati: tuttavia, osservando le concordanze, si può notare che in nove casi si tratta dello stesso testo copiato in diverse pagine internet; in realtà, la forma ha quindi solo 3 occorrenze. Questo caso permette di riflettere

|             |             |    |     |   |   |                             |                             |
|-------------|-------------|----|-----|---|---|-----------------------------|-----------------------------|
| capotecnico | capotecnici | 7  | 316 | 0 | 0 | capotecnici,<br>capitecnici | capitecnici,<br>capotecnici |
|             | capitecnici | 66 |     | 0 |   |                             |                             |

Tabella 8. Composti con *capo-*: casi di discrasia tra dati empirici e dizionari

Il primo caso è quello di *capocannoniere*, del quale i due dizionari riportano due forme di plurale differenti: il Gradit registra la forma con flessione esterna *capocannonieri*, il DO quella con doppia flessione *capicannonieri*; i dati estratti da itWaC mostrano invece un quadro molto chiaro in cui la forma con flessione esterna è la più frequente nell'uso dei parlanti.

Il secondo caso è quello di *capocomico* e *capotecnico*: per entrambi i composti i due dizionari registrano due plurali, laddove i dati di itWaC mostrano la netta prevalenza di una delle due forme, quella con flessione esterna nel primo caso (*capocomici*), quella con doppia flessione nel secondo (*capitecnici*).

Diverso è il caso di *caporedattore*, del quale itWaC attesta ben tre forme di plurale in concorrenza nell'uso dei parlanti<sup>27</sup>: tale oscillazione non è colta dai dizionari che si limitano a registrare una sola forma, su cui peraltro non concordano (per il Gradit il plurale di *caporedattore* è *caporedattori*, per il DO è *capiredattori*).

## 7.2 I composti *basso-* + *Nome*

I composti *basso*+*Nome* sembrano costituire una serie non più produttiva in italiano contemporaneo: nei repertori di neologismi di Adamo e Della Valle (2003b, 2006) non è registrata nessuna parola composta con *basso* come primo costituente. Insieme alle forme con *alto*, questo tipo di composti viene segnalato dalle grammatiche<sup>28</sup> perché ammette delle oscillazioni nella formazione del plurale.

| Lemma      | Forme plurali attestate | itWaC         |                 | Corpus del NVdB |                 | Plurale Devoto Oli        | Plurale Gradit                     |
|------------|-------------------------|---------------|-----------------|-----------------|-----------------|---------------------------|------------------------------------|
|            |                         | Freq. plurale | Freq. singolare | Freq. plurale   | Freq. singolare |                           |                                    |
| bassadanza | bassedanze              | 2             | 27              | 0               | 0               | bassadanze,<br>bassedanze | bassadanze                         |
| bassofondo | bassofondi              | 6             | 198             | 0               | 0               | bassifondi                | bassifondi,<br>bassofondi<br>(ant) |
|            | bassifondi              | 1.513         |                 | 13              |                 |                           |                                    |
| bassopiano | bassopiani              | 44            | 157             | 0               | 2               | bassopiani,<br>bassipiani | bassipiani                         |
|            | bassipiani              | 71            |                 | 0               |                 |                           |                                    |

su uno dei problemi con cui si deve fare i conti quando si utilizza itWaC: essendo costituito da pagine web scaricate automaticamente dalla rete, esso contiene molte copie di testi che possono falsare anche in modo significativo il dato quantitativo sulle occorrenze di una forma.

<sup>27</sup> Questi dati riflettono, e quindi confermano, la situazione delineata dal grafico prodotto da GNV (paragrafo 3).

<sup>28</sup> È ad esempio segnalato in Serianni (1989: 155).

|              |              |       |       |    |    |              |   |
|--------------|--------------|-------|-------|----|----|--------------|---|
| bassorilievo | bassorilievi | 1.272 | 1.327 | 10 | 13 | bassorilievi | bassorilievi,<br>bassirilievi<br>(ant.) |
|              | bassirilievi | 42    |       | 0  |    |              |   |

Tabella 9. Composti con *basso-*: casi di discrasia tra dati empirici e dizionari

I dati riportati in Tabella 9 permettono di osservare come, anche per questo tipo di forme, le indicazioni dei dizionari non sempre coincidano e i dati estratti dai corpora possano aiutare a chiarire la situazione.

Della forma *bassadanza* itWaC attesta una sola forma plurale con doppia flessione, *bassedanze*; i due dizionari si comportano in modo diverso: il Gradit registra la forma con flessione esterna, il DO riporta due plurali.

*Bassopiano* presenta nell'uso due forme plurali concorrenti, entrambe attestate in itWaC con un significativo numero di occorrenze; tale oscillazione è segnalata dal DO, ma non dal Gradit, che si limita a registrare la forma con doppia flessione.

Nei casi di *bassofondo* e *bassorilievo*, i dati di itWaC e del Corpus del NVdB indicano che i parlanti utilizzano una forma plurale in particolare, nettamente più frequente rispetto alla concorrente e registrata dal DO come unico plurale; in questi casi il Gradit sceglie di riportare anche l'altra variante, specificando però che si tratta di una forma più antica.

### 7.3 I composti Verbo + Nome

I composti Verbo+Nome rappresentano la tipologia su cui per più tempo e in modo più significativo si è concentrata l'attenzione degli studiosi: l'aspetto su cui si è più a lungo dibattuto è la natura del primo costituente (cfr. Bisetto 1999; Floricic 2008); i contributi più recenti ne hanno invece analizzato la produttività e le proprietà semantiche e morfosintattiche (cfr. Magni 2010; Ricca 2005, 2010, 2015; Von Heusinger, Schwarze 2011).

Per analizzare la flessione di questo tipo di forme è fondamentale poter disporre di un corpus che consenta lo spoglio delle concordanze per osservare i contesti d'uso. Questo tipo di composti può infatti essere variabile, ossia presentare una forma per il singolare e una per il plurale, o invariabile, quando compare nell'uso in un'unica forma: per discriminare tra i due casi è quindi necessario osservare il contesto sintattico in cui il composto è utilizzato. La tabella 10 riporta un campione di composti Verbo+Nome<sup>29</sup> che, sulla base dei dati estratti dal corpus del NVdB, possono essere classificati come "variabili": si tratta di composti che al singolare sono costituiti da un elemento verbale e un sostantivo singolare (ad esempio *grattacielo*); al plurale essi vengono sempre flessi modificando il secondo elemento (quindi *grattaciel*). Queste forme non sono quindi mai attestate nel corpus come invariabili: il composto *passaporto*, ad esempio, è sempre attestato come singolare; la forma *passaporti* indica sempre un plurale.

| Lemma | Forme plurali attestate | Corpus del NVdB | itWaC |  | Plurale Gradit |
|-------|-------------------------|-----------------|-------|--|----------------|
|-------|-------------------------|-----------------|-------|--|----------------|

<sup>29</sup> Per questioni di spazio si riportano solo i primi otto composti VN variabili in ordine di frequenza; per un quadro più ampio rimando a Micheli (2016: 245-52).

|               |               | Occ.<br>singolare | Occ.<br>plurale | Occ.<br>Tot<br>lemma | Occ.<br>tot<br>lemma | Plurale<br>Devoto<br>Oli |             |
|---------------|---------------|-------------------|-----------------|----------------------|----------------------|--------------------------|-------------|
| battibecco    | battibecco    | 20                | 0               | 39                   | 1.391                | Variabile                | Variabile   |
|               | battibecchi   | 0                 | 19              |                      |                      |                          |             |
| batticuore    | batticuore    | 18                | 0               | 21                   | 699                  | Invariabile              | Variabile   |
|               | batticuori    | 0                 | 3               |                      |                      |                          |             |
| giravolta     | giravolta     | 11                | 0               | 16                   | 786                  | Variabile                | Variabile   |
|               | giravolte     | 0                 | 5               |                      |                      |                          |             |
| grattacielo   | grattacielo   | 55                | 0               | 104                  | 6.978                | Variabile                | Variabile   |
|               | grattacieli   | 0                 | 49              |                      |                      |                          |             |
| parapetto     | parapetto     | 39                | 0               | 44                   | 2.534                | Variabile                | Variabile   |
|               | parapetti     | 0                 | 5               |                      |                      |                          |             |
| parasole      | parasole      | 12                | 0               | 14                   | 539                  | Invariabile              | Variabile   |
|               | parasoli      | 0                 | 2               |                      |                      |                          |             |
| passaporto    | passaporto    | 158               | 0               | 181                  | 19.141               | Variabile                | Variabile   |
|               | passaporti    | 0                 | 23              |                      |                      |                          |             |
| portabandiera | portabandiera | 13                | 0               | 16                   | 917                  | Invariabile              | Invariabile |
|               | portabandiere | 0                 | 3               |                      |                      |                          |             |

Tabella 10. Composti Verbo+Nome: forme variabili più frequenti nel corpus del NVdB

Come si può notare osservando i dati sulla frequenza totale dei lemmi, queste forme registrano in itWaC un altissimo numero di occorrenze, che rende di fatto impossibile controllarne tutti i contesti: in questo caso, le dimensioni di itWaC costituiscono quindi un forte limite a tale risorsa. Diversa è la situazione del corpus del NVdB, sulla base del quale è invece possibile analizzare il fenomeno e confrontarlo con le due opere lessicografiche.

In questo caso è soprattutto il DO a riportare informazioni discordanti con i dati empirici: *batticuore*, *parasole* e *portabandiera* vengono infatti registrati come invariabili, nonostante nell'uso siano attestate le rispettive forme flesse; più preciso sembra essere il Gradit che sbaglia solo nel caso di *portabandiera*.

## 8. Conclusioni

Con questo lavoro si è cercato di contribuire alla riflessione sul rapporto tra dati empirici e risorse lessicografiche, con particolare riferimento a un fenomeno morfologico spesso oggetto di dubbi da parte dei parlanti: la formazione del plurale delle parole composte.

Nella prima parte del lavoro (§ 1-3) si è discussa la possibilità di utilizzare il web come fonte di dati, mettendo in luce i numerosi problemi metodologici che ne inficiano la scientificità e ne scoraggiano quindi l'uso nell'ambito di studi linguistici, nei quali d'altra parte il contributo dei corpora tradizionali risulta insostituibile. Sono stati poi analizzati alcuni aspetti quantitativi e



qualitativi che distinguono le parole composte dai lessemi semplici, in relazione ai quali sono stati proposti dei criteri metodologici per la scelta del campione e del corpus da utilizzare per un'indagine che intenda osservarne la formazione del plurale. I principali problemi ancora irrisolti riguardano in particolare due questioni: le caratteristiche che un corpus dovrebbe presentare per permettere lo studio delle parole composte e la loro effettiva applicabilità a una risorsa reale. Dimensioni e contenuto del corpus costituiscono gli aspetti su cui è necessaria una riflessione più ampia rispetto a quanto è stato fatto finora. In particolare, da un lato occorre valutare con più precisione l'estensione del corpus rispetto alla frequenza delle entità lessicali oggetto di studio, dall'altro costituiscono requisiti imprescindibili il bilanciamento del corpus e la possibilità di consultarne le concordanze. Riuscire a garantire l'estensione necessaria e la qualità del contenuto di un corpus rappresenta un problema di non facile soluzione a livello pratico; questo per almeno due motivi: perché la raccolta di corpora molto estesi, come ad esempio itWaC, implica l'utilizzo di una procedura automatica che non permette alcun bilanciamento rispetto alla tipologia testuale e perché le grandi dimensioni di un corpus impediscono, o quantomeno limitano, la possibilità di consultarne le concordanze. Sul rapporto tra dimensioni e qualità del contenuto devono quindi concentrarsi le riflessioni future.

Nella seconda parte del contributo (§ 4-7) sono stati presentati la metodologia di indagine e i risultati di una ricerca condotta su due corpora di italiano contemporaneo, itWaC e il corpus del Nuovo Vocabolario di Base: attraverso la descrizione di alcuni *case studies*, relativi a tre particolari tipologie di composti, si è messo in luce in che modo, e in che misura, i dati estratti da corpora possono contribuire a migliorare le informazioni di natura morfologica fornite da due dizionari, il Devoto Oli 2014 e il Gradit, di cui si sono messi in evidenza i limiti. L'analisi dei composti con *capo-* e *basso-* come primo costituente ha permesso di individuare almeno tre casi in cui quanto riportato dai dizionari non riflette i dati quantitativi estratti da corpora e necessita quindi di integrazioni o correzioni: (a) Gradit e Devoto Oli registrano forme di plurale diverse, di cui una assente, o attestata con valori di frequenza bassissimi, dai corpora; (b) uno dei due dizionari, o entrambi, registra più forme plurali, laddove nell'uso ne risulta attestata una in particolare, con alti valori di frequenza; (c) uno dei due dizionari, o entrambi, registra un solo plurale, nonostante nell'uso vi siano più forme in concorrenza, con valori di frequenza molto vicini. Il caso dei composti Verbo Nome ha permesso di mettere in luce l'importanza di osservare i contesti d'uso per individuare quando la forma è invariabile: anche rispetto alla variabilità/invariabilità di tali composti si sono registrate incongruenze tra dati e dizionari.

Rispetto a tali problemi, il contributo dei due corpora si è dimostrato differente, in una certa misura complementare, e strettamente legato alle caratteristiche strutturali delle due risorse: in particolare, si è potuto osservare che un corpus di medie dimensioni e di contenuto variato, come il corpus del NVdB, permette di estrarre dati quantitativi e qualitativi relativi soltanto a un numero limitato di composti, ma garantisce la possibilità di osservarne i contesti d'uso e coglierne la variabilità/invariabilità; d'altra parte, un corpus di grandi dimensioni come itWaC fornisce dati quantitativi più consistenti e relativi a un maggior numero di forme, ma ne limita l'analisi qualitativa.

Il presente lavoro ha dimostrato che, seppur ancora limitato da questioni teoriche e pratiche, il contributo dei dati empirici in lessicografia appare irrinunciabile anche per quanto riguarda le informazioni morfologiche che accompagnano i lemmi; è d'altra parte necessario portare avanti la riflessione teorica sulla realizzazione di corpora e sull'estrazione di dati rappresentativi e consistenti che permettano di rendere proficuo il dialogo tra dizionari e dati.

---

**BIBLIOGRAFIA**

- Adamo, G., Della Valle, V. (2003a), *L'Osservatorio neologico della lingua italiana: linee di tendenza nell'innovazione lessicale dell'italiano contemporaneo*, in Idd. (a c. di), *Innovazione lessicale e terminologie specialistiche*, Firenze, Olschki, pp. 83-105.
- Adamo, G., Della Valle, V. (2003b), *Neologismi quotidiani: un dizionario a cavallo del millennio (1998-2003)*, Firenze, Olschki.
- Adamo, G., Della Valle, V. (2006), *2006 parole nuove. Un dizionario di neologismi dai giornali*, Milano, Sperling & Kupfer.
- Baroni, M., Bernardini S., Ferraresi A., Zanchetta E. (2009), *The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora*, in "Journal of Language Resources and Evaluation", 43(3), pp. 209-226.
- Baroni, M., Guevara, E., Pirrelli, V. (2007), *NN compounds in Italian: Modelling category induction and analogical extension*, in V. Pirrelli (a cura di), *Psycho-Computational Issues in Morphology Learning and Processing* (Special Issue of "Lingue e Linguaggio", 6.2), Bologna, Il Mulino, pp. 263-290.
- Baroni, M., Kilgarriff, A. (2006), *Large linguistically-processed web corpora for multiple languages*, in *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*. Association for Computational Linguistics, pp. 87-90.
- Baroni, M., Guevara, E., Pirrelli, V. (2006), *Sulla tipologia dei composti NN in italiano: principi categoriali ed evidenza distribuzionale a confronto*, in G. Ferrari, R. Benatti, M. Mosca (a cura di), *Linguistica e Modelli tecnologici della ricerca. Atti del XL Congresso Internazionale di Studi della Società di Linguistica Italiana*, Roma, Bulzoni, pp. 21-38.
- Beißwenger, M., Chanier, T., Chiari, I., Erjavec, T., Fisser, D., et al. (2016) *Integrating corpora of computer-mediated communication into the language resources landscape: Initiatives and best practices from French, German, Italian and Slovenian projects. CLARIN Annual Conference 2016*, Oct 2016, Aix-en-Provence, France. <https://www.clarin.eu/content/programme-clarin-annual-conference-2016>.
- Bisetto, A. (1999), *Note sui composti VN dell'italiano*, in P. Benincà, A. Mioni, L. Vanelli (a c. di), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia*, Roma, Bulzoni, pp. 503-38.
- Brysbaert, M., New B. (2009), *Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English*, in "Behavior research methods" 41.4, pp. 977-90.
- Chiari, I., (2016) (a c. di), *Capirsi e fraintendersi al computer. La negoziazione del senso nella conversazione sui nuovi media*, Roma, Carocci (in stampa).
- Chiari, I., De Mauro, T. (2012), *The new basic vocabulary of Italian: problems and methods*, in "Statistica Applicata. Italian Journal of Applied Statistics", 22 (1), pp. 21-35.
- Crystal, D. (2006), *Language and the internet*, Cambridge, Cambridge University Press.
- D'Achille, P., Grossmann, M. (2009), *Stabilità e instabilità dei composti aggettivo + aggettivo in italiano*, in E. Lombardi Vallauri, L. Mereu (a c. di), *Spazi linguistici. Studi in onore di Raffaele Simone*, Roma, Bulzoni, pp. 143-71.
- Devoto G., Oli G. C. (2014), *Il Devoto Oli. Vocabolario della lingua italiana*, a cura di Luca Serianni e M. Trifone, Firenze, Le Monnier.
- Favretti, R. Rossini, Tamburini F., De Santis C. (2002), *CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model*, in *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*. Munich: Lincom-Europa".
- Florjic, F. (2008), *The Italian Verb-Noun anthroponymic Compounds at the Syntax / Morphology Interface*, in "Morphology", 18.2, pp. 167-93.
- Gomez Gane, Y. (2008), *Google ricerca libri e la linguistica italiana: vademecum per l'uso di un nuovo strumento di lavoro*, in "Studi linguistici italiani", 2, pp. 1000-19.
- Gradi = *Grande dizionario italiano dell'uso*, diretto da Tullio De Mauro, Torino, Utet, 2007.
- Grandi, N. (2015) (a c. di), *La grammatica e l'errore. Le lingue naturali tra regole, loro violazioni ed eccezioni*, Bologna, Bononia U.P.
- Grossmann, M., Rainer F. (2009), *Italian adjective-adjective compounds: Between morphology and syntax*, in "Italian Journal of Linguistics", 21.1, pp. 71-96.
- Hathout, N., Montermini, F., Tanguy, L. (2008), *Extensive data for morphology: using the World Wide Web*, in "French Language Studies", 18, pp. 67-85.

- Iacobini, C., Thornton, A. M. (2016), *Morfologia e formazione delle parole*, in S. Lubello (a c. di), *Manuale di linguistica italiana*, Berlin/Boston, De Gruyter Mouton, pp. 190-220.
- Iannàccaro, G. (2000), *Per una semantica più puntuale del concetto di 'dato linguistico': un tentativo di sistematizzazione epistemologica*, in "Quaderni di semantica: rivista internazionale di semantica teorica e applicata", 21.1, pp. 51-80.
- Kilgarriff, A. (2007), *Googleology is bad science*, in "Computational linguistics", 33.1, pp. 147-51.
- Kilgarriff, A., Grafenstette, G. (2003), *Introduction to the special issue on the web as corpus*, in "Computational linguistics", 29.3, pp. 333-47.
- Lehmann, C. (2004), *Data in linguistics*, in "The Linguistic Review", 21.3-4, pp. 175-210.
- Libber, G., Jarema, G. (2007) (a c. di), *The Representation and Processing of Compound Words*, Oxford University Press.
- Lüdeling, A., Evert, S., Baroni, M. (2007), *Using web data for linguistic purposes*, in M. Hundt, N. Nesselhauf, C. Biewer (a c. di), *Corpus linguistics and the Web*, Amsterdam, Rodopi, pp. 7-24.
- Magni, E. (2010), *From the periphery to the core of Romance [VN] compounds*, in "Lingue e Linguaggio", 9.1, pp. 3-39.
- Masini, F., Scalise, S. (2012), *Italian compounds*, «PROBUS», 24, pp. 61-91.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Brockman, W. (2012), The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. Lieberman Aiden, *Quantitative Analysis of Culture Using Millions of Digitized Books*, in "Science", pp. 176-82.
- Micheli, M. S., (2016), *Sul plurale delle parole composte nell'italiano contemporaneo*, in "Studi di lessicografia italiana", XXXIII, pp. 227-55 (in stampa).
- Montermini, F. (2015), *Regole (e irregolarità) nella formazione delle parole*, in Grandi (2015), pp. 63-83.
- Montermini, F. (2008), *La composition en italien dans un cadre de morphologie lexématique*, in D. Amiot, *La composition dans une perspective typologique*, Artois Presses Université, pp. 161-87.
- Ricca, D. (2015), *Verb-noun compounds in Romance*, in Peter O. Müller, I. Ohnheiser, S. Olsen, F. Rainer (a c. di), *Word-formation. An International Handbook of the Languages of Europe*, Berlin/Boston, De Gruyter Mouton, pp. 688-707.
- Ricca, D. (2010), *Corpus data and theoretical implications: With special reference to Italian VN compounds*, in S. Scalise, I. Vogel (a c. di), *Cross-disciplinary Issues in Compounding*, Amsterdam/Philadelphia, John Benjamins, pp. 237-54.
- Ricca, D. (2005), *Al limite tra sintassi e morfologia; i composti aggettivali VN nell'italiano contemporaneo*, in M. Grossmann, A. Thornton, (a c. di) *La formazione delle parole. Atti del XXXVII Congresso di Studi della Società di Linguistica Italiana*, Roma: Bulzoni, pp. 1000-22.
- Scalise, S., Bisetto, A. (2009), *The classification of compounds*, in R. Lieber e P. Stekauer (a c. di), "The Oxford handbook of compounding", OUP Oxford, pp. 49-82.
- Serianni, L. (2014), *Giusto e sbagliato: dove comincia il territorio dell'errore?*, in S. Lubello (a c. di), *Lezioni d'italiano. Riflessioni sulla lingua del nuovo millennio*, Bologna, Il Mulino, pp. 235-46.
- Serianni, L. (2004), *Il sentimento della norma linguistica nell'Italia di oggi*, in "Studi Linguistici Italiani", XXX, pp. 85-103.
- Serianni, L. (1989), *Grammatica italiana*, Torino, UTET.
- Sgroi, S. C. (2016), *Grammatica "clericale" vs. grammatica "laica"*, in "Rivista Italiana di Dialettologia. Lingue dialetti società", XXXIX, pp. 169-85.
- Sgroi, S. C. (2010), *Per una grammatica «laica». Esercizi di analisi linguistica dalla parte del parlante*, Torino, UTET.
- Thornton, A. M. (2012), *Reduction and maintenance of overabundance. A case study on Italian verb paradigms*, in "Word Structure", 5, pp. 183-207.
- Von Heusinger, K., Schwartz, C. (2011), *Italian V+N compounds, inflectional features and conceptual structure*, in "Morphology", 23.3, pp. 325-50.

**M. SILVIA MICHELI** • PhD Student in Linguistic Science (University of Pavia, University of Bergamo, Italy); scientific interests: morphology, word-formation, corpus linguistics.

**E-MAIL** • s.micheli@outlook.it



# DAL DATO ALLO STRUMENTO

Quando l'errore è un valore:  
questioni teoriche e pratiche nell'allestimento del LISSICS  
(Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico)

---

Luca CIGNETTI, Silvia DEMARTINI

**ABSTRACT** • In this paper we introduce the project Tlscrivo (and Tlscrivo2.0), providing an overview of its main features and focusing particularly on the preparation of LISSICS (*Dictionary of Italian Language Written in Italian Switzerland at School*): a resource designed to be useful both to scholars and to teachers. The ongoing process of lemmatisation and PoS tagging is facing many challenges, especially in managing orthographic errors. Also, we will illustrate some of the most relevant orthographic mistakes in the corpus, composed of texts written by primary and lower secondary school students.

**KEYWORDS** • Italian Learner Corpus, Natural Language Processing, Written Italian Language, Orthography, Language Teaching, Language Acquisition, Children's Writing Competences

## 1. I progetti di ricerca Tlscrivo e Tlscrivo2.0

L'analisi di corpora dedicati a specifici tipi di testo in cui la lingua presenta caratteristiche spesso non conformi alla lingua *standard* (su cui Berruto 2010) è una possibilità ricca e complessa. È ricca in quanto permette di addentrarsi nell'esame e nello studio di fenomeni che non sempre emergono da raccolte di testi dalle caratteristiche complessivamente vicine alla norma; è complessa perché testi che presentano, a più livelli, molti tratti devianti rispetto alla norma di riferimento pongono non pochi problemi a livello di trattamento automatico.

I testi raccolti nell'ambito delle ricerche Tlscrivo (DoRe 13DPD3\_136603 La scrittura oggi, tra parlato e lingua mediata dalla rete. Aspetti teorico-descrittivi, diagnosi e interventi didattici) e Tlscrivo2.0 (FNS 100012\_156247 Scrivere a scuola nel terzo millennio. Descrizione della varietà e del vocabolario dell'italiano scritto in contesto scolastico ticinese e implicazioni didattiche)<sup>1</sup> sono un caso esemplare in questo senso. Infatti, il corpus in esame – d'ora in poi corpus "DFA-Tlscrivo" –, raccolto nel 2012, cioè nella fase iniziale del progetto, è composto da 1.735 testi scritti a scuola da bambini e ragazzi tra gli 8 e i 14 anni, in Canton Ticino (cfr. par. 1.1) ed è perciò rappresentativo di una particolare varietà di scrittura: quella dei giovani scriventi in contesto scolastico.

L'obiettivo di fondo dei lavori è quello di rinnovare la riflessione intorno alla didattica della scrittura sulla base di dati di prima mano che contribuiscano a delineare il quadro linguistico

---

<sup>1</sup> Si tratta di due fasi triennali di uno stesso progetto in continuità (2011-2014 e 2014-2017), finanziato dal Fondo Nazionale Svizzero per la Ricerca Scientifica e condotto dal *Centro di Didattica dell'italiano e delle lingue nella scuola* del DFA (Dipartimento Formazione e Apprendimento) della SUPSI (Scuola Universitaria Professionale della Svizzera Italiana), a Locarno, sotto la responsabilità di Simone Fornara.

attuale, permettendo di elaborare proposte didattiche innovative per la scuola primaria e secondaria di primo grado. Il primo triennio della ricerca, terminato nel 2014, ha portato alla pubblicazione di un volume di saggi dedicati a fenomeni e ad aspetti specifici della lingua usata dai bambini e dai ragazzi (Cignetti *et al.* 2016a), oltre che all'applicazione dei risultati delle analisi dei testi nell'ambito della formazione di base e continua degli insegnanti. Invece, la seconda fase dei lavori ha tra gli obiettivi principali quello di realizzare il LISSICS, cioè il *Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico*: uno strumento che vorrebbe essere utile tanto agli studiosi di linguistica quanto agli insegnanti, che in esso vedrebbero delineate e sistematizzate alcune delle tendenze di scrittura che emergono quotidianamente nelle produzioni degli allievi<sup>2</sup>.

### 1.1. Il corpus: i dati in analisi

Come si è accennato, il corpus è costituito da testi di tipo narrativo-riflessivo redatti a mano in contesto scolastico in risposta alla consegna di scrittura descritta in Fornara *et al.* (2015) e in Cignetti *et al.* (2016b). Analoga per i diversi ordini di scolarità, la consegna sollecitava lo stesso tipo di azione cognitiva e di produzione testuale, ma era differenziata dalla proposta di due diversi testi-stimolo di partenza: la favola *La tartaruga e la lepre* di Esopo per i bambini di scuola elementare e il racconto *Il giardino segreto* di Italo Calvino per i ragazzi di scuola media. Eccone la formulazione:

Dopo aver letto e analizzato in classe il racconto di Calvino/la favola di Esopo, ti è stato chiesto di pensare a un episodio che hai vissuto o a cui hai assistito dal quale hai ricavato un insegnamento. Raccontalo ora in forma scritta (minimo una pagina, massimo due pagine) e spiega che cosa ti ha insegnato.

Il campione di testi raccolti in modo omogeneo sul territorio del Canton Ticino risulta così composto (SE sta per scuola elementare, SM sta per scuola media<sup>3</sup>):

| Ordine scolastico | SE                       | SM                      |
|-------------------|--------------------------|-------------------------|
| Numero istituti   | 35                       | 21                      |
| Numero classi     | 48 (24 di III, 24 di IV) | 51 (25 di II, 26 di IV) |
| Numero testi      | 742                      | 993                     |
| Totale testi      | 1.735                    |                         |

Tab. 1 – Numeri di istituti, classi e testi raccolti del corpus “DFA-TIscrivo”.

<sup>2</sup> Date le dimensioni del corpus “DFA-TIscrivo”, si tratterà di uno strumento comparabile, per estensione, al LIP di De Mauro *et al.* (1993), al LIPSI di Pandolfi (2009) e al *Lessico elementare* di Marconi *et al.* (1994). Quest'ultimo è lo studio nell'insieme più affine al LISSICS, in quanto offre l'analisi di due corpora da 500.000 parole ciascuno (uno è costituito da componimenti di un campione bambini di scuola elementare, l'altro da una selezione delle più diffuse letture per l'infanzia). Per il LISSICS si prevede tuttavia un formato non solo cartaceo, ma anche consultabile in formato digitale.

<sup>3</sup> Secondo le denominazioni in uso in Canton Ticino, dove la scuola media ha durata quadriennale.

I testi, trascritti in formato elettronico (e successivamente controllati), codificati<sup>4</sup> e predisposti per il trattamento con software specifici, offrono circa 391.250 parole grafiche<sup>5</sup>. Questo numero, che potrà considerarsi definitivo solo al termine del lavoro di pulizia e trattamento del corpus, include «ogni sequenza di caratteri separata dalle altre da uno spazio bianco o da un segno di interpunzione» (De Mauro 2005: 14), e tiene anche conto della tokenizzazione separata delle forme elise (*c'era* = due token). La punteggiatura è conteggiata a parte. Le peculiarità dei dati ai diversi livelli linguistici (ortografia, morfologia e sintassi, ma anche semantica e organizzazione testuale) rendono necessaria una progettazione mirata del lessico di frequenza (LISSICS). In particolare, il *Lessico*, oltre a raccogliere le parole dei testi con relativa annotazione per parti del discorso, vorrebbe non perdere la dimensione dell'errore (in particolare grafico e morfologico) e proporre un lemmario quanto più possibile aderente alla realtà linguistica (cioè, per esempio, che lemmatizzi le polirematiche). Il raggiungimento di questi obiettivi va considerato in relazione alle possibilità offerte dal trattamento automatico del linguaggio e, in questa prospettiva, si stanno cercando gli strumenti più efficienti a disposizione per la lingua italiana, con l'obiettivo di focalizzare l'intervento manuale solo alle fasi e ai casi in cui resta necessario; tenendo conto che, limitando l'osservazione ai tagger, spesso la qualità delle performance cala quando operano su testi lontani dall'italiano standard, ci si propone di migliorarne, se possibile, le funzionalità<sup>6</sup>.

## 2. Dai dati alle informazioni: gli obiettivi delle ricerche e i software in uso

L'esplorazione di dati linguistici pone sempre la sfida dell'estrazione di informazioni a partire da dati grezzi, in modo da renderle accessibili e consultabili in modo esaustivo e agevole, perdendo meno informazione possibile. In estrema sintesi, le prospettive scientifiche e applicative delle ricerche *Tiscrivo* e *Tiscrivo2.0* (alcune delle quali già in parte sviluppate<sup>7</sup>) si possono riassumere nei punti seguenti:

- delineare i tratti tipici della scrittura delle giovani generazioni in contesto scolastico su diversi livelli di analisi (dall'ortografia alla testualità), sulla base del più grande corpus di testi scolastici finora raccolto nella Svizzera Italiana;

<sup>4</sup> Sono a disposizione i seguenti metadati: sesso, ordine scolastico e classe, ubicazione della scuola, informazioni circa la provenienza linguistica dell'allievo (fornite dai docenti).

<sup>5</sup> La scuola elementare contribuisce con il 28% delle parole (circa 109.000) e la scuola media con il 72% (circa 281.000). Analisi sulla lunghezza media di testi e frasi, sulla leggibilità dei testi e sulla caratterizzazione lessicale sono in corso con Corrigelit (<http://www.corrige.it/>) e con READ-IT (<http://www.italianlp.it/demo/read-it/>). Ad esempio, una prima analisi con READ-IT del subcorpus di 3SE (per la quale si ringraziano molto Felice Dell'Orletta e Giulia Venturi) rileva, tra le altre cose, che i testi dei bambini più piccoli sono composti mediamente da 6,5 periodi, che ogni periodo è composto mediamente da 25 token e che ogni testo conta in media 136 token. Si tratta di dati interessanti da confrontare in prospettiva evolutiva, mettendoli in relazione con altri aspetti della competenza scrittoria al crescere dell'ordine e del grado di scolarità.

<sup>6</sup> Dal 2017 l'équipe di ricerca si avvale di una collaborazione col Dipartimento Tecnologie Innovative della SUPSI (Manno), cosa che permetterà un più approfondito e specializzato supporto informatico per cercare e sperimentare soluzioni ai fini del progetto.

<sup>7</sup> Oltre al già citato Cignetti, Demartini, Fornara (2016a), che raccoglie un'ampia serie di contributi dedicati ai vari oggetti di studio approfonditi nella prima fase della ricerca, si vedano Fornara *et al.* (2015), Cignetti *et al.* (2016b) e Demartini (2016).

- rilevare e studiare gli eventuali influssi della Comunicazione Mediata dal Computer sulla scrittura manuale a scuola, cioè su un tipo di scrittura legata a un mezzo e a un contesto diversi (in cui possono però verificarsi frequenti mescolamenti impropri di codici e di registri, soprattutto nei testi dei ragazzi più grandi);
- costruire il LISSICS (*Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico*), nella duplice prospettiva di affinare l'analisi automatica del tipo di testi in esame e di offrire, poi, uno strumento di consultazione agile, utile sia per gli specialisti, sia per un pubblico più vasto (come può essere quello degli insegnanti);
- fornire indicazioni didattiche mirate a partire dal quadro emerso dalle analisi, in particolare dai principali errori individuati.

Per lavorare a questi obiettivi, al momento il corpus è trattato e analizzato principalmente con l'ausilio di tre software, destinati a operazioni diverse: Atlas.ti, T-LAB e TreeTagger. Il primo strumento è usato per etichettare e codificare fenomeni di interesse presenti nei testi (vale a dire aspetti specifici – a diversi livelli – studiati dai ricercatori, come, per citare qualche esempio, l'uso del *che* relativo, l'uso del *ma*, l'impiego di similitudini e metafore, i titoli dei testi ecc.): il programma permette di consultarli ed esaminarli tutti insieme, mantenendo, al contempo, il riferimento alle produzioni del corpus in cui essi si trovano; il secondo (su cui Lancia 2004) è in uso prevalentemente per effettuare analisi tematiche e alcune analisi lessicali (ad esempio concordanze e contesti elementari); il terzo è quello individuato per la costruzione del lessico: effettua infatti PoS tagging e lemmatizzazione<sup>8</sup>.

### **2.1 L'allestimento del LISSICS: difficoltà e possibilità nel trattamento del corpus**

Come si è accennato, l'italiano dei testi del corpus “DFA-Tiscrivo” presenta una varietà di tratti significativi per chi si occupa di studiare la lingua scritta da un particolare tipo di apprendenti (cioè da bambini e ragazzi prevalentemente italofoeni L1). È fondamentale non perderli, ma, anzi, riuscire a renderne conto sia a fini descrittivi (per tracciare un quadro documentato della scrittura sui banchi di scuola), sia in termini di accesso agevole alle informazioni. Nella pratica, come mostrano anche altre recenti analisi che presentano alcune affinità con quella qui presentata (in particolare Spina 2014 e Barbagli *et al.* 2015), si tratta di un'operazione tutt'altro che semplice.

Per fornire un quadro d'insieme del corpus, il gruppo di ricerca ha preso in considerazione tutti i livelli linguistici, dapprima attraverso studi qualitativi mirati (cioè dedicati a specifici fenomeni ricorrenti nei testi). In questa sede, limitiamo le considerazioni ad alcune questioni pratiche legate all'allestimento del LISSICS, sintetizzando così le principali difficoltà operative: non perdere niente dei dati dal punto di vista grafico e morfologico, e renderli conformi a uno standard completo e funzionale, agevole, poi, per la consultazione (cartacea e digitale). Nei prossimi paragrafi, limitandoci a ciò che è di stretta pertinenza del LISSICS (cioè al PoS tagging e alla lemmatizzazione), ci soffermeremo su alcuni esempi di casi critici e poi in modo specifico sugli errori ortografici, particolarmente caratteristici e significativi nel corpus (pertanto si è scelto di non normalizzarli, bensì di tenerne traccia).

<sup>8</sup> Al momento si stanno testando anche altri tagger per valutarne le prestazioni.



## 2.1.1. Alcuni esempi di casi critici

Il software attualmente in uso per annotare le forme per parti del discorso e ricondurle a lemma è TreeTagger<sup>9</sup>, con i parametri per l'italiano di Marco Baroni. Il primo output del programma, cioè la prima restituzione del corpus (il cui testo si può leggere nella colonna di sinistra) con etichette grammaticali e riconduzione a lemma, si presenta come nelle tabelle seguenti (Tabb. 2 e 3):

| Forma        | PoS Tag   | Lemma        |
|--------------|-----------|--------------|
| rimasimo     | VER:fin   | <unknown>    |
| solo         | ADV       | solo         |
| in           | PRE       | in           |
| pantaloncini | NOUN      | pantaloncino |
| corti        | ADJ       | corto        |
| .            | SENT      | .            |
| Ci           | CLI       | ci           |
| siamo        | AUX:fin   | essere       |
| rimasti      | VER:ppast | rimanere     |
| un'ora       | VER:fin   | <unknown>    |
| in           | PRE       | in           |
| piscina      | NOUN      | piscina      |

Tab. 2 - La prima restituzione di TreeTagger: esempi di "unknown".

|       |          |       |
|-------|----------|-------|
| Io    | PRO:pers | io    |
| avevo | VER:fin  | avere |
| paura | NOUN     | paura |
| ma    | CON      | ma    |
| non   | NEG      | non   |

<sup>9</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>, Schmid 1994.

|           |              |          |
|-----------|--------------|----------|
| osavo     | VER2:fin     | osare    |
| dirlo     | VER:infi:cli | dire     |
| perché    | WH           | perché   |
| mi        | CLI          | mi       |
| avrebbero | AUX:fin      | avere    |
| presa     | VER:ppast    | prendere |
| in        | PRE          | in       |
| giro      | NOUN         | giro     |

Tab. 3 – La prima restituzione di TreeTagger: l’etichetta *WH* e le espressioni polirematiche.

Questi stralci mostrano alcuni casi rispetto ai quali è stato necessario operare delle scelte e intervenire, quando possibile, in fase di input del corpus, affinando il processo di tokenizzazione (ad esempio delle parole apostrofate). La tabella 2 mostra infatti due casi di “unknown”, cioè di forme di fronte alle quali il programma non riesce a fornire una risposta e commette errori: in un caso, si tratta di un passato remoto errato, *rimasimo* (che il software etichetta correttamente ma non lemmatizza), nell’altro, dell’elisione corretta *un’ora*. La tabella 3 mostra invece un caso di etichettatura “WH” (*Wh words*) applicata in modo non pertinente (qui *perché* è una congiunzione causale, e non introduce un’interrogativa) e una polirematica (*avrebbero presa in giro > prendere\_in\_giro*) che, nell’idea di lessico che si vorrebbe realizzare, andrebbe considerata come un’unica unità polilessicale, e che invece il software tratta parola per parola, seguendo un altro criterio.

L’individuazione e la tipologizzazione di casi critici ricorrenti come quelli citati, rispetto ai quali non vi è una restituzione corretta di TreeTagger (o non pienamente idonea al risultato finale desiderato), rende necessario progettare misure correttive. Al momento, si stanno valutando e testando le seguenti possibilità: orientare da subito la lemmatizzazione agli obiettivi finali (ad esempio a livello di identificazione delle polirematiche, che possono essere almeno in parte individuate modificando i parametri del software); procedere con il controllo manuale dell’annotazione di una parte di corpus, in modo da avere un gold standard con cui confrontare le prestazioni di tagger diversi; provare ad automatizzare quanto più possibile l’azione di verifica della distanza tra forme grafiche errate e forme corrette, associando al programma degli algoritmi di controllo<sup>10</sup>. Quest’ultimo intervento sarebbe particolarmente efficace per migliorare il trattamento degli errori ortografici.

### 2.1.2 Un caso specifico: gli errori ortografici nel processo di lemmatizzazione

Gli errori di ortografia sono estremamente diffusi nel corpus “DFA-Tiscrivio”, com’è lecito aspettarsi dal tipo di testi che lo compongono, e rappresentano uno scoglio non semplice da superare nell’allestimento del LISSICS. Infatti, essi sono un caso esemplare di contrasto,

<sup>10</sup> Che considerino per esempio la *distanza di Levenshtein* o di *editazione* (l’unità di misura che considera il numero minimo di modifiche elementari che consentono di trasformare una stringa in un’altra, cfr. ad esempio Naldi 2014: 46-49).

significativo e complesso, fra le possibilità di trattamento automatico del linguaggio e la realtà linguistica, cioè fra lo strumento e il dato. Anzi, sarebbe più esatto dire tra i dati e *gli* strumenti, intendendo con “strumenti” sia quelli usati in fase di analisi, sia il prodotto finale (lo strumento di consultazione che, si auspica, verrà ricavato dal trattamento dei dati).

Errori ortografici come *stó, fù, pò, finche, iniziato, soprattutto, pichiarci, gnente* ed errate segmentazioni del tipo *linsegnamento, laveva, perterra, glie lo dato, emmezo, or'mai* sono solo pochissimi esempi delle molte varianti che caratterizzano il materiale in analisi. Si tratta di forme di indubbio interesse e valore per uno studio che si occupa dei tratti dell'italiano scritto dei bambini e dei ragazzi (prevalentemente italofofoni), forme che sovente rivelano con evidenza alcune radicate difficoltà nell'acquisizione e nell'affinamento del codice scritto: una grafia come *ce lo* per *ce l'ho* adombra ad esempio un profondo problema di riflessione grammaticale<sup>11</sup> (su questo tipo di errori, detti *ortografico-morfologici*, cfr. Cignetti e Demartini 2016: 75-76), cui conseguono, nella costruzione di uno strumento come il LISSICS, problemi di attribuzione dell'esatta parte del discorso e di lemmatizzazione. Come mantenere la forma grafica (*lo*) e, allo stesso tempo, ricondurre correttamente a lemma (“*lo*” = *l'ho* = *lo* + *avere*)? E lo stesso vale per un caso di concrezione dell'articolo come quello nella tabella 4 (*laveva*):

|            |           |            |
|------------|-----------|------------|
| il         | ART       | il         |
| mio        | DET:poss  | mio        |
| allenatore | NOUN      | allenatore |
| laveva     | VER:fin   | <unknown>  |
| sgridato   | VER:ppast | sgridare   |

Tab. 4 – L'univerbazione errata *laveva* e la risposta del software.

Com'è lecito aspettarsi, TreeTagger non può gestire automaticamente un'univerbazione simile, che andrebbe correttamente rappresentata secondo questa intenzione<sup>12</sup>, conservando la forma realmente scritta ma segmentandola anche nelle parole (clitico + verbo) che la compongono:

```
1-2 laveva
  1  l      > lo
  2  aveva  > avere
```

Due parole (1 e 2) rappresentate in un unico token grafico improprio: si tratta di una delle sfide sottese alla realizzazione di un lessico come quello qui presentato, interessato al mantenimento del repertorio di forme errate; una sfida che mette in luce come le parole vadano intese sia nella loro componente grafica (parola come sequenza di caratteri, per cui *laveva* è una sola parola), sia in quella analitica profonda (per cui *laveva* sono due parole). Nessuna di queste

<sup>11</sup> Non è sempre semplice distinguere e rubricare un errore come meramente grafico oppure, invece, grammaticale in senso esteso. Pur non ambendo, per ora, a una tipologizzazione estremamente dettagliata degli errori, si tiene come riferimento (per l'affinità del materiale in analisi) l'elenco di errori di Barbagli *et al.* (2016: 93).

<sup>12</sup> Il modello di rappresentazione di riferimento è il formato CoNLL-U di Universal Dependencies (<http://universaldependencies.org/format.html>).

due componenti può essere lasciata da parte in un lavoro il cui fine principale è orientato all'individuazione delle difficoltà degli scriventi.

Analogamente, anche errori ortografici come quelli prima citati (uso delle doppie lettere, accenti sbagliati, *h* mancanti o in eccesso ecc.) compromettono il trattamento automatico dei dati, inducendo il software a restituire un “unknown” accanto alla parola contenente un errore (Tab. 5).

|       |          |           |
|-------|----------|-----------|
| e     | CON      | e         |
| anchè | VER:fin  | <unknown> |
| mio   | DET:poss | mio       |
| papà  | NOUN     | papà      |

Tab. 5 – L'accento improprio su *anchè* induce in errore il tagger e causa la mancata lemmatizzazione.

È proprio nel caso dei moltissimi errori grafici come questo, per i quali la revisione manuale rischia di essere non solo molto lunga, ma anche passibile di imprecisioni, che si mostrerebbe particolarmente utile e fruttuoso automatizzare quanto più possibile la riconduzione al lemma di riferimento di forme relativamente poco distanti da quella corretta (*anchè-anche, senò-se no, picchiarci-picchiarci* ecc.).

### 3. Le principali tipologie di errore ortografico nel corpus DFA-TIscrivo

Al fine di illustrare le possibilità di estrazione di specifiche tipologie di errore offerte dal corpus “DFA-TIscrivo”, mostreremo di seguito alcuni esempi relativi al livello dell'ortografia<sup>13</sup>.

In primo luogo, in tutti e quattro i sottocorpora (3a e 5a SE, 2a e 4a SM) sono stati riscontrati numerosi errori d'uso dell'accento diacritico sui monosillabi; entro questa categoria, vale da esempio l'errore di accento sulla *e* con valore copulativo, diffuso nei sottocorpora di SE ma attestato anche nel sottocorpus di 4a SM:

- (1) il primo tiro l'ho fatto io però non *e* stato un bel tiro. [3a SE]
- (2) *e* stato un addio piuttosto doloroso. [4a SM]

Ben distribuiti sono anche gli errori relativi ad altri monosillabi con accento diacritico, come *là ~ la, sé ~ se, sì ~ si, dà ~ da* o *lì ~ li*, quest'ultimo presente nell'esempio (3):

- (3) Un giorno sono andato a un pranzo, e *li* c'era un fiume. [3a SE]

Ampiamente attestato è anche l'impiego dell'accento sui monosillabi che non ne richiedono mai la presenza, e anche in questi casi la distribuzione si concentra soprattutto nei sottocorpora di SE, benché non manchino occorrenze in quelli di SM:

- (4) 2 settimane *fà* io e la mia clase siamo andati allo stadio di Cornaredo. [3a SE]
- (5) Un paio di giorni *fà* ho fatto una gara di ginnastica. [5a SE]

<sup>13</sup> Per approfondimenti circa i dati qui sinteticamente presentati, cfr. Cignetti (2016).

- 
- (6) l'anno prossimo giocherò con i due difensori di due anni *fà* che hanno lasciato il Lugano. [4a SM]
  - (7) terra ovunque e qualche filetto verde *quà* e là. [2a SM]
  - (8) Quando sono arrivata *quà* non sapevo niente. [4a SM]
  - (9) A *mè* non è mai capitato di avere veri problemi con gli amici. [2a SM]
  - (10) non *stò* qui ha raccontare tutto perché sono veramente tante cose. [3a SE]

Per quanto riguarda gli errori d'impiego dell'apostrofo, risultano invece attestate in tutti i 4 sottocorpora le forme *po* e *pò* e le forme *qual'è* e *qual'era*:

- (11) la strada era un *po* in pendenza. [4a SM]
- (12) ogni minuto si fermava per chiedere *qual'era* la nota. [3a SE]

Numerosi sono anche i casi di apostrofo dopo l'articolo indeterminativo maschile *un*:

- (13) Il custode era *un'uomo* sulla cinquantina di statura media, *un'uomo* molto severo e specialmente cattivo con i bambini e con i ragazzi. [2a SM].

Inoltre, limitatamente alla SE, si contano anche alcuni casi di apostrofo dopo *ad*:

- (14) *Ad' un certo punto* mia sorella disse. [3a SE]

In merito all'uso della *h*, è emersa la tendenza all'impiego di più grafie alternative per il verbo *avere* (sono per esempio attestate in 3a SE le forme *ò*, *o*, *à*no e *anno*):

- (15) Quindi *o preso* l'ancora e sono risalito su in cima. [3a SE]
- (16) Adesso *ò imparato* che se passo nelle scorciatoie che non so bene non ci devo andare. [3a SE]

E anche di questi errori non mancano i casi nei sottocorpora di SM:

- (17) non bisogna mentire perché le bugie *anno* le gambe corte. [2a SM]

Né mancano gli impieghi di *h* in eccesso (es. 18) o gli errori dovuti alla collocazione del digramma *ch* prima di *a* e di *o* (ess. 19-21):

- (18) Grazie *ha* questo episodio ho imparato che non bisogna mai sottovalutare. [5a SE]
- (19) E loro mi hanno detto di *giochare* a calci di rigore. [5a SE]
- (20) ho giocato a calcio con dei miei amici, nei *parcho* giochi dei pallazzi. [5a SE]
- (21) le ossa mi dolevano e la testa sembrava che dovesse *schoppiarmi*. [4a SM]

Una categoria di errore diffusa quasi esclusivamente nel sottocorpus di 3 SE è invece l'errata segmentazione delle parole, il cui tipo più frequente riguarda l'errata discrezione dell'articolo, tale da dare luogo a forme come *linsegnamento*, *lostesso* e *laltro*:

- (22) E se non rispetto *linsegnamento* rischio di perdermi ancora. [3a SE]
- (23) con la porta ben chiusa ma sentivo *lostesso* mia mamma che mi diceva: smettila! [5a SE]
- (24) se te lo vedi sbalzare fuori da un angolo lo spavento te lo prendi *lostesso*. [4a SM]
- (25) Passammo ai rigori un rigore dopo *laltro* lui inizio tiro palo gol. [3a SE]

Sono inoltre attestati esempi di errata fusione dei costituenti di sintagma, come *perquello e dasola*:

- (26) lei era intenta a giocare *perquello* non mi ascoltò. [3a SE]
- (27) C'era un giardino medio, per andare con il bus *dasola*. [3a SE]

Altra categoria di errore molto comune riguarda l'impiego delle *doppie*, che può realizzarsi come errato scempiamento (tra gli esempi le forme *abiamo*, *arabbiata*, *tapeto*, *arivata*: cfr. ess. 28-31) o come errato raddoppiamento (ess. 32-34):

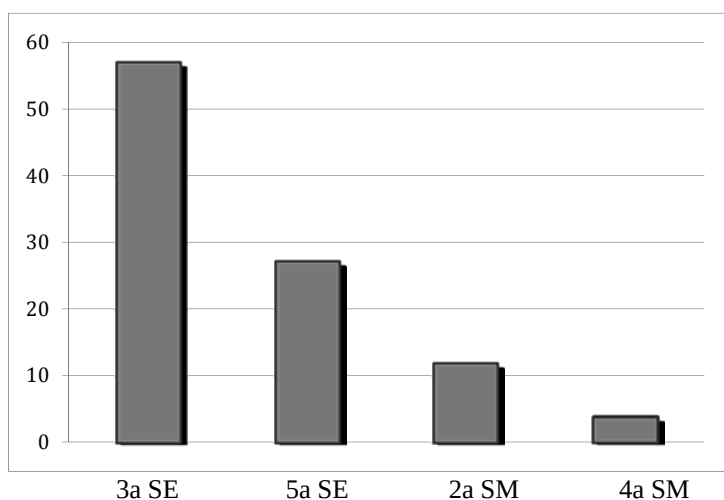
- (28) siamo andati al Penz a fare un pic-nic *abiamo* grigliato la carne, le salciccie, i sevelà arancioni e bianchi (3a SE)
- (29) Mi ero *arabbiata* tanto e buttavo in giro le cose perché ero furiosa. [3a SE]
- (30) Pero sto sempre sdraiata sul *tapeto* perché mi fa sempre calmare. [3a SE]
- (31) Quando sono *arivata* a scuola, la maestra mi ha *avertito* di non ritornare a casa con il motorino. [5a SE]
- (32) Io mi sono *comportatto* bene solo che lui. [3a SE]
- (33) A scuola feci *lezzione* bagnando tuti i fogli che passavano sotto il nio naso. [5a SE]
- (34) L'amica di mia madre mi ha raccontato di una vicenda molto *corraggiosa*. [2a SM]

Questa tipologia di errore ortografico è evidentemente dovuta all'influenza della pronuncia regionale ticinese, assimilabile, per questo tratto, a quella dell'Italia settentrionale<sup>14</sup>.

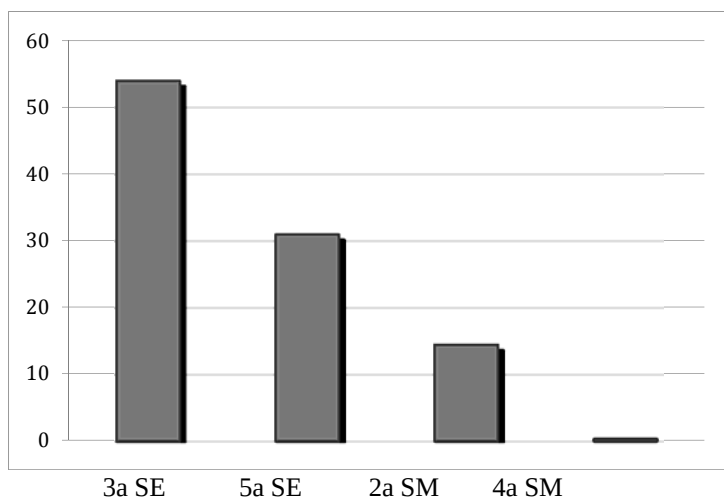
#### 4. Distribuzione degli errori ortografici nei diversi livelli scolastici

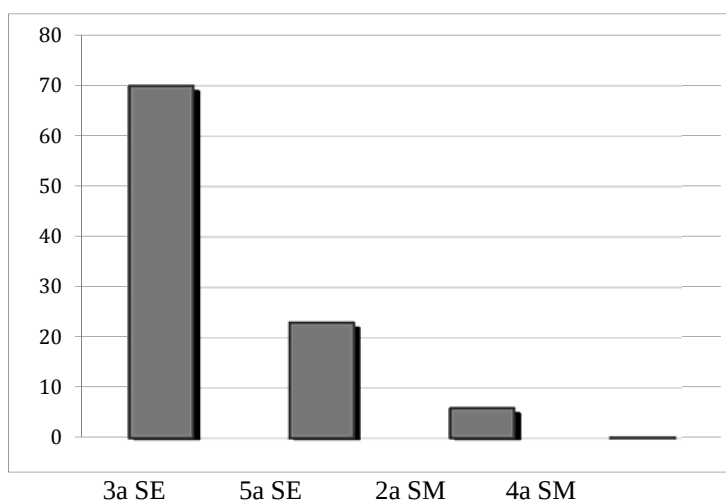
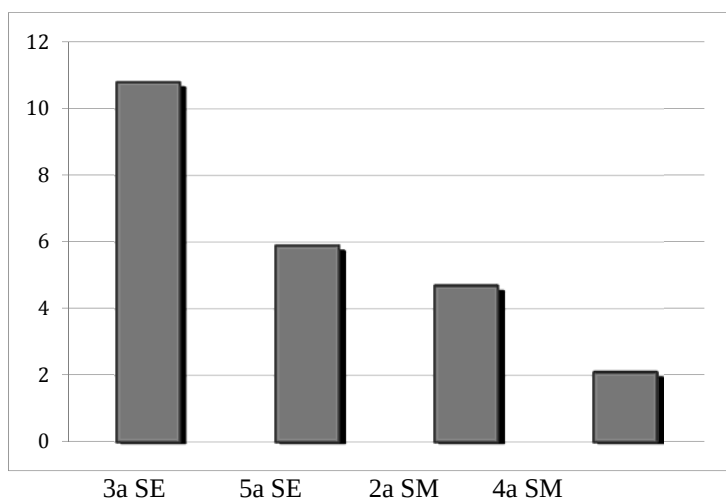
Proponiamo di seguito i risultati relativi ad alcune voci-campione in merito alla distribuzione delle diverse categorie di errore nei quattro sottocorpora. Un primo esempio riguarda l'errore di accento nelle parole *perché*, *però* e *può*. In questo caso, i dati, normalizzati, mostrano che oltre la metà degli errori si concentra nel corpus di 3a SE, con una riduzione piuttosto graduale nelle classi successive (Tab. 5).

<sup>14</sup> Cfr. De Blasi (2014: 65), dove, a proposito dei "fenomeni fonetici presenti in generale in tutta l'Italia settentrionale", si cita "lo scempiamento delle consonanti intense (per es. *afetare* invece di *affettare*)". Sul problema della norma ortografica, cfr. Cignetti & Demartini (2016).

Tab. 5 – Occorrenze delle forme *\*perche, \*pero, \*puo*.

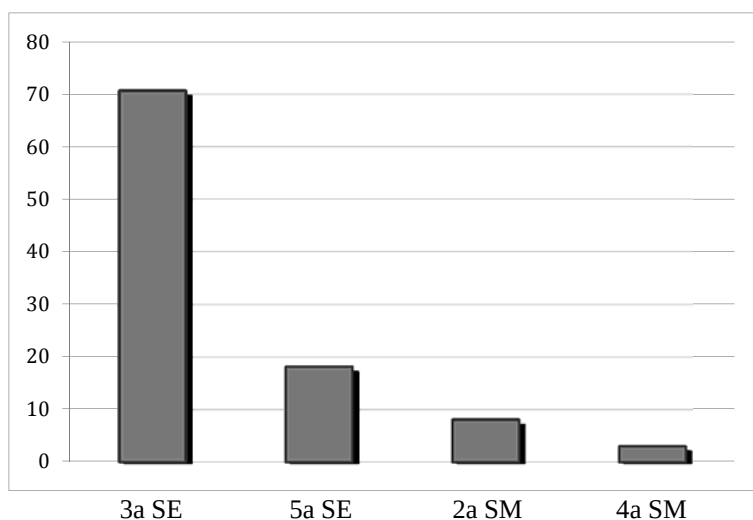
I dati disaggregati permettono di illustrare tale progressione con maggiore precisione, mostrando come la riduzione sia graduale con le voci *perché* e *può*, mentre nel caso di *può* il miglioramento nella fase di passaggio tra SE e SM non appare completamente soddisfacente, un fatto che può essere interpretato come segnale della necessità di un migliore coordinamento tra i diversi cicli scolastici (Tabb. 6-8):

Tab. 6 – Occorrenze della forma *\*perche*

Tab. 7 – Occorrenze della forma *\*pero*.Tab. 8 – Occorrenze della forma *\*puo*.

Ulteriori dati esemplificativi di questa fase della ricerca, relativi agli accenti in eccesso su alcuni monosillabi, sono riprodotti nelle Tab. 9 e 10: nella prima, i dati aggregati di dieci forme campione consentono di osservare una riduzione significativa degli errori tra la 3a SE e la 5a SE, mentre nella seconda l'osservazione della distribuzione degli errori nei diversi cicli di studio permette di comprendere meglio il fenomeno.



Tab. 9 – Occorrenze delle forme *pò, sù, stò, chè, dò, ò, lò, và, hà, nò*.

|     | 3 SE | 5 SE | 2 SM | 4 SM |
|-----|------|------|------|------|
| pò  | V    | V    | V    | V    |
| sù  | V    | V    | V    | -    |
| stò | V    | V    | V    | -    |
| chè | V    | V    | -    | -    |
| dò  | V    | V    | -    | -    |
| ò   | V    | -    | -    | -    |
| lò  | V    | -    | -    | -    |
| và  | V    | -    | -    | -    |
| hà  | V    | -    | -    | -    |
| nò  | V    | -    | -    | -    |

Tab. 10 – Distribuzione delle forme *pò, sù, stò, chè, dò, ò, lò, và, hà, nò*.

Se in 4a SM i problemi (tra quelli selezionati in queste dieci forme) si limitano dunque, sostanzialmente, al solo *po'* accentato, in 2a SM risultano invece attestate, oltre a numerosi casi di *pò*, anche diverse occorrenze di *sù* e di *stò*; in 5a SE compaiono molti casi di *dò* e di *chè* (pronomi e complementatore) e in 3a SE risultano molto comuni altri problemi d'uso dell'accento, come mostra la presenza delle forme *ò, lò, và, hà, nò*.

## 5. Conclusioni

Le osservazioni sin qui raccolte permettono di cogliere le potenzialità euristiche offerte dalla realizzazione del LISSICS nel campo del trattamento automatico del linguaggio; contestualmente, fanno emergere le potenzialità diagnostiche dello strumento, in quanto anche questi soli primi dati offrono molti elementi utili alla comprensione delle reali competenze ortografiche degli studenti della scuola dell'obbligo del Canton Ticino, che spesso si discostano non poco dagli obiettivi dichiarati, forse un po' ottimisticamente, nei piani di studio ufficiali. La consapevolezza della

reale distribuzione degli errori tra i diversi livelli scolastici resa possibile da questo studio fornirà informazioni preziose per l'elaborazione di proposte didattiche coerenti con lo sviluppo delle singole competenze degli apprendenti.

## BIBLIOGRAFIA

- Barbagli, A., Lucisano, P., Dell'Orletta, F., Montemagni, S., Venturi, G. (2015), *CIItA: un Corpus di Produzioni Scritte di Apprendenti l'Italiano L1. Annotato con Errori*, in Bosco, C., Tonelli, S., Zanzotto, F.S. (a c. di), *Proceedings of the Second Italian Conference on Computational Linguistics, CLiC-it 2015*. Torino, Accademia University Press, pp. 31-35.
- Barbagli, A., Lucisano, P., Dell'Orletta, F., Montemagni, S., Venturi, G. (2016), *CIItA: an L1 Italian Learner Corpus to Study the Development of Writing Competence*, in Calzolari, N. et al. (a c. di), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, ELRA, pp. 88-95.
- Berruto, G. (2010), "Italiano standard", in *Enciclopedia dell'Italiano Treccani*, diretta da Raffaele Simone, Roma, Istituto dell'Enciclopedia, pp. 729-731.
- Cignetti, L. (2016), *Tipologie e frequenza degli errori di ortografia nella scrittura degli apprendenti*, in Cignetti, L., Demartini, S. e Fornara, S. (a c. di) (2016a), pp. 19-36.
- Cignetti, L., Demartini, S. (2016), *L'ortografia*, Roma, Carocci.
- Cignetti, L., Demartini, S., Fornara, S. (2016a) (a c. di), *Come TIscrivo? La scrittura a scuola tra teoria e didattica*, Roma, Aracne.
- Cignetti, L., Demartini S., Fornara, S. (2016b), *Il lessico di TIscrivo. Caratterizzazione del vocabolario e osservazioni in prospettiva didattica*, in *Atti del Workshop SLI-Giscel* svoltosi durante il XLVII Congresso Internazionale SLI 2013, "Sviluppo della competenza lessicale. Acquisizione, apprendimento, insegnamento", Salerno, 27 settembre 2013, pp. 43-60.
- Demartini, S. (2016), *La grammatica nei testi scritti a scuola. Rilievi dall'analisi del corpus TIscrivo*, in Benedetti M., Bruno C., Dardano P., Tronci L., (a c. di) *Grammatiche e grammatici: teorie, testi e contesti*, Atti del XXXIX Convegno della Società Italiana di Glottologia, Roma, pp. 197-202.
- De Blasi, N. (2014), *Geografia e storia dell'italiano regionale*, Bologna, Il Mulino.
- De Mauro, T. (2005), *La fabbrica delle parole. Il lessico e problemi di lessicologia*, Torino, UTET.
- De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. (1993), *LIP. Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.
- Fornara, S., Cignetti, L., Demartini, S., Guaita M., Moretti A. (2015), *Costruzione del testo e punteggiatura tra norma, uso e didattica negli elaborati del corpus TIscrivo*, in "Bulletin Suisse de Linguistique Appliquée", Actes du colloque VALS-ASLA 2014 (Lugano, 12-14 février 2014), No spécial 2015, t. 1, pp. 71-94.
- Lancia, F. (2004), *Strumenti per l'analisi dei testi. Introduzione all'uso di T-LAB*, Milano, FrancoAngeli.
- Naldi, M. (2014), *Traduzione automatica e traduzione assistita*, Bologna, Esculapio.
- Pandolfi, E.M. (2009), *LIPSI. Lessico di frequenza dell'italiano parlato nella Svizzera italiana*, Bellinzona, Osservatorio Linguistico della Svizzera Italiana.
- Marconi, L., Ott M., Pesenti E. (1994), *Lessico elementare. Dati statistici sull'italiano scritto e letto dai bambini delle elementari*, Bologna, Zanichelli.
- Schmid, H. (1994), *Probabilistic Part-of-Speech Tagging Using Decision Trees*, in *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>).
- Spina, S. (2014), *Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione*, in Basili, R., Lenci, A., Magnini B. (a c. di), *Proceedings of the First Italian Conference on Computational Linguistics, CLiC-it 2014*, Pisa, Pisa University Press, pp. 354-359.

**LUCA CIGNETTI** • Teacher-researcher in Didactics of Italian language at the University of Applied Sciences and Arts of Southern Switzerland (SUPSI). His main fields of research are Didactics of Italian language, Didactics of writing and Textual linguistics. His recent publications include *L'ortografia* (Roma, Carocci, 2016; with S. Demartini), *Il piacere di scrivere. Guida all'italiano del terzo millennio* (Roma, Carocci, 2014; with S. Fornara) and *L'Inciso. Natura linguistica e funzioni testuali* (Alessandria, Edizioni dell'Orso, 2011).

**E-MAIL** • [luca.cignetti@supsi.ch](mailto:luca.cignetti@supsi.ch)

**SILVIA DEMARTINI** • Researcher in Didactics of Italian language at the University of Applied Sciences and Arts of Southern Switzerland (SUPSI). Her main fields of research are Didactics of Italian language, History of Italian grammar and Children's writing competences. Her recent publications include *L'ortografia* (Roma, Carocci, 2016; with L. Cignetti), *Grammatica e grammatiche in Italia nella prima metà del Novecento. Il dibattito linguistico e la produzione testuale* (Firenze, Cesati, 2014) and *La punteggiatura dei bambini. Uso, apprendimento e didattica* (Roma, Carocci, 2013; with S. Fornara).

**E-MAIL** • [silvia.demartini@supsi.ch](mailto:silvia.demartini@supsi.ch)



# VERSO UN LESSICO DI VALENZA DEL LATINO EMPIRICAMENTE MOTIVATO

---

Berta GONZÁLEZ SAAVEDRA, Marco PASSAROTTI

**ABSTRACT.** Despite a centuries-long tradition in lexicography, Latin lacks state-of-the-art computational lexical resources. This situation is strictly related to the still quite limited amount of linguistically annotated textual data for Latin, which can help the building of new lexical resources by supporting them with empirical evidence. However, projects for creating new language resources for Latin have been launched over the last decade to fill this gap. In this paper, we present Latin Vallex, a valency lexicon for Latin built in mutual connection with the semantic and pragmatic annotation of two Latin treebanks featuring texts of different eras. On the one hand, such a connection between the empirical evidence provided by the treebanks and the lexicon allows to enhance each frame entry in the lexicon with its frequency in real data. On the other hand, each valency-capable word in the treebanks is linked to a frame entry in the lexicon.

**KEYWORDS.** Valence, Latin, Treeban

## 1. Introduzione

Nonostante una secolare tradizione lessicografica, la lingua latina manca ancora di risorse lessicali di tipo computazionale aggiornate allo stato dell'arte. Ciò è strettamente connesso alla limitata disponibilità di corpora testuali latini annotati linguisticamente, sulla cui base empirica possano essere costruite nuove risorse lessicali. Tuttavia, una serie di progetti mirati allo sviluppo di treebank a dipendenze per il latino è stata avviata nel corso dell'ultimo decennio. Egualmente, ha preso avvio la realizzazione di risorse lessicali fondamentali come *Latin WordNet* (Minozzi, 2010).

Questo articolo presenta Latin Vallex, un lessico di valenza per il latino costruito in connessione con l'annotazione semantico-pragmatica di due treebank latine comprensive di testi di epoche e generi diversi.

L'articolo è organizzato nel modo seguente. La sezione 2 riporta lo stato dell'arte relativo ai lessici di valenza, concentrandosi particolarmente su quelli per la lingua latina. La sezione 3 presenta *Latin Vallex*, descrivendone la struttura delle entrate lessicali e l'interrogazione dei dati. La sezione 4 conclude il lavoro e abbozza i prossimi sviluppi della risorsa.

## 2. Valenza e lessici

La nozione di valenza è generalmente intesa come il numero di complementi obbligatori richiesti da una parola: essi sono usualmente nominati 'argomenti', mentre i complementi non obbligatori sono detti 'aggiunte', o 'satelliti'. Il precursore dell'idea moderna di valenza è

considerato essere Lucien Tesnière in virtù del suo testo *Eléments de syntaxe structurale*, pubblicato postumo nel 1959 (Tesnière, 1959).

Approcciare la semantica lessicale attraverso il concetto di valenza è una pratica diffusa nella ricerca linguistica. Esso è, infatti, presente in diversi contesti teorici, tra i quali spicca la *Frame Semantics* di Charles Fillmore (1982), in base alla quale il significato di alcune parole può essere compreso appieno solo conoscendo gli elementi che fanno parte del *frame* evocato dalla parola stessa. Alcuni di questi elementi sono obbligatori (*core frame element*), mentre altri sono occasionali (*not core frame element*)<sup>1</sup>.

In base al numero dei propri argomenti, le parole sono considerate essere zerovalenziali (ad esempio, *piovere*), monovalenziali (*camminare*), bivalenziali (*mangiare*), trivalenziali (*dare*), tetravalenziali (*spostare*) etc. Le posizioni argomentali possono venire arricchite con ruoli semantici, ossia etichette che specificano la relazione semantica che intercorre tra l'argomento e la parola di cui quest'ultimo è complemento (obbligatorio). Ad esempio, nel caso del verbo trivalenziale *dare* i ruoli semantici dei tre argomenti sono rispettivamente Agente (1), Paziente (2) e Destinatario (3): "(1) dà (2) a (3)".

I criteri a supporto della distinzione tra argomenti e aggiunte non sono rigidamente definiti e unanimemente accettati (al punto che la distinzione stessa è talvolta messa in dubbio); ciò richiede che essi vengano esplicitati ogni volta che la nozione di valenza è utilizzata e applicata a dati empirici (si veda 3.1).

Sono oggi disponibili per molte lingue descrizioni del lessico fondate sulla valenza. Esse sono rappresentate da risorse lessicali che recitano un ruolo importante nel trattamento automatico del linguaggio (TAL) grazie alla loro ampia applicabilità in ambiti come il *semantic role labeling*, la *word sense disambiguation*, l'acquisizione di restrizioni preferenziali e la realizzazione di trebank (Urešová, 2004).

Così come altre risorse linguistiche, anche i lessici di valenza possono essere costruiti in modalità *intuition-based* o *corpus-driven*, in base all'importanza del ruolo giocato dall'intuizione umana e dall'evidenza empirica estratta da corpora testuali nel corso della realizzazione della risorsa. Ad esempio, lessici come *PropBank* (Kingsbury & Palmer, 2002), *FrameNet* (Ruppenhofer et al., 2006) e *PDT-Vallex* (Hajič et al., 2003) sono stati inizialmente sviluppati in modalità *intuition-based* per venire successivamente controllati e perfezionati attraverso il confronto con dati estratti da corpora. Esempi di lessici acquisiti automaticamente dall'evidenza testuale sono, invece, *VALEX* (Korhonen et al., 2006) e *LexShem* (Messiant et al., 2008).

Mentre sono numerosi i lessici di valenza compilati per le lingue moderne, molto lavoro resta da fare per la realizzazione di risorse simili per le lingue antiche e, in particolare, per il latino e il greco. In merito al latino, Happ riporta una lista di verbi associati alle proprie valenze (Happ, 1976: 480-565). Bamman & Crane (2008), invece, descrivono un *dynamic lexicon* estratto automaticamente dalla Perseus Digital Library utilizzando la Latin Dependency Treebank (Bamman & Crane, 2006) come evidenza su cui addestrare un PoS tagger e un parser a dipendenze. Questa risorsa associa a ciascuna entrata lessicale informazione qualitativa e quantitativa in merito ai suoi modelli di sottocategorizzazione e alle sue restrizioni preferenziali. Infine, *IT-VaLex* (McGillivray & Passarotti, 2009) è un lessico di sottocategorizzazione le cui entrate (verbal) sono indotte automaticamente a partire dal livello di annotazione sintattica della *Index Thomisticus* Treebank (Passarotti, 2011). La medesima struttura di *IT-VaLex* è ripresa da un lessico sviluppato sulla base della Latin Dependency Treebank e descritto da McGillivray (2013: 31-60).

<sup>1</sup> Un'ampia descrizione del lessico basata sulla *Frame Semantics* è fornita dalla risorsa lessicale *FrameNet* sviluppata alla University of California (Berkeley, USA) dal gruppo di ricerca che faceva capo a Charles Fillmore: <https://framenet.icsi.berkeley.edu/fndrupal/>.

### 3. Latin Vallex

#### 3.1 La struttura di Latin Vallex

*Latin Vallex* (LV) è stato realizzato congiuntamente allo sviluppo del livello di annotazione semantico-pragmatica di due treebank latine a dipendenze: la *Index Thomisticus* Treebank (IT-TB), che include testi di Tommaso d’Aquino, e la Latin Dependency Treebank (LDT), che riporta estratti di testi (sia in prosa che in poesia) dell’età classica. Ciascuna parola valenziale che occorre nella porzione delle due treebank annotata a livello semantico-pragmatico è associata a una *frame entry* in LV.

La struttura di LV richiama quella del lessico di valenza per la lingua ceca *PDT-Vallex*, prodotto nel contesto teorico della *Functional Generative Description* (FGD; Sgall et al., 1986). La FGD è la teoria linguistica che motiva lo stile di annotazione semantico-pragmatica anche delle treebank latine, corrispondente al cosiddetto livello ‘tectogrammaticale’ della *Prague Dependency Treebank* del ceco (PDT). Questo livello viene realizzato a partire da uno precedente di tipo sintattico, chiamato ‘analitico’, e include l’annotazione dei ruoli semantici e la risoluzione delle ellissi e delle anafore/catafore. Il *dialogue test* di Panevová (1974-1975) e i criteri descritti da Mikulová *et alii* (2005: 100-102, 116-162) sono utilizzati per distinguere tra argomenti e aggiunte. Sia il livello analitico che quello tectogrammaticale sono graficamente rappresentati attraverso alberi a dipendenze.

A livello macroscopico, il lessico è diviso in entrate lessicali. Un’entrata lessicale consiste in una sequenza di *frame entry*, ciascuna delle quali tende a corrispondere a uno dei sensi della parola in questione. La *frame entry* contiene una descrizione del *valency frame* e dei suoi attributi (*frame attribute*). Un *valency frame* è una sequenza di posizioni argomentali (*frame slot*), ciascuna delle quali rappresenta un complemento della parola ed è associata ad alcuni suoi tratti morfologici di superficie (con alcune deviazioni: si veda 3.2). Gli attributi corrispondono a nomi di ruoli semantici (chiamati *functor* in FGD) utilizzati per esprimere i tipi di relazioni semantiche che intercorrono tra la parola e i suoi complementi.

La struttura di un’entrata di LV può essere riassunta come segue:

Nome dell’entrata lessicale (lemma) – PoS

- Frame Entry 1:
  - Valency Frame:
    - Frame slot 1
    - Frame slot *n*
  - Frame Attribute:
    - Functor 1
    - Functor *n*
- Frame Entry *n*:...

I ruoli semantici riportati nelle *frame entry* di LV sono quelli per gli argomenti (chiamati *inner participant*), che, in base alla FGD, corrispondono ai complementi a cui sono assegnati i seguenti *functor*: Agente (ACT[or]), Paziente (PAT[ient]), Destinatario (ADDR[essee]), Risultato (EFF[ect]) e Origine (ORIG[o]). Anche alcune aggiunte (*free modification*) possono rientrare nelle *frame entry*, venendo registrate come posizioni opzionali. L'insieme dei valori per i ruoli semantici utilizzato nella IT-TB e nella LDT è il medesimo descritto nel manuale di annotazione tectogrammaticale della PDT (Mikulová et al., 2005).

La sola differenza esistente tra LV e *PDT-Vallex* è conseguenza del fatto che il cosiddetto *argument shifting* non viene applicato nell'annotazione tectogrammaticale della IT-TB e della LDT. L'*argument shifting* (Mikulová et al., 2005: 103-105) è un criterio utilizzato per determinare il tipo di argomento e, dunque, assegnare il *functor* più confacente alla posizione argomentale in questione: esso stabilisce che al primo argomento va sempre assegnato il *functor* ACT, mentre il secondo argomento riceve sempre il *functor* PAT. Tutti gli altri *functor* argomentali (ADDR, EFF e ORIG) diventano (*to shift*) ACT e PAT nel caso in cui occorranza come primo o secondo argomento<sup>2</sup>.

Ad esempio, in base all'*argument shifting*, se un verbo ha un argomento di tipo ORIG ma non include un PAT nel proprio *frame* argomentale, la posizione del PAT (seconda posizione del *frame*) viene occupata dall'argomento ORIG, a cui viene assegnato il *functor* PAT. Ciò si riflette nei dati di *PDT-Vallex*, in cui infatti non occorrono *frame entry* costituite da due posizioni i cui attributi siano, ad esempio, ACT e ORIG, in quanto all'argomento con attributo ORIG sarebbe assegnato il *functor* PAT proprio in virtù dell'*argument shifting*. Invece, una situazione del genere può accadere in LV, come risultato dell'annotazione tectogrammaticale della IT-TB e della LDT. Ad esempio, l'entrata del verbo *resulto* ("risultare") in LV include una *frame entry* con due posizioni, i cui attributi sono rispettivamente ACT e ORIG.

Un'occorrenza testuale di questa *frame entry* di *resulto* è nella seguente frase tratta dalla IT-TB (*Summa contra Gentiles*, libro 1, capitolo 27, numero 4):

- (1) "ex unione formae et materiae resultat aliquid compositum" ("dall'unione della forma e della materia risulta qualcosa composto").

Nella frase (1), gli argomenti per il verbo *resultat* sono *aliquid* ("qualcosa") ed *ex unione* ("dall'unione"). Alla parola *aliquid* viene assegnato il *functor* ACT, mentre *ex unione* ha *functor* ORIG. Se l'*argument shifting* fosse stato applicato, *ex unione* avrebbe ricevuto *functor* PAT<sup>3</sup>.

Come anticipato, oltre ai *functor* per gli argomenti, anche alcuni *functor* per le aggiunte possono essere presenti nelle *frame entry*. Questi *functor* sono registrati come opzionali. I *functor* opzionali più frequenti nelle *frame entry* sono quelli di tipo spaziale e direzionale, che sono per lo più impiegati per i verbi di movimento (Mikulová et al., 2005: 503-514). Ad esempio, la *frame entry* prototipica del verbo *venio* ("venire") include tre posizioni, i cui *functor* sono rispettivamente ACT, DIR1 (Direzione-Da) e DIR3 (Direzione-A).

<sup>2</sup> Il livello di annotazione tectogrammaticale della IT-TB e della LDT non include l'*argument shifting*, in quanto nelle treebank latine questo livello è inteso come meno direttamente connesso alla struttura sintattica rispetto alla PDT e più orientato all'analisi semantica.

<sup>3</sup> In base a una interpretazione più agentiva o più risultativa del soggetto sintattico del verbo *resulto*, alle sue posizioni argomentali potrebbero venire assegnati rispettivamente i *functor* ACT ed EFF invece che ACT e ORIG. Nell'esempio discusso, *aliquid* riceverebbe *functor* EFF, mentre (*ex*) *unione* sarebbe ACT: *aliquid* sarebbe così inteso come il risultato dell'azione rappresentata dal nome deverbale *unio*. Seppur semanticamente motivata, questa interpretazione scarta, tuttavia, dallo stile di annotazione tectogrammaticale della PDT e ad essa è dunque preferita quella che fa uso dei *functor* ACT e ORIG.



Un altro esempio è l'entrata del verbo *termino* che, in base alla struttura delle entrate lessicali di LV, include due *frame entry*, corrispondenti a due diversi sensi della parola: (a) “marcare il confine di qualcosa” e (b) “limitare qualcosa a qualcos'altro”. La *frame entry* per il primo senso è costituita da un *valency frame* con due posizioni, la prima delle quali è rappresentata da un nome al nominativo (n1) mentre la seconda è un nome all'accusativo (n4). Gli attributi di queste due posizioni sono rispettivamente ACT e PAT. La *frame entry* per il secondo senso, invece, consiste in un *valency frame* con tre posizioni: un nome al nominativo (n1), un nome all'accusativo (n4) e un sintagma preposizionale governato dalla preposizione *in* (in+n4). I *functor* per questi tre argomenti sono rispettivamente ACT, PAT e DIR3.

L'entrata lessicale di *termino* in LV appare, dunque, come segue.

*termino* – V

- Frame Entry 1 (“marcare il confine di qualcosa”):
  - Valency Frame:
    - Frame slot 1: n1
    - Frame slot 2: n4
  - Frame Attribute:
    - Functor 1: ACT
    - Functor 2: PAT
- Frame Entry 2 (“limitare qualcosa a qualcos'altro”):
  - Valency Frame:
    - Frame slot 1: n1
    - Frame slot 2: n4
    - Frame slot 3: in+n4
  - Frame Attribute:
    - Functor 1: ACT
    - Functor 2: PAT
    - Functor 3: DIR3

I tratti morfologici (PoS e caso) riportati nei frame slot risultano dal confronto con l'evidenza testuale fornita dalle due treebank latine sulla cui base LV è costruito.

### 3.2 La realizzazione di Latin Vallex

Ciascuna parola valenziale che gli annotatori incontrano nel corso dell'annotazione tectogrammaticale della IT-TB e della LDT viene collegata a una *frame entry* di LV<sup>4</sup>. Le parole valenziali possono essere verbi (*do* - “dare”), aggettivi (*contrarius* - “contrario”), nomi (*descriptio* - “descrizione”) e avverbi (*similiter* - “similmente”)<sup>5</sup>.

Al momento, LV include 1.373 entrate lessicali e 3.406 *frame entry*: nello specifico, 1.049 verbi (2.903 *frame*), 236 nomi (394 *frame*), 86 aggettivi (106 *frame*) e 2 avverbi (3 *frame*). Queste entrate risultano dall'annotazione tectogrammaticale delle prime 2.000 frasi della *Summa contra Gentiles* di Tommaso d'Aquino (IT-TB), dell'intera *De Catilinae coniuratione* di Sallustio (701 frasi) e di 100 frasi tratte dal *De bello gallico* di Cesare e dalle *Orationes in Catilinam* di Cicerone (LDT)<sup>6</sup>.

Dal momento che la IT-TB e la LDT non sono bilanciate in modo tale da essere sufficientemente rappresentative della lingua latina (o di una sua specifica variante), nel corso della realizzazione di LV alle entrate lessicali realizzate in modalità *corpus-driven* è stata accostata una serie di entrate prodotte per intuizione. Nello specifico, per fini di rappresentatività della risorsa, LV include le entrate di tutte le parole valenziali presenti tra le 1.000 più frequenti della lingua latina riportate da Delatte *et alii* (1981). Benché la maggior parte di queste parole siano già presenti in LV, in quanto risultanti dal lavoro di annotazione tectogrammaticale, 163 di esse non sono state ancora trovate nei testi annotati. Dunque, le entrate per queste parole sono state inserite in LV sulla base dell'intuizione del lessicografo (e non sono, dunque, collegate ad alcuna occorrenza nelle *treebank*), che assegna ad esse quella che è considerata essere la loro *frame entry* prototipica, registrando solo i *functor* per le posizioni argomentali e non anche i tratti morfologici, indisponibili in quanto risultanti dal confronto con l'evidenza fornita dalle *treebank*. Benché la maggior parte delle entrate *intuition-based* di LV riceva una sola *frame entry* (prototipica), ci sono casi in cui più di una *frame entry* viene loro assegnata: in totale, il numero di *frame entry* per queste 163 entrate è 168. Nessuna di queste *frame entry* è connessa ad alcuna occorrenza testuale nelle *treebank* fino al momento in cui gli annotatori non ne incontrino la prima nel corso del lavoro di annotazione: a quel punto, la *frame entry* viene modificata in base all'evidenza empirica, aggiungendo i tratti morfologici.

La figura 1 mostra una porzione dell'albero tectogrammaticale della seguente frase tratta dalla IT-TB (*Summa contra Gentiles*, libro 1, capitolo 5, numero 2):

<sup>4</sup> Essendo le *treebank* e le *frame entry* di LV biunivocamente collegate (tranne che per le 163 entrate prodotte in modalità *intuition-based* descritte più avanti nella sezione), le modifiche eventualmente apportate a una risorsa vengono automaticamente applicate anche all'altra. Una serie di query scritte in linguaggio SQL (*Structured Query Language*) consentono altresì di indurre alcune *frame entry* di LV a partire dall'annotazione tectogrammaticale delle *treebank*. Tuttavia, questo metodo di creazione del lessico di valenza non è esaustivo in quanto ancora manchevole nell'estrazione dell'informazione da strutture orizzontali, come ad esempio le costruzioni coordinate. Le query SQL sono al momento utilizzate per fini di verifica della qualità di LV, confrontando il loro risultato con quanto prodotto dagli annotatori.

<sup>5</sup> L'annotazione tectogrammaticale viene prodotta da due annotatori, che lavorano in fasi diverse. Il primo costruisce l'albero tectogrammaticale di una frase sulla base dell'output di una serie di script di conversione automatica dal livello analitico a quello tectogrammaticale (González Saavedra & Passarotti, 2014). Il secondo annotatore verifica il lavoro del primo. I casi di disaccordo tra le scelte dei due annotatori vengono raccolti e discussi regolarmente, risultando in periodiche revisioni delle regole di annotazione descritte in Mikulová *et alii* (2005).

<sup>6</sup> Le frasi tratte dai testi di Cesare e di Cicerone sono le prime 100 disponibili di essi nella LDT.

- (2) “[...] christianae religioni [...], quae singulariter bona spiritualia et aeterna promittit” (“[...] alla religione cristiana [...], che unicamente promette beni spirituali ed eterni”).

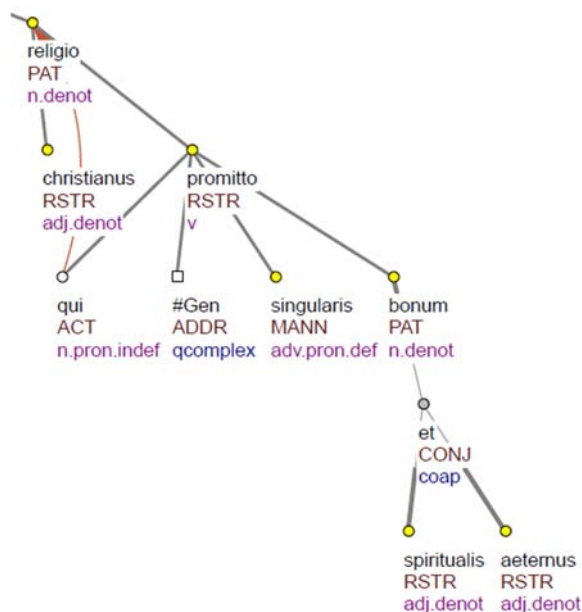


Figura 1: Un albero tectogrammaticale<sup>7</sup>

Nel corso della costruzione della porzione di albero tectogrammaticale riportato nella figura 1, gli annotatori incontrano una occorrenza della parola valenziale *promitto* (“promettere”) e la associano alla corrispondente *frame entry* in LV o costruiscono la *frame entry* ex novo, qualora essa non sia già presente nella risorsa.

La *frame entry* per questa occorrenza di *promitto* include tre posizioni argomentali. La prima è rappresentata da un pronome al caso nominativo (*quae*; nodo *qui*). La seconda è un nome all’ accusativo (*bona*; nodo *bonum*). La terza posizione risulta dalla risoluzione dell’ ellissi di un argomento che non può essere identificato contestualmente ed è, quindi, considerato un argomento “generico” (#Gen): dal momento che questo argomento non ha una realizzazione testuale, ad esso non è associato alcun tratto morfologico (PoS e caso) nel *frame slot*. I *functor* per i tre *slot* sono rispettivamente i seguenti: ACT, PAT e ADDR.

Oltre ai tre nodi che entrano a far parte della *frame entry*, nella porzione di albero riportata nella figura 1, *promitto* governa anche un quarto nodo, che corrisponde alla parola *singulariter* nella frase (nodo *singularis*) e riceve *functor* MANN[er] (Modo). Questo nodo non rientra nella *frame entry* di *promitto* perché il *functor* MANN è assegnato ad aggiunte non riportate nelle entrate lessicali di LV.

Più del 60% delle *frame entry* di LV presentano un *valency frame* con due posizioni argomentali. Nella maggior parte di questi *valency frame*, i ruoli semantici sono ACT e PAT. Il

<sup>7</sup> Nella usuale visualizzazione degli alberi tectogrammaticali, le forme di parola sono sostituite dal corrispondente lemma. Ad esempio, *qui* è il lemma della forma *quae*. Per le specifiche sui *functor* che occorrono negli alberi riportati nelle Figure 1 e 5, si rimanda a Mikulová et alii (2005).

secondo più frequente tipo di *valency frame* in LV (circa il 20% del totale dei *frame*) include tre argomenti. Diversamente dalle *frame entry* bivalenziali, quelle trivalenziali presentano una configurazione piuttosto varia dei ruoli semantici.

La figura 2 rappresenta le *frame entry* trivalenziali di LV attraverso un network indotto automaticamente dalla risorsa<sup>8</sup>. I nodi colorati in rosso sono quelli per i *functor*, mentre quelli in bianco sono per le *frame entry*, il cui nome è formato dal lemma dell'entrata lessicale a cui appartengono e una lettera assegnata alla specifica *frame entry*. Ad esempio, il nodo nominato *amo-a* corrisponde alla *frame entry* 'a' del lemma *amo* ("amare"). Nel network, un nodo per un *functor* è connesso a un nodo per una *frame entry* attraverso un ramo se quel *functor* è presente in almeno una posizione argomentale della *frame entry*.

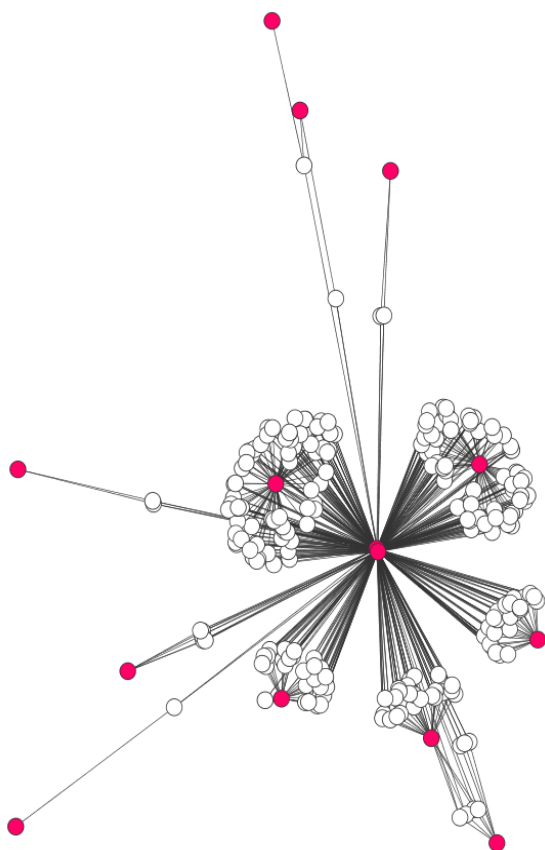


Figura 2: Network delle *frame entry* trivalenziali di *Latin Vallex*

Il centro della figura 2 è occupato da due nodi rossi, che corrispondono ai *functor* ACT e PAT: la maggior parte dei nodi per le *frame entry* sono connessi ad essi. Ciò significa che la maggior parte delle *frame entry* trivalenziali in LV hanno un ACT e un PAT tra i propri *functor*.

Intorno al centro del network sono visibili cinque gruppi principali di nodi. Ruotando in senso orario a partire dal gruppo in alto a sinistra, essi sono rispettivamente i gruppi per i *functor*

<sup>8</sup> Il network è stato creato con il software *Cytoscape* (Shannon et al., 2003; Saito et al., 2012). Nodi e relazioni sono organizzati in base al *setting* nominato *Edge-weighted Spring Embedded* (Kohl et al., 2011).

ADDR, (il più numeroso), EFF, ORIG, DIR3 e LOC[ative] (Locativo). Questi sono i *functor* assegnati più frequentemente alla terza posizione argomentale delle *frame entry* trivalenziali (le prime due posizioni essendo occupate da ACT e PAT rispettivamente). Dunque, ad esempio, i nodi raccolti attorno al nodo per il *functor* ADDR rappresentano quelle *frame entry* trivalenziali i cui ruoli semantici sono ACT, PAT e ADDR (come nel caso del verbo *attribuo* - “attribuire”).

I nodi più periferici nel network rappresentano, invece, *functor* assegnati alla terza posizione argomentale con bassa frequenza. Ad esempio, il nodo per il *functor* ACMP (*Accompaniment* - *Accompagnamento*), posizionato in basso a sinistra nel network, è connesso a soli tre nodi, che corrispondono alle *frame entry* (del tipo ACT-PAT-ACMP) per i verbi *admisceo* (“mischiare”), *coniungo* (“congiungere”) e *unio* (“unire”).

I tratti morfologici riportati nei *frame slot* non riflettono i dati testuali nel caso di tre tipi di costruzioni: (a) proposizioni passive, (b) proposizioni infinitive e (c) ablativo assoluto. Ciò è motivato da due ragioni. La prima è legata alla necessità di non moltiplicare eccessivamente il numero di *frame entry*, raccogliendo più forme testuali in una *frame entry* comune. La seconda dipende dal fatto che LV è strettamente legato al livello di annotazione tectogrammaticale delle *treebank*: questo livello ha il fine di rappresentare la sintassi soggiacente di una frase (*underlying syntax* in FGD), altresì considerata il suo significato letterale, attraverso un modello il più possibile indipendente da quello della sintassi di superficie.

Benché i tratti morfologici delle posizioni argomentali di queste tre costruzioni non siano registrati nelle entrate di LV riflettendo in pieno il livello testuale, essi possono sempre essere estratti dal livello morfologico di annotazione delle *treebank*. Più precisamente, i *frame slot* per le occorrenze di parole valenziali nelle tre costruzioni menzionate sono costruiti come descritto nelle sezioni che seguono.

### 3.2.1 *Proposizioni passive*

Le proposizioni passive vengono trasformate nella corrispondente forma attiva prima di assegnare una *frame entry* al loro verbo principale (o costruirne una *frame entry* ex novo).

Si prenda ad esempio la seguente frase tratta dalla IT-TB (*Summa contra Gentiles*, libro 1, capitolo 1, numero 2):

- (3) “*sapientes dicantur qui res recte ordinant*” (“[che] saggi siano detti coloro che ordinano le cose in modo retto”).

Nella frase (3), il verbo principale (*dicantur*) è alla forma passiva. La *frame entry* di LV per il lemma *dico* (“dire”) cui questa occorrenza della forma *dicantur* è connessa riflette la forma attiva della frase: “che [un soggetto generico] dica saggi coloro che ordinano le cose in modo retto”. Dunque, la *frame entry* associata a questa occorrenza del lemma *dico* include un *valency frame* con tre posizioni argomentali cui sono associati i seguenti attributi:

- (1) un ACT generico, ossia non espresso testualmente e non identificabile dal contesto;
- (2) un PAT rappresentato da un pronome: “coloro (che ordinano le cose in modo retto)”;
- (3) un EFF, che è il *functor* assegnato ai complementi predicativi obbligatori: “saggio”.

Questa soluzione consente di assegnare la medesima *frame entry* all'occorrenza del verbo *dico* in una frase come quella d'esempio indipendentemente dal fatto che essa sia espressa in forma attiva o passiva.

### 3.2.2 *Proposizioni infinitive*

In latino, una proposizione infinitiva (nota come “accusativo con l'infinito”: AcI) è una costruzione formata da un verbo all'infinito il cui soggetto è flessso al caso accusativo.

La *frame entry* di LV per un AcI riflette l'equivalente costruzione con la forma verbale finita. Nelle costruzioni attive, l'ACT di un AcI riceve, dunque, il caso nominativo (invece dell'accusativo, che appare a livello testuale); ciò avviene anche in quelle passive, ma a fronte della previa trasformazione della costruzione da passiva ad attiva (secondo quanto detto in 3.2.1).

Ad esempio, si consideri la seguente frase tratta dalla LDT (*De Catilinae coniuratione*, XX):

- (4) “quis mortalium [...] tolerare potest [...] illos binas aut amplius domos continuare [...]?”  
 (“chi tra i mortali può tollerare che loro costruiscano di seguito due o più palazzi?”).

Nella frase (4), la parola *illos* (“loro”) è un pronome all'accusativo plurale che recita il ruolo di soggetto della forma verbale all'infinito *continuar*e (“assommare”, “costruire di seguito”).

La *frame entry* associata a questa occorrenza del verbo *continuo* include due posizioni argomentali:

- (1) un ACT espresso da un pronome al nominativo: *illos* (accusativo) → *illi* (nominativo);
- (2) un PAT rappresentato da un nome all'accusativo: *domos* (“palazzi”).

In questo modo, la medesima *frame entry* viene assegnata all'occorrenza del verbo *continuo* indipendentemente dal fatto che essa compaia in un AcI o in una costruzione con il verbo in forma finita, quest'ultima essendo usualmente rappresentata da una proposizione introdotta da una congiunzione subordinativa, come ad esempio *quod* (“che”).

### 3.2.3 *Ablativo assoluto*

L'ablativo assoluto è una costruzione in cui un nome e un participio formano una proposizione che è disgiunta rispetto alla struttura sintattica del resto della frase in cui occorre. Sia il nome che il participio sono flessi al caso ablativo; il nome ha il ruolo di soggetto del participio.

Nell'assegnare le *frame entry* di LV, gli ablativi assoluti sono trattati nello stesso modo delle proposizioni passive e degli AcI. Nella *frame entry*, il nome che funge da soggetto riceve *functor* ACT e caso nominativo nel caso di un ablativo assoluto alla forma attiva (ossia con il participio al tempo presente). Invece, in caso di ablativo assoluto passivo (participio perfetto), prima il participio è trasformato nella forma attiva e, poi, il nome che aveva originariamente il ruolo di soggetto riceve *functor* PAT e caso accusativo<sup>9</sup>.

<sup>9</sup> Rispetto all'attribuzione del *functor* PAT al nome in ablativo con ruolo di soggetto possono esserci eccezioni, come nel caso del verbo *doceo* (“insegnare”), dove il *functor* non è PAT ma ADDR in conseguenza della sua costruzione con il doppio accusativo.

Ad esempio, si veda la seguente frase tratta dalla IT-TB (*Summa contra Gentiles*, libro 1, capitolo 43, numero 10):

(5) “[...] qualibet quantitate finita data [...]” (“essendo stata data una qualsiasi quantità finita”).

La parola *data* è un participio perfetto del verbo *do* (“dare”). Essendo passivo, l’ablativo assoluto viene innanzitutto reso alla forma attiva (“avendo [un soggetto generico] dato una qualsiasi quantità finita”); quindi, il nome che fa da soggetto del participio (*quantitate*) riceve il caso accusativo nella *frame entry*.

La *frame entry* associata a questa occorrenza di *do* include tre posizioni argomentali:

- (1) un ACT generico;
- (2) un PAT espresso da un nome all’ accusativo: *quantitate* (ablativo) → *quantitatem* (accusativo);  
un ADDR generico.

### 3.3 Interrogare Latin Vallex

*Latin Vallex* ed entrambe le treebank del latino sono liberamente disponibili presso il sito Internet della IT-TB (<http://itreebank.marginalia.it/view/download.php>)<sup>10</sup> e accessibili attraverso un’implementazione del motore di ricerca PML-TQ (*Prague Markup Language – Tree Query*) (Štěpánek & Pajas, 2010; <http://itreebank.marginalia.it/view/resources.php>).

Le query vanno dapprima scritte in un apposito box secondo il formato del linguaggio PML-TQ; successivamente, la corrispondente forma grafica può essere visualizzata in un grafo ad albero. Ad esempio, la seguente query cerca tutte le *frame entry* (*v-frame*) che abbiano (*child*) una posizione argomentale (*v-element*) rappresentata da un Destinatario (*functor* = “ADDR”):

```
v-frame [ child v-element [ functor = “ADDR” ] ]
```

La figura 3 mostra la forma grafica di una query disegnata secondo il linguaggio PML-TQ. La query cerca quelle entrate lessicali di LV (nodo \$n0) che includano una *frame entry* (\$n1) con almeno tre posizioni argomentali, cui siano associati rispettivamente i seguenti *functor*: ADDR (\$n4), PAT (\$n2) e ACT (\$n3). Inoltre, la query impone che l’argomento con *functor* ADDR sia rappresentato da una parola flessa al caso dativo (*case* = “3”).

<sup>10</sup> Della LDT è accessibile la porzione annotata a livello tectogrammaticale e la corrispondente controparte di livello analitico. L’intera LDT è disponibile presso la Perseus Digital Library: [https://perseusdl.github.io/treebank\\_data/](https://perseusdl.github.io/treebank_data/).

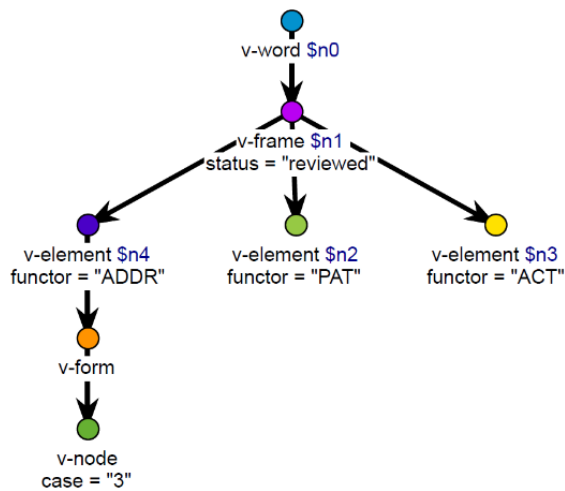


Figura 3: Una query grafica di PML-TQ su *Latin Vallex*

La figura 4 presenta uno degli output prodotti dalla query sopra descritta. In particolare, riporta una *frame entry* del verbo *confero* (“conferire”).

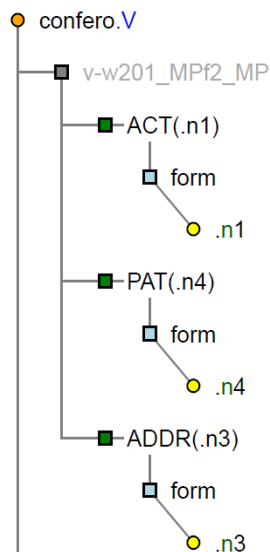


Figura 4: Una *frame entry* del verbo *confero*

In accordo con quanto stabilito nella query, questa *frame entry* include tre posizioni argomentali: un Agente, un Paziente e un Destinatario. Le posizioni argomentali sono ulteriormente specificate dalla PoS e dal caso (nodi “form”): l’Agente è un nome al nominativo (n1), il Paziente è un nome all’accusativo (n4) e il Destinatario è un nome al dativo (n3).



Data la connessione biunivoca tra il lessico e i dati delle treebank, è possibile spostarsi da una specifica *frame entry* di LV alle sue occorrenze nei dati testuali attraverso una query come la seguente<sup>11</sup>:

```
t-node $t := [val_frame.rf v-frame $v := [ id = "v-w201_MPf2_MP"]]
```

Questa query cerca nel livello tectogrammaticale di annotazione delle treebank quei nodi (t-node \$t) cui sia associato un identificativo del *valency frame* (val\_frame.rf) che connetta il nodo tectogrammaticale con la *frame entry* di LV che ha id "v-w201\_MPf2\_MP", ossia la *frame entry* riportata nella figura 4. La figura 5 presenta uno degli output di questa query.

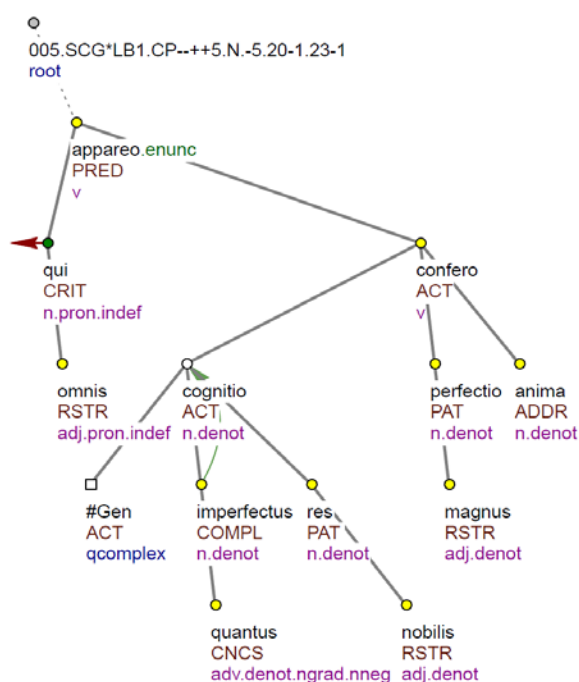


Figura 5: Una occorrenza testuale di una *frame entry*

La figura 5 mostra l'albero tectogrammaticale della seguente frase della IT-TB (*Summa contra Gentiles*, libro 1, capitolo 5, numero 5):

- (6) “ex quibus omnibus apparet quod de rebus nobilissimis quantumcumque imperfecta cognitio maximam perfectionem animae confert” (“in base a tutte le qual cose, risulta che la conoscenza delle cose più nobili, per quanto imperfetta, conferisce la massima perfezione all’anima”).

<sup>11</sup> Ciò è possibile a condizione che la *frame entry* in questione non rientri tra quelle costruite in modalità *intuition-based*, nel qual caso i nodi “form” della *frame entry* ricevono il valore “typical” e la *frame entry* non è connessa ad alcuna occorrenza nelle treebank.

In questo albero, l'occorrenza del verbo *confero* (*confert*) governa un Agente rappresentato da un nome al nominativo (*cognitio*), un nome all'accusativo come Paziente (*perfectionem*; nodo *perfectio*) e un nome al dativo con il ruolo di Destinatario (*animae*; nodo *anima*).

Richiamando l'esempio discusso in 3.2.2, la prossima query collega una *frame entry* di LV con sue occorrenze testuali che presentano specifiche proprietà morfologiche. L'esempio in questione concerne una occorrenza di un AcI la cui testa verbale è una forma del verbo *continuo*. L'entrata lessicale di *continuo* in LV può essere richiamata attraverso la seguente query, che cerca la *v-word* (entrata lessicale) il cui attributo *lemma* ha valore "continuo":

```
v-word [lemma = "continuo" ]
```

La figura 6 mostra l'entrata di *continuo*. L'identificativo dell'unica *frame entry* associata a *continuo* informa che questa è l'entrata numero 508 di LV (w508) e che quella riportata è la sua prima (e unica) *frame entry* (f1). La *frame entry* include un Agente espresso da un pronome al nominativo (u1) e un Paziente rappresentato da un nome all'accusativo (n4).

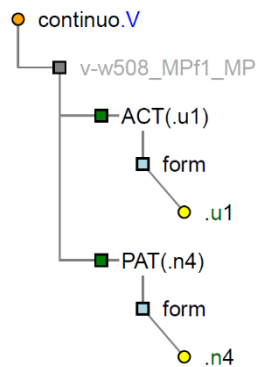


Figura 6. L'entrata lessicale del verbo *continuo*

Tra le occorrenze testuali della *frame entry* di *continuo* riportata nella figura 6, la query seguente cerca quelle dove *continuo* governa una costruzione AcI.

```

t-node $n0 :=
[ val_frame.rf v-frame $n3 :=
[ id = "v-w508_MPf1_MP" ],
a/lex.rf a-node $n1 :=
[ (m/tag ~ "^3.[HQ]" or m/tag ~ "^v...n"), a-node $n2 :=
[ afun = "Sb", (m/tag ~ ".....[DM]" or m/tag ~ ".....a") ] ] ];
  
```

La figura 7 presenta la medesima query in formato grafico.

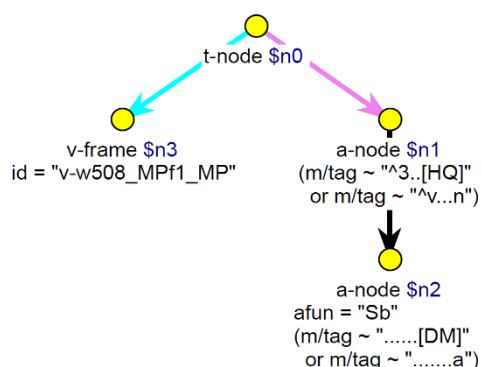


Figura 7. Una query sulle treebank in formato grafico

Questa query cerca nelle treebank i nodi tectogrammaticali (t-node \$n0) la cui *frame entry* in LV abbia id uguale a "v-w508\_MPF1\_MP" (v-frame \$n3). Il t-node \$n0 corrisponde a un nodo nel livello analitico di annotazione delle treebank (a-node \$n1) i cui codici morfologici (m/tag) sono quelli per le forme verbali al modo infinito. L'a-node \$n1 governa un altro nodo analitico (a-node \$n2), che ha funzione sintattica di soggetto (afun = "Sb") ed è una parola flessa al caso accusativo<sup>12</sup>.

La figura 8 mostra uno degli output di questa query.

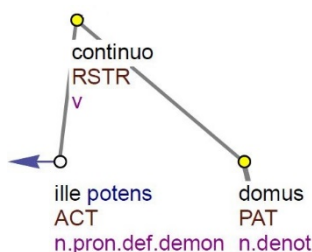


Figura 8. Una porzione di albero risultante da una query

La figura 8 presenta la porzione di un albero tectogrammaticale corrispondente alla proposizione "illos [...] domos continuare" (si veda la frase (4) in 3.2.2). Il nodo per la parola *continuare* (*continuo* nella porzione di albero in figura) governa un Agente rappresentato da un

<sup>12</sup> Dal momento che la IT-TB e la LDT usano *tagset* morfologici diversi, la query cerca due diverse sequenze di codici sia per i verbi all'infinito che per le parole accusativo. In entrambi i casi, la prima sequenza di codici nella query (ossia quella che precede l'operatore or) usa il *tagset* della IT-TB, mentre la seconda è costruita in base a quello della LDT. La documentazione relativa ai *tagset* della IT-TB e della LDT è disponibile a questa URL: <http://itreebank.marginalia.it/view/documentation.php>. Entrambe le treebank sono state recentemente rese disponibili nell'ambito del progetto *Universal Dependencies* (<http://universaldependencies.org/>; McDonald et al. 2013) con codici morfologici comuni in accordo con il *Google Universal PoS tagset* (Petrov et al., 2012).

pronomi all' accusativo (*illos*; nodo *ille*) e un Paziente espresso da un nome all' accusativo (*domos*; nodo *domus*)<sup>13</sup>.

Dunque, benché la costruzione AcI non sia riportata come tale nelle *frame entry* di LV, le occorrenze testuali in forma di AcI delle entrate lessicali di LV possono sempre essere recuperate.

#### 4. Conclusioni

In questo articolo abbiamo presentato *Latin Vallex*, un lessico di valenza per il latino costruito in connessione con l'annotazione semantico-pragmatica di due treebank latine che includono testi di epoche e generi diversi. Da un lato, questa connessione tra l'evidenza testuale e la descrizione lessicale consente di assegnare a ciascuna *frame entry* di LV (costruita in modalità *corpus-driven*) la frequenza delle sue occorrenze nelle treebank. Dall'altro, ogni parola valenziale che occorre nei testi delle treebank è connessa a una *frame entry* in LV.

Al fine di rendere LV sufficientemente rappresentativo del lessico latino, abbiamo inserito in esso anche un certo numero di entrate costruite in modalità *intuition-based*. La relazione tra le due strategie di realizzazione delle entrate lessicali è uno degli aspetti più delicati nello sviluppo di LV. Infatti, se un lessico completamente *corpus-driven* ha il pregio di essere empiricamente motivato in virtù di una mutua relazione con l'evidenza testuale che lo supporta, uno svantaggio di questo approccio consiste nel fatto che i testi possono non essere sufficientemente rappresentativi della lingua oggetto, essendo possibile che *frame* valenziali prototipici non compaiano nel lessico semplicemente perché essi non occorrono nei testi usati come base empirica dello stesso.

D'altra parte, un lessico realizzato in modalità totalmente *intuition-based* presenta il rischio di includere solo quei *frame* valenziali che il lessicografo crede essere i più prototipici per una specifica parola. Ciò è particolarmente problematico nel momento in cui si ha a che fare con una lingua antica e, dunque, non si hanno a disposizione parlanti nativi.

Dunque, è necessario un confronto serrato con l'evidenza fornita da un numero sempre maggiore di testi, sia per aumentare la copertura lessicale realizzata in modalità *corpus-driven* che per valutare la qualità dei contenuti di LV costruiti sulla base dell'intuizione del lessicografo.

Come menzionato nell'Introduzione, un lessico di valenza può avere numerose applicazioni nell'area del TAL. In questo senso, LV rientra in un gruppo di risorse lessicali per il latino di cui fanno parte anche l'analizzatore morfologico *LEMLAT* (Passarotti, 2004), il lessico di sottocategorizzazione sintattica *IT-VaLex* e *Latin WordNet*. La nostra speranza è di poter integrare tutte queste risorse al fine di sfruttare al meglio i diversi tipi di informazione lessicale che esse portano a supporto sia del TAL che di ricerche di tipo linguistico teorico.

#### BILIOGRAFIA

- Bamman, D. and Crane, G. (2006). The design and use of a Latin dependency treebank. In J. Nivre and J. Hajič (Eds.), *Proceedings of the Fifth Workshop on Treebank and Linguistic Theories (TLT2006)*. Prague, Czech Republic: ÚFAL, pp. 67--78.
- Delatte, L., Evrard, E., Govaerts, S. and Denooz, J. (1981). *Dictionnaire fréquentiel et Index inverse de la langue latine*. Université de Liège: Laboratoire d'analyse statistique des langues anciennes.

<sup>13</sup> Come detto, il livello tectogrammaticale di annotazione include anche la risoluzione delle anafore/catafore. Nella figura 8, il pronome *ille* rimanda al lemma *potens* (“[le persone] potenti”), che occorre nella precedente frase del testo. Questa connessione è graficamente rappresentata dalla freccia che punta a sinistra rispetto al nodo di *ille*.

- 
- Fillmore, C. (1982). *Frame semantics*. *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., pp. 111--137.
- González Saavedra, B. and Passarotti, M. (2014). Challenges in Enhancing the *Index Thomisticus* Treebank with Semantic and Pragmatic Annotation. In V. Enrich, E. Hinrichs, D. De Kok, P. Osenova and A. Prepiórkowski (Eds.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT-13)*. Tübingen, Germany: Department of Linguistics, University of Tübingen, Germany, pp. 265--270.
- Hajič, J., Panevová, J., Uřešová, Z., Bémová, A., Kolárová-Rezníčková, V. and Pajas, P. (2003). PDT-VALLEX: Creating a Large Coverage Valency Lexicon for Treebank Annotation. In J. Nivre and E. Hinrichs (Eds.), *TLT 2003 – Proceedings of the Second Workshop on Treebank and Linguistic Theories*. Volume 9 of *Mathematical Modelling in Physics, Engineering and Cognitive Sciences*, Växjö, Sweden: Växjö University Press, pp. 57--68.
- Happ, H. (1976). *Grundfragen einer Dependenz-Grammatik des Lateinischen*. Goettingen, Germany: Vandenhoeck & Ruprecht.
- Kingsbury, P. and Palmer, P. (2002). From Treebank to Propbank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*. Las Palmas - Gran Canaria, Spain: ELRA, pp. 1989--1993.
- Kohl, M., Wiese, S. and Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Methods in Molecular Biology*, 696, pp. 291--303.
- Korhonen A., Krymolowski, Y. and Briscoe, T. (2006). A Large Subcategorization Lexicon for Natural Language Processing Applications. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy: ELRA, pp. 1015--1020.
- McDonald, R.T., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zang, H., Täckström, O., Bedini, C., Castelló, N.B. and Lee, J. (2013). Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA, USA: ACL, pp. 92--97.
- McGillivray, B. and Passarotti, M. (2009). The Development of the *Index Thomisticus* Treebank Valency Lexicon. In *Proceedings of LaTeCH-SHELT&R Workshop 2009*. Athens, Greece: ACL, pp. 43--50.
- McGillivray, B. (2013). *Methods in Latin Computational Linguistics*. Leiden: Brill.
- Messiant, C., Korhonen, A. and Poibeau, T. (2008). LexSchem: A Large Subcategorization Lexicon for French Verbs. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakech, Morocco: ELRA, pp. 533--538.
- Mikulová, M. et alii. (2005). Annotation on the tectogrammatical layer in the Prague Dependency Treebank. The Annotation Guidelines. Prague, Czech Republic: ÚFAL.
- Minozzi, S. (2010). The Latin WordNet project. In P. Anreiter and M. Kienpointner (Eds.), *Latin Linguistics Today*. *Latin Linguistics Today*. Akten des 15. Internationalen Kolloquiums zur Lateinischen Linguistik. Innsbruck, Austria: Innsbrucker Beiträge zur Sprachwissenschaft, pp. 707--716.
- Panevová, J. (1974-1975). On Verbal Frames in Functional Generative Description. Part I, *Prague Bulletin of Mathematical Linguistics*, 22, pp. 3--40; Part II, *Prague Bulletin of Mathematical Linguistics*, 23, pp. 17--52.
- Passarotti, M. (2004). Development and perspectives of the Latin morphological analyser LEMLAT. In A. Bozzi, L. Cignoni and J.L. Lebrave (Eds.), *Digital Technology and Philological Disciplines*. *Linguistica Computazionale*, XX-XXI, pp. 397--414.
- Passarotti, M. (2011). Language Resources. The State of the Art of Latin and the *Index Thomisticus* Treebank Project. In M.S. Ortola (Ed.), *Corpus anciens et Bases de données*, «ALIENTO. Échanges sapiientiels en Méditerranée», N°2. Nancy, France: Presses universitaires de Nancy, pp. 301--320.
- Petrov, S., Das, D., and McDonald, R. (2012). A Universal Part-of-Speech Tagset. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, Turkey: ELRA, pp. 2089--2096.
- Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R. and Scheffczyk, J. (2006). *FrameNet II. Extendend Theory and Practice*. E-book available at [http://framenet.icsi.berkeley.edu/index.php?option=com\\_wrapper&Itemid=126](http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126).
-

- 
- Sgall, P., Hajičová, E. and Panevová, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht, NL: D. Reidel.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), pp. 2498--504.
- Štěpánek, J. and Pajas, P. (2010). Querying Diverse Treebank in a Uniform Way. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*. Valletta, Malta: ELRA, pp. 1828--1835.
- Tesnière, L. (1959). *Éléments de syntaxe structural*. Paris, France: Editions Klincksieck.
- Urešová, Z. (2004). The Verbal Valency in the Prague Dependency Treebank from the Annotator's Point of View. Bratislava, Slovakia: Jazykovedný ústav Ľ. Štúra, SAV.

**BERTA GONZÁLEZ SAAVEDRA** • graduated in Classical Philology at the Universidad Complutense of Madrid (Spain). She has been working at the Index Thomisticus Treebank since 2015, collaborating with Marco Passarotti in the annotation of Latin texts. She also graduated in Italian Philology at the Universitat of València (Spain). Her research fields are Historical Linguistics, Semantics and Indo-European Linguistics.

**E-MAIL** • berta.gonzalezsaavedra@unicatt.it

**MARCO PASSAROTTI** • is researcher at Università Cattolica del Sacro Cuore (Milan) in the area of computational linguistics. A pupil of one of the pioneers of humanities computing, father Roberto Busa SJ, his main research interests deal with developing and disseminating language resources and NLP tools for Latin. Since 2006, he heads the "Index Thomisticus" Treebank project. In 2009, he founded the CIRCSE research centre on computational linguistics at Università Cattolica. Currently, he is Principal Investigator of a FIR-2013 funded project and Coordinator of a Marie Skłodowska-Curie Individual Fellowships. He organized and chaired several international scientific events. He co-chairs the series of workshops on 'Corpus-based Research in the Humanities' (CRH). He is author of one book and of about seventy papers published in scientific reviews and proceedings of national and international conferences.

**E-MAIL** • marco.passarotti@unicatt.it

# T-PAS: COSTRUIRE UNA RISORSA PER L'ITALIANO BASATA SULL'ANALISI DI UN CORPUS

---

*Anna* FELTRACCO

**ABSTRACT** • This contribution discusses some features of the T-PAS resource, a repository of typed predicate argument structures for Italian, acquired from corpora. We briefly describe T-PAS and the process of acquiring these structures and annotating the corpus. We then highlight the main characteristics of the resource that derive from the methodology used for its creation. We conclude by citing the first works that use T-PAS and proposing further developments of the resource.

**KEYWORDS** • Predicate Argument Structure, Lexical Resource, Corpus Analysis

## 1. Introduzione

La risorsa T-PAS (*Typed Predicate Argument Structures*) è una raccolta di strutture argomentali per verbi della lingua italiana (Jezek et al., 2014). In T-PAS le strutture (nel prosieguo *T-PAS*) sono motivate semanticamente e sono acquisite seguendo la procedura lessicografica CPA, *Corpus Pattern Analysis* (Hanks, 2004; Hanks and Pustejovsky, 2005). Le strutture *T-PAS* sono identificate esaminando esempi del verbo in un corpus. Gli esempi sono quindi associati alle strutture così identificate. Un importante riferimento per il progetto T-PAS è PDEV, *Pattern Dictionary of English Verb* (Hanks and Pustejovsky, 2005)<sup>1</sup>, la più importante risorsa prodotta seguendo la procedura CPA per la lingua inglese. Per ciò che riguarda la lingua italiana, una risorsa complementare a T-PAS è LexIt (Lenci et al., 2012) che riporta informazioni acquisite in modo totalmente non-supervisionato su verbi, aggettivi e nomi. A differenza di T-PAS, in LexIt le strutture non sono distinte con criteri semantici.<sup>2</sup>

Il presente contributo si articola in sei sezioni. Nella Sezione 2 descriviamo la procedura di acquisizione delle *T-PAS* e lo stato dell'arte della risorsa alla quale l'autrice ha contribuito per la parte lessicografica. Nella Sezione 3 evidenziamo le principali caratteristiche di T-PAS: in particolare, ne presentiamo le specificità unitamente ad alcune limitazioni. Nella Sezione 4 riportiamo i primi esperimenti in cui la risorsa è stata utilizzata e i risultati ottenuti e nella Sezione

---

<sup>1</sup> <http://pdev.org.uk/>

<sup>2</sup>Tra i recenti progetti che si occupano di annotazioni verbali per la lingua italiana citiamo anche il progetto IMAGACT (Moneglia et al. 2012) il cui obiettivo è la costruzione di una ontologia interlinguistica dell'azione le cui entrate sono costituite da scene che rappresentano un prototipo dell'azione. IMAGACT si basa sull'analisi di verbi di azione ad alta frequenza estratti da corpora del parlato di lingua italiana e inglese. Citiamo inoltre il progetto COMBINET (<http://combinet.humnet.unipi.it/>), il cui obiettivo è lo studio delle combinazioni di parole nella lingua italiana e la costruzione di una risorsa lessicale di combinazioni.

5 ne proponiamo alcuni sviluppi. Infine, nella Sezione 6 riportiamo le conclusioni del contributo.

## 2. La risorsa T-PAS

In questa Sezione viene descritta la risorsa T-PAS e la procedura per l'acquisizione e la definizione delle strutture *T-PAS*. Vengono inoltre presentati alcuni dati relativi alla versione attuale della risorsa.

### 2.1 La metodologia di acquisizione delle T-PAS

La procedura per l'identificazione delle *T-PAS* prevede varie fasi, che sono di seguito descritte. Come accennato nell'Introduzione, lo sviluppo della risorsa T-PAS avviene solo attraverso l'analisi di un corpus. La prima fase, infatti, prevede l'esame di un campione di 250 concordanze del verbo nel corpus. Successivamente il lessicografo individua le *T-PAS* per il verbo in esame identificando gli elementi rilevanti per la disambiguazione del verbo nel contesto, sia a livello di struttura argomentale sia a livello di elementi lessicali per ciascuna posizione argomentale. Ogni *T-PAS* presenta quindi informazioni lessicali, sintattiche e semantiche.

Attraverso lo strumento di editing<sup>3</sup>, il lessicografo definisce la struttura della *T-PAS* marcando la presenza di *subject*, *object*, *indirect object*, *complement*, *adverbial* e *clausal*<sup>4</sup>. Queste posizioni argomentali vengono marcate dal lessicografo solo se la loro presenza è ritenuta essenziale per la definizione della struttura della *T-PAS* e quindi per la definizione del senso del verbo. È possibile inoltre precisare che l'argomento, seppure obbligatorio, non è sempre realizzato nei testi esaminati (per esempio nel caso di argomenti impliciti). Inoltre è possibile marcare esplicitamente che la struttura prevede l'assenza di un complemento oggetto diretto ([[NO OBJ]]).

Per ogni *T-PAS* il lessicografo registra anche informazioni di tipo lessicale e semantico, annotando i *Tipi Semantici* (ST) per ogni posizione argomentale. I Tipi Semantici rappresentano delle generalizzazioni degli elementi lessicali più frequentemente osservati nel corpus per un certo verbo<sup>5</sup>. L'esempio 1 riporta la *T-PAS#2* del verbo *divorare* dove [[Document]] rappresenta la generalizzazione di {*documento*, *rivista*, *libro*, *romanzo* ...}:

- (1) *T-PAS#2* del verbo *divorare*:  
 [[Human]-subj] *divora* [[Document]-obj]  
 Esempio: L'amica divorò il romanzo.

Il lessicografo sceglie i ST più appropriati da una lista di 224 Tipi creata precedentemente operando una generalizzazione a partire dal set di elementi lessicali osservati per le posizioni

<sup>3</sup> Lo strumento di editing è in lingua inglese come per la risorsa PDEV (Hanks and Pustejovsky, 2005). Tuttavia, rispetto alla versione inglese sono state apportate alcune modifiche. Ad esempio, mentre nella versione inglese è possibile annotare *that-clause*, nella versione italiana è possibile annotare *che-clause*. Inoltre, è stata aggiunta la possibilità di annotare l'alternanza -che [[CLAUSE]] | -di [[INF[V]]] (ciò permette, ad esempio, per il verbo *dire*, di associare alla stessa *T-PAS* i seguenti esempi: "Per questo dicono che la vita è artificiale.", "La commissione dice di aver inviato..."; *T-PAS#2* del verbo *dire*: [[Human | Human Group] -subj] *dire* che [[CLAUSE]] | di [[INF[V]]]).

<sup>4</sup> Per un approfondimento su questi ruoli si rimanda a Cinková and Hanks (2010) e El Maarouf (2013).

<sup>5</sup> Per convenzione i Tipi Semantici vengono indicati tra parentesi quadre e con l'iniziale maiuscola.



argomentali dei verbi in un corpus<sup>6</sup>. In questa lista i ST sono organizzati in una gerarchia i cui elementi sono legati da una relazione “IS-A”; ad esempio, nella Tabella 1 [[Plane]] IS-A [[Vehicle]], [[Vehicle]] IS-A [[Machine]] e così via.

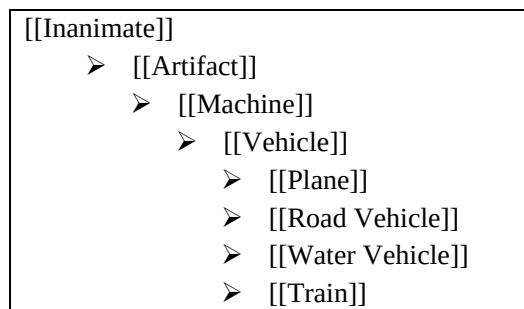


Tabella 1: Sezione della gerarchia dei Tipi Semantici.

Se la generalizzazione possibile è massima, il lessicografo può scegliere il ST [[Anything]] che rappresenta la radice della gerarchia. Viceversa, se nessuna generalizzazione è possibile per la posizione argomentale, il lessicografo annota la lista dei singoli elementi lessicali, come nell'esempio 2 che riporta la T-PAS#3 del verbo *eleggere*<sup>7</sup>.

- (2) T-PAS#3 del verbo *eleggere*:  
 [[Human]-subj] *elegge* { domicilio | recapito | residenza }

Come detto, la distinzione tra T-PAS è principalmente semantica, per cui differenti realizzazioni sintattiche dello stesso argomento (per esempio, un nome e una frase subordinata) afferiscono a una sola T-PAS, purché il senso del verbo rimanga invariato, come nell'esempio 3 tratto da Jezek et al. (2014).

- (3) [[Human]-subj] finisce [[Event]-obj | di INF [V]]  
 a. Finisce *l'allenamento*.  
 b. Non faccio in tempo a finire *di bere la mia birra*.

Anche l'alternanza attivo-passivo viene registrata in un'unica T-PAS: nell'esempio che segue, 4a e 4b sono entrambi associati alla T-PAS#2 di *abbattere*.

- (4) T-PAS#2 di *abbattere*:  
 [[Human | Event]-subj] *abbatte* [[Building]-obj]  
 a. Mi parve di sentire muratori occupati ad *abbattere* muraglie o a praticare un'apertura nella roccia.  
 b. Perché il muro fu *abbattuto*?

Al contrario, l'alternanza causativo-incoativo viene rappresentata con due distinte T-PAS (si veda esempio 5 che riporta T-PAS#1 e T-PAS#2 di *cuocere*).

<sup>6</sup> Questa lista è stata creata analizzando le concordanze per circa 1.500 verbi in Inglese, Italiano, Spagnolo in un lavoro precedente. Si veda Jezek et al. (2014). La stessa lista è attualmente utilizzata per la risorsa inglese PDEV.

<sup>7</sup> L'alternanza di più elementi è indicata con una barra verticale |.

- (5) T-PAS#1 di cuocere:  
[[Human-subj] *cuoce* [[Food]-obj]  
Esempio: Pulite gli spinaci e *cuoceteli*.

T-PAS#2 di cuocere:  
[[Food]-subj] *cuoce*  
Esempio: Mentre il sugo *cuoce* a fuoco lento [..]

Anche l'alternanza del tipo "spray-load" viene registrata con due *T-PAS* distinte, come nell'esempio in 6 per il verbo *caricare*.

- (6) T-PAS#1 di caricare:  
[[Human]-subj] *carica* [[Animate | Inanimate]-obj] (su | in [[Vehicle]] | su | in {spalle | schiena})  
Esempio: [..] controlliamo che le valige siano state *caricate* sul pullman [..]

T-PAS#2 di caricare:  
[[Human]-subj] *carica* [[Vehicle]-obj] (con | di [[Animate]] | con | di [[Inanimate]])  
Esempio: Stanno *caricando* camion e ambulanze di medicinali destinati a Najaf.

Dopo aver definito le *T-PAS*, il lessicografo associa gli esempi del corpus alla *T-PAS* corrispondente (annotando il numero identificativo della *T-PAS*). L'esempio associato alla *T-PAS* diventa quindi esempio d'uso del verbo per quel determinato senso (Popescu, 2012).

Casi particolari riguardano gli esempi che chiaramente afferiscono a una *T-PAS* ma in cui uno degli elementi lessicali in una determinata posizione argomentale non sembra essere ricompreso nel ST scelto per quella posizione: il lessicografo può allora decidere di aggiungere un ST per la posizione argomentale (qualora molti esempi presentino questa caratteristica) o di associare l'esempio alla *T-PAS* annotandolo come "anomalo". In 7 viene presentato il caso di un esempio associato alla *T-PAS#1* di *divorare* e annotato come "anomalo". La *T-PAS* infatti seleziona il ST [[Food]] come oggetto, dove [[Food]] comprende {carne, provviste, anatre, corpi, vittime, uova, ecc.}: il termine *piatto* è stato annotato come "anomalo".

- (7) Esempio con argomento "anomalo":  
T-PAS#1 di divorare:  
[[Animate]-subj] *divora* [[Food]-obj]  
Ti ho rivisto alla festa dell'Unità, mangiare quel piatto di pasta con i funghi, mi guardavi e *divoravi* quel piatto, stavo andando via quando sei arrivata tu.

Per il lessicografo, è possibile inoltre marcare esempi che contengono errori (come in 8) o esempi che risultino non chiari o non classificabili (ad esempio, perché il contesto è insufficiente per disambiguare il senso del verbo e identificare la *T-PAS* corrispondente, come in 9).

- (8) Esempio di errore per il verbo *abitare*:  
[..] oggi a Milano Iosposa, quattro giorni di sfilate dedicate agli *abiti* nuziali.
- (9) Esempio non classificabile per il verbo *divorare*:  
In fondo se avessero smesso di rapinare i viandanti, avrebbero anche smesso di *divorarli*, pensava fiduciosamente.

Infine, per ogni T-PAS viene creata una *implicature*, che consiste in una spiegazione del senso della T-PAS. Questa *implicature* è “ancorata” alla T-PAS perché riporta, per quanto possibile, gli stessi Tipi Semantici nelle posizioni argomentali (Cinková and Hanks, 2010).<sup>8</sup>

Risulta chiaro dunque che sia la struttura identificata sia i Tipi Semantici scelti dal lessicografo per ogni posizione argomentale sono rilevanti per la disambiguazione del verbo, come si vede nella Tabella 2 che riporta tre T-PAS del verbo *abbattere*, le *implicature* per le T-PAS e alcuni esempi tratti dal corpus.

| Verbo: abbattere |  |
|------------------|--|
| ➤ T-PAS 2:       | [[Human Event]] <i>abbattere</i> [[Building]]  |
| IMPLICATURE:     | [[Human Event]] demolisce, distrugge, butta giù [[Building]]   |
| CORPUS:          | Dovranno essere <i>abbattuti</i> in tutto altri tre immobili.<br>... muratori occupati ad <i>abbattere</i> muraglie... |
| ➤ T-PAS 3:       | [[Human]] <i>abbattere</i> [[Animate]]   |
| IMPLICATURE:     | [[Human]] uccide, toglie la vita a [[Animate]]   |
| CORPUS:          | ...Kenai, il più giovane, <i>abbatte</i> l'orso...<br>...un bracconiere <i>abbatteva</i> un coniglio...                |
| ➤ T-PAS 8:       | [[Weather Event]] <i>abbattersi</i> [[NO OBJ]] su [[Location]]   |
| IMPLICATURE:     | [[Weather Event]] <i>si rovescia</i> [[NO OBJ]] su [[Location]]  |
| CORPUS:          | ...un acquazzone <i>si abbatte</i> sull'isola...<br>Una pioggia insistente <i>si abbatte</i> sulla città.              |

Tabella 2: T-PAS#2, T-PAS #3 e T-PAS#8 del verbo abbattere.

Le tre componenti della risorsa T-PAS (Figura 1) sono dunque: la lista delle T-PAS, la lista dei Tipi Semantici, la lista di esempi del corpus associata alle T-PAS.

<sup>8</sup> Per la composizione della *implicature* il lessicografo spesso ricorre all'uso di sinonimi del verbo; tuttavia, può decidere di formulare una *implicature* anche utilizzando delle perifrasi. La *implicature* dunque non rappresenta sistematicamente un elenco di termini sinonimici, ma ha l'obiettivo di chiarire il senso del verbo nella T-PAS.

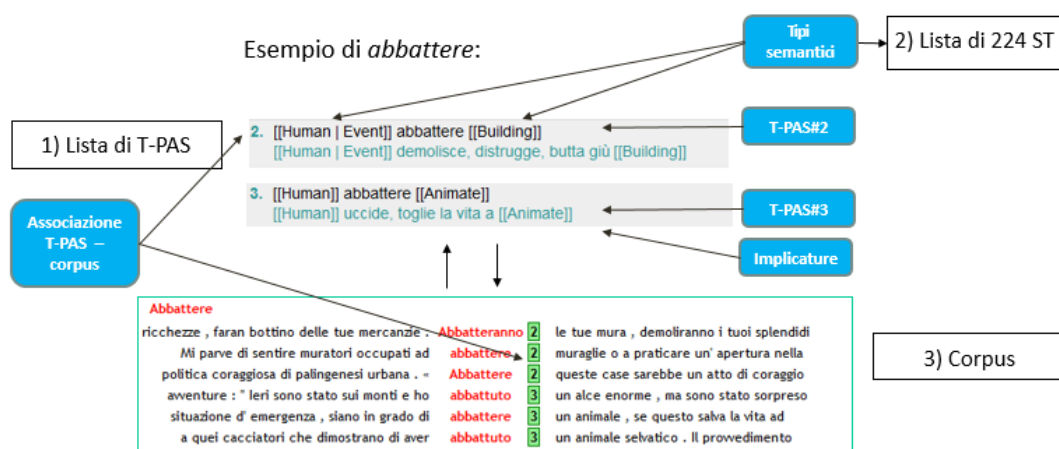


Figura 1: Le componenti di T-PAS.

## 2.2 La versione corrente di T-PAS

T-PAS è stata sviluppata presso il Dipartimento di Studi Umanistici dell'Università di Pavia, in collaborazione con il gruppo Human Language Technology della Fondazione Bruno Kessler (FBK), Trento e il supporto tecnico della Facoltà di Informatica presso la Masaryk University a Brno (CZ). La risorsa è distribuita sotto una licenza Creative Commons Attribution 3.0<sup>9</sup>.

La prima versione di T-PAS (Jezek et al., 2014) è composta da 1.000 verbi analizzati. Questi verbi sono stati estratti in modo casuale dalla lista totale di verbi nella risorsa Sabatini Coletti 2008 (Sabatini e Coletti, 2007), tenendo conto della loro polisemia nella risorsa e secondo le seguenti proporzioni: il 10% dei verbi presenta 2 sensi nel Sabatini Coletti 2008, il 60% dai 3 ai 5 sensi e il 30% dai 6 agli 11 sensi. Il numero totale delle T-PAS identificate nella risorsa è di 4241; il numero minimo di T-PAS per verbo è 1 e il numero massimo è 20.

Il corpus di riferimento per la risorsa è un versione ridotta di ItWAC (Baroni and Kilgarriff, 2006)<sup>10</sup>.

|                                   |      |
|-----------------------------------|------|
| Numero totale verbi               | 1000 |
| Numero totale T-PAS               | 4241 |
| Numero minimo di T-PAS per verbo  | 1    |
| Numero massimo di T-PAS per verbo | 20   |

Tabella 3: Dati quantitativi di T-PAS.

## 3. Caratteristiche salienti di T-PAS

In T-PAS vengono fornite informazioni linguistiche per ogni T-PAS a livello lessicale (ossia la selezione del Tipo Semantico o del set lessicale per ogni posizione argomentale), a livello sintattico (la struttura della T-PAS fornisce informazioni sintattiche relativamente alle relazioni grammaticali degli argomenti) e a livello semantico (le T-PAS vengono distinte semanticamente).

<sup>9</sup> [tpas.fbk.eu](http://tpas.fbk.eu)

<sup>10</sup> Per chiarimenti sulla preparazione del corpus si veda Jezek et al. (2014).

T-PAS è dunque una risorsa importante per studi di linguistica italiana: l'analisi delle *T-PAS* può, ad esempio, agevolare lo studio delle strutture predicato-argomento e lo studio degli argomenti precedentemente definiti “anomali”.

Un'altra caratteristica peculiare di T-PAS è legata alla metodologia di acquisizione delle singole *T-PAS*: diversamente da altre risorse lessicali, le *T-PAS* sono definite partendo dall'analisi di un corpus. Di conseguenza, in T-PAS sono registrati solo sensi del verbo effettivamente utilizzati. Un altro aspetto fondamentale della procedura seguita per lo sviluppo della risorsa riguarda l'associazione degli esempi alle rispettive *T-PAS*, il che permette di eliminare la necessità di creare esempi artificiali o di cercare esempi *ad-hoc* in un corpus. Inoltre il lessicografo, nell'assegnare un esempio alla *T-PAS*, non solo contribuisce ad arricchire la risorsa ma disambigua di fatto il senso del verbo nell'esempio. Infine, visto che gli esempi nel corpus rappresentano evidenze empiriche delle scelte del lessicografo, il processo di assegnazione gli permette di verificare le scelte operate per la definizione della *T-PAS* ed eventualmente di rivedere alcuni elementi della struttura, correggere le *implicature* e rivedere la scelta dei Tipi Semantici.

T-PAS è come si è visto una risorsa ricca di informazioni (utili sia nel campo della ricerca linguistica, sia nel campo del Trattamento Automatico del Linguaggio). L'annotazione manuale garantisce un buon livello di affidabilità, ma vi sono alcune limitazioni insite nel procedimento stesso. Infatti, l'annotazione manuale degli esempi richiede tempistiche elevate e sono possibili errori di disattenzione del lessicografo in tutte le fasi descritte, dalla costruzione delle *T-PAS* all'assegnazione degli esempi. Inoltre, è possibile che si registri un disaccordo tra più lessicografi in merito, ad esempio, al numero delle *T-PAS* identificate (diversa granularità) o alla scelta del ST per una posizione argomentale. Nella Tabella 4 viene riportato un esempio di disaccordo nell'identificazione delle *T-PAS* tra due lessicografi, nominati Annotatore 1 e Annotatore 2, per il verbo *colpire*. Mentre l'Annotatore 1 distingue due *T-PAS* di *colpire* (ossia *T-PAS#1* per “battere, percuotere, picchiare volontariamente”, come in “Condannato ad altri quattro anni di carcere per aver *colpito* un altro detenuto, Ron comincia a ...” e *T-PAS#2* per “toccare, centrare violentemente” come in “una delle parti del chakram *colpisce* in pieno il petto della creatura ...”), l'Annotatore 2 crea una sola *T-PAS* per i due casi.

|                       |   |
|-----------------------|---|
| Verbo: <i>colpire</i> |   |
| Annotatore 1:         |   |
| T-PAS 1               | [[Human1]] <i>colpire</i> ([[Human2]]<br>[[Human1]] batte, percuote, tocca con violenza, picchia volontariamente ([[Human2]])   |
| T-PAS 2:              | [[Animate1   Energy   Particle1]] <i>colpire</i> [[Animate2   Inanimate   Particle2]]<br>[[Animate1   Energy   Particle1]] <i>centra violentemente</i> [[Animate2   Inanimate   Particle2]] |
| Annotatore 2:         |   |
| T-PAS 1:              | [[Animate1   Physical Object1]] <i>colpire</i> [[Animate2   Physical Object2]]<br>[[Animate1   Physical Object1]] <i>picchia, tocca violentemente</i> [[Animate2   Physical Object2]]       |

Tabella 4: Esempio di disaccordo nell'identificazione delle *T-PAS* tra due annotatori.

#### 4. Primi esperimenti con T-PAS

Le informazioni raccolte in T-PAS rendono la risorsa estremamente interessante per la sua applicazione nel campo del Trattamento Automatico del Linguaggio, per esempio nei task in cui è necessario disambiguare il senso del verbo. Popescu et al. (2014) propongono un sistema di

*Word Sense Disambiguation* sviluppando per la disambiguazione del verbo un modello probabilistico che utilizza le *T-PAS* e, in particolare, i Tipi Semantici identificati per ogni posizione argomentale. I risultati mostrano una buona precisione.

In Jezeq et al. (2014) viene presentato un esperimento che valuta l'accordo tra due annotatori che hanno contribuito alla creazione della risorsa nell'operazione di raggruppamento degli esempi nel corpus. Viene richiesto ai due annotatori di creare le *T-PAS* per gli stessi verbi (50) analizzando e annotando per ogni verbo gli stessi esempi (250). Ne risultano due distinti set di *T-PAS*, uno per ogni annotatore. L'accordo viene determinato calcolando il grado di similarità tra le due distribuzioni. I risultati mostrano un buon grado di sovrapposizione tra le due distribuzioni, comprovando l'affidabilità della risorsa (Jezeq et al. 2014)<sup>11</sup>.

*T-PAS* è stata inoltre utilizzata in un esperimento pilota per testare uno schema di annotazione per i tipi di relazioni di opposizione tra frame verbali (Feltracco et al., 2015). L'esperimento coinvolge due annotatori a cui viene richiesto di esaminare coppie di verbi e rispettive *T-PAS*. Per ogni coppia di verbi, gli annotatori devono identificare la presenza o l'assenza di una relazione di opposizione tra le *T-PAS* dei due verbi ed eventualmente marcare il tipo di opposizione secondo lo schema proposto<sup>12</sup>. Per svolgere questo compito, entrambi gli annotatori hanno ritenuto essenziale che per ogni posizione argomentale fossero espressi i Tipi Semantici, al fine di disambiguare i vari sensi del verbo e conseguentemente annotare la relazione di opposizione solo tra sensi verbali opposti; era comunque possibile - ed è stata ritenuta utile per svolgere il task - anche l'analisi degli esempi associati alle *T-PAS*. Questo esperimento pilota rappresenta il primo tentativo di annotazione di una relazione semantica tra le *T-PAS* della risorsa e ha permesso di evidenziare alcuni aspetti importanti per l'annotazione delle relazioni semantiche tra *T-PAS*<sup>13</sup>.

*T-PAS* può inoltre trovare applicazione nei task che richiedono di identificare automaticamente gli argomenti del verbo, come per esempio nel *Semantic Role Labeling* (Gildea and Jurafsky, 2002) in cui i ruoli semantici sono individuati anche sulla base della struttura sintattica della frase.

## 5. Possibili sviluppi della risorsa

Possibili sviluppi della risorsa *T-PAS* riguardano innanzitutto l'incremento del numero di verbi analizzati. Inoltre, l'annotazione degli argomenti nelle istanze del corpus (per il momento sono taggati nel corpus solo i verbi), inclusi gli elementi considerati "anomali", potrebbe favorire, ad esempio, lo studio delle alternanze più frequenti ed essere un punto di partenza per l'estensione automatica della copertura della risorsa.

<sup>11</sup> Gli autori riportano valori medi di BCubed di 0,77 (F1) e di Purity di 0,60 (F1).

<sup>12</sup>La valutazione dell'esperimento prevede il calcolo dell'accordo tra i due annotatori che hanno annotato 25 coppie di verbi. Dai risultati emerge un alto livello di accordo sia nell'identificazione delle relazioni di opposizione (Coefficiente di Dice = 0,98) sia nell'identificazione del tipo di relazione di opposizione (accordo nel riconoscere lo stesso tipo di relazione o nel riconoscere l'assenza di una relazione di opposizione, Coefficiente di Dice = 0,97). Per maggiori dettagli si veda Feltracco et al. (2015).

<sup>13</sup> Ad esempio, è emersa la necessità di annotare la relazione anche in presenza di Tipi Semantici non condivisi dalle *T-PAS* che instaurano la relazione; ad esempio, la *T-PAS#2* di *abbattere* [[Human | Event]] *abbattere* [[Building]] e la *T-PAS#1* di *costruire* [[Human | Institution]] *costruire* ([[Building | Route]]) sono in relazione di opposizione se [[Human]] è considerato soggetto (la *T-PAS#1* di *costruire* non seleziona [[Event]] come soggetto). Inoltre, è risultato necessario permettere agli annotatori di associare una *T-PAS* di un verbo a più di una *T-PAS* dell'altro verbo.

Un primo tentativo di annotazione automatica degli elementi lessicali in posizione argomentale è stato condotto in via sperimentale in Feltracco et al. (2016) con risultati positivi in termini di precisione. L'esperimento si avvale di una mappatura tra i Tipi Semantici e i lemmi della risorsa MultiWordNet (Pianta et al., 2002) e rappresenta anche il primo lavoro che combina l'utilizzo di T-PAS con un'altra risorsa per l'italiano.

Infine, la risorsa potrebbe essere arricchita con l'annotazione delle relazioni semantiche tra T-PAS; il primo lavoro in questa direzione è descritto alla Sezione 4.

## 6. Conclusioni

In questo contributo abbiamo presentato le caratteristiche principali di T-PAS, una risorsa consistente in una raccolta di strutture predicato-argomento per i verbi della lingua italiana. In T-PAS, la definizione delle strutture e la selezione dei Tipi Semantici per ogni posizione argomentale è basata sull'osservazione del verbo in esempi tratti da un corpus, esempi che vengono poi associati alle strutture T-PAS di pertinenza. Abbiamo evidenziato, come elementi maggiormente caratterizzanti della risorsa, il fatto che, proprio perché basata sull'analisi di un corpus, T-PAS registra per ogni verbo solo i sensi effettivamente utilizzati e che gli stessi esempi del corpus agevolano la pratica lessicografica di definizione delle T-PAS. Scopo del progetto T-PAS è quindi la costruzione di una risorsa per l'italiano basata su dati empirici. Il risultato è una risorsa ricca di informazioni lessicali, sintattiche e semantiche utili sia per l'indagine linguistica sia per la ricerca in campo computazionale.

## BIBLIOGRAFIA

- Baroni, M., Kilgarriff, A. (2006), "Large Linguistically - Processed Web Corpora for Multiple Languages". In Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations, pages 87-90. Association for Computational Linguistics.
- Cinková, S., Hanks, P. (2012), Validation of Corpus Pattern Analysis. - Assigning pattern numbers to random verb samples. [http://ufal.mff.cuni.cz/spr/data/publications/annotatio\\_n\\_manual.pdf](http://ufal.mff.cuni.cz/spr/data/publications/annotatio_n_manual.pdf).
- El Maarouf, I. (2013), Methodological aspects of Corpus Pattern Analysis. In *Proceedings of the 34th ICAME conference (ICAME 2013)*, Santiago de Compostela, Spain.
- Feltracco, A., Jezek, E., Magnini, B. (2015), Opposition relations among verb frames. In *Proceedings of The Third Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 16-24, Denver, Colorado, June. Association for Computational Linguistics.
- Feltracco, A., Gatti L., Magnolini S., Magnini, B., Jezek, E. (2016), Using WordNet to Build Lexical Sets for Italian Verbs. In *Proceedings of the Eighth Global WordNet Conference (GWC'16)*, Bucharest, Romania, January.
- Gildea D., Jurafsky D. (2002), Automatic labeling of semantic roles. *Computational linguistics*, 28(3), 245-288.
- Hanks, P., Pustejovsky, J. (2005), A Pattern Dictionary for Natural Language Processing. In *Revue française de linguistique appliquée*, 10(2).
- Hanks, P. (2004), Corpus Pattern Analysis. In *Proceedings of the XI EURALEX International Congress*. Lorient, France, July.
- Jezek, E., Magnini, B., Feltracco, A., Bianchini A., Popescu, O. (2014), T-PAS: A resource of corpus-derived Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May.

- 
- Lenci A., Lapesa, G., Bonansinga, G. (2012), LexIt: A Computational Resource on Italian Argument Structure. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3712-3718, Istanbul, Turkey, May.
- Moneglia, M., Monachini, M., Calabrese, O., Panunzi, A., Frontini, F., Gagliardi, G. and Russo, I. (2012), The IMAGACT Cross-linguistic Ontology of Action. A new infrastructure for natural language disambiguation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2606-2613, Istanbul, Turkey, May.
- Pianta, E., Bentivogli, L., Girardi, C. (2002), MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet (GWC02)*, Mysore, India, January.
- Popescu, O., Vo, N.P.A., Feltracco, A., Jezek, E., Magnini, B. (2014), Toward Disambiguating Typed Predicate Argument Structures for Italian. In *Proceedings of the First Italian Conference on Computational Linguistics (CLIC-it 2014)*, Pisa, Italy, December.
- Popescu, O. (2012), Building a resource of patterns using semantic types. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2999-3006, Istanbul, Turkey, May.
- Sabatini F., Coletti V. (2007), *Dizionario della lingua italiana 2008*. Milano: Rizzoli Larousse.

**ANNA FELTRACCO** • Ph.D. student in Linguistic Sciences at the University of Pavia and the University of Bergamo in partnership with Fondazione Bruno Kessler, Trento, Italy where she is pursuing her Ph.D. within the Human Language Technologies Group. She has collaborated in the T-PAS project as annotator, participating in the interannotator agreement evaluation and other experiments using the resource which are cited in this contribution. Her research interests cover corpus linguistics and natural language processing.

**E-MAIL** • feltracco@fbk.eu



# ItINERARI

---



# RIFRAZIONI DI UN MONDO CALEIDOSCOPICO

Per una lettura dell'opera di V. Majakovskij

---

*Francesca* GALFIONE

**ABSTRACT** • A formal approach has been chosen to investigate Vladimir Mayakovsky's poetic credo: the functional analysis of the literary process constitutes the guiding principle for the description and classification of the analysed issues. A small selection of poems has been investigated focusing on rhythmic elements and stylistic procedures: the texts are divided into their component parts, so that the constructional processes can be analysed according to their artistic function. Mayakovsky's "ripped" verse reminds us of Cubism, being made up of a continuous succession of microtmesis and palindromes; it arouses a sort of "shift" in the literary language and generates a fuzzy semantic system, in which the semasiological weight of each lexeme undergoes a change due to its disposition in the verse line. Once the formal analysis on the selected poems was completed, it appeared clear how specific rhythmic-syntactic devices and formal procedures actually carry with them not just a solid poetic vision, but also a subliminal sense of humanity.

**KEYWORDS** • Mayakovsky, Ripped verse, Formal analysis, Poetry, Displacement

## **Un futurismo indefesso**

Se si dovesse indicare un motivo che unisce e spiega tutti gli aspetti dell'opera majakovskiana, si dovrebbe forse indicare il "salto del presente" (Buttafava 1977: 28). Majakovskij è ossessionato dal futuro, è un utopista inguaribile, egli vede e annuncia il "grande" a venire e, contemporaneamente, accetta e vede il reale, la concretezza del corpo. Non vede, non accetta, odia il presente, il quotidiano.

Il tormento nell'angustia del limite posto e la volontà di superare i confini statici sono un motivo costantemente variato da Majakovskij. Nessuna tana al mondo può contenere il poeta e l'orda sfrenata dei suoi desideri. [...] L'io del poeta è un ariete che rimbomba contro un futuro proibito, è la volontà [...] di incarnare il Futuro, di attingere l'assoluta pienezza dell'essere<sup>1</sup>. (Jakobson 1975: 8-9)

Alla tensione creativa verso il futuro è sempre contrapposta l'insofferenza per la staticità dell'inerte presente, per il consumarsi della vita in asfittici schemi immutabili; così, tutta la poesia di Majakovskij è percorsa dalla smania di gettare lo sguardo oltre il tempo, in una tormentata ricerca che, in forme e con mezzi sempre diversi, tende alla costruzione di una cultura poetica rivoluzionaria.

L'intento che persegue nei suoi versi sembra implicare, sul piano del linguaggio letterario, non una semplice modifica del materiale artistico ricevuto in eredità, ma una vera e propria rottura con il passato; ai suoi occhi, è tempo che la poesia compia un salto rivoluzionario e che scioglia i legami con la tradizione, nel rifiuto di canoni ormai fossilizzati e vincolanti. A tal fine, l'autore vuole trarre dalla parola tutte le possibilità e le potenzialità nascoste e lo fa con bizzarre

---

<sup>1</sup> Nel presente articolo si è optato per l'inserimento di citazioni in traduzione, anche in virtù del fatto che il testo italiano di riferimento si è rivelato a tratti più completo di quello sovietico.

associazioni linguistiche, con una straniante – eppure ricercatissima – resa grafica del testo, con giochi verbali e minuziosi esperimenti sulle possibilità combinatorie di prefissi, radici e suffissi verbali. Esauriente è l’affermazione di A. M. Ripellino:

Majakovskij ha un senso vivissimo dei giochi verbali, dei neologismi, dei calembours e, con estrema insofferenza per gli stampi consueti, rimescola e intreccia i vocaboli in modo da ravvivarne i colori e la sostanza semantica, affinché il verso acquisti una nuova tangibile concretezza. (Ripellino 1979: 58)

La sua è una passione – una fede – reale, quasi esasperante, quasi un’ossessione.

La figura poetica centrale di Majakovskij è lui stesso, il poeta-tribuno. In tutta la prima parte della sua produzione, l’io del poeta è un’immagine complessa e a tratti farsesca, che riassume in sé l’idea dell’eroe ribelle e del buffone. Egli è comunque, sempre, una figura reale, concreta, non una maschera lirica; al di là delle pose da provocatore, da “bellimbusto”, da istrione si nota una percezione vivissima dei mutamenti sociali prodottisi con lo sviluppo industriale e tecnico, una consapevolezza del cambiamento avvenuto nello status stesso di “uomo”. In questo senso, “il narcisismo spinto all’eccesso diventa azione sociale” (Buttafava 1977: 14). Senza la sua caratteristica carica eversiva e il suo gigantismo, “l’io majakovskiano non sarebbe che l’incarnazione di uno sfogo lirico, personale e asociale” (Buttafava 1977: 14). In questo quadro rientra anche l’unicità della sua produzione poetica, tesa a un lavoro costante sul verso, che deve essere calcolato sulla voce e adeguato quindi alle rotture, agli sbalzi, alla libertà della lingua parlata.

Alla fine del 1920 il poeta e critico letterario russo K. Čukovskij, che si era interessato del futurismo ancor prima della rivoluzione del 1917, tenne nel Palazzo delle Arti di Pietrogrado una conferenza dal titolo *Achmatova e Majakovskij*, in cui riflette sulla peculiarità personale, poetica e stilistica di Majakovskij:

Majakovskij è poeta delle catastrofi e delle convulsioni. [...]. Majakovskij è poeta del movimento, del dinamismo, della bufera. [...] Proprio per questi temi è necessario lo stile iperbolico, il gigantismo, la tendenza al colossale che organicamente gli sono propri. [...] Egli è poeta del suono e del fracasso, di ruggiti e stridori di ogni tipo, incapace di silenzio. [...] Strutturalmente i suoi versi sono appelli alla folla. [...] Egli è Isaia vestito da apache. Dalla sua gola di mille voci ruggisce la strada moderna in rivolta, e non è colpa sua se a volte è volgare come una bestemmia, ed elementare come uno sparo. [...] Il bello in Majakovskij sono le metafore pungenti e precise. [...] Da esse davvero sprizza l’allegria provocante della strada, la mordacità del mercato. [...] (Cit in Worosyliski 1990: 305 – 306)

Il suo verso è oratorio e perentorio, orientato verso l’azione concreta immediata, che consiste nella trasformazione della psiche dell’ascoltatore attraverso l’appello alla rivolta contro le forme alienanti del *byt*, il quotidiano stereotipato; ciò porta l’autore a concepire la parola come lo strumento privilegiato di una nuova rivoluzione, la rivoluzione spirituale (o culturale) che doveva, a suo avviso, completare le trasformazioni politiche ed economiche del Paese. La sua poesia rappresenta sempre una lotta determinata contro ciò che minaccia di canonizzare e di istituzionalizzare la vita.

Il poeta che supera e sprona il tempo è un’immagine costante di Majakovskij. [...] Egli brama conoscere la propria vita in anticipo e la riconosce nel suo proprio romanzo. (Jakobson 1977: 26)

Eroe dei propri versi, Majakovskij si muove in un ambiente insolito per la poesia: è un mondo cupo, popolato di uomini disperati, di folle urbane che si contorcono e non trovano pace tra lembi di palazzi, stralci di rotaie, strade-labirinto dalla prospettiva incongrua. L’irrequietezza degli uomini avviliti è resa ancora più tragica dalla contemporanea sommossa delle cose che insorgono ribelli, come dotate di vita propria. La natura appare “storpiata”, sembra seguire leggi dinamiche proprie e sconosciute alla logica umana, tanto che gli oggetti ne emergono frantumati,

scomposti alla maniera cubista e in perenne tumulto, e sta al lettore/osservatore il compito di raccostarne le diverse componenti (Ambrogio 1978: 16). L'immagine risultante è sempre diversa, sempre imprevedibile e sempre spaventevole, in quanto illogica.

Le operazioni dei futuristi nel campo della semantica, della sintassi, della versificazione realizzano in vario modo il programma dell'“odio per la lingua” espresso con veemenza sin dalla pubblicazione di *Poščečina obščestvennomu vkusu* (*Schiaffo al gusto corrente*). In Majakovskij si tratta della scoperta delle possibilità inesplorate del linguaggio poetico piuttosto che di una sua immotivata distruzione. I mezzi preferiti dall'autore sono l'iperbole, lo smottamento dei legami semantici comunemente percepiti, assunti come validi dalla ragione e istituzionalizzati dalle convenzioni linguistiche, la modificazione straniante dei nessi tra oggetti e concetti, l'eliminazione degli “accessori” tradizionali della poesia e della sua “aureola” estetica, la comparazione “indecente”, la privazione della staticità di ogni nome e concetto, la costruzione di neologismi e di versi liberi. Majakovskij vuole rompere l'automatismo della percezione, opponendosi a opinioni e gusti livellati e inerti. Sua intenzione è “rinnovare interamente un materiale poetico, che ha coscienza di non essere interamente nuovo, partendo da una rifondazione radicale della lingua e della struttura” (Woroszyliński 1990: 293). Il poeta esige la trasfigurazione istantanea del mondo tramite la parola: i suoi puzzle verbali decostruiscono il retaggio di ogni visione precedente del mondo e i suoi versi esercitano un'azione d'urto sulle facoltà percettive del lettore.

Nel fare ciò, Majakovskij mostra il desiderio di farsi capire, di spiegare che la nuova poesia non è soltanto una spiritosa trovata, ma possiede un proprio ruolo imprescindibile nella vita di ciascuno. Parte della poesia del futurismo è poesia delle grandi città. La città sostituisce la natura e le forze naturali, diventa elemento attivo e dinamico che porta alla nascita del nuovo uomo urbano, in grado di esperire un nuovo tipo di vita e, pertanto, nuove modalità emozionali. I telefoni, gli aeroplani, gli ascensori, le ciminiere delle fabbriche, i palazzi sono le componenti della bellezza della natura urbana. Cambia il ritmo della vita e la poesia non può esimersi dalla rappresentazione di tali trasformazioni: essa deve restituire la nuova psiche della città moderna, compito per cui la letteratura del passato più prossimo, quella simbolista, sembra essere inadeguata.

La “rivoluzione poetica” auspicata dai futuristi non può però limitarsi all'ampliamento dei propri temi, accettando e inglobando in sé la nuova tematica urbana: obiettivo primario è dare una nuova autonomia alla parola, la quale “non deve descrivere, ma esprimere di per sé” (cit. in Katanjan 1956: 60). Majakovskij si dedica così alla creazione di un nuovo linguaggio poetico, in grado di esprimere la collera, la disperazione o l'esaltazione della folla. Per contrastare il mutismo e l'incapacità della letteratura precedente nel giungere a una veritiera e sintetica espressione del reale, il nuovo linguaggio deve essere sonoro, naturale, democratico (non nel senso dell'accessibilità, ma del contenuto), colorito, esagerato per risultare immediato.

Il poeta può cantare, ma, se si limita al canto, non è moderno: bisogna fischiare, urlare, tuonare: la stessa poesia contiene l'esigenza di una cacofonia assordante. (Etkind 1990: 329)

Majakovskij esige non tanto la semanticità, quanto la percettività dell'opera, e riesce nel suo intento creando una inaspettata “faccia sonora” della stessa: egli sottomette “poeticamente ciò che dianzi era definito cacofonico”, rendendo possibile “un allargamento della percezione del mondo” (Šklovskij 1940: 117). Nella sua ricerca di una pratica linguistica inedita, il poeta concentra l'attenzione sulla *faktura* del materiale verbale, termine che in pittura indica la scabra rozzezza della superficie dell'opera, la densità grumosa del colore; tale tecnica porta a prediligere consonanti sonore e a favorire, nella produzione testuale, una concentrazione di fonemi fricativi e affricati. Le consonanti si fanno “portatrici del colorismo pittorico in poesia” (Kručenyč, Chlebnikov 1913: 12).

### Il rapporto con la città: le “fantasie cubiste”

Il tema costante, ma infinitamente variato della città percorre l’esperienza futurista con frequenza quasi ossessiva, in una descrizione della metropoli come alienante città infernale, turbolenta ragnatela di strade. Immagini di città già popolavano la letteratura del primo Novecento, in diverse declinazioni; nelle opere dei simbolisti, primi fra tutti A. Blok e A. Belyj, l’ambientazione urbana si era fatta man mano dominante. Tuttavia, anche in relazione a questo tema, i futuristi si allontanano dalla tradizione e tendono alla rappresentazione di un “immaginario urbano sintetico e astratto” (Casari 2008: 281), impersonale, pur muovendo da una realtà ben definita.

Soggetto predominante della poesia giovanile di Majakovskij è la città, totalmente trasformata dall’avvento dell’epoca industriale; in questo contesto il dialogo tra pittura e poesia si fa particolarmente serrato. La “città del futuro” proposta dall’autore è costituita di elementi verbali e non, di segni grafici atti a trasmetterne una visione essenziale, ma di grande impatto psicologico. “Un cinetismo frenetico, convulso, debordante squassa la città” (Casari 2008: 283), mentre il punto di vista cambia continuamente e la visione si scompone, fino a risultare in una sorta di prospettiva rovesciata. I celebri versi di *Adišče goroda (L’infernaccio della città, 1913)* ben esemplificano tale tendenza:

Адище города окна разбили  
 На крохотные, сщущие светами адки.  
 Рыжие дьяволыб вздымались автомобили,  
 Над самым ухом взрывая гудки.<sup>2</sup>  
 (Majakovskij 1955: 55)

La città si tormenta, si strugge, offrendo ad ogni passo visioni lubriche:

Лысый фонарь  
 сладострастно снимает  
 с улицы  
 черный чулок.<sup>3</sup>  
 (Majakovskij 1955: 38)

Le cose si personificano e si rivoltano, lo spazio è popolato da una folla di reietti in un clima di cupa disperazione; i veri protagonisti sono i camini delle case, le grondaie, le insegne, le automobili, mentre l’artista si fa apostolo e martire, ma anche pagliaccio blasfemo.

Majakovskij è in fondo ferito e oppresso dalla città; i suoi mezzi espressivi sono diversi e opposti a quelli dei simbolisti, ma i sentimenti sono simili. La città è un’entità sgradevole, ostile, eppure vivissima, con cui il poeta deve necessariamente fare i conti popolando di sé piazze e strade. Egli risponde all’angoscia causata dalla città asfittica e soffocante opponendovi il suo carisma, la sua forza di ribellione sotto forma di gigantismo. A suon di urla sembra voler svegliare gli oggetti torpidi e inerti del tanto odiato presente, tenta di dar loro la vita affinché si dirigano verso il futuro, condividendo l’ansia di ciò che verrà che corrode l’autore stesso dall’interno. Egli è il poeta più urbano del cubofuturismo, ma “sta stretto nella città e nel presente”, in questo “oscuro regno di crocicchi” (Vitale 1979: LV) che lo pervadono di inquietudine. La città è macabra, è il regno della turpitudine, di un sonno – di una morte – paralizzante.

<sup>2</sup> Trad. it.: “Le finestre frantumarono l’infernaccio della città / in minuscoli infernucci succhianti con le luci. / Rossicci diavoli, si impennavano le automobili, / facendo esplodere le trombe proprie sull’orecchio”. In Carpi 2014: 129.

<sup>3</sup> V. Majakovskij, *Iz ulicy v ulicu* [Da una strada all’altra], vv. 40 – 43. Trad. it.: “Un lampione calvo / sfilava lussurioso / una calza nera / alla strada”. [Tutte le traduzioni sono mie, ove non diversamente indicato. – F.G.]

Il suo è un urbanismo saturo di orrore e intriso di senso di disfacimento, ma anche un urbanismo contraddittorio, in cui affreschi cittadini dalle immagini a tinte forti si alternano a visioni lascive e ad “urli di protesta contro la città-prigione” (Vitale 1979: LXXIII), a dimostrazione della nevrosi che pervade le strade, gli elementi cittadini e, inevitabilmente, la vita degli uomini. La città sembra avere la meglio sul singolo, la cui personalità si disgrega: ora a predominare sono l'anomia e l'inquietudine.

Le prime poesie di Majakovskij sono “fantasie cubiste” (Nicholls 2000: 193): la vivida immaginazione urbana dell'autore è legata all'impiego massiccio di metafore “pittoriche” per creare un effetto altamente emotivo. L'autore scompone inoltre la parola in modo da farle assumere una sorprendente concretezza:

У-  
лица.  
Лица  
у  
догов  
годов  
рез-  
че.  
Че-  
рез [...]⁴  
(Majakovskij 1955: 38)

Majakovskij è il poeta della città per eccellenza, il cantore del movimento, della velocità, del suono; nella sua produzione gli elementi urbani appaiono in veste antropomorfa e sono dotati di capacità di movimento.

La poesia *Iz ulicy v ulicu* (*Da una strada all'altra*, 1913) sembra evocare la visuale che si ha da un tram in movimento; l'unicità della poesia risiede nel fatto che i rapidi slittamenti di prospettiva vengono incorporati nel gioco linguistico della versificazione, con “effetti palindromici che traducono la struttura delle immagini in modo equivalente a un'intersezione di piani” (Nicholls 2000: 194). L'autore effettua una vera e propria dissezione della realtà al fine eliminarne le caratteristiche comunemente percepite e porne in risalto altre, spesso deformate e quasi ripugnanti.

Majakovskij reifica oggetti e parole e ne fa delle armi per distruggere, frammentare e ricostruire la realtà secondo la propria visione dell'arte e del mondo.

Le automobili, su una delle quali sembra muoversi l'autore, sono “*železnych konej*”, le case “fuggono”, le ciminiere, “*žirafij risunok*”, si stagliano nel cielo come dotati di vita propria; il tram è vorace, è dotato di fauci da cui escono i binari. La forza degli elementi urbani, moderni, ha vinto sugli abitanti stessi della città (verso 25, *my zavoevany!*), che assistono a una sfilata di vasche da bagno, docce, ascensori. L'immagine finale è quella di un lampione che, voluttuoso, “*snimaet / s ulicy / černyj čulok*” e così facendo ne mette in evidenza le parti più nascoste, esponendole alla sua luce calda in una simbolica scena di seduzione.

Il titolo della lirica è già di per sé indicativo e costituisce il principio base di costruzione dell'opera: la struttura della lirica ricorda esattamente quella della città, che si sviluppa “da una strada all'altra” in modo contorto e in un movimento incessante. La poesia è un moto continuo, uno spostarsi avanti e indietro, un intersecarsi e tornare su se stessa. E gli eroi della poesia sono infatti quelli della città, i veicoli - i tram, le automobili, persino un ascensore. Ovunque regna il movimento vorticoso; la sensazione di spostamento, di trasporto da un punto all'altro è resa anche a livello lessicale, tramite l'uso di prefissi, preposizioni e verbi indicanti movimento: *attraverso* (*čerez*, v. 9-10), *fuggire* (*begat'*, v. 12), *tirare fuori* (*tjanut'*, v. 23), *strappare via* (*vyryvat'*, v. 38).

⁴ V. Majakovskij, *Iz ulicy v ulicu* [Da una strada all'altra]. Trad. it.: “Stra- / da. / I volti / dei / cani / mostrano gli anni / in modo più nitido. / Verso / ...” [F.G.].

All'intonazione retorica si unisce un approccio più visivo: ciò risulta evidente dalla resa concreta dell'opera, nella sua presentazione grafica (e pittorica) e nella sua sonorità verbale, chiaramente percepibile al momento dell'esecuzione. Mai come allora la poesia era stata così aperta nell'accogliere ogni tipo di intrusione: si dimostrava essere "assetata di concretezza", mentre "in essa avveniva la guerra civile della forma" (Šklovskij 1940: 32-33). Così, in poesia fece irruzione la pittura: già nel 1913, infatti, lo stesso Majakovskij, durante un intervento in una conferenza presso l'Unione della Gioventù, propose la formula "Cubismo nella parola. Futurismo nella parola".

Tutto il mondo civile del nostro tempo si sta trasformando in un'immensa, gigantesca città. [...] La città stessa diventa una forza naturale, nel cui seno nasce il nuovo uomo di città. (cit. in Carpi 2014: 14)

Majakovskij avverte che il cambiamento essenziale consiste nel fatto che tutto è diventato "fulmineo". "La febrilità è il simbolo del nostro ritmo di vita.": i ritmi sognanti e pacati della vecchia letteratura non possono corrispondere all'attuale dinamismo febbrile delle cose e la poesia, secondo i futuristi e per loro tramite, deve adeguarsi al proprio tempo. Tale adeguamento non può esaurirsi in un rinnovamento di contenuti, ossia nella celebrazione della tematica urbana, né in una riforma puramente stilistica; la parola poetica, secondo Majakovskij, deve "esprimere di per sé", tramutandosi nella "parola come tale" (cit. in Carpi 2014: 14-15). Il poeta vuole ridare corporità al tessuto verbale, riconsegnare alla parola la propria autonomia espressiva; su tale linea di ricerca si colloca tutta una serie di procedimenti tecnico-formali inediti: l'arricchimento del bagaglio lessicale mediante l'invenzione di parole nuove, l'inserzione di voci impoetiche, termini colloquiali, la disgregazione dei nessi semantici, la negazione dell'ortografia ordinaria e dei canoni metrico-ritmici, i calembours, la scomposizione grafico-volumetrica dei segni verbali. Di questi ultimi si vuole privilegiare la "caratterizzazione grafica e fonetica" (cit. in Carpi 2014: 16).

Le rime di Majakovskij sono fatte secondo il loro suono: Trenin, nel libro *Nel laboratorio del verso di Majakovskij*, descrive come la peculiarità rimica dell'autore consista nel fatto che egli

[...] rinuncia decisamente alla forma ortografica tradizionale della parola e tenta di fissarla approssimativamente così come viene pronunciata (ossia crea una sorta di trascrizione fonetica scientifica). (cit. in Šklovskij 1940: 126)

Spingendosi oltre nel processo di sperimentazione, Majakovskij scompone la frase e ne rimescola le parti con procedimenti inediti, negando la classica gerarchia tra parole significative e parole ausiliarie: l'accento è posto sui termini linguistici secondari, che acquistano una nuova rilevanza e, soprattutto una nuova semantica.

I poeti avevano concentrato tutti i loro sforzi su flebili variazioni delle sfumature semantiche, avevano già lavorato sulle entità ritmico-sintattiche spostandole, e questo fu spazzato via da Majakovskij, il quale sconvolse la salda banchisa di ghiaccio delle parole e [...] costruì sul linguaggio oratorio una nuova poesia tattile. (Šklovskij 1940: 132)

I cubofuturisti lottano contro la melodicità dei simbolisti, esaltando la "testura della parola" (*faktura slova*), fondata sulle iterazioni e sull'accumulo di consonanti dal timbro sonoro. Per questo nel Majakovskij del primo periodo si ritrovano innumerevoli allitterazioni e ripetizioni di fonemi insoliti in poesia, con insistenza su fricative e affricate. L'acustica ha spesso in questo autore ugual risalto della pittura: accanto a metafore pittoriche vi sono immagini acustiche e strumenti (in *Iz ulicy v ulicu* compare e "si sente" il "rullo dei tamburi"). "Majakovskij orchestra la poesia e il suo mondo" con una "orchestrazione di strumenti fragorosi e di rumori" (Ripellino 1988: 117). Nella poesia presa in esame i riferimenti a manifestazioni di carattere sonoro/rumoroso sono molte: *campana* (*kolokol'nyj*, v. 14), *tram* (*tramvaj*, v. 23), *che tu strilli o*



non strilli (*kriči, ne kriči*, v. 31), il vento tagliente / ... / strappa (*veter koljučij / ... / vyryvaet*, vv. 36-38).

Altrettante volte vengono menzionati nel testo cliché tipici del credo e dell'attività avanguardista: ai versi 21 – 23 il tema circense è applicato alla realtà urbana (*Fokusnich / rel'sy / tjanet iz pasti tramvaja*: “Il prestigiatore tira fuori i binari dalle fauci del tram”), mentre al verso 17 notiamo l'uso del verbo *vypestrit'*, “rendere variopinto”, che non solo costituisce un ulteriore richiamo alla pittura e al cubismo, ma è anche un termine utilizzato in senso del tutto inconsueto, da intendersi come “liberare verso l'esterno”, con significato di “spargere / gettare d'intorno colori variopinti”. Al verso 31 (*Kriči, ne kriči*: “Che tu strilli o non strilli”) predomina il tema dell'urlo, del grido, indicativo della violenta volontà iconoclastica dei futuristi, mentre al verso 13 (*prygnuli pervye kuby*: “sono saltati giù i primi cubi”) l'autore menziona indirettamente se stesso e i suoi compagni d'avanguardia. Nella lirica compaiono inoltre l'immagine dell'automobile, simbolo della modernità, e quella del fuoco ardente (verso 30, *Telo žgut ruki*: “Le mani bruciano il corpo”).

L'avvento della nuova epoca industriale e le trasformazioni urbane che ne conseguono portano quasi necessariamente l'uomo a percepire il mondo che lo circonda come spaventoso e straniante: l'artista comprende che una sua fedele rappresentazione può avvenire soltanto mediante contrasti, iperboli, metafore fantasmagoriche. La percezione inconscia della città invade le sue poesie e lo porta a illuminare i dettagli più stranianti, a volte macabri, dello spazio urbano. Secondo l'opinione del critico Efim Etkind, la città di Majakovskij è un incubo, soprattutto perché l'artista ignora i concetti astratti: ciascuno di loro diventa entità materiale, un mostro visibile. La concretizzazione dell'astratto è in effetti uno dei procedimenti fondanti dell'estetica dell'autore, ravvisabile nell'arco della sua intera produzione; d'altro canto, ogni concetto concreto diventa ancora più “fisico”, materiale e mira a raggiungere il massimo grado di percettibilità; la tangibilità espressiva è portata al limite estremo, in una alterazione costante della realtà e degli elementi urbani, soggetti umani compresi. Inoltre, Majakovskij rappresenta il mondo interiore dell'uomo come una proiezione ancor più deforme e angosciante del mondo esteriore, in quanto nell'animo umano tutto avviene come all'esterno, ma in modo molto più spaventoso; perciò, il poeta conferisce ai pensieri l'intensità e la concretezza tipica degli incubi (Etkind 1990: 332-333). Per tutti questi motivi, la città majakovskiana è il regno del terrore e ci si imbatte, in essa, in un susseguirsi di ferite, lacerazioni, vuoti, frammenti e brandelli, in elementi rotti, spezzati, sfilacciati e in piena convulsione: “sono gli effetti di una eccessiva tensione dello spazio, di una morbosa ipertrofia” (Buttafava 1977: 14). È Čukovskij il primo a comprendere lo spasmodico conflitto che si genera nell'animo del poeta:

Majakovskij è un illusionista, un visionario, un impressionista, ci è del tutto alieno, si trova tra noi per puro caso [...]. La città per lui non è delizia, non inebriante gioia, ma crocefissione, Golgota, corona di spine, e ogni visione della città è un chiodo che gli infliggono nel cuore. Egli piange, è preso da isterismo [...]. Non sa neppure, poveretto, confondersi con la sua folla, per cui singhiozza ancor più amaramente. (cit. in Šklovskij 1940: 65 – 66)

Quella del primo Majakovskij è una poesia in cui, soprattutto nei notturni urbani, deformità, cupezza e nausea minano e distruggono canoni e convenzioni: è questo uno dei modi in cui Majakovskij esplicita la tendenza futurista alla de-estetizzazione del reale e alla canonizzazione del Brutto (Vitale 1979: L).

### Una poesia dipinta

Nel 1912 Majakovskij debutta nel volume collettaneo *Poščečina obščestvennomu vkusu* (*Schiaffo al gusto corrente*), con liriche che dimostrano fin da subito la peculiarità della sua produzione; già la sua prima poesia stampata, *Noč* (*Notte*, 1912), è una sorta di quadro futurista, in cui assistiamo allo spostamento dinamico degli oggetti, alla compenetrazione dei piani che fa

entrare una cosa dentro l'altra (la città si insidia dentro ai personaggi e agli oggetti), alla vera e propria commistione *fisica* tra esseri umani e immagini della città.

L'autore unisce un'immaginazione vivida e impressionistica all'impiego reiterato della metafora, che presto diventerà la colonna portante del verso nella generalità della sua produzione lirica.

Бульварам и площади было не странно  
увидеть на зданиях синие тоги.  
И раньше бегущим, как желтые раньб  
огни обручали браслетами ногу.<sup>5</sup>  
(Majakovskij 1955: 33)

L'opera majakovskiana associa un immaginario urbano all'uso di ampie immagini pittoriche, il cui carattere iperbolico, oltre ad anticipare in qualche modo le forme della città surrealista, le rende particolarmente concrete, totalizzanti, portatrici di un effetto altamente emotivo.

Le immagini di Majakovskij [...] sono goffe, non perfettamente calzanti, fanno rumore, invertono. Sono contraddittorie, nei suoi versi scorrono polle di temperatura variabile. (Šklovskij 1940: 93)

L'orientamento verso l'immagine inattesa, in un susseguirsi di visioni urbane vivificate, popolate da insegne di negozi, auto sfreccianti, uomini "disumanizzati" e oggetti personificati, è chiaro sintomo non solo dell'influenza della pittura coeva, ma anche dell'interesse dell'autore per il cinema, arte del movimento e commistione di procedimenti artistici per eccellenza. La poesia è infatti per Majakovskij un'arte viva, pulsante ed essenzialmente teatrale, da proporre in procedimenti grafici originali, tesi a offrire al lettore/ascoltatore la possibilità di un'irripetibile simultaneità di percezione. L'aspirazione del lettore è creare una nuova "forma d'arte totalizzante, arte della parola, del gesto" (Carpi 2014: 18), del segno grafico, del colore, del sentimento.

Il rapporto con la pittura non è soltanto esteriore, ma intrinseco; la rivoluzione tipografica viene propugnata sin dalle prime raccolte futuriste, nelle quali la disposizione di emistichi è "dislocata", in analogia a quanto avviene nei piani della pittura cubista. Come i pittori avanguardisti si servono di sezioni di oggetti e di prospettive accavallate, i poeti futuristi usano parole spezzate, dimezzate e intersezioni simultanee di diversi punti di vista. Il primo Majakovskij traduce nella lingua della poesia il metodo pittorico con immagini generalmente urbane.

Un esempio classico di trasferimento dei principi del cubismo nella fattura del verso è la già citata *Iz ulicy v ulicu*, che secondo Malevič si può considerare come il migliore "esperimento di cubismo in versi"<sup>6</sup> (cit. in Ripellino 1988: 147). La scomposizione delle parole, l'inversione del loro ordine interno corrisponde alle scomposizioni volumetriche dei cubisti. L'effetto visivo è ottenuto dalla segmentazione delle parole sulla pagina, e da una successione di frasi e vocaboli che quasi si leggono allo stesso modo da destra a sinistra e viceversa. La frammentazione e la distorsione di immagini e parole nel verso di Majakovskij richiama qui chiaramente la pittura cubista, dove le immagini vengono frazionate e disposte sulla tela in frammenti e su molteplici piani.

<sup>5</sup> "Per i viali e la piazza non era strano / vedere edifici togati d'azzurro. / E a chi correva avanti, come gialle ferite, / le luci adornavano le gambe di braccialetti". [F.G.]

<sup>6</sup> Citazione originale: "Dice Chardžiev che Malevič riteneva la poesia *Iz ulicy v ulicu* il più riuscito esperimento di cubismo poetico". Ripellino intende qui rinviare all'articolo di N. Chardžiev, "Majakovskij i živopis", in N. Chardžiev, V. Trenin, *Poetičeskaja kul'tura Majakovskogo*, Moskva, Iskusstvo, 1970, p. 48.

*Iz ulicy v ulicu* presenta in realtà una costruzione molto classica, contrariamente a ciò che può sembrare, e lo si nota dalle rime. La poesia mette a nudo i procedimenti iconico-pittorici della poesia majakovskiana: nell'ambito di versi ancora in parte legati al metro sillabo-tonico e in toni e immagini che rimandano alla poesia simbolista (la concezione metaforica del linguaggio, la visione esoterica della parola poetica intesa come chiave d'accesso a una realtà superiore), improvvisamente si assiste alla scomposizione della parola in analogia alla scomposizione della figura nello spazio di ispirazione cubista. La lirica tende all'immagine straniante in un quadro cupo e mobile di ambientazione urbana, popolato da oggetti che paiono vivere di vita propria (*železnye koni* – cavalli di ferro [presumibilmente, tram], le ciminiere - *žirafij risunok*, un disegno di giraffe, *Beguščie doma* – case in corsa, *lysyj fonar'* - un lampione caldo, ecc).

Il brano segue la norma del verso accentuativo, ma il verso è smembrato e disposto in modo peculiare. Violano la regolarità metrica solo i versi 30 - 35, che costituiscono in realtà un unico verso:

|   |   |
|---|---|
| Тело жгут руки.<br>Кричи, не кричи:<br>«Я не хотела!» -<br>резок<br>жгут<br>муки. | Тело жгут руки. / Кричи, не кричи: // «Я не хотела!» - // резок / жгут / муки.<br>2 piedi                    + 2 piedi                    // 2 piedi                    // 2 piedi<br><hr style="width: 80%; margin: 0 auto;"/> = Dattilo 8 piedi |
|---|---|

Troviamo versi molto brevi in apertura (versi 1-10), mentre i versi successivi sono tutti dattili da 3 piedi (l'eccezione è costituita dal verso 16, un dattilo da 4 piedi).

Propongo di seguito una divisione della lirica in dattili graficamente regolare:

|   |   |           |
|---|---|-----------|
| У- / лица. / Лица / у / догов / годов / | - уу / - уу / - уу / - уу /                       | 4р.       |
| рез- / че.                              | - у (у) Verso a sé; dattilo irregolare di 1 piede |           |
| Че- / рез / железных коней /            | - уу / - уу / - уу /                              | 3р.       |
| с окон бегущих домов /                  | - уу / - уу / - уу /                              | 3р.       |
| прыгнули первые кубы. /                 | - уу / - уу / - уу /                              | 3р.       |
| Лебеди шей колокольных, /               | - уу / - уу / - уу /                              | 3р.       |
| гнитесь в силках проводов! /            | - уу / - уу / - уу /                              | 3р.       |
| В небе жирафий рисунок готов /          | - уу / - уу / - уу / - уу /                       | 4р.       |
| выпестрить ржавые чубы. /               | - уу / - уу / - уу /                              | 3р.       |
| Лифт. / Лиф души расстегнули,           | - уу / - уу / - уу                                | 3р. } 5р. |
| Тело жгут руки.                         | - уу / - уу                                       | 2р. }     |
| Кричи, не кричи: «Я не хотела!» -       | - уу / - уу                                       | 2р. } 4р. |
| резок / жгут / муки.                    | - уу / - уу                                       | 2р. }     |
| Ветер колючий / трубе / вырывает        | - уу / - уу                                       | 2р.       |
| дымчатой шерсти клоч.                   | - уу / - уу / - уу / - уу /                       | 4р.       |
| Лысый фонарь / сладострастно снимает    | Dattilo imperfetto                                |           |
| с улицы /черный чулок.                  | - уу / - уу / - уу / - уу /                       | 4р.       |
|   | - уу / - уу / - уу                                | 3р.       |

L'andamento della lirica è tendenzialmente tripodico e tetrapodico con variazioni. Tutti i versi composti da dattili da due piedi possono essere associati al verso da due dattili immediatamente successivo o precedente, con la conseguente formazione di dattili da quattro piedi con rime interne. Con “*vanny.*”, al verso 26, si interrompe la regolarità del dattilo, che riprende poi alla fine del verso 29, con “*rasstegnuli*”. La lettura regolare del verso subisce una variazione ed è come se in questa sezione (versi 26-29) venissero inserite parole in modo casuale, parole interessanti solo come “portatrici di suoni”, riprendendo in tal modo l'artificio iniziale: nei primi 10 versi troviamo infatti rime relativamente semplici e/o classiche (versi 1-3, *U- / lica. / Lica*, il cui primo membro *U- / lica* si presenta in tmesi, a formare una rima perfetta e identica con *Lica*), e rime il cui effetto è incentrato su una specifica resa fonetica. Si vedano come esempio i versi 5 - 6, rimanti con i termini in anagramma *dogov - godov*, o i versi 7 - 10, anch'essi

realizzati con termini in anagramma posti in tmesi, secondo una costruzione a chiasmo (*rez- / če. / Če- rez*), a creare un palindromo perfetto, grafico e sonoro.

L'inserzione di termini "a sé stanti" o di dattili da un piede soltanto, che non sembrano riconducibili a una precisa struttura metrico-compositiva, ha probabilmente la funzione di variare la cadenza intontiva della lirica, mettendone in rilievo determinati elementi e fungendo al tempo stesso da "pausa musicale". Il parallelismo tra la parte iniziale e quella finale della poesia, con il conseguente ripetersi di artifici sonori specifici, dimostra come la struttura sia studiata e appositamente creata.

Nel corpo centrale troviamo rime paronomastiche: si vedano a titolo di esempio i versi 12, 15 e 16, con le forme *domov, provodov* e *gotov* (il cui senso è rafforzato anche a livello lessicale), o la rima *pašni-vašni* ai versi 20 e 24, il cui suono è ulteriormente ripreso, in assonanza, dal termine *vanny* al verso 26.

Nelle sue ricerche sulla rima russa (Samojlov 1982), il poeta David Samojlov nota come nella poesia di Majakovskij la rima rappresenti la colonna portante del verso. Per questo essa è percepita da Majakovskij come un fattore di importanza fondamentale, costitutivo del verso, come una possente fonte di associazioni. A prevalere sono le rime improprie, ma al tempo stesso ricche: non soltanto rimano i suoni o le parole finali, ma versi interi, nei quali vi è coincidenza di gruppi consonantici. Un ruolo notevole è giocato dalle rime composte, un tempo usate nella poesia russa per suscitare un effetto comico. Esse, deformando la parola, conferiscono ai singoli termini un'espressività maggiorata: una tale messa in rilievo di parti di vocaboli rende le parole stesse straniate e "fantastiche".

A una analoga deformazione dei termini linguistici porta anche la rima imparisillaba, basata sulla riduzione delle vocali atone propria della lingua russa, ma anche sull'adattamento delle parole che rimano tra loro.

Majakovskij fa a brandelli le parole e le schiaccia, ne ridistribuisce i confini; una parte, un fonema non ha ormai meno valore della parola stessa. I morfemi entrano in un nuovo rapporto con il "tutto" della parola e tra loro, e al tempo stesso agiscono in modo indipendente. Le unità sonore cominciano a organizzarsi autonomamente, senza dar conto dei significati lessicali tradizionali: le parole vengono ora lette in qualsiasi direzione e in qualsiasi ordine, da sinistra a destra, dalla fine all'inizio, dal centro, in una continua permutazione delle parti. Proprio in questo senso la parola majakovskiana è legata al cubismo: il poeta permette, tramite gli artifici artistici impiegati, di osservarla simultaneamente da ogni lato.

L'autore scompone i segni verbali come fossero cose, oggetti in sé, "volumi", e ne evidenzia la trama fonica, il profilo grafico, quasi per saggiare la loro struttura e resistenza, il dinamismo interno, e condurre al limite estremo la loro espressività. (Ambrogio 1976: 15)

L'intera lirica è cosparsa di gags verbali<sup>7</sup> (Ripellino 1988: 108-109), in cui Majakovskij gioca con significati e accostamenti inaspettati, rimescolamenti di uguali fonemi a formare parole diverse, anagrammi, ripetizioni sonore e palindromi. Oltre ai già citati anagrammi, si vedano i versi 1 – 4, che propongono una ripetizione sonora in chiasmo (*U- / lica. / Lica / u*) e i versi 18 – 19 e 22 (*forel', / syn / ... / ...rel'sy*), che realizzano una rima perfetta paronomastica franta, all'occhio non subito percepibile proprio perché in tmesi. Sono presenti svariate ripetizioni lessicali in rima equivoca, mentre la rima imperfetta ai versi 25 – 26 (*My zavoevany! / Vanny*) risulta particolarmente sofisticata, poiché suona come identica grazie alla difficile distinzione, in lingua russa, tra consonanti semplici e geminate (le consonanti doppie hanno infatti durata minore rispetto a quella caratteristica della lingua italiana).

<sup>7</sup> "Spesso il funambolismo delle immagini-gags di Majakovskij ha una corrispondenza nel burlesco e nella strutturazione visuale delle comiche mute. [...] Il pianista che si impiglia nella tastiera in *Vladimir Majakovskij* [...] è una gag verbale [...]."

Tale orchestrazione fonica è ulteriormente rafforzata lungo l'intera lirica da allitterazioni, anafore (di cui una in costruzione palindroma: *kriči, ne kriči*, verso 31) e assonanze interne ai versi che rendono evidenti i ripetuti richiami sonori.

Majakovskij imprime una svolta pressoché unica alla visione cubista del linguaggio futurista, dimostrando grande abilità nell'inserire il tessuto verbale in un tessuto pittorico in liriche estremamente innovative; anche quando la frammentazione è solo apparente e il poeta compone versi seguendo norme tradizionali, la disposizione del materiale verbale sulla pagina è straniante e presenta un'impostazione grafica non comune, indice di una ricercata volontà di sperimentazione. L'autore esige la dissezione della realtà per esaminarne le caratteristiche in maniera individuale, ma al tempo stesso nel complesso; ogni parte della raffigurazione poetica (o, per meglio dire, pittorico-poetica) è così a se stante, ma contemporaneamente diventa parte del tutto. Dal suddetto procedimento emergono l'acuta coscienza del materiale posseduta da Majakovskij e la sensazione potente della parola e dell'architettura dell'opera verbale.

### BIBLIOGRAFIA

- Majakovskij V. (1955), *Polnoe sobranie sočinenij v trinadcati tomach*, Moskva, Chudožestvennaja literatura.
- Ambrogio I. (a cura di) (1978<sup>2</sup>), *Majakovskij – antologia lirica*, Milano, Edizioni Accademia.
- Ambrogio I. (1976), *Majakovskij*, Roma, Editori Riuniti.
- Buttafava G. (a cura di) (1977), *Per conoscere Majakovskij*, Milano, Arnoldo Mondadori Editori.
- Carpi G. (a cura di) (2014), *Poesie – Vladimir Majakovskij*, Milano, BUR.
- Carpi G. (2010), *Storia della letterature russa. Da Pietro il Grande alla rivoluzione d'Ottobre*, Roma, Carocci editore.
- Casari R. (2008), *La città-macchina e il laboratorio futurista russo*, in Leone M. (a cura di), *La città come testo. Scritture e riscritture urbane*, Santa Rufina di Cittaducale (Ri), Aracne.
- Etkind E. (1989-1991), *La poetica di Majakovskij*, in Etkind E., Nivat G., Serman I., Strada V. (a cura di), *Storia della letteratura russa. Il Novecento*, vol. 2: *La rivoluzione e gli anni Venti*, Torino, Einaudi, pp. 327-339.
- Jakobson R. (1975), *Una generazione che ha dissipato i suoi poeti. Il problema Majakovskij*, a cura di Vittorio Strada, Torino, Einaudi.
- Katanjan V. (1956<sup>3</sup>), *Majakovskij. Literaturnaja chronica*, Moskva, Gosudarstvennoe Izdatel'stvo Hudožestvennoj Literatury.
- Kručnych A., Chlebnikov V. (1913), *Slovo kak takovoe*, Moskva, Tipolit. “Ja. Dankin i Ja. Chomutov” (Digitalizzazione dell'originale).
- Nicholls P. (2000), *La forma e le scritture. Una lettura critica del Modernismo*, trad. di S. Mambrini, Roma, Armando.
- Ripellino A.M. (1988<sup>2</sup>), *L'arte della fuga*, Napoli, Edizioni Guida.
- Ripellino A.M. (1979<sup>4</sup>), *Poesia russa del Novecento*, Milano, Feltrinelli.
- Samojlov D. (1982<sup>2</sup>), *Kniga o russkoj rifme*, Moskva, Chudožestvennaja literatura.
- Šklovskij V. (1940), *Majakovskij*, trad. di Maria Olsoufieva, Milano, Il Saggiatore.
- Vitale S. (1979), *Per conoscere l'avanguardia russa*, Milano, Mondadori.
- Woroszyliński W. (1989-1991), *Vladimir Majakovskij*, in Etkind E., Nivat G., Serman I., Strada V. (a cura di), *Storia della letteratura russa. Il Novecento*, vol. 2: *La rivoluzione e gli anni Venti*, Torino, Einaudi, p. 289-326.

**FRANCESCA GALFIONE** • ha studiato presso l'Università degli Studi di Torino, dove ha discusso una tesi di letteratura russa dal titolo “Rifrazioni di un mondo caleidoscopico: per una lettura stilistica ‘altra’ di V. Majakovskij nel contesto del futurismo russo”. Sta proseguendo il suo percorso di studi presso lo stesso Dipartimento di Lingue e Letterature Straniere e Culture Moderne nel Corso di Laurea Magistrale in Traduzione.

**E-MAIL** • francesca.galfione@gmail.com



# LO SPECCHIO DEL REALISMO

Conoscenza e affettività del testo letterario

---

Pino MENZIO

**ABSTRACT** • An analysis of the metaphor of the mirror confirms that literary realism is not a mere and optically objective reflection of reality, as meant by the more naïve versions of philosophical realism. On the contrary, it's an hermeneutical interpretation of the world, driven by affectivity and emotions.

**KEYWORDS** • Philosophy of Literature, Literary Realism, Hermeneutics, New Realism

Gli ultimi anni, segnati da varie discussioni mediatico-culturali su un possibile ritorno del realismo in filosofia, hanno registrato un certo interesse per la questione anche in sede letteraria. In tal senso, saggi come *Realismo e letteratura* di Federico Bertoni (2007) e *Il realismo è l'impossibile* di Walter Siti (2013) hanno contribuito a ribadire la convinzione che il realismo letterario è una specifica, ben individuata forma della più generale funzione di conoscenza della letteratura; e hanno segnalato che tale conoscenza "realistica" si configura, con sempre maggior chiarezza nel corso dell'evoluzione storica del realismo, quale progressiva decostruzione dell'idea di una riproduzione servile, di un'operazione banalmente e piattamente imitativa: l'opera letteraria, il romanzo o racconto realista non è (non può e non deve essere) una copia diretta della realtà di fatto, ma una sua ri-creazione interpretativa e conoscitiva – il che, come vedremo, implica le più raffinate strategie di dissimulazione della costitutiva eterogeneità del testo letterario rispetto alla realtà fattuale, capaci di rendere apparentemente (illusoriamente) coestensivi l'universo verbale e il mondo extralinguistico. Se quindi, con Siti (2013: 59), «il realismo oppone la realtà alla Realtà», cioè mostra che dietro ai fatti quali si danno usualmente vi è una verità ulteriore, un significato più ricco che la letteratura è chiamata a illuminare, il realismo letterario ha poco a che fare con il realismo filosofico, che guarda invece alla materialità diretta, agli oggetti concreti e alla percezione sensibile come unica forma di conoscenza estetica "vera" (Menzio 2016a).

Non stupisce che Bertoni e Siti, per descrivere l'incrocio tra la funzione conoscitiva del realismo letterario e la sua progressiva, inesorabile presa di distanza dalla realtà empirica, dal darsi dei fatti evenemenziali, immediati e concreti, siano ricorsi anche a un'analisi della metafora dello specchio. Questa immagine ha infatti accompagnato, con declinazioni diverse, l'intero percorso storico del realismo, in ossequio a una tendenza generale che vede la *mimesis* letteraria legarsi preferibilmente alle metafore ottiche e visive, in quanto tramite di una conoscenza netta e trasparente: tra le quali, oltre appunto allo specchio, si segnalano lo strumento ottico o la lente attraverso cui, per Proust, il lettore di un romanzo può leggere in se stesso, ma anche la finestra, il vetro, lo schermo perspicuo, il microscopio, la camera oscura, la macchina fotografica (Bertoni 2007: 91-92). Restando però allo specchio e alla sua valenza descrittiva e interpretativa del realismo letterario, secondo Siti (2013: 13) «bisogna ammettere che si tratta di uno strano specchio, costruito senza vetro e senza pellicola argentata. Con materiali del tutto inadatti allo scopo»: ciò che indubbiamente facilita l'operazione di decostruire il suo impiego ingenuo, quale

---

strumento, immagine o paradigma di una conoscenza obiettiva, neutrale, esente da alterazioni o limiti.

Già nell'Ottocento, infatti, si percepisce chiaramente che la metafora dello specchio ha innanzitutto una funzione retorico-strumentale. Ad esempio, quando Stendhal (nella premessa di *Armance*, nella prefazione a *Lucien Leuwen* o in vari passi de *Il rosso e il nero*) equipara il romanzo ad uno specchio, oppure loda gli artisti olandesi in quanto «pittori-specchi» capaci di riprodurre esattamente la natura, agisce con un intento strategico, per ritagliarsi una zona franca o un alibi preventivo: se la sua opera mostra il fango dell'esistenza, la malvagità degli uomini o la corruzione del mondo, la colpa non è sua (di Stendhal), ma dell'esistenza stessa che egli è tenuto a rispecchiare (Bertoni 2007: 180-181). Tale difesa, per quanto suggestiva e abilmente argomentata, mostra però le corde nel momento in cui si rifletta sul carattere intensivo e selettivo della conoscenza letteraria: quello della letteratura è uno specchio che, inevitabilmente, ritaglia e seleziona specifici nuclei del reale, sin quasi a deformarlo rispetto alla percezione usuale; se si vuole, è uno specchio che presenta al lettore dei frammenti di visione, delle scaglie o nodi densi di significato, *partes pro toto* capaci di incarnare la verità di un personaggio, una società o un'epoca, e con ciò di trasmettere una precisa conoscenza e interpretazione del mondo.

Proprio in ragione di tale carattere selettivo, e quindi conoscitivo, già nell'Ottocento si era intuito che la metafora dello specchio doveva essere rivista e modificata: «Non uno specchio piatto che si limita a riflettere la realtà ma uno “specchio ustorio” (secondo Hugo nella prefazione al *Cromwell*), uno specchio concavo che concentri in una fiamma i raggi sparsi della realtà» (Siti 2013: 50-51): un *miroir de concentration* che, lungi dal restituire un'immagine uniforme, servile e scolorita del mondo, metta a fuoco ciò che in esso è veramente significativo. Nello stesso senso anche Balzac, nella prefazione alla *Peau de chagrin*, parla dell'arte come di uno «specchio concentrico»: non un dispositivo ottico trasparente, neutro e imparziale, ma una sorta di seconda vista che permette all'autore, grazie alla sua fantasia, di «cogliere la verità in tutte le possibili situazioni», se non addirittura di «inventare il vero» (Bertoni 2007: 200, che segnala che tale immagine dello *speculum concentrationis* proviene da Leibniz). Quasi due secoli dopo tali prese d'atto, come ha osservato Orhan Pamuk, risulta a maggior ragione illusoria l'idea «che lo “specchio” dello scrittore (metafora antiquata, questa, del modo in cui il romanzo ritrae, o “riflette”, la realtà) sia perfetto e naturale». Uno specchio letterario davvero conforme, inappuntabile e obiettivo non esiste: «Ci sono solo quelli che soddisfano perfettamente le nostre aspettative, e ogni lettore e lettrice che decide di leggere un romanzo sceglie quello che si confà ai suoi gusti» (Pamuk 2012: 36).

Non solo nella letteratura l'immagine convenzionale, materiale e “realistica” dello specchio, che lo vede come uno strumento otticamente perfetto, desta perplessità, in quanto ne risulta un dispositivo parallelo od omologo, se non addirittura meccanicamente ausiliario, rispetto alla realtà. Certo, l'operazione del rispecchiamento fedele non ha nulla a che vedere con l'incorporazione materiale della realtà nel testo letterario, opzione piuttosto curiosa e tuttavia assai cara ai seguaci italiani di Searle – il che, ancora una volta, mette in crisi qualsiasi tentativo di applicare il realismo filosofico ingenuo alla letteratura (Menzio 2016a). Tuttavia è indubbio che lo specchio così inteso, otticamente fedele e perfetto nel restituire il mondo, non aggiunge nulla al mondo stesso: in particolare, non aggiunge quell'aspetto squisitamente letterario di montaggio, e quindi di conoscenza e interpretazione, del reale che permette di andare oltre la sua materialità oggettiva, di osare di più in termini ermeneutici, di portare alla luce quegli elementi del mondo che non si possono cogliere direttamente, a prima vista. La semplice duplicazione speculare, l'ostensione dell'immediata evidenza, per quanto lateralmente (otticamente) invertita, rischia anzi di ostacolare quella comprensione della realtà che pure è implicita nella “riflessione”: che non è solo, percettologicamente, un'esperienza ottica, ma è anche, metaforicamente, un'esperienza

---



mentale, «il tornare a se stesso del pensiero dopo che si è posato su cose e su concetti e idee di cose» (Rigotti 2007: 51) – operazione che implica un “di più”, quell’elemento interpretativo, combinatorio o creativo che non solo nella letteratura, ma anche nella filosofia, mette capo a un ampliamento del reale, se non a una vera e propria invenzione. In effetti nella storia della filosofia, almeno fino all’affermarsi della modernità, ha largamente dominato l’idea per cui «la nostra mente riflette, come specchio, la natura delle cose, sicché basta affinare lo strumento, renderlo lucido, levigato, *sine macula*, per approssimarsi sempre più alla verità fino a raggiungerla» (Rigotti 2007: 59). Secondo tale realismo tradizionale, che ha tra i suoi nuclei più salienti la scolastica e Cartesio (e che, come ha segnalato Rorty, è entrato in crisi a partire da Kant, configurandosi sempre più irrimediabilmente come realismo ingenuo), «quando la nostra mente avrà riflesso perfettamente le cose, allora conosceremo perfettamente la verità, che non potrà che mostrarcisi in tutta la sua evidenza» (Rigotti 2007: 59): circostanza in cui la filosofia diventerebbe propriamente una scienza, come si augurano Searle e i suoi seguaci.

Tale realismo ingenuo, però, trascura che nessun processo conoscitivo, tanto meno quello filosofico così delineato, può essere davvero integralmente oggettivo e scientifico, in quanto esso: a) è accompagnato, mediato e condizionato dalle modalità cognitive della mente umana, tra cui vi è anche, inevitabilmente, l’adesione a determinati presupposti storico-culturali, quelli appunto in cui tale processo si inserisce; b) la percezione materiale da cui esso parte fornisce dati puramente grezzi ed elementari, sin dall’inizio intrecciati a progetti, speranze, timori, aspettative; e soprattutto c) esso, come qualsiasi altro processo conoscitivo, è inevitabilmente mediato dal linguaggio, come ben sanno i cultori della letteratura: la quale appunto e non casualmente, se guardiamo all’evoluzione della metafora dello specchio in chiave soggettiva (o de-oggettivante), si propone ancora una volta come esempio o modello di ciò che accade in altri campi del sapere. Come è stato sottolineato, infatti, nella creazione letteraria lo specchio è collocato non all’esterno, bensì «*nella mente* dello scrittore: è lo strumento di una percezione mediata e filtrata dal linguaggio, tradotta negli schemi della fantasia poetica, calata nei dispositivi retorici e simbolici della costruzione narrativa» (Bertoni 2007: 208). L’uso di questo specchio, grazie a cui ci si allontana dall’empiria per conoscerla meglio, porta ad una ricostruzione del reale che confina necessariamente con la sua invenzione: in tal senso, per Filippo La Porta (2010: 52), «lo scrittore realista “inventa” sempre una realtà, che però è un’invenzione in cui altri si riconoscono», in cui riescono a cogliere «la logica stessa delle cose, altrimenti opaca o nascosta». Tale ri-creazione del mondo «ci parla di noi, di qualcosa che abbiamo tutti in comune, della verità sommersa del nostro tempo e della nostra condizione»: il che accade di necessità attraverso un “di più” rispetto alla mera riproduzione speculare, alla duplicazione otticamente perfetta o all’immagine fotografica (ammesso che quest’ultima non sia già interpretativa, perché comunque individualizzata e selettiva rispetto alla totalità del mondo). Ovvero, come afferma Gadamer in *Verità e metodo*:

Che cosa sia il riconoscimento [proprio dell’esperienza artistico-letteraria], nella sua essenza più profonda, non lo si capisce se ci si limita a osservare che in esso viene conosciuto di nuovo qualcosa che già si conosce, che il conosciuto viene riconosciuto. Il piacere del riconoscimento consiste piuttosto nel fatto che in esso si conosce *più* di ciò che già si conosceva (Gadamer 1983: 146).

Grazie a questa operazione creativa (o trasposizione conoscitiva) la realtà riceve un incremento d’essere, si presenta con una verità che non possiede nell’esperienza comune. Qualsiasi argomento, tratto dal mondo concreto, dalle vicende quotidiane, dalla storia o dalla mitologia, viene innalzato dalla rappresentazione artistico-letteraria, quando è riuscita, alla sua verità più piena e definitiva: «Dal punto di vista della conoscenza del vero, l’essere della rappresentazione è più che l’essere del materiale rappresentato, l’Achille omerico più che il suo

modello» (Gadamer 1983: 146-147)<sup>1</sup>. Come è evidente, questo “di più” mette fuori gioco qualsiasi idea di *mimesis* intesa come riproduzione speculare otticamente perfetta, come restituzione pedissequa e servile. Anzi, per Gadamer lo specchio ha uno statuto ontologico ancora inferiore a quello della copia. Nell’essenza della copia, infatti, «è implicito che essa non ha altro compito che di uguagliare il modello. Il criterio della sua adeguatezza è che nella copia si riconosca l’originale». L’immagine speculare, viceversa, non è altro che «una pura apparenza, non ha nessun essere effettivo e si può concepire nella sua effimera esistenza solo come dipendente dal fatto di specchiarsi» (Gadamer 1983: 172, 174): se si vuole, è un’immagine fatta apposta per dar ragione al realismo filosofico ingenuo, ponendo tutta l’importanza sull’oggetto di partenza.

Per contro, l’immagine estetica nel vero senso della parola ha un suo essere proprio. Questo suo essere di rappresentazione, cioè appunto ciò in cui essa non si identifica con il raffigurato, le dà, in confronto alla pura copia, il carattere positivo di immagine. Persino le tecniche moderne dell’immagine [fotografia, cinema, video-arte] possono venire usate artisticamente nella misura in cui permettono di trar fuori dal raffigurato qualcosa che non è dato nel suo aspetto comune, nel suo modo di presentarsi puro e semplice. Una immagine di questo tipo non è una pura copia, giacché rappresenta qualcosa che senza di essa non si presenterebbe in quel modo. Essa dice qualcosa di più circa l’originale (Gadamer 1983: 174).

Restando però alla metafora dello specchio, alla sua storia e al suo ruolo nella teoria e nella scrittura letteraria, è ovvio che tale immagine può anche comparire all’interno di un testo narrativo o poetico, acquisendo una specifica valenza poetologica, autoreferenziale o metaletteraria. In tale veste, essa può operare non solo nel senso “realistico” che abbiamo descritto sinora, cioè come indice e tramite del realismo letterario; ma può anche operare nel senso di una particolare astrazione, come accade ad esempio nel *Sonetto in -yx* di Mallarmé. Come è noto, questa poesia disegna un contesto di progressivo, inesorabile svuotamento di qualsiasi significato umano. In una mezzanotte segnata dall’angoscia più pura e tagliente, dopo il fuoco di un tramonto (in buona sostanza, la morte di Dio di Nietzsche) che ha consumato tutti i sogni personali, gli ideali storici, le convinzioni politico-religiose di un’epoca (sogni e illusioni i cui resti non sono raccolti da alcuna «anfora cineraria»), la poesia descrive una «stanza vuota» in cui non si danno né oggetti visibili (la misteriosa «piega»), né suoni, né persone: se è vero che il proprietario della stanza, il Maestro o poeta, è disceso agli Inferi come Orfeo, recando con sé la propria poesia in quanto «solo oggetto di cui si onora il Nulla». Al culmine di questo svuotamento, combustione o nullificazione, come nucleo prospettico e chiave risolutiva di questa impeccabile descrizione poetica del nichilismo contemporaneo, compare appunto, accanto ad una «finestra deserta», l’immagine di uno specchio: la cui cornice dorata, davvero ottocentesca nel suo fregio di nuvole fiammanti, ninfe e liocorni, a propria volta descrive (o simboleggia) il tramonto del sole. Tale cornice rinserta l’«oblio», cioè una nuova nullificazione; e tuttavia nello specchio si riflette qualcosa, lo «splendente settetto» conclusivo, ovvero le sette stelle dell’Orsa Maggiore. Questa immagine in chiusa ha anzitutto un carattere autoreferenziale, giacché il «settetto» sta appunto ad indicare le sette coppie di rime del sonetto stesso; ma l’incidersi delle stelle nel testo indica anche la capacità della poesia di farsi eco, segno o testimonianza di un’arcana armonia del cosmo – il suo compito di rispecchiare la legge che governa l’universo, di sillabare il ritmo misterioso e nitido dei rapporti esistenti nel mondo.

Le sue pure unghie l’onice altissime dedicando,  
L’Angoscia, in questa mezzanotte, lampadofora,

<sup>1</sup> Sulla categoria di verosimile in Aristotele, e sulla re-interpretazione della *mimesis* aristotelica in senso hegeliano da parte di Gadamer, cfr. Menzio 2016a: 31, 41, 48.

Sogni di vespro regge arsi dalla Fenice  
Che anfora cineraria non raccoglie

Sopra gli stipi, nella sala vuota: nessuna piega,  
Abolito gingillo d'inerzia sonora (piani  
Il Maestro andò a attingere allo Stige  
Con quel solo oggetto di cui si onora il Nulla).

Ma presso la finestra a nord deserta, un oro  
Agonizza forse lungo il fregio di liocorni  
Che fuoco avventano contro una ninfa d'acqua,

Ella, defunta nuda nello specchio,  
Pur se nell'oblio che la cornice serra  
S'incide splendente il setto in quell'istante. (Mallarmé 1980: 63)<sup>2</sup>

Come si vede, in questo sonetto la metafora dello specchio, lungi dall'aver una valenza realistica, ottiene un peculiare effetto di astrazione, in quanto l'operazione riflessiva affidata alla poesia non rinvia alla realtà storica, quotidiana e concreta ma alla legge strutturale del cosmo, al ritmo tra i rapporti formali che costituiscono il mondo, al puro insieme delle relazioni fra il tutto. Qualcosa di analogo, anche se in termini inizialmente più concreti, sembra accadere in una poesia giovanile di Franco Fortini, *Giardino allagato* (1950). Qui l'immagine speculare è data da alcune pozzanghere nelle quali, dopo un forte piovasco, si riflette un cielo ancora tempestoso in cui il vento, tra squarci (o «aureole») di luce, muove rapidamente nuvole sature di pioggia. Che questo specchio delle pozzanghere abbia a che fare con la poesia è indicato dal fatto che l'acqua, oltre a riflettere le nuvole, le aperture di luce e il volto del poeta, «rispecchia» (simbolizza, metaforizza) anche «i miei vecchi / segni»: ovvero le parole poetiche, gli elementi linguistici che propriamente costituiscono e attivano lo sguardo della poesia. Tale sguardo e tali parole, con costitutiva *pietas*, «portano / ai colori» e a nuova vita luoghi perduti nella memoria, eventi ormai velati dal trascorrere degli anni; e soprattutto, in virtù di imitazione o *mimesis*, creano una verità poetica che si affianca a quella, tangibile, delle cose concrete («i miei vecchi / segni [...] quali altri veri a fianco / delle vere cose imitano?»). È dunque vero il cielo in tempesta, ed è vera la sua immagine poetica («Doppia figura, aperto // segreto e come credo / egualmente a due cieli?»).

E tuttavia, come per le stelle di Mallarmé, viene da domandarsi se anche questo cielo fortiniano battuto dal vento non possa avere un significato ulteriore, una valenza simbolica o metaforica più ampia: se, ad esempio, non rinvii a quella tempesta del progresso economico, politico e sociale che, nella *Nona Tesi* di Benjamin (1995: 80), travolge l'angelo della Storia impedendogli di sostare; o se invece, con maggiore ottimismo e plausibilità filologica, non simbolizzi un processo storico che attraverso lacerazioni, distruzioni e traumi («Le fronde brune rotte giù / a capogiro in fondo / alle pozze di pioggia») porta a un rasserenamento, ad un cielo finalmente sgombro di nubi – a propria volta immagine dell'utopia politica realizzata, ovvero pegno di una salvezza a carattere religioso.

Le fronde brune rotte giù  
a capogiro in fondo  
alle pozze di pioggia  
dove il vento dev'ia  
il volumi dei nuvoli

<sup>2</sup> La traduzione è leggermente modificata.

e l'acqua che rispecchia  
tra le aureole i miei vecchi  
segni e le pupille

quali luoghi persi portano  
ai colori quale velo  
quali altri veri a fianco  
delle vere cose imitano?  
Pende intanto questo viso  
sulla greggia in transito. E vedo  
me sul limite e il riso  
mio che un fiato incrina è incerto.  
Doppia figura, aperto

segreto e come credo  
egualmente a due cieli?  
Questo è il mio bene? È questo  
che dai primi anni ad ora  
m'accora e mi consola?  
La sera vola, scrolla  
fronde, una goccia annulla  
l'attimo, ombra è il giardino.  
E riprendo il cammino. (Fortini 1978: 98)

La metafora dello specchio, nel suo impiego in ambito letterario, risulta quindi particolarmente flessibile, indicando sia nel senso di una resa realistica del mondo, sia nel senso dell'astrazione. Da un altro e connesso punto di vista, guardando in generale ai due possibili estremi semantici di ogni messaggio, l'immagine (il simbolo, la metafora) dello specchio può innanzitutto rinviare al nulla. Questa ipotesi, non troppo distante dalla «sala vuota», la «finestra deserta» e l'«oblio che la cornice serra» di Mallarmé, si verifica ad esempio nelle arti figurative, in particolare nelle cosiddette *vanitates* barocche: nature morte in cui, insieme ad altre allegorie della caducità umana, vi è spesso anche uno specchio che non riflette nulla, ovvero riflette il nulla che è l'osservatore. Questo effetto di abolizione o nullificazione richiama alla mente la memorabile chiusa de *Gli anelli di Saturno* di Sebald – dove, dopo un intenso percorso di esplorazione, insieme geografica (l'estremo sud-est dell'Inghilterra) e spirituale, dell'entropia del mondo moderno, nel segno della malinconia e di una partecipazione costante e sobria, l'autore ritorna circolarmente al punto di partenza, ovvero agli scritti del trattatista barocco Thomas Browne. Anche qui compare l'immagine di uno specchio muto, oscurato mediante un velo di seta nera: con gesto di estrema *pietas* nei confronti dell'anima morente, perché non possa cogliere il riflesso del proprio dissolversi nel nulla.

E Thomas Browne, [...] in un passo del suo libro *Pseudodoxia Epidemica* – passo che al momento non mi riesce di ritrovare – ricorda come ai suoi tempi in Olanda, nelle case in cui era morto qualcuno, fosse usanza coprire con un crespo di seta nera tutti gli specchi e tutti i dipinti dove erano raffigurati paesaggi, persone o i frutti dei campi, affinché l'anima in procinto di lasciare il corpo non venisse distratta, durante il suo ultimo viaggio, dalla vista della propria immagine o da quella della sua terra natale, ormai perduta per sempre (Sebald 2010: 306).

Sempre in termini semantici generali, il contrario di tale nulla dovrebbe essere il tutto – o meglio, nei nostri anni, quella particolare variante tardo-moderna del tutto che è l'infinitamente

molteplice. Anche in questo caso la metafora dello specchio esercita il proprio ruolo esplicativo e interpretativo, con quel peculiare effetto di derealizzazione che si ha quando la riproduzione speculare è multipla: ovvero quando lo scrittore, facendo sembrare del tutto vera una cosa, ma al contempo suggerendo il sospetto (o la certezza) che sia falsa, crea un meccanismo in cui «la gibbigiana delle menzogne può essere un corridoio di specchi senza fine» (Siti 2013: 37). Non necessariamente tale impianto narrativo dischiude una prospettiva postmodernista, se è vero che il «potenziamento del riflesso» che si ha quando gli specchi sono plurimi e si moltiplicano tra loro, producendo un numero di immagini potenzialmente infinito, ha viceversa operato, «in Plotino, Porfirio, Macrobio, nello Pseudo-Dionigi come pure in Nicolò Cusano, come modello mentale di un'ontologia neoplatonica, in cui gli ordini di realtà si presentano come riflessi sempre più tenui dell'ineffabile luce divina» (Rigotti 2007: 54-55).

Ancora una volta, riassuntivamente, la metafora letteraria dello specchio appare particolarmente flessibile nella sua capacità semantica, se è vero che può indicare verso la realtà storica concreta o verso l'astrazione, verso il nulla, il tutto o l'infinitamente molteplice. È però escluso che possa rinviare, in modo ingenuo, ad una rappresentazione-copia del reale, a una riproduzione scientifica ed esatta del mondo esterno – che è appunto quanto affermava Lukács nell'atto di distinguere la *Widerspiegelung* artistica da quella scientifica: «Al pari di quello gnoseologico, il rispecchiamento estetico vuole comprendere, scoprire e riprodurre con i suoi mezzi specifici la totalità della realtà nella sua ricchezza dispiegata di contenuti e di forme. Modificando decisamente [...] il processo soggettivo, esso suscita mutamenti qualitativi nell'immagine riflessa del mondo» (Lukács 1957: 146-147). Ovvero ancora una volta, per ritornare alle osservazioni di Siti (2013: 15-16), nell'esperienza letteraria «il realismo non è la storia di un rispecchiamento piatto e subalterno, ma di uno svelamento impossibile»: è un'impresa conoscitiva che respinge i luoghi comuni, le rappresentazioni stereotipate, i personaggi scontati o prevedibili, le vicende di maniera, i cliché di genere e sottogenere, il sentimentalismo come contraffazione dell'affettività – e che al contrario, mirando sostanzialmente all'impossibile, cerca di cogliere la realtà integrale, quale essa è per davvero. Proprio in tal senso di conoscenza totale, per Siti, il realismo letterario rappresenta una sfida alla natura e alla storia, e per questo non può allontanarsene troppo: non può cioè trasgredire le leggi naturali, i vincoli cronologici, la coerenza materiale o logica degli eventi, e deve invece giocare sul terreno della realtà di fatto, sovvertendola o eludendola sottilmente, a fini conoscitivi, secondo un principio di scostamento minimo. Il che inganna i lettori meno raffinati, convinti che la narrazione realista “che funziona” è quella che riprende (o incorpora) i fatti così come si danno direttamente, nella loro concreta immediatezza – in altri termini, quei lettori o filosofi che non avvertono che l'incorporazione letteraria è una metafora, e non un fatto reale.

Se da un lato, da Saussure in avanti, la scrittura letteraria ha un carattere eterogeneo, non omologo rispetto al reale; e se, dall'altro lato, è una forma di conoscenza e interpretazione del reale stesso, se ne deduce ovviamente che tale conoscenza sarà non-percettologica: «Tra cogliere le percezioni in contropiede e ingannare i sensi, il passo è breve: l'occhio è più primitivo della mente» (Siti 2013: 14-15), con la conseguenza che attraverso immagini analogiche o simboliche, cioè attraverso rappresentazioni letterarie fittizie, percettologicamente (inevitabilmente) non corrette, si può far comprendere qualcosa del reale che rimarrebbe altrimenti nascosto – e che rimane nascosto al percettologo *hard*, al realista ingenuo convinto che la conoscenza estetica “vera” è solo quella direttamente trasmessa dai sensi. In questa prospettiva la crisi epistemica otto-novecentesca, cioè il progressivo declino del paradigma mimetico tradizionale, che vuole l'arte come copia esatta della natura, mette capo a un'idea dell'opera letteraria come raffinata, consapevole illusione ottica. «La magia della letteratura è racchiusa in ciò che i vecchi artigiani fiorentini ancora chiamano “pittura d'inganno”, e i francesi da sempre *trompe-l'œil*» (Siti 2013:

15): è racchiusa in una rappresentazione del mondo fittizia, non coincidente con l'oggetto, interpretativa, creativa, metaforica, simbolica – e tuttavia diretta alla conoscenza del mondo stesso: una conoscenza inevitabilmente puntiforme, non integrale né esaustiva, diretta a «veder brillare nei freddi sterrati della realtà qualche fiore di Realtà» (Siti 2013: 61).

Tale coesistenza di eterogeneità e conoscenza dà corpo, in Siti, a un apparato pirotecnico di esempi che tengono insieme i due aspetti, con affermazioni (piccoli capolavori di eleganza e finezza) che sembrano pensate apposta per innervosire i percettologi *hard*: ad esempio, «quando Vulcano forgia le armi di Achille, l'illusione della vita è così perfetta che Omero descrive gli altorilievi metallici con lo stesso grado di finzionalità con cui descrive le battaglie reali» (Siti 2013: 14). Né manca una celere casistica di tutti gli strumenti letterari adoperati per tale finzionalità conoscitiva: stile, ritmo, meccanismi di contenuto e di contesto (*in primis* l'insistenza sui dettagli concreti e insignificanti, che appunto inganna i lettori più ingenui), incroci semantici, giochi col paratesto ecc. Tali strumenti determinano una serie di “effetti di reale” in senso barthesiano: in altri termini, sono procedimenti o convenzioni che danno l'impressione di un legame diretto tra referente e significante, ma che, al contrario, propriamente ri-creano la realtà di fatto, la traspongono e dislocano nell'universo verbale, con finalità conoscitive rispetto ad essa. Al contempo funzione diretta e modalità (inevitabilmente) indiretta di conoscenza del mondo, «la tecnica realista è un inseguimento infinito a rappresentare zone sempre più nascoste e proibite della realtà, impiegando artifici sempre più sofisticati e illusionistici» (Siti 2013: 22): ovvero usando quelli che lo strutturalismo ha variamente definito come operatori realisti, connotatori di mimesi, tramiti di illusione realista, referenziale, mimetica ecc. (Bertoni 2007: 331-332). In termini non strutturalisti, se si preferisce, come ha affermato Mario Vargas Llosa (2011: 30), «la letteratura è una rappresentazione fallace della vita che, tuttavia, ci aiuta a capirla meglio, a orientarci in quel labirinto in cui nasciamo, viviamo e moriamo. [...] Grazie a essa riusciamo a decifrare, per lo meno parzialmente, quel geroglifico che è l'esistenza per la maggior parte degli esseri umani».

Come si sarà notato, Siti sembra quasi suggerire che, in questa raffinata e indispensabile impresa conoscitiva, gli scrittori realisti siano più “finzionali” dei simbolisti – o lo siano ugualmente, ma in un modo più sottile e inappariscente: appunto per l'abilità, il virtuosismo, il talento tecnico e la perizia combinatoria con cui riescono ad occultare i propri artifici espressivi, trasmettendo al lettore l'impressione (o meglio la certezza) di avere a che fare con un mondo saldo, coerente e del tutto omologo rispetto a quello reale: mentre invece gli oggetti, personaggi, azioni, eventi, ambienti, situazioni e problemi che lo costituiscono sono integralmente ricreati e rielaborati – ancora una volta, non sono dati di fatto immediati, vicende che sembrano essersi scritte da sé, ma costruzioni verbali che dissimulano sapientemente le tracce della loro impervia, defatigante produzione. Non stupisce che il campione di questa consapevolezza sia considerato Flaubert, che più volte, nel suo epistolario, sottolinea come non vi sia nulla di più macchinoso e “costruito” della naturalezza letteraria. La realtà (o l'effetto di realtà) del romanzo è infatti il risultato di un *tour de force* tecnico-formale, di un lavoro strenuamente metodico e applicativo: è un'illusione, suscitata innanzitutto dall'impersonalità dell'opera, cioè da quell'apparente eclissi del narratore, fittizia e simulata, in cui la mano dell'autore resta invisibile. Il che, per Flaubert, si compendia nel principio secondo cui «la prima qualità dell'Arte è l'*illusione*», che è peraltro «la vera verità» a cui l'opera narrativa può giungere (Bertoni 2007: 244-246).

Sulla falsariga di Flaubert, il più coerente dei suoi discepoli, Maupassant, affermava che «i Realisti di talento dovrebbero chiamarsi piuttosto Illusionisti», in quanto “rendere vero” ciò che viene descritto è una sorta di gioco di prestigio, una suggestione mesmerica che consiste appunto nel dare l'illusione completa del vero attraverso una composizione così abile, dissimulata e apparentemente semplice da rendere impossibile percepirne e individuarne il disegno (Siti 2013:

14). Con intenzione analoga, sempre secondo la ricostruzione di Siti, anche Chesterton, nel lodare le descrizioni dickensiane, notava che in esse le cose paiono più reali e presenti di quanto non siano abitualmente, al punto da dar corpo ad un grado di realismo quasi onirico, che non esiste nella realtà di fatto. Quest'ultima rischia anzi di risultare, dal punto di vista artistico-letterario, poco credibile: così, secondo un aneddoto verosimilmente spurio ma molto illuminante, il pubblico dell'ultima recita del *Malato immaginario* di Molière avrebbe lamentato che questi in scena "era morto male", quando aveva avuto uno sbocco di sangue che l'avrebbe fatto morire per davvero, poche ore dopo. Secondo Virginia Woolf, il realismo propriamente efficace è quello che si dispone ad evocare uno scatto, un gesto fulmineo, un'illuminazione improvvisa capace di dare senso al Tutto: è l'epifania di un significato che senza ragione, in circostanze anche banali, investe le persone e le trasmuta in simboli. Anzi, quanto più una cosa, persona o circostanza è descritta con nitore e precisione, tanto più è idonea ad acquisire un carattere simbolico (Siti 2013: 49, 25, 53). Il realismo letterario compie così una sorta di "miracolo della presenza", con quelle parole folgoranti che sono tramite del *blitzhaftes Erkenntnis* di Benjamin (2000: 510 [N 1, 1]): parole che, cogliendo dettagli altrimenti perduti nel flusso della consuetudine, azzerano le banali distinzioni percettive e illuminano così il mistero, la realtà profonda, la ragione ultima delle cose.

Ancora una volta, il realismo letterario è irrevocabilmente legato alla conoscenza, si definisce come una specifica forma o funzione conoscitiva del reale, nel *continuum* dei suoi livelli di significazione ma anche con un implicito rinvio alla sua ulteriorità: «La rappresentazione della realtà è efficace se sembra nascondere sempre un altro strato della realtà», più profondo o verticale – o se quella descritta è «una realtà frugata per rivelarne la mancanza, l'inadeguatezza a una luce superiore» (Siti 2013: 25, 60), il suo disporsi in uno spazio segnato dall'orma vuota di Dio. Ciò è vero al punto che, per Siti, si può cogliere una traccia di realismo anche in quelle scritture non mimetiche (cioè dichiaratamente finzionali o irrealistiche) che tuttavia mirano alla conoscenza, vogliono raggiungere e afferrare un nucleo profondo della realtà. Dietro il loro presentarsi, all'uno o all'altro estremo delle possibilità letterarie, come raffinate creazioni autoreferenziali (si pensi al *Sonetto in -yx* di Mallarmé) o come affabulazioni ludiche, evasive o di puro intrattenimento, esse colgono spesso nodi individuali o dilemmi collettivi difficili da circoscrivere con gli strumenti della mimesi più tradizionale, appunto perché ancora informi, inesplorati o impropriamente situati – e così si propongono come «scritture che, proprio per questo, spesso appaiono profetiche» (Siti 2013: 37). Per Siti, un esempio di tale dislocamento della funzione conoscitiva anche nei contesti artistici più strani, imprevedibili e capovolti è dato dalla pittura iperrealista, che ad un primo livello (quello che guarda all'effetto simil-fotografico) sembra un trionfo dell'artificio fine a se stesso, di una capacità di raffigurare il reale così talentuosa e virtuosistica, in termini visivi, da rendere l'immagine indistinguibile dal reale stesso. Ad uno sguardo più approfondito, però, si nota che gli oggetti rappresentati hanno contorni troppo nitidi, sono troppo delineati e precisi per essere davvero credibili. Nel suo proporre una realtà più vera del vero, quindi, tale pittura riflette piuttosto l'irrealtà del mondo contemporaneo: che va appunto compresa attraverso l'arte, e non rimossa fuggendo grottescamente verso gli oggetti, la cosalità e la percettologia *hard*.

A proposito della funzione conoscitiva della letteratura, abbiamo già avuto occasione di suggerire che quella trasmessa dai romanzi, i racconti e le poesie è una conoscenza affettivamente connotata – e quindi, come tale, strettamente connessa alla soggettività dell'individuo, sia esso l'autore o il lettore del testo: è una conoscenza che si esplica in modi partecipativi e comprensivi, attraverso una specifica vicinanza sentimentale ed emotiva nei confronti delle persone, cose ed eventi raffigurati (Menziò 2010). Di solito, nell'illustrare questa tesi, si rinvia al pensiero di Martha Nussbaum, che nell'importante volume *Il giudizio del poeta* (1995) ha appunto sottolineato il carattere empatico e identificativo della conoscenza letteraria, che porta il lettore a

mettersi nei panni dei vari personaggi, rivivendo dall'interno le loro esperienze affettive ed emozionali: il che promuove la sua apertura umana, l'eliminazione delle sue unilateralità ideologiche, la sua pluralizzazione come soggetto. Ma la tesi che la letteratura sia una forma di conoscenza affettivamente connotata non emerge solo nel pensiero anglosassone, ma è ben radicata anche nella tradizione continentale. Così ad esempio, nella *Critica del Giudizio* di Kant, che pure non attribuisce all'esperienza estetica un valore direttamente conoscitivo, il bello (ovvero l'opera letteraria riuscita) suscita «un sentimento d'amore e intima inclinazione»; in altri termini, grazie alla propria connotazione affettiva, «il bello ci prepara ad amare qualche cosa, anche la natura, senza interesse» (Kant 1997: 217, 209). Nell'*Estetica* di Hegel, viceversa, è esplicito il riconoscimento che l'opera artistica, in quanto manifestazione sensibile dell'idea, è una forma di conoscenza del reale; ma è altrettanto evidente che tale conoscenza ha carattere affettivo e partecipativo, in quanto «agli occhi della concezione e configurazione poetica ogni parte, ogni momento deve essere per sé interessante, per sé vivente, ed essa indugia quindi con gioia nel singolo, lo dipinge con amore e lo tratta come una totalità per sé» (Hegel 1997: II, 1097). In ciò l'esperienza artistica e letteraria si contrappone al pensiero filosofico, che, trascurando le particolarità individuali delle cose, accorpa tutte le manifestazioni della vita, della realtà e del mondo in categorie generali e astratte, e con ciò le volatilizza. Sempre in ambito continentale, un'analogia valorizzazione della letteratura come conoscenza affettivamente connotata si ha anche nello storicismo tedesco di fine Ottocento, con la sua distinzione tra il *Verstehen* delle scienze umane (in cui rientra, ovviamente, anche la letteratura) e l'*Erklären* delle scienze naturali. Mentre il metodo di queste ultime richiede il distanziamento dell'osservatore, il *Verstehen* umanistico-letterario è una comprensione di indole partecipativa e immedesimativa, fondata sulla *sympatheia*, il coinvolgimento umano, l'apertura di un orizzonte dialogico in cui l'Altro (il mondo, le persone, le cose) non può mai essere un oggetto a disposizione, liberamente manipolabile (Menzio 2010).

Il carattere affettivo della conoscenza letteraria è, ovviamente, anche al centro delle riflessioni degli scrittori e dei poeti. Il che vale, non da ultimo, per Siti, se è vero che in lui la simpatia del narratore realista per gli oggetti concreti è «la devozione-per-il-mondo di chi nel mondo non sa vivere» (Siti 2013: 49): è un tratto esplicitamente affettivo che testimonia un'incertezza esistenziale, una vera e propria mancanza d'essere, che riesce a trovare negli oggetti un paradossale sollievo. Per l'affettività della letteratura, però, uno dei paradigmi poetico-culturali più rilevanti è costituito da Rilke: che, in alcuni punti cruciali del suo epistolario, analizza la pittura di Cézanne in quanto esempio più forte e persuasivo di un amore artistico, e quindi poetico, rivolto indistintamente a tutte le forme del reale, senza operare preferenze o distinzioni di sorta, né appellarsi a gerarchie valoriali collettive o personali. Tale adesione al mondo nella sua interezza, con il suo corollario di attenzione, devozione micrologica e cura creaturale, rappresenta per Rilke il principio (cioè il momento iniziale e il nucleo concettuale primario) di una poetica della vicinanza: è «la semplice vita di un amore che si è costituito, che, senza affatto vantarsene, si accosta a tutto, senza accompagnamenti, discreto, silenzioso. Il lavoro vero e proprio, la pienezza dei compiti, tutto inizia solo dietro a questo costituirsi» (Rilke 1950: I, 208)<sup>3</sup>: nel senso più stretto, la creazione letteraria può iniziare *solo dopo* l'attivarsi della dimensione affettiva e partecipativa della conoscenza. Appunto in questi termini, la letteratura si propone come modello di una posizione etica più ampia, che eccede i ruoli di scrittore o lettore, ed opera nella realtà complessiva della vita, dei rapporti quotidiani e dell'esperienza sociale (Menzio 2010).

Nell'enunciare questo modello etico-letterario in termini generici, c'è senza dubbio il rischio che esso sembri una posizione latamente retorica, oppure il frutto di un automatismo del pensiero,

<sup>3</sup> Lettera a Clara Rilke del 19 ottobre 1907. Per un'analisi più ampia di queste posizioni rilkiane cfr. Menzio 2013: 258-260.



la trasposizione un po' rapida di principi teorici individuati in un contesto assai diverso. Per evitare questi pericoli, merita chiarire tale posizione etico-letteraria più nel dettaglio. Essa suggerisce innanzitutto di porsi di fronte al mondo in un atteggiamento di conoscenza – o meglio, dato che non è possibile conoscere *tutto* il mondo, la realtà, gli eventi, le persone ecc., suggerisce di porsi in un atteggiamento di costitutiva apertura verso la possibilità della conoscenza. Questa disposizione indebolisce o esclude il giudizio immediato, e consente di attivare (o di ammettere come sempre attivabile) un percorso di approfondimento, che porti ad una integrazione, correzione o scomparsa dei propri inevitabili pregiudizi – o che eventualmente li confermi, ma solo come il risultato di una loro mobilitazione o verifica. In secondo luogo, questo modello etico-letterario suggerisce che tale conoscenza (meglio, che ogni conoscenza) non sia mai separata dalla dimensione affettiva, sentimentale ed emozionale: e ciò tanto dal lato della persona conoscente, chiamata a tener conto della propria soggettività nel conoscere il mondo, quanto dal lato delle persone (etnie, società, religioni, classi ecc.) eventualmente conosciute, di cui occorre sforzarsi di comprendere gli affetti, le ragioni, le paure, le motivazioni personali – il che approfondisce e arricchisce ogni processo conoscitivo, e soprattutto lo allontana da quella tipica veste “oggettiva” delle tecnoscienze e dell'economia, che in realtà mira a funzionalizzare ogni aspetto del mondo in chiave produttiva. Infine, il modello proposto dalla letteratura suggerisce che la tonalità affettiva fondamentale di ogni conoscenza sia la *pietas*, cioè la consapevolezza di quella radicale caducità creaturale che accomuna gli umani, la natura e il mondo: il che dovrebbe portare ad un atteggiamento di cura, premura, dedizione e amore, e quindi ad una diminuzione della violenza individuale e collettiva.

Viene senz'altro spontaneo obiettare che quella così delineata è una posizione buonista, irenica e consolatoria, di facile presa retorica e di comoda applicazione saggistica: ma non è affatto detto che le cose stiano così. Infatti, come abbiamo avuto occasione di notare in merito a Robert Walser (Menziò 2016b), è assai probabile che un atteggiamento del genere aumenti di molto la vulnerabilità del soggetto, in quanto lo espone nel modo più immediato ai traumi, le ferite e i disinganni di una realtà che, molto spesso, sembra compiacersi di infrangere ogni possibile illusione sulla bontà del mondo. Così, risulta assai più difeso dai fallimenti della storia l'individuo scettico e disincantato, il conservatore che vede nell'hegeliana «immane potenza del negativo» una conferma dei propri presupposti pessimistici, piuttosto che il progressista che viene periodicamente, se non quotidianamente, smentito nella propria fiducia (o ingenuità) storicistica nel migliorare delle umane sorti: il che porta a supporre che il cinismo sia, in fondo, una forma di difesa preventiva rispetto agli urti e alle ferite che l'esperienza storica porta di regola con sé. E tuttavia la vulnerabilità (la “debolezza” in senso esistenziale, e non solo filosofico) implicita nel modello che stiamo analizzando, può implicare un peculiare arricchimento, ottenuto, come spesso accade, a un certo prezzo: quello di una conoscenza che, per dar conto delle contraddizioni del reale, tiene conto delle contraddizioni proprie, cioè della necessità di rimanere (o di rifarsi ogni volta, secondo la poetica ungarettiana) *consapevolmente ingenua* di fronte all'esperienza – se si vuole, della necessità di pagare il prezzo delle proprie ferite per rimanere a contatto con il mondo, in tutti i modi in cui esso si esplica, e per averne una comprensione più profonda.

## BIBLIOGRAFIA

- Benjamin, W. (1995), *Tesi di filosofia della storia*, in *Angelus Novus. Saggi e frammenti*, a cura di Renato Solmi, Torino, Einaudi: 75-86
- Benjamin, W. (2000), *I «passages» di Parigi*, a cura di Rolf Tiedemann e Enrico Ganni, tr. it. di Renato Solmi, Antonella Moscati, Massimo De Carolis, Giuseppe Russo, Gianni Carchia, Francesco Porzio, Torino, Einaudi

- 
- Bertoni, F. (2007), *Realismo e letteratura. Una storia possibile*, Torino, Einaudi
- Fortini, F. (1978), *Una volta per sempre. Poesie 1938-1973*, Torino, Einaudi
- Gadamer, H. G. (1983), *Verità e metodo*, a cura di Gianni Vattimo, Milano, Bompiani
- Hegel, G. W. F. (1997), *Estetica*, tr. it. di Nicolao Merker e Nicola Vaccaro, Torino, Einaudi
- Kant, I. (1997), *Critica del Giudizio*, tr. it. di Alfredo Gargiulo, Roma-Bari, Laterza
- La Porta, F. (2010), *Meno letteratura, per favore!*, Torino, Bollati Boringhieri
- Lukács, G. (1957), *Prolegomeni a un'estetica marxista. Sulla categoria della particolarità*, tr. it. di Fausto Codino e Mazzino Montinari, Roma, Editori Riuniti
- Mallarmé, S. (1980), *Sonetti*, a cura di Cosimo Ortosta, Milano, Guanda
- Menzio, P. (2010), *Nel darsi della pagina. Un'etica della scrittura letteraria*, Torino, Libreria Stampatori
- Menzio, P. (2013), *Della poesia come amore per il mondo. Riflessioni a partire da Letteratura ed etica di Tzvetan Todorov*, in "Enthymema", VIII: 258-272, <http://riviste.unimi.it/index.php/enthymema/article/view/2884>
- Menzio, P. (2016a), *Realismo letterario e realismo filosofico. Per il chiarimento di un equivoco*, in "Enthymema", XIV: 26-50, <http://riviste.unimi.it/index.php/enthymema/article/view/6954>
- Menzio, P. (2016b), *Una forma del pensiero. Il saggio come spazio comune tra filosofia e letteratura*, in corso di stampa
- Pamuk, O. (2012), *Romanzieri ingenui e sentimentali*, tr. it. di Anna Nadotti, Torino, Einaudi
- Rigotti, F. (2007), *Il pensiero delle cose*, Milano, Apogeo
- Rilke, R. M. (1950), *Briefe*, Wiesbaden, Insel-Verlag
- Sebald, W. G. (2010), *Gli anelli di Saturno*, tr. it. di Ada Vigliani, Milano, Adelphi
- Siti, W. (2013), *Il realismo è l'impossibile*, Roma, Nottetempo
- Vargas Llosa, M. (2011), *Elogio della lettura e della finzione*, tr. it. di Paolo Collo, Torino, Einaudi

**PINO MENZIO** • He has written about Greek tragedy (*Prometeo, sofferenza e partecipazione*, 1992), the metaphor of travelling (*Il viaggio dei filosofi*, 1994) and the relationship between ethics and aesthetics (*Orientarsi nella metropoli. Walter Benjamin e il compito dell'artista*, 2002; *Da Baudelaire al limite estetico. Etica e letteratura nella riflessione francese*, 2008; *Nel darsi della pagina. Un'etica della scrittura letteraria*, 2010). He is editor of the website «Etica e letteratura».

**E-MAIL** • [pino.menzio@fastwebnet.it](mailto:pino.menzio@fastwebnet.it)

# FRASEOLOGIA ITALIANA PER APPRENDENTI L<sub>2</sub> E SPECIALISTI

La pratica dell'approccio costruzionista\*

---

Riccardo IMPERIALE, Elmar SCHAFFROTH

**ABSTRACT** • Mastering phraseology is a key issue to both native and non-native speakers of a language. Teaching lexical combinations proves to be necessary if learners aim at developing their communicative language skills, especially from B1 to C2 levels.

This paper proposes a possible way of presenting the Italian phraseology based on the principles of Construction Grammar. In particular, the Piattaforma per chi impara l'italiano offers a wide range of pieces of information, which are not usually provided by traditional lexicographic references, especially as far as the pragmatic and semantic aspects of phraseology are concerned.

**KEYWORDS** • Phraseology, Italian L2, Lexicography, Language Teaching

## 1. Introduzione

Ettinger (2013) ci riporta le sue esperienze da studente in Francia quando cercava di utilizzare i fraseologismi del francese che aveva imparato. Parla di un'impresa difficile e di un'arma a doppio taglio. Alcuni dei suoi interlocutori erano stupiti, altri piuttosto infastiditi. Tali sentimenti misti sono dovuti a vari fattori. Da una parte sorprende il fatto che uno straniero padroneggi la fraseologia di una lingua seconda<sup>1</sup>, dal momento che questo settore linguistico è considerato una delle più grandi difficoltà in una lingua. Inoltre, il fatto che un parlante non madrelingua si cimenti con i segreti fraseologici di una lingua straniera è spesso giudicato secondo il principio del *Quod licet Iovi non licet bovi*. D'altra parte è probabile che una vera e propria padronanza dei fenomeni fraseologici sia estremamente rara, anche perché gli stessi parlanti nativi hanno un sapere linguistico limitato delle locuzioni idiomatiche e di altri tipi di fraseologismi<sup>2</sup> della propria lingua madre, in relazione al livello socioculturale e al talento retorico individuale.

---

\* Ringraziamo la prof.ssa Carla Marelli per suggerimenti importanti e per averci dato la possibilità di pubblicare il nostro articolo in questa rivista, la dott.ssa Elisa Corino per la collaborazione alla realizzazione del testo, e i due revisori per le loro osservazioni e sollecitazioni, inoltre Erica Di Blanca e Silvia Caravaggi per la consulenza linguistica.

1 Di seguito il termine lingua seconda o L2 comprende i concetti 'lingua straniera' e 'lingua seconda'.

2 Per una classificazione dei fraseologismi si veda Makkai (1972), Mel'čuk (1995, 2012, 2013), Baranov/Dobrovol'skij (1996), Fleischer (1997), Burger (2015), Gries (2008) e Schafroth (2015), per la terminologia eterogenea si veda Burger et al. (2007), e per la definizione del concetto 'espressione idiomatica' Casadei (1995a) e, in sintesi, Giacoma (2012). A proposito della differenziazione delle espressioni idiomatiche in *idiomatically combining expressions* (ad es. *to spill the beans*) e *idiomatic phrases* (ad es. *to kick the bucket*) si consulti Nunberg/Sag/Wasow (1994) e Croft (2001: 179-185, pass.), con osservazioni critiche Schafroth (2014).

---

Due domande si impongono:

1. A che cosa serve allora una fraseodidattica per una lingua straniera o una lingua seconda?
2. Come si dovrebbe articolare tale fraseodidattica?

La risposta alla prima domanda è facile. Siccome in una lingua il numero delle combinazioni lessicali imprevedibili (o non interamente prevedibili) e non liberamente formabili è più grande del numero di “costruzioni libere”<sup>3</sup> e di conseguenza in ogni testo, in ogni discorso e in ogni conversazione si trova qualche tipo di ‘discorso ripetuto’ non-composizionale, è indispensabile una competenza linguistica passiva dei più frequenti e più importanti fraseologismi in L2. Se, in più, le esigenze degli apprendenti L2 avanzati li portano a volersi esprimere in modo competente e *near-native*, una buona conoscenza della fraseologia dovrebbe anche far parte della loro competenza linguistica attiva. Tenendo conto del potenziale retorico e pragmatico delle locuzioni idiomatiche si capisce che il settore della fraseologia, inteso in senso lato, non può essere escluso dall’insegnamento di una L2 dal livello B1 al C2<sup>4</sup>.

La risposta alla seconda domanda è molto più difficile. È vero che non mancano tentativi di evidenziare i principali problemi e proporre soluzioni (cfr. Ettinger 2007, Meunier/Granger (a c. di) 2008, Hallsteinsdóttir/Winzer-Kiontke/Laskowski 2011), ma finora non esiste alcun manuale didattico di fraseologia per una lingua europea che sia adatto all’apprendimento. Libri come Götz/Lorenz (2002) per l’inglese, Bárdosi/Ettinger/Stölting (2003) per il francese, Ettinger/Nunes (2006) per il portoghese e Hessky/Ettinger (1997) per il tedesco<sup>5</sup> sono importanti passi nella direzione giusta, ma la descrizione del sapere linguistico a proposito delle locuzioni idiomatiche ivi trattate è troppo scarsa. Ciò vale ancora di più per i tanti dizionari e raccolte fraseologici<sup>6</sup>, fra cui – e ora passiamo alla lingua italiana – Lurati (2001), Pittàno (2009) e Quartu/Rossi (2012) sono i più noti<sup>7</sup>, ma nessuno di questi è in grado di descrivere la complessità linguistica dei fraseologismi nell’uso quotidiano.

Il concetto dell’uso quotidiano, certo, non è privo di problemi teorici e metodologici. È ovvio che i dati che i linguisti ricavano dai corpora debbano essere trattati con cautela, non solo riguardo alla lingua parlata (cfr. De Mauro 1994, Sornicola 1994, Cresti/Panunzi 2013, Lemnitzer/Zinsmeister 2015), ma anche a quella scritta, soprattutto della stampa quotidiana. Non tutte le attestazioni che le analisi di fonti come *Cosmas II*, *Corpus CORIS*, *La Repubblica Corpus* ci forniscono sono affidabili: si può naturalmente sempre trattare di errori linguistici o tipografici,

---

3 Cfr. *l’idiom principle* di Sinclair (1991: 110ff.), il detto di Hausmann (1993: 471) della non-apprendibilità del lessico di una lingua straniera, la convinzione di Gross (1988: 22) che le “phrases figées” in una lingua sono “plus nombreuses que les phrases libres” e Nuccorini (2007a: 698) secondo cui “non-compositionality is a general and widespread characteristic of language and it is not confined to idiomatic expressions and complex lexemes only as their most relevant defining feature”.

4 Il *Quadro comune europeo di riferimento per la conoscenza delle lingue* (QCER) non si esprime sul ruolo della fraseologia nell’insegnamento di una lingua straniera. La soluzione di questa questione fondamentale è uno degli obiettivi del progetto di ricerca FRAME all’Università degli Studi di Milano. Non possiamo escludere che già al livello A2 si possano (e si debbano) insegnare alcune collocazioni e formule di saluto e di altre situazioni quotidiane. Nella ricerca linguistico-didattica, ad ogni modo, non esistono ancora risultati e raccomandazioni a proposito dell’uso di fraseologismi nell’insegnamento di una L2.

5 Il libro è esaurito. Sul sito di Ettinger (2009b) si trova una versione fortemente ampliata.

6 Fenati/Rovere/Schemann (2009) è un’opera pubblicata in Italia. È il più ricco dizionario fraseologico tedesco-italiano.

7 Lo Cascio (2013) e Blumenthal/Rovere (2016) invece non sono dizionari fraseologici, ma forniscono molta fraseologia in contesto e danno indicazioni importanti sull’uso quotidiano.

o di occasionalismi e di creazioni spontanee (cfr. Dardano 1974, 2012, Beccaria 2002, Adamo/Della Valle 2005, De Mauro 2005, Loporcaro 2006). Con particolare attenzione devono essere inoltre verificate le attestazioni rilevate in rete, come in *WEBBIT*. Gli autori di questo articolo sono consci delle trappole che costituiscono i corpora e cercano di evitare gli scogli dell'aleatorio, dell'effimero, della confusione e della contaminazione. L'*uso quotidiano* si può quindi solo riflettere 1°: nella frequenza attestata nei più grandi e in vari tipi di corpora di una lingua (la soglia di frequenza essendo ancora da definirsi); 2°: nella conferma, in casi di dubbio, ricevuta tramite una verifica tra parlanti madrelingua; 3°: (il più debole criterio però) attraverso la lessicografia monolingue. Il GRADIT, la grande opera lessicografica di 350.000 lemmi, porta nel suo titolo ciò che intende descrivere: l'*italiano dell'uso*. Questo obiettivo è prevalentemente realizzato sulla base di due criteri, tramite i quali vengono analizzate fonti scritte e "parlate" (parlate-scritte, per essere esatti): quello della frequenza e della disponibilità (cfr. Corda/Marello 2004: 31ss.). Il risultato di questo lavoro permette di categorizzare una grande parte del lessico descritto nel GRADIT secondo le marche d'uso FO (fondamentale), AU (di alto uso), AD (di alta disponibilità), CO (comune), BU (di basso uso) o OB (obsoleto) (De Mauro 2000: XVII; cfr. Nuccorini 2007a: 700)<sup>8</sup>. Il fatto però che a quasi tutte le espressioni idiomatiche sia assegnata la categoria *comune* ("sono così marcati i vocaboli che sono usati e compresi indipendentemente dalla professione o mestiere che si esercita o dalla collocazione regionale e che sono generalmente noti a chiunque abbia un livello mediosuperiore di istruzione") ci lascia insoddisfatti perché dalle analisi di vari corpora dell'italiano emerge un quadro di grande eterogeneità, sia in riferimento alla frequenza, ma anche al contesto in cui i fraseologismi vengono usati. Questo quadro non identifica un pari livello di "usualità" per espressioni idiomatiche come *mettere la/una pulce nell'orecchio*, *farsi le ossa*, *scendere in campo*, *arrampicarsi sugli specchi*, *menare il can per l'aia*, *prendere in giro*, *avere un diavolo per capello*, *prendere fischi per fiaschi*, o *perfino prendere per il culo*, *farsi il culo* – tutte *comuni* in De Mauro (2000). Spogliare corpora per ottenere risultati differenziati a proposito della frequenza di un'espressione idiomatica ci pare, di conseguenza, uno strumento metodologico indispensabile (cfr. infra, nota 12).

La lessicografia generale italiana rivela un quadro simile per quanto riguarda il livello linguistico della descrizione e la quantità di informazioni (cfr. Schafroth 2013a). Chi ci dice in maniera affidabile quali sono le varianti di una locuzione idiomatica (per es. *trovare* e *avere oltre dare del filo da torcere*)? Chi ci informa su elementi facoltativi (per es. *del* nel nostro esempio) o su altri elementi che si possono interporre (come *dare più/molto/non poco filo da torcere*)? Dove possiamo trovare informazioni sugli argomenti sintattici di queste locuzioni idiomatiche e dei loro ruoli semantici? Dov'è menzionato con quali lessemi un fraseologismo si combina prevalentemente (*essere in grado/capace di*, *potere*, *sapere*, *riuscire a dare del filo da torcere*)?<sup>9</sup> E quali sono i "significati contestuali", la motivazione (pragmatica) per utilizzarlo, le particolarità

8 Lasciamo a parte marche d'uso come TS (tecnico-specialistico), LE (di uso solo letterario) e RE (regionale), anch'esse specchio indiretto del grado di frequenza.

9 Ci riferiamo esplicitamente alla lessicografia e fraseografia e non alla linguistica o alla fraseologia. Nuccorini (2007a: 701), ad esempio, prende in considerazione l'aspetto di collocatori combinabili con fraseologismi, come *patata bollente* o *dare retta*. L'autrice fa anche riferimento all'importanza di elementi "connotativi", qui nel caso di *dar retta*, come una parola negativa, l'inserimento in una costruzione ipotetica o l'uso frequente di *bisogna*. In definitiva, Nuccorini denuncia la mancanza di studi che concentrino l'attenzione su, appunto, aspetti sintagmatici e semantico-pragmatici come vengono descritti nella *Piattaforma*.

stilistiche e retoriche? La lessicografia convenzionale non è in grado di fornire tali informazioni<sup>10</sup>, che possono essere utili anche per madrelingua e specialisti come insegnanti d'italiano nelle scuole superiori, traduttori e interpreti, studiosi d'italianistica, in particolar modo di linguistica italiana e altri.

Siccome il criterio dello spazio limitato vale anche per i dizionari retrodigitalizzati, cioè le versioni digitali di dizionari cartacei già esistenti (cfr. Marengo 1996, Aprile 2008, Schafroth 2009, Cacia/Papa/Verdiani 2013, Della Valle/Patota 2016), non c'è da aspettarsi una rivoluzione lessicografica nel trattamento della fraseologia. Il dizionario fraseologico ideale di cui avrebbero bisogno tanto madrelingua quanto apprendenti L2, deve dunque essere caratterizzato dalla possibilità di mettere a disposizione un minimo di informazioni necessarie non solo per capire ma anche per utilizzare un fraseologismo in modo linguisticamente competente ed efficace, in relazione con gli obiettivi pragmatici della comunicazione, e il numero più alto possibile di esempi e di attestazioni, tutti estratti dall'uso linguistico autentico, scritto e parlato. Il formato di una tale descrizione può solo essere digitale – e con *digitale* intendiamo *neodigitalizzato*, cioè ideato in modo nuovo secondo le moderne tecniche informatiche e messo in rete per gli utenti.

A questo punto sorge la domanda: quale teoria linguistica può essere in grado di costituire la base metodologica per una tale impresa? In questo articolo descriviamo il nostro approccio all'analisi olistica dell'uso di fraseologismi italiani. Essi sono ampiamente illustrati con esempi, studi di caso e spiegazioni approfondite, poiché ciascuna delle locuzioni idiomatiche trattate viene descritta sia nella “modalità base” sia nella “modalità dettagliata” (cap. 4). Con ciò ci proponiamo di corrispondere il più possibile alle esigenze linguistiche e didattiche di apprendenti di italiano L2. Il nostro metodo si iscrive nel progetto di linguistica applicata *Piattaforma per chi studia italiano* dell'Istituto di romanistica dell'università di Düsseldorf, che verrà delineato nel capitolo 3. Nel capitolo 2 presentiamo la proficua applicazione dell'approccio costruzionista, cioè della Grammatica delle Costruzioni, alla fraseologia, che è alla base delle nostre ricerche. Nell'ultimo capitolo (5) daremo una breve sintesi.

## 2. La fraseologia e la Grammatica delle Costruzioni

Il trattamento dei fraseologismi nella Grammatica Generativa (GG) come fenomeni anomali può essere considerato uno dei principali motivi per la nascita della Grammatica delle Costruzioni (CxG). Secondo la GG, per poter spiegare perché *mettersi l'anima in pace* significa ‘rassegnarsi’ e *menare il can per l'aia* ‘temporeggiare’, ci vogliono regole particolari di interpretazione per tener conto del loro significato non compositivo o del loro comportamento sintattico irregolare, una specie di appendice alla grammatica, quindi. Nel loro articolo indicativo del 1988 Fillmore/Kay/O'Connor caldeggiavano un approccio completamente diverso: l'appendice non è l'eccezione, ma la regola. Propongono dunque un tipo di grammatica funzionalista “olistica”, che includa tutto il sapere linguistico necessario affinché un parlante possa capire e produrre un numero potenzialmente infinito di locuzioni linguistiche, tra cui anche i fraseologismi. Tra queste informazioni linguistiche ci sono non solo quelle morfologiche, sintattiche e semantiche, ma anche quelle prosodiche, pragmatiche e discorsive (cfr. Croft 2001):

---

<sup>10</sup> Questo non è un rimprovero, ma si vuole qui semplicemente sottolineare che questo medium ha i suoi limiti naturali, o, come giustamente affermato da uno dei revisori: “il lessicografo che sulla base di analisi di attestazioni estratte da corpora elettronici (con un'analisi, cioè, corpus-driven), ambisse a una descrizione di decine di migliaia di fraseologismi, pur potendo lavorare in condizioni ideali, dovrebbe avere una prospettiva di vita assai lunga. In tal caso, però, i mutamenti linguistici che interverrebbero tra l'inizio e la fine dell'impresa lessicografica, finirebbero per inficiare la validità di quanto elaborato”.

[...] in knowing how to use their language, speakers know how to create and recognize associations between semantically interpreted sentences and particular types of situations. Such PRAGMATIC knowledge uses but does not contribute to semantic interpretation (Fillmore/Kay/O'Connor 1988: 502).

Il compito di una grammatica è pertanto la descrizione di tutto ciò che è caratterizzato da peculiarità formali e/o semantico-pragmatiche che sono in qualche modo idiosincratiche, cioè né prevedibili né deducibili da altri segni linguistici o da principi generali.

It can be hoped that the structure-building principles of the so-called core and the machinery for building the phraseological units of the kind discussed in this paper may be of a uniform type, the former being a degenerate instance of the latter (Fillmore/Kay/O'Connor 1988: 534).

Le unità fraseologiche sono dunque fin dall'inizio al centro dell'interesse della CxG<sup>11</sup>. I principi della loro descrizione hanno carattere strutturale, nel senso che essi valgono sia per le strutture regolari (*core*) sia per quelle irregolari (*peripheral constructions*).

Descrivere una “costruzione periferica”, cioè combinazioni lessicali che, in qualche modo, non siano né composizionali, né deducibili né prevedibili, significa quindi indicare tutto il sapere linguistico necessario per capire interamente una tale costruzione e per utilizzarla in modo competente ed efficace. I diversi filoni della CxG (cfr. Ziem/Lasch 2013, Hoffmann/Trousdale (a c. di) 2013) hanno descritto modelli teorici, ma non hanno prodotto alcuna metodologia sistematica per applicare tali modelli alla prassi linguistica. È vero che gli approcci *corpus-based* (cfr. Goldberg 1995, 2006, Croft 2001, riassunti da Ziem/Lasch 2013) hanno illustrato la loro teoria con esempi pertinenti, oramai famosi, quello che manca però è un'attuazione approfondita della CxG con l'applicazione di tutti i principi di cui parlano Fillmore/Kay/O'Connor (1988) a un settore della grammatica o del lessico (cfr. Schafroth 2014).

Colmare questa lacuna è l'obiettivo del gruppo di ricerca sulla fraseologia italiana all'Istituto di romanistica dell'Università di Düsseldorf. Una parte delle sue attività, descritte nel capitolo successivo, consiste nella descrizione contrastiva, secondo i principi della CxG, delle 600 locuzioni idiomatiche più frequenti e correnti della lingua italiana<sup>12</sup>. A una tale “fraseologia complessiva” contribuiscono ovviamente anche i dizionari monolingui, bilingui e specializzati

---

11 Anche se nel frattempo la CxG non si focalizza più sulla fraseologia (Dobrovolskij 2011: 119), ma assegna la priorità all'analisi di strutture sintattiche (Ziem 2015: 45), un'elaborazione dell'approccio costruzionista al livello fraseologico in rapporto alla prassi lessicografica non è mai stata effettuata.

<sup>12</sup> L'analisi di **frequenza** è effettuata sulla base dei cinque corpora PAISÀ, WEBBIT, *La Repubblica*, *CORIS* e *itTenTen*: per ognuna delle 2030 espressioni idiomatiche (EI), rilevate dal repertorio fraseologico di Quartu/Rossi (2012), viene calcolata la media aritmetica. Siccome il 60% di queste EI dispone solo di un significato (fraseologico) (o di più significati fraseologici), come *arrampicarsi sugli specchi*, e il 40% comprende anche un significato “letterale” (come *alzare un polverone*), questa relazione viene considerata per il calcolo delle più frequenti EI sottoposte all'esame quantitativo (360 della categoria esclusivamente fraseologica, 240 della categoria “mista”). Il secondo criterio è la **disponibilità**, che chiamiamo, seguendo Hallsteinsdóttir/Šajánková/Quasthoff (2006), *Geläufigkeit* (giudizio metalinguistico sull'uso corrente di un'EI). Le 600 EI vengono valutate, prima dal gruppo del progetto (10 madrelingua) secondo i criteri seguenti: la EI in questione è conosciuta o sconosciuta, viene utilizzata attivamente o no, ha una o più varianti. In tal modo si potrà evitare che il tipo di testo dominante dei corpora (quello dei testi giornalistici) possa falsare il “valore comunicativo” di un'EI tanto da far contraddire frequenza e consapevolezza del fraseologismo da parte dei parlanti. Nel caso in cui ci fossero tali contraddizioni bisognerà consultare un gruppo più grande di madrelingua.

(fraseologici e pedagogici), ma la maggior parte del sapere linguistico proviene dalle analisi effettuate sulla base di corpora linguistici dell'italiano (cfr. Crocco 2015), come PAISÀ, WEBBIT, Corpus "La Repubblica" e CORIS.

L'idea di esaminare attentamente l'uso linguistico autentico, com'è documentato ad esempio nella stampa, non è nuova<sup>13</sup>. Lo troviamo ad esempio in Ettinger (2009a), oppure in Rovere (2003) e (2009), due analisi, basate su corpora, di espressioni idiomatiche italiane e tedesche effettuate in prospettiva contrastiva, con una proposta metodologica elaborata (Rovere 2003) che mira a inglobare in modo adeguato il potenziale semantico di fraseologismi utilizzati in testi autentici, per contribuire a migliorare la qualità degli articoli in dizionari bilingui<sup>14</sup>.

Le riflessioni di Mel'čuk (1995: 168) vanno già nella direzione del nostro approccio. Egli si chiede: "What should be stated about the given idiom (or phraseme) in its linguistic description for this idiom to be correctly selected and used in speech?" La risposta però viene data solo dal punto di vista della teoria "senso-testo": come si può riuscire a esprimere i "frasemi" sulla base delle funzioni lessicali? Siccome queste funzioni, come *Magn* (intensificazione) o *Oper<sub>1</sub>* (presenza di un verbo supporto in una collocazione), sono in rapporto semantico-lessicale paradigmatico o sintagmatico con un dato "vocabolo" (lemma), dovrebbe essere possibile descrivere un'espressione idiomatica come *coucher en joue* ('appoggiare il calcio del fucile alla guancia (per mirare)'/ 'puntare') tramite una funzione lessicale, partendo dal vocabolo *fucile*. Secondo Mel'čuk (2008) questa funzione sarebbe *Prepar<sup>II</sup>Real<sub>1</sub>*: il fucile viene preparato per realizzare l'obiettivo (di sparare). Sarebbe auspicabile che il *Lexique actif du français* (Mel'čuk/Polguère 2007) esplicitasse in linea di principio questi nessi tra il significato del lemma (o della "lexie") e quello del fraseologismo, ma si limita a elencare le forme dei frasemi nella microstruttura<sup>15</sup>.

Un progetto fraseologico quindi, con analisi sistematiche, integrali e dettagliate, basate su corpora linguistici, prima non esisteva. Per modellizzare il sapere linguistico in riferimento alla fraseologia, bisognava trovare un formato lessicografico digitale che permettesse di integrare tutti i livelli linguistici considerabili e di connettere informazioni tra di loro. Il formato, ideato secondo i principi della CxG e della *semantics of understanding* di Fillmore (1985), è chiamato "fraseoframe". Esso prevede tutti i descrittori ritenuti necessari, organizzandoli secondo i *feature structures* di un diagramma di costruzioni generico (Fried/Östman 2004: 26, 30), vale a dire 'attributi' e 'valori' (*attributes* e *values*), per es. "funzione sintattica / frase" o "campi semantici / contrarietà, impazienza" per l'esempio di *Cosa me lo chiedi a fare?*<sup>16</sup> Inoltre sono elencate le attestazioni raccolte nei corpora per illustrare le particolarità linguistiche di una locuzione idiomatica.

Il termine *frame* è ispirato da Fillmore (1982) e si riferisce all'inclusione di tutte le caratteristiche linguistiche di un fraseologismo in un "quadro epistemico", chiamato *fraseoframe*. In analogia all'evocazione di un *frame* (attraverso un lessema) si può dire che, con la realizzazione fonica o grafica di un fraseologismo, vengono attivati anche gli elementi epistemici semantici e

<sup>13</sup> Sulla variabilità lessico-sintattica delle espressioni idiomatiche nell'italiano parlato si veda Casadei (1995b).

<sup>14</sup> Sui fraseologismi in alcuni dizionari pedagogici si veda Nuccorini (2007b), sui dizionari pedagogici (italiani) da varie prospettive Giacomini/Rovere (2009) e Schafroth (2011).

<sup>15</sup> Per le collocazioni invece questi nessi vengono spiegati sulla base di funzioni lessicali: ad es. in Polguère (2003) e Mel'čuk (2011).

<sup>16</sup> La maggior parte dei descrittori viene però rappresentata in maniera dettagliata, come i principali significati del fraseologismo (secondo l'uso linguistico), la motivazione pragmatica, la sintassi interna e esterna e le note d'uso.



pragmatici di cui dispongono in genere i madrelingua, ma non gli apprendenti L2. Questi elementi epistemici meritano particolare attenzione e sono elaborati nei *fraseoframe* (v. cap. 4). Inoltre, non solo i *fraseoframe* appartenenti a uno stesso campo semantico – come RIMPROVERO, IMPAZIENZA, RABBIA – possono essere connessi (tramite link) l'uno con l'altro, ma anche gli elementi del “thesaurus lessicale” e del “thesaurus fraseologico”, cioè rappresentazioni onomasiologiche in base al concetto del fraseologismo in questione, possono essere collegati con altri *fraseoframe* in cui pure appaiono<sup>17</sup>.

Un esempio elaborato di *fraseoframe*, con riferimento al fraseologismo a schema fisso *Che me lo chiedi a fare?*<sup>18</sup>, è accessibile online su Schafroth (2013b); un altro esempio tratta la locuzione idiomatica *saperla lunga* e può essere consultato, pure *online*, in Imperiale (2016). Un numero di circa 100 locuzioni idiomatiche italiane, analizzate in modo dettagliato tramite il modello del *fraseoframe*, è disponibile sulla *Piattaforma per chi studia italiano* (Schafroth (dir.) 2014–), nella rubrica *Phraseme*.

### 3. Obiettivi della Piattaforma per chi studia italiano

Obiettivo primario della *Piattaforma* è la creazione di una banca dati, utilizzabile nel tempo, su fenomeni della lingua italiana (come L2) rilevanti e specifici del processo di insegnamento e apprendimento, con cinque diverse “interfacce” o componenti a cui corrispondono progetti parziali diversificati, ma riconducibili a un progetto d'insieme.

Questa banca dati costituisce una documentazione estesa, e unica in questa forma, di una pluralità di fenomeni della lingua italiana ripartiti nei seguenti cinque ambiti:

1. aspetti di *grammatica* italiana rilevanti per l'insegnamento o apprendimento e particolarmente complicati dal punto di vista tedesco, per es. pronomi, comparativo o congiuntivo;
2. casi di dubbio sulla *norma* linguistica dell'italiano (un oggetto d'analisi in genere del tutto trascurato nel percorso di studi in italianistica);
3. *lessico*: “articoli modello” di un Dizionario per apprendenti di italiano L2;
4. sapere linguistico sui *fraseologismi* dell'italiano (con i *fraseoframe*);
5. sapere linguistico sulle particolarità comunicative e pragmatiche dei *segnali discorsivi* italiani (con considerazione delle problematiche traduttive) – teoria e metodo basati sugli approcci della grammatica delle costruzioni.

Gli studenti imparano a scoprire e possono apprendere l'italiano come lingua straniera non solo da una fonte di informazione (per es. una grammatica o un dizionario come leo.org, pons.de o lingostudy.de), bensì tracciando in chiave *compilativa* – consultando cioè tutte le opere più importanti relative al tema trattato –, un quadro d'insieme su un ampio spettro di forme di

17 Un campo semantico costituisce un *frame*. Gli elementi linguistici e extralinguistici di un tale *frame* sono riuniti nel sapere linguistico dei parlanti, per es. RIMPROVERO, basato sulle loro esperienze; essi sono rappresentati da lessemi, fraseologismi, strutture sintattiche, conoscenze enciclopediche, ecc.

18 Altri esempi sarebbero *cosa ci va a fare?* o *cosa lo compri a fare?*, che dimostrano i cambi richiesti nei pronomi dalla costruzione di ciascun verbo.

rappresentazione metalinguistica e su differenti approcci grammaticografici e lessicografici, nonché opinioni contrastanti in merito a fenomeni specifici.

Riguardo al fenomeno considerato (per es. il congiuntivo in italiano) il lavoro compilativo aiuta, per di più, a delineare una sorta di “metagrammatica” o “grammatica della grammatica”, nonché a mettere gli studenti a conoscenza della *discussione sulla norma* dell’italiano e del dibattito critico della linguistica in Italia.

Questa visione d’insieme è garantita innanzitutto dall’analisi di portali di consultazione linguistica (cfr. Accademia della Crusca, Zanichelli o Treccani), di *blog* sulla lingua o *mailing list* come italiano\_L2.

Gli studenti imparano inoltre come descrivere, in modo olistico, fenomeni grammaticali e lessicali (per es. locuzioni idiomatiche o “phraseotemplates” cioè costruzioni schematiche con una struttura sintattica abbastanza fissa, ma lessicalmente aperta)<sup>19</sup>, attraverso le metodologie della *linguistica dei corpora*, contribuendo con ciò a realizzare una descrizione rilevante, comprensibile, e finora mai documentata in questa modalità, di come funziona un fenomeno – anche in senso pragmatico e discorsivo.

Imparano per di più, sulla base del lavoro compilativo e del sapere linguistico, approfondito attraverso ricerche sui corpora, a formulare *proposte autonome centrate sul soggetto che apprende* (per es. articoli di dizionari e aspetti grammaticali).

*Tutti gli utenti* possono trarre profitto inoltre dall’approccio olistico-descrittivo della piattaforma, poiché mai prima d’ora sono stati presentati *fraseologismi, segnali discorsivi e variazione normativa* della lingua italiana in modo così dettagliato. Chiunque desiderasse sapere in quali situazioni e con quale motivazione e scopo i madrelingua italiani fanno uso di idiomi come *arrampicarsi sugli specchi* o con quali effetti comunicativi e pragmatici un segnale discorsivo come *capirai* venga usato in italiano, non rimarrà deluso dalla *Lernerplattform Italienisch*.

Nel corso degli ultimi anni è stata data una maggiore importanza alla descrizione dei fraseologismi, con il risultato di un progetto di ricerca presentato alla Deutsche Forschungsgemeinschaft in settembre 2016.

Nel prossimo capitolo spieghiamo come funziona l’applicazione dell’approccio olistico della Grammatica delle Costruzioni alla fraseologia. Presentiamo descrizioni esaustive per i più importanti descrittori del *fraseoframe*, differenziando per ciascuna delle locuzioni idiomatiche trattate tra “modalità base” e “modalità dettagliata”.

#### **4. Alcuni esempi: locuzioni idiomatiche (modalità base e modalità dettagliata)**

##### **4.1. Modalità base e modalità dettagliata**

La modalità base e la modalità dettagliata si differenziano principalmente per la rispettiva presenza o assenza di un approccio didattico. Categorie tipiche della modalità base come *consigli per l’uso, motivazione e sintassi esterna* si rivolgono, infatti, a chi, non essendo madrelingua,

---

<sup>19</sup> Si veda l’esempio di *Cosa me lo chiedi a fare?* I *phraseotemplates*, in tedesco *Phraseoschablonen* o *Phrasem-Konstruktionen*, che in questo articolo sono anche menzionati come *fraseologismi a schema fisso*. A proposito dei fraseologismi comparativi, anch’essi “Phrasemkonstruktionen” (ad es. *essere contento come una pasqua*), si veda Mollica/Schafroth (in corso di stampa).

desidera sapere con che intenzione illocutiva e con quali effetti perlocutivi usare un determinato fraseologismo, oppure come costruirlo sintatticamente. La modalità dettagliata invece si distingue principalmente per un ulteriore approfondimento delle informazioni e per la trattazione di concetti linguistici come quello dei ruoli semantici e la teoria degli atti linguistici. Si rivolge quindi a chi voglia specificare le sue conoscenze su un determinato fraseologismo anche da un punto di vista teorico.

In seguito, distinguendo le due modalità d'uso e i principali livelli linguistici (sintattico, semantico, pragmatico), saranno presentati alcuni esempi di fraseologismi italiani analizzati secondo il concetto del *fraseoframe*, esposto nel capitolo precedente. Alla fine di ogni capitolo sarà proposto un *fraseoframe* esemplare. Le analisi si basano su dati empirici tratti da alcuni dei principali corpora dell'italiano: PAISÀ, WEBBIT, Corpus "La Repubblica", CORIS.

## 4.2. Sintassi interna ed esterna

### 4.2.1 Modalità base

La modalità base dispone di diverse categorie per fornire all'apprendente informazioni sulla variabilità della sintassi interna e sulle caratteristiche principali della sintassi esterna di una locuzione idiomatica.

Per 'variabilità della sintassi interna' si intendono fondamentalmente i seguenti fattori:

- (1) a. la sostituibilità degli elementi strutturali (=ES)
- b. la declinabilità degli ES
- c. l'obbligatorietà/la facoltatività degli ES
- d. la combinabilità degli ES

Un esempio di sostituibilità è dato dal fraseologismo *alzare un polverone*, in cui il verbo finito *alzare* può essere sostituito con diverse varianti. Oltre all'altrettanto frequente *sollevare*, infatti, sono altresì ammissibili (seppur un po' meno frequenti<sup>20</sup>) *fare*, *suscitare*, *montare* e *creare*. Anche l'articolo indeterminativo *un* può essere sostituito: per esempio con la forma determinativa *il* o con l'aggettivo *questo*. Il fraseologismo è inoltre variabile a livello morfologico (declinabilità), e non solo per quanto riguarda il verbo finito: è attestata, infatti, la forma plurale *polveroni* che a livello semantico sembra implicare una generalizzazione (l'*alzare polveroni* viene descritto come un'azione tipica dell'agente)<sup>21</sup>.

L'ES *polverone/i* è da considerarsi al contempo l'unico ES obbligatorio della locuzione idiomatica. Vengono considerati obbligatori quegli ES che non possono essere omessi senza che l'uso diventi non-idiomatico o sgrammaticato. Nel caso di *alzare un polverone* è possibile l'omissione del verbo finito e quella dell'articolo (in caso di uso ellittico anche contemporaneamente), cfr.: "*Sembra di capire che tra legale e assistito sia stato stretto un patto chiaro: niente **polveroni**, niente argomenti che non riguardino direttamente la sua difesa.*" (La

20 Differenze riguardo alla frequenza delle singole varianti d'uso sono molto rilevanti e vanno segnalate, per esempio attraverso l'impiego di parentesi tonde per le varianti meno frequenti, cfr.: [alzare/sollevare/(fare)/(suscitare)/(montare)/(creare)] *un polverone*.

<sup>21</sup> Cfr. "E anche se - come d'abitudine - faranno un gran casino e **alzeranno polveroni**, è evidente per tutti che sono fra i principali responsabili della mancata vittoria dell'Unione: [...]" (WEBBIT; grassetto degli autori).

Repubblica; grassetto degli autori<sup>22</sup>). Il verbo finito e l'articolo sono da considerarsi quindi ES facoltativi<sup>23</sup>.

Un caso di variabilità molto pronunciata è rappresentato da *tenere il piede in due staffe*, di cui pressoché ogni ES è sia sostituibile che declinabile (solo *piede* non è sostituibile). Inoltre la locuzione è rilevante in termini di combinabilità in quanto i suoi ES non sono liberamente combinabili tra loro. Ad esempio, usando l'ES *staffe* (oppure la variante moderna *scarpe*) nella forma singolare (*una staffa/scarpa*), l'ES *piede* deve assumere necessariamente la forma plurale ed essere preceduto dal numerale *due* (*tenere due piedi in una staffa/scarpa*): la variante *tenere \*i piedi in una staffa/scarpa* non è attestata.

Infine per quanto riguarda la modalità base la sintassi interna comprende la categoria dei *modificatori*: si tratta di elementi come *tirare troppo la corda, non stare più nella pelle* etc. che servono a modificare semanticamente uno o più degli ES e che sono fortemente ancorati alla struttura interna del fraseologismo.

Prima di passare alla sintassi esterna, è necessario soffermarsi su un altro aspetto che si trova, per così dire, a metà tra sintassi interna ed esterna, e che, visto l'alto grado di idiosincrasia, è molto rilevante in ottica didattica: le valenze grammaticali e lessicali<sup>24</sup>. Si tratta di informazioni indispensabili per sapere come integrare una locuzione idiomatica nella frase. In generale è importante la differenziazione tra valenze obbligatorie e valenze facoltative; queste ultime possono essere rappresentate attraverso l'impiego di parentesi: *farsi le ossa (su qualcosa)*. A volte l'uso di una valenza è caratteristico di una specifica accezione idiomatica, come nel caso di *fare cilecca con qualcuno* che corrisponde quasi sempre all'accezione 'manifestare un'occasionale impotenza sessuale' (GRADIT 2000).

Per quanto riguarda la sintassi esterna, aspetti interessanti dal punto di vista di un apprendente L2 possono essere i seguenti:

- (2) • la modificazione degli ES di una locuzione idiomatica attraverso attributi (per es. *attaccare bottoni infiniti, chilometrici a qlcu*) o complementi di specificazione (per es. *rompere il ghiaccio del silenzio*)  
 • la presenza di collocatori o di elementi cooccorrenti (per es. *è inutile menare il can per l'aia, la verità verrà a galla, farsi le ossa sul campo*)  
 • un frequente uso causativo (cfr. "Matt deve fare una perizia psicologica per la scuola, cioè lo fa andare in tilt." (PAISÀ)) o negativo

<sup>22</sup> Di qui in avanti evidenziazioni di questo tipo saranno sempre degli autori, a meno che non venga segnalato esplicitamente il contrario.

<sup>23</sup> Particolare è il caso di *trattare con i guanti (di velluto) qlcu/qlcs*. L'unico ES obbligatorio di questa locuzione idiomatica è *guanti/o*. Usato da solo però, *guanti/o* perderebbe ogni senso idiomatico e va dunque combinato necessariamente o con *di velluto* o con il verbo *trattare (con)*.

<sup>24</sup> Si applica qui la distinzione tra 'valenze grammaticali' (per es. *saperla lunga su, di qlcs*) e 'valenze lessicali' (per es. *saperla lunga in fatto di, in quanto a qlcs*) seguendo la classificazione di Schafroth (2011: 43), secondo cui le valenze lessicali si distinguono da quelle grammaticali in quanto indicano "un rapporto piuttosto libero tra due unità" e sono "utilizzabil[i] con vari lessemi" (ibid.). A differenza di Schafroth (2011), però, che restringe il termine valenze grammaticali a casi in cui il verbo determina non più di due preposizioni (cfr. ibid.), il termine viene utilizzato qui per comprendere anche verbi, o meglio locuzioni idiomatiche, che determinano l'uso di tre (o potenzialmente anche più di tre) valenze, come nel caso di *sparare a zero*: stando alle attestazioni trovate nei corpora, *sparare a zero* ammette le valenze *a, su* e *contro qlcu/qlcs*. Il criterio distintivo è dunque quello del suddetto "rapporto piuttosto libero tra due unità", che nel caso di *sparare a zero a/su/contro qlcu/qlcs* non sembra sussistere.

- tipi di costruzione caratteristici, come per es. 'Se VOLERE cercare (proprio) il pelo nell'uovo', 'FARE X è scoprire l'acqua calda' (cfr. "Dire che si tratta di attività dove esistono difficoltà oggettive è scoprire l'acqua calda." (WEBBIT) ecc.

Si tratta quindi di descrivere il contesto sintattico tipico di una locuzione idiomatica.

#### 4.2.2 Modalità dettagliata

La modalità dettagliata comprende due rubriche dedicate alla sintassi interna, quali *la variabilità degli ES* e *gli elementi facoltativi*, più una rubrica di carattere piuttosto generico, denominata *peculiarità sintattiche*.

La rubrica *variabilità degli ES* serve sostanzialmente ad approfondire e ad arricchire le informazioni riguardanti gli aspetti elencati sotto (1): si danno esempi per l'uso delle singole varianti formali degli ES sulla base di attestazioni trovate nei corpora, si forniscono indicazioni più esplicite circa la frequenza d'uso, e si segnalano altre possibili varianti formali, più specifiche, come nel caso di *scoprire l'acqua calda*, che in certi contesti al posto del verbo finito ammette l'uso di forme di conversione come *la scoperta dell'acqua calda* o *scopritori dell'acqua calda*. In aggiunta si discutono altre questioni specifiche: per quanto riguarda gli ES con funzione di sintagma nominale, ad esempio, si tratta la questione della possibilità di un uso ellittico (*stavolta niente polveroni!*) o attributivo (*è una donna con i piedi per terra*). Per quanto concerne invece il predicato, la domanda che spesso sorge è se alcune varianti del verbo finito siano da considerarsi ancora parte idiomatica o fissa della locuzione idiomatica. Nel caso di *andare con i piedi di piombo*, per esempio, varianti come *procedere*, *camminare* o *muoversi con i piedi di piombo* sembrano ancora appartenere al livello idiomatico della locuzione, mentre nella frase "Ma sono questioni delicate, che vanno affrontate coi piedi di piombo" (La Repubblica) il predicato *vanno affrontate* ha un significato compositivo e quindi non-idiomatico. Verbi di questo tipo, aggiunti, per così dire, "liberamente" al nucleo semantico della locuzione sono da considerarsi varianti del verbo finito solo a livello di costruzione, e non a livello idiomatico.

Un altro fattore che determina il grado di variabilità della sintassi interna di una locuzione idiomatica è la possibilità o l'impossibilità di integrare elementi facoltativi. Questi ultimi in genere sono sintagmi avverbiali o aggettivali o, a volte, anche frasi parentetiche. Per *tenere il piede in due staffe*, ad esempio, si sono trovati i seguenti elementi facoltativi:

- (3)
- *tenuto* [un po'] *il piede in due scarpe*
  - *piazzare lo* [stesso] *piede in due scarpe*
  - *un solo piede infilato in* [tutte e due le] *staffe*
  - *il* [proverbiale] *piede in due staffe*
  - inoltre è attestato l'inserimento di parentesi, cfr.: "Non le è stato infatti difficile scoprire, ahimé troppo tardi, che il suo 'principe azzurro' teneva, **come suol dirsi**, *il piede in due scarpe*." (La Repubblica)<sup>25</sup>.

Vista la varietà di elementi integrati, *tenere il piede in due staffe* può considerarsi un fraseologismo abbastanza flessibile sotto questo punto di vista. Altri fraseologismi sembrano esserlo molto meno, come per esempio *dire il fatto suo a qlcu* e *dirne quattro a qlcu* per i quali non si sono trovate simili ampie variazioni.

<sup>25</sup> I risultati, riportati in maniera riassuntiva e schematica, sono tratti dalle attestazioni trovate nei corpora e hanno carattere rappresentativo e non esaustivo.

Infine, la rubrica *peculiarità sintattiche* serve a raccogliere tutte le caratteristiche sintattiche non attribuibili alle categorie fin qui presentate. Si tratta di informazioni piuttosto specifiche, in genere non più rilevanti (o quantomeno non indispensabili) in ottica didattica. I fenomeni più frequentemente individuati sono i seguenti:

- (4) • un ordine sintattico marcato, per es. causato da dislocazioni a sinistra (cfr.: “*Buon incontro comunque, magari una pulce nell’orecchio gliel’ho messa, chissà [...]*.” (La Repubblica)) o da un’inversione degli attanti (cfr.: “*È inutile affannarsi a spaccare in quattro il capello della politica; talvolta si tratta semplicemente di soldi.*” (La Repubblica))
- l’uso passivo del verbo finito (cfr.: “*Non pochi bastoni tra le ruote sono stati messi in questi anni all’azione della magistratura [...]*.” (La Repubblica))
- la possibilità di collocare il complemento oggetto al centro della sintassi interna (cfr.: “*Su questi temi noi laici possiamo mettere i clericali alle corde.*” (PAISÀ))
- l’uso sostantivato del verbo finito (e quindi dell’intera locuzione) (cfr.: “*Ma c’è anche chi vede fosco e comincia a pensare che il tirare troppo la corda potrebbe essere pericoloso.*” (CORIS))<sup>26</sup>.

#### 4.2.3 Fraseoframe di “alzare un polverone” (sintassi)<sup>27</sup>

- (5) **alzare un polverone**  
**forma**<sup>28</sup>:      alzare il polverone  
                       alzare polverone/polveroni  
                       sollevare un/il polverone  
                       fare un polverone / fare polverone  
                       creare/montare/suscitare un polverone

##### valenze grammaticali:

- alzare un polverone (**su** qualcuno/qualcosa): “*Dopo Vallettopoli, in realtà, a sollevare un nuovo polverone sulla Gregoraci una settimana fa ci aveva pensato Novella 2000, mostrando un incontro che doveva essere segretissimo, [...]*.” (PAISÀ); “*In tarda serata la riunione è stata smentita ma la notizia è servita ad alzare un ennesimo polverone su una vicenda già abbastanza confusa.*” (La Repubblica)

##### valenze lessicali:

- alzare un polverone (**attorno a** qualcosa): “*Trovo sconcertante che in Italia si sia alzato un polverone enorme attorno al caso Englaro [...]*.” (PAISÀ)
- alzare un polverone (**a proposito di** qualcosa): “*Cara Repubblica, politici, amministratori e giornalisti della mia Palermo in questi giorni hanno sollevato un vergognoso polverone a proposito dell’ormai famosa vignetta di Giorgio Forattini.*” (La Repubblica)

##### sintassi interna: ES obbligatori:

[polverone/(polveroni)]<sup>29</sup>

##### variabilità degli ES obbligatori:

26 Se una di queste caratteristiche dovesse rivelarsi particolarmente rilevante e quindi tipica dell’uso di un fraseologismo anche in generale, ciò andrebbe precisato nell’apposita rubrica *consigli per l’uso* (vedasi cap. 4.4) per avvisarne l’apprendente.

27 Di qui in avanti le rubriche appartenenti alla modalità dettagliata saranno evidenziate in blu.

28 Il campo forma offre un quadro riassuntivo di tutte le varianti formali attestate.

<sup>29</sup> Tra parentesi quadre vengono posti gli elementi che fanno parte della sintassi interna del fraseologismo o che vi possono essere integrate. Le parentesi tonde indicano varianti d’uso meno frequenti.

- è attestata la forma plurale *polveroni* (che è meno frequente di *polverone*, ma non insolita), cfr.: “*Che poi i fatti di Genova servano ad alzare **polveroni** politici, beh questo era nelle aspettative.*” (WEBBIT)
- l'ES *polverone/i* può essere usato senza verbo finito (e persino senza articolo), cfr.: “[...] *le infelici sortite di Buttiglione, Tremaglia e Speroni, ormai cadute nell'oblio, dopo qualche giorno di **polverone** mediatico, altro non sono che l'espressione di un sentimento ancora largamente diffuso nel nostro provincialissimo paese [...].*” (CORIS); è attestato anche l'uso ellittico, cfr.: “*Sembra di capire che tra legale e assistito sia stato stretto un patto chiaro: niente **polveroni**, niente argomenti che non riguardino direttamente la sua difesa.*” (La Repubblica)

**sintassi interna: ES facoltativi:**

[alzare/sollevare/(creare)/(fare)/(montare)/(suscitare)/...] [un/(il)/(questo)/...] *polverone* [su qualcosa]

**variabilità degli ES facoltativi:**

- il verbo finito può essere variato in diversi modi: oltre ai più comuni *alzare* e *sollevare* risultano attestati più volte anche *creare*, *fare*, *montare* e *suscitare*; inoltre si possono usare verbi non-idiomatici, cfr.: “*Essere europeisti deve significare anche questo: abituarsi, una volta per tutte, a ragionare sui fatti, **evitando** accuratamente ogni *polverone* all'italiana.*” (WEBBIT); “*Era semplicemente stato scoperto dal modo in cui **aveva reagito** nel *polverone*.*” (WEBBIT); “***Si è trattato** di un *polverone* di estate.*” (CORIS)
- l'articolo **un**, a seconda dei contesti, può essere omissivo (cfr. “*Ma sì: continuiamo il discorso ideologico della corruzione degli scienziati, che **fa** sempre *polverone*.*” (PAISÀ)) o sostituito con vari elementi, per es. con la forma determinativa **il** o con l'aggettivo **questo**, cfr. risp.: “*È stato Silvio ad **alzare il** *polverone* per avere una scusa per attaccare il presidente [...].*” (PAISÀ); “*Il coordinatore della Margherita, Dario Franceschini, dice che è patetico, da parte della Cdl, **«sollevare questo** *polverone* per coprire il devastante comportamento del ministro leghista.»*” (CORIS); inoltre è attestato l'impiego di aggettivi o avverbi come **troppo**, **nessuno** e **ogni**, cfr. “*È che quando si **alza troppo** *polverone* un po' di *polvere* rimane addosso a tutti.*” (La Repubblica); “[...] non volevo **sollevare nessun** *polverone*.” (WEBBIT)

**sintassi interna: elementi facoltativi:**

- **ALZARE**<sup>30</sup> un [gran] (3x), [grosso], [piccolo], [altro], [vero e proprio] *polverone*
- *alzare* [tutto] *questo polverone*; *ha alzato* [tutto] *il polverone*; *si alza il* [solito] *polverone*; *non si alza* [certo] *un polverone*
- **SOLLEVARE** un [gran(de)] (4x), [altro], [nuovo], [utile], [inutile], [vergonoso] *polverone*
- **SOLLEVARE** [subito], [ultimamente]; [tutto] *un polverone*; *stava per sollevarsi l'ennesimo* *polverone*; *non volevo sollevare* [nessun] *polverone*
- **FARE** un [gran] (2x) *polverone*; *fa* [sempre] *polverone*
- **CREARE** un [bel], [gran], [altro] *polverone*
- *è stato montato un* [grande] *polverone*
- *ha suscitato* [ulteriormente] un [bel] *polverone*
- *un polverone* [giustamente] *sollevato*
- inoltre è attestato l'inserimento di sintagmi avverbiali davanti alla valenza su qualcosa, cfr.: “*In questi giorni si è alzato un vero e proprio *polverone* **in rete soprattutto** sul significato da dare alla parola 'influenza' [...].*” (CORIS)

30 In questo campo la scrittura maiuscola indica che il lessema si trova alla forma base, anche se nel discorso è stato usato almeno una volta in forma flessa. La scrittura minuscola significa invece che il lessema viene riportato nella forma in cui effettivamente è stato usato.

**sintassi esterna:**

molto caratteristico è l'impiego di attributi e complementi di specificazione, cfr.: “Un ragazzo ammette di essere omosessuale [...] e [...] si alza il solito polverone **mediatico**.” (PAISÀ); “Sulla privacy è stato sollevato ultimamente un polverone **immondo**.” (WEBBIT); “La canzone *Moja štikla* solleva un altro polverone di polemiche.” (PAISÀ); inoltre è attestato l'uso di frasi relative con funzione attributiva, cfr.: “Epperò, se l'avesse fatto per sollevare un polverone **che adesso sarà la destra a dover gestire?**” (WEBBIT)

**peculiarità sintattiche:**

1. alcune attestazioni mostrano l'uso di complesse frasi relative che segmentano la sintassi interna della locuzione idiomatica confermandone la variabilità, cfr.: “In **questo gran polverone** di proposte e di richieste **che** si sta rapidamente alzando, va detta subito una parola di chiarezza: [...]” (La Repubblica); “Sono questi i temi che potremo affrontare in un prossimo articolo: ma prima, bisognava fare un po' di chiarezza su tutto **il polverone che** la Tv tradizionalmente intesa ha **sollevato** e sta **sollevando**.” (CORIS)

2. è documentato inoltre l'uso passivo, cfr.: “Le mie saranno sicuramente ipotesi, ma **si è fatto** un polverone per un'intervista che aveva il solo difetto di non avere il contraddittorio.” (PAISÀ)

**4.3. Semantica****4.3.1. Modalità base**

Nella modalità base l'analisi semantica consiste quasi esclusivamente nella definizione del significato della locuzione idiomatica. L'unica altra categoria presente in questa modalità è quella del campo semantico, il cui scopo è quello di offrire agli apprendenti alternative linguistiche per esprimere fatti o stati di cose simili tra loro, come nel caso di *non stare più nella pelle* e *non vedere l'ora*, entrambi appartenenti al campo semantico dell'impazienza<sup>31</sup>.

Da un punto di vista tecnico, la definizione del significato idiomatico dovrebbe tenere conto fondamentalmente dei seguenti criteri:

- (6) (i) La parafrasi dovrebbe includere tutti i tratti semantici costitutivi di un'accezione.
- (ii) La parafrasi dovrebbe chiarire metalinguisticamente il rapporto tra i tratti semantici costitutivi.
- (iii) La definizione dovrebbe tenere conto della potenziale pluralità semantica di una locuzione idiomatica e dei contesti d'uso cui le varianti semantiche sono legate.
- (iiii) Ogni parafrasi deve essere illustrata con degli esempi di uso tipico.

Il punto (i) potrebbe sembrare un'ovvietà, invece rappresenta un criterio che i dizionari (persino quelli idiomatici) a volte non osservano<sup>32</sup>. Nel caso di avere la coda di paglia, ad esempio, si leggono parafrasi come:

- (7) *essere in colpa con la paura continua di essere scoperto* (De Agostini 1990)
- (8) *sentirsi in difetto e quindi adombrarsi e reagire male di fronte a discorsi che vengono erroneamente interpretati come una velata accusa* (Capire l'antifona 1995)

31 La classificazione dei campi semantici si basa sul modello di Bárdosi/Ettinger/Stölting (2003), *Redewendungen Französisch/Deutsch*. Il modello è stato adattato all'italiano e viene ampliato con definizioni proprie nei casi in cui le categorie predefinite non risultano sufficienti.

32 Va tenuto in conto, naturalmente, che si parla di dizionari non concepiti per fini didattici.



Entrambi i dizionari citati (il secondo è specialistico) escludono rispettivamente un tratto semantico costitutivo di questa locuzione idiomatica. In (7) manca il riferimento al comportamento tipico dell'agente: chi ha la coda di paglia, infatti, di solito si distingue per un comportamento eccessivamente nervoso (per es. reagendo in maniera impulsiva a presunte critiche) oppure, anche, eccessivamente passivo (per es. parlando o agendo meno del solito). La definizione (8) invece non esplicita il sema '[voler] nascondere il motivo del proprio senso di colpa' che è da considerarsi fondamentale in quanto rappresenta il motivo del comportamento particolare dell'agente. Inoltre, tale comportamento non deve essere necessariamente sospettoso, come suggerito da (8), ma, come accennato sopra, può avere la coda di paglia anche chi si comporta in maniera stranamente passiva<sup>33</sup>.

Si propone dunque la seguente parafrasi semantica per *avere la coda di paglia*:

- (9) sentirsi in colpa, sentirsi in difetto per qlcs o sim. e temere di essere scoperto dagli altri, comportandosi pertanto in maniera stranamente nervosa o, anche, stranamente passiva (*per es. reagendo in maniera molto impulsiva a critiche, scusandosi senza essere stato accusato, parlando poco o non agendo quando si dovrebbe agire ecc.*)

Una tale parafrasi include i tre semi costitutivi del fraseologismo ('sentirsi in colpa', '[voler] nascondere il motivo del proprio senso di colpa', 'comportarsi in maniera strana o insolita'), che sono al contempo indispensabili per un apprendente L2.

Inoltre la parafrasi proposta, e qui si passa al secondo dei criteri sopra elencati, chiarisce il rapporto semantico tra i tratti costitutivi. Vengono cioè distinti il livello più generico della stranezza del comportamento dell'agente e il livello più concreto dell'espressione di tale comportamento (la parte scritta in corsivo e tra parentesi). In altre parole: si distinguono i tratti semantici costitutivi da quelli marginali, il che per un apprendente può essere di importanza fondamentale. Si veda, a tale proposito, quest'altro esempio:

- (10) *rassegnarsi, accettare razionalmente un fatto sgradevole; rendersi conto dell'inevitabilità di un evento negativo o doloroso* (Quartu/Rossi 2012)

Questa definizione di *farsi una ragione di qualcosa* potrebbe fuorviare un potenziale apprendente perché rischia di suggerire che *rendersi conto dell'inevitabilità di un evento negativo o doloroso* costituisca un'accezione idiomatica a se stante, mentre in realtà essa sta in una relazione causale con la prima parte della definizione: chi si fa una ragione di qualcosa, si rassegna emotivamente a un fatto sgradevole e lo accetta razionalmente [rendendosi conto che è inevitabile e irreversibile].

Tuttavia, il criterio più importante di quelli elencati sotto (6) è il terzo, quello della potenziale pluralità semantica. Come dimostrato da Stathi (2006)<sup>34</sup> sulla base di dati empirici tratti dai corpora, tante locuzioni idiomatiche possono assumere varie accezioni a seconda del contesto

33 Si veda a riguardo la seguente attestazione trovata nel corpus CORIS: Se un presunto padre ritiene che quel bambino di cui si chiede il riconoscimento non sia suo, ha una grande arma a suo favore: l'esame del Dna, prova sicura. Se rifiuta quest'analisi, nel 99 per cento dei casi, vuol dire **che ha la coda di paglia** e quindi o non è certo del responso oppure nè [sic!] talmente certo che vuole fuggire dalle sue responsabilità (CORIS)

34 Pur limitandosi ai fraseologismi del tedesco, lo studio di Stathi (2006) e i risultati emersi sono senz'altro applicabili universalmente. Per la lingua italiana si può rinviare, ad esempio, al lavoro di Nuccorini (2007a).

semantico-pragmatico in cui vengono usate; tra l'altro, si tratta spesso di accezioni non considerate dalla lessicografia. In italiano un caso significativo in questo senso è rappresentato dal fraseologismo *saperla lunga*. Sommando le definizioni tratte dai vari dizionari dell'italiano (specialistici e non) risultano documentate le seguenti accezioni:

- (11) a. conoscere molto bene un argomento, essere competente [in un campo] o essere esperto (cfr. Quartu/Rossi 2012, *Devoto-Oli* 2014)  
 b. essere scaltro, furbo o astuto (cfr. Quartu/Rossi 2012, *GRADIT* 2000)  
 c. sapere più di quanto si ha intenzione di dire (cfr. *Treccani* 2013)

L'analisi dei corpora ha fatto emergere quattro altre varianti:

- (12) d. essere ben informato su qlcs, essere a conoscenza di certi fatti, cfr.: “*Sandro Federico, il capo della squadra mobile è ottimista: il pentito dimostra di **saperla lunga**, forse è la chiave giusta per violare gli impenetrabili santuari della camorra.*” (La Repubblica)  
 e. avere vissuto già molte esperienze prevalentemente negative in un certo ambito, cfr.: “*Terribile, lo so – fece il carabiniere con l'aria di chi **la sa troppo lunga** per farsi ferire da quelle immagini.*” (WEBBIT)  
 f. tenere un atteggiamento supponente credendosi più abili o scaltri degli altri, cfr.: “*L'accelerazione da fermo ci sembra buona, la progressione non ha la 'cattiveria' di altri motori ma ci permette di lasciarci alle spalle chi ci aveva lanciato la classica occhiata di chi **la sa lunga**.*” (PAISÀ)  
 g. mostrarsi disinvolto e molto sicuro di sé, cfr.: “*Adoro le due foto che ho messo...soprattutto quella con le due ragazze... cappottone, sigaretta da parte ed espressione di chi **la sa lunga!***” (WEBBIT)

Queste ulteriori accezioni individuate sono indispensabili ai fini di una rappresentazione complessiva del significato<sup>35</sup> di *saperla lunga*<sup>36</sup>.

Il criterio della pluralità semantica ha ripercussioni anche sulla categoria del campo semantico: se una locuzione idiomatica può avere varie accezioni, dovrebbe essere attribuibile anche a diversi campi semantici. Tuttavia, è opportuno includere nella categoria solo i campi semantici più dominanti. La differenziazione tra campi semantici dominanti e meno dominanti, peraltro, può rivelarsi a sua volta abbastanza opinabile. Nel caso di *saperla lunga*, ad esempio, non è facile decidere se le accezioni f. e g. siano sufficienti per aggiungere il campo semantico *presunzione* e se le accezioni a. e e. giustifichino l'attribuzione del campo semantico *esperienza*.

#### 4.3.2. Modalità dettagliata

Le categorie semantiche presenti nella modalità dettagliata sono i *ruoli semantici* e le *peculiarità semantiche*. Quest'ultima prevede la documentazione di caratteristiche semantiche abbastanza specifiche, di solito non più rilevanti in ottica didattica. Ecco due esempi riguardanti i fraseologismi *non vedere l'ora* e *fare le ore piccole*:

35 Ovviamente, ci sono anche casi, o meglio contesti d'uso, in cui il valore semantico di una locuzione idiomatica sembra quasi indefinibile in quanto sospeso tra due semi diversi. In questo caso si parla di *vaghezza semantica* (cfr. Stathi/Hümmer 2006). La vaghezza semantica è da considerarsi tuttavia irrilevante in ottica didattica, dato che non fornisce informazioni aggiuntive sull'uso di un fraseologismo.

36 Per una proposta di definizione completa si veda il capitolo 4.3.

## (13) Peculiarità semantiche

*non vedere l'ora*

L'uso dell'avverbio di negazione *non* nella locuzione *non vedere l'ora* si è ormai talmente grammaticalizzato che l'avverbio ha praticamente perso la sua funzione negativa. Lo dimostra il fatto che *non* non può essere sostituito con elementi di valore negativo come nemmeno, nessuno ecc. (*Non vedo l'ora!* - \**Nemmeno io!*), ma solo con elementi di valore affermativo (*Non vedo l'ora!* - *Anch'io!*)<sup>37</sup>.

*fare le ore piccole*

La locuzione (*le*) *ore piccole* (usata cioè senza il verbo *fare*) può essere interpretata sia in senso strettamente temporale, designando il periodo dalla mezzanotte all'alba (vedasi es. a.), oppure in senso non-temporale, riferendosi piuttosto all'attività svolta dall'agente nel periodo dopo la mezzanotte (vedasi es. b.), cfr.: a. "*Ancora, bisogna dire che pubblichiamo i menù del giorno di Natale: in alcune zone, è considerata più importante la cena della vigilia, che spesso si conclude a **ore piccole** [...]*" (WEBBIT); b. "*Il problema era che avevamo deciso di codificare il tutto in C++ e quindi finimmo per ridisegnare quasi completamente il progetto in analisi Object Oriented durante l'ultimo paio di settimane [...] questo portò a molte ore piccole [...]*" (WEBBIT)<sup>38</sup>

L'analisi dei ruoli semantici si basa sulle classificazioni di Christian Lehmann<sup>39</sup> e di Peter von Polenz (cfr. 2008: 167-174). Vengono analizzati solo gli attanti non-idiomatici di una locuzione<sup>40</sup>, mentre si tiene conto dei criteri dell'(in)animatezza e dell'astrattezza dei referenti extralinguistici, considerati molto importanti in quanto a volte forniscono informazioni sul significato e sui contesti d'uso di un fraseologismo. Si veda l'esempio di *fare le ore piccole* che, seppur raramente, in certi contesti ammette un soggetto inanimato e astratto, cfr.: *Si sa che è difficile sfuggire al fascino di Fidel, quando si è ospiti delle sue chilometriche **conversazioni che fanno le ore piccole*** (WEBBIT). A volte la questione del ruolo semantico di un costituente si presenta abbastanza complessa. Il soggetto di *chinare il capo* per esempio, a seconda del contesto d'uso, sembra attribuibile sia al ruolo semantico dell'agente, sia a quello del paziente: se la rassegnazione è frutto di una scelta prevalentemente autonoma, si tratta di un agente<sup>41</sup>, se invece

37 Dato che in questo caso l'impossibilità di sostituire *non* con elementi di valore negativo rappresenta un'informazione rilevante anche in ottica didattica, essa va riportata in maniera più dettagliata nell'apposita categoria *consigli per l'uso* (che sarà descritta nel capitolo seguente).

<sup>38</sup> Si tenga conto che la modalità dettagliata conserva comunque un approccio contrastivo, rivolgendosi in primis a madrelingua tedeschi. Perciò vengono fornite anche informazioni che a un madrelingua italiano potrebbero sembrare ovvie.

39 Cfr. Christian Lehmann, Semantische Rollen, [www.christianlehmann.eu/ling/lg\\_system/sem/index.html?http://www.christianlehmann.eu/ling/lg\\_system/sem/semant\\_rollen.php](http://www.christianlehmann.eu/ling/lg_system/sem/index.html?http://www.christianlehmann.eu/ling/lg_system/sem/semant_rollen.php).

40 Questa scelta, il cui scopo è quello di evitare un approccio troppo teorico, non vuole mettere in dubbio che in alcuni casi anche gli attanti idiomatici di un fraseologismo siano attribuibili a un ruolo semantico. In *venire a galla* ad esempio, *a galla* sarebbe interpretabile come *meta* (*Ziel*), dato che la frase preposizionale *a galla* conserva il significato 'in superficie' anche al di fuori della locuzione idiomatica.

41 Cfr.: "Come mai francesi e russi hanno **chinato** così docilmente **la testa**? Dopo la fine della guerra contro Saddam Hussein, Parigi e Mosca avevano negato per un mese e mezzo a Stati Uniti e Gran Bretagna ogni legittimazione Onu sull'Iraq occupato." (PAISA).

la rassegnazione viene imposta da qualcuno, si tratta più di un soggetto paziente<sup>42</sup>. Può essere determinante anche il contesto grammaticale: in caso di negazione (per es. *Questa è gente che non china il capo a nessuno*) l'arrendevolezza diventa resistenza e quindi non ci sono dubbi sul ruolo dell'agente<sup>43</sup>. Infine, il concetto dei ruoli semantici serve anche a tenere conto della differente semantica delle varianti del verbo finito di una locuzione idiomatica, come nel caso di *dare e trovare del filo da torcere*: il verbo *dare* comporta il ruolo dell'agente, *trovare* invece quello del paziente.

#### 4.3.3. Fraseoframe di saperla lunga (semantica)

##### (14) *saperla lunga*

**significato: 1** conoscere molto bene un argomento, essere competente [in un campo] o essere esperto di qlcs (per es. della vita): “Tra gli artisti più colti del nostro tempo, cioè di quelli che **la sanno** assai **lunga** sul fatto dell’arte, Sassu per me è pittore soprattutto d’istinto.” (La Repubblica); “L’atmosfera non è quella dell’ufficio di pubbliche relazioni di una multinazionale, ma in quanto a organizzazione questi ragazzi **la sanno lunga**.” (La Repubblica); “‘Aver fiducia in se stessi e volontà a non cedere sono alla base delle grandi e delle piccole imprese’, dice con l’aria di chi **la sa lunga**, Jacek Palkiewicz, giornalista italo-polacco, [...]” (La Repubblica) ▪ (anche) essere ben informato su qlcs, essere a conoscenza di certi fatti: “Che i servizi segreti **la sapessero lunga** sulla stagione altoatesina delle bombe, del resto, lo sostengono da tempo i gruppi di oltranzisti sudtirolesi.” (La Repubblica); “Sandro Federico, il capo della squadra mobile è ottimista: il pentito dimostra di **saperla lunga**, forse è la chiave giusta per violare gli impenetrabili santuari della camorra.” (La Repubblica)

**2** avere fatto già molte esperienze prevalentemente negative in un certo ambito [e quindi parlare per esperienza di qlcs oppure non farsi impressionare più da certe situazioni]: “È contrario ai nostri principi, e di ingerenze straniere **la sappiamo lunga**: dall’era di Stalin all’era di Breznev, [...]” (La Repubblica); “Terribile, lo so – fece il carabiniere con l’aria di chi **la sa troppo lunga** per farsi ferire da quelle immagini.” (WEBBIT)

**3** essere scaltro, furbo o astuto: “Fate attenzione -- sono stato tentato di scrivere loro -- se in mezzo alla notte vi telefona un certo D’Avanzo meglio che facciate rispondere che siete fuori a pranzo perché quello lì **la sa lunga** e rischia di mettervi nei guai.” (La Repubblica); “Andrea ‘il biondo’ dimostra più dei suoi 17 anni. È molto alto, il fisico da atleta, lo sguardo freddo di uno che **la sa lunga** [...]” (La Repubblica) ▪ (anche) dare l’impressione di svelare agli altri volutamente meno di quanto si sa di un fatto, o fingere di essere meno astuto di quanto si è realmente: Dice sempre che non gli interessa molto quello che succede nei piani alti dell’azienda, che non conosce bene la situazione finanziaria, ma io non gli credo: secondo me **la sa lunga** su quello che succede e non vuole svelarcelo. Insomma, non mi fido di lui. (R. I.)

**4** tenere un atteggiamento supponente credendosi più intelligente, più abile o più scaltro degli altri: “Ah: la differenza fra infertilità e subfertilità e sterilità la conosco bene, non ti preoccupare. E siccome non c’entra niente con gli argomenti che ho trattato, è evidente che sei tu a non sapere che cosa sono. Non fare il giochino di quello che **la sa lunga**.” (PAISÀ); “L’accelerazione da fermo ci sembra buona, la progressione non ha la “cattiveria” di altri motori ma ci permette di lasciarci alle spalle chi ci aveva lanciato la classica occhiata di chi **la sa lunga**.” (PAISÀ) ▪ (più raramente) mostrarsi disinvolto e molto

42 Cfr.: “È ancora il Sudafrica a rendere la vita l’unico inferno in cui vivere e che lui è costretto ad accettare. Così anche una violenza carnale può (se non deve) essere accettata, per continuare a vivere, in silenzio, **chinando il capo**, per sopravvivere.” (PAISÀ).

43 Com’è ben noto, un altro caso in cui il contesto grammaticale risulta determinante è quello dell’uso passivo.

sicuro di sé: “Adoro le due foto che ho messo... soprattutto quella con le due ragazze... cappottone, sigaretta da parte ed espressione di chi **la sa lunga!**” (WEBBIT)

**campo semantico:** sapere; abilità; esperienza; presunzione

**ruoli semantici:**

1. il soggetto è **agente**: *solitamente* [+animato] o con referenza generica (cfr. “Terribile, lo so – fece il carabiniere con l’aria **di chi la sa troppo lunga per farsi ferire da quelle immagini.**” (WEBBIT)); *anche* personificazione metonimica (cfr. “Una società seria **la sa più lunga** dei suoi singoli membri.” (La Repubblica)) o, *raramente*, personificazione di referenti [+inanimati] (astratti) (cfr. “Ma è un libro che **la sa molto lunga**: esemplare della disperazione siciliana, totalmente disincantato.” (La Repubblica))
2. i complementi oggetto delle valenze *su qualcosa* e *di qualcosa* sono **tema**: [+inanimato]

**peculiarità semantiche:** /

#### 4.4. Pragmatica

##### 4.4.1. Modalità base

Per fornire all’apprendente le informazioni necessarie su come usare una locuzione idiomatica da un punto di vista pragmatico-discorsivo, la modalità base dispone delle rubriche *motivazione*, *contesto situazionale* e *consigli per l’uso*. Quest’ultima serve a descrivere alcune caratteristiche fondamentali come ad esempio:

- (15) a. la frequenza d’uso  
 b. il registro  
 c. un possibile effetto ironico, spregiativo, scherzoso, retorico o sim.  
 d. il contesto sociale adeguato o tipico dell’uso  
 e. gli effetti perlocutivi sull’interlocutore

Inoltre la rubrica prevede la distinzione tra fraseologismi simili tra loro a livello formale, semantico o pragmatico<sup>44</sup> così come la distinzione tra alcuni fraseologismi italiani e i loro (presunti) equivalenti tedeschi<sup>45</sup>. Infine, rientrano in questa categoria tutte le altre peculiarità (non necessariamente di tipo pragmatico) che possono servire all’apprendente tedescofono, come per esempio nel caso di *dire il fatto suo a qlcu.* Qui è importante segnalare che il pronome possessivo si riferisce sempre all’oggetto indiretto, visto che un madrelingua tedesco, forse per motivi di interferenza determinati dall’equivalente *jdm. seine Meinung sagen* dove il pronome *seine* si riferisce al soggetto, probabilmente supporrebbe l’esatto contrario.

44 Un esempio significativo in questo senso è rappresentato dalle varianti *attaccar bottone (con qlcu)* e *attaccare un bottone a qlcu*: nonostante quest’ultima si differenzi dalla prima attraverso tratti semantici come [+fastidioso], [+noioso] ecc., le due varianti non vengono trattate separatamente dalla maggior parte dei dizionari. Invece, differenze di questo tipo vanno segnalate all’apprendente in maniera esplicita, dato che sono indispensabili per evitare un uso inappropriato.

45 Si veda il caso di *mettere la pulce nell’orecchio a qlcu* e del traduce *jdm. einen Floh ins Ohr setzen* che si distinguono in parte per quanto riguarda il significato e soprattutto per quanto riguarda l’uso. Il fraseologismo tedesco infatti, diversamente da quello italiano, di solito designa l’insinuare un’idea in qualcuno che al parlante o a terzi sembra difficile da realizzare, irragionevole, senza senso e sim. L’uso tipico consiste quindi nel rinfacciare al paziente di essersi messo in testa una tale idea.

Per quanto riguarda il punto a. dei criteri sopra elencati, è indispensabile sapere che, ad esempio, *non vedere l'ora e prendere in giro qlcu* sono presenti nell'uso di tutti i giorni, mentre un fraseologismo come *mettere il carro davanti ai buoi* sta entrando sempre più in disuso. Quanto al punto b., differenze di registro sono invece più rari; la maggior parte dei fraseologismi corrisponde a un registro colloquiale (*darci dentro, gettare fango addosso*) o neutro (*mettere in riga qlcu, rimboccarsi le maniche*). Un esempio di registro variabile riguarda invece le due varianti *chinare il capo* e *abbassare la testa*: la prima comporta un registro molto elevato (da evitare nell'italiano parlato), mentre la seconda corrisponde a un registro neutro.

Il criterio c. caratterizza l'uso di molti fraseologismi. Spesso il parlante implica un giudizio negativo (*spaccare il capello in quattro, menare il can per l'aia* ecc.), altre volte si tratta di un uso ironico-spregiativo (*scoprire l'acqua calda*). Inoltre le peculiarità possono essere legate a singole varianti (pragmatiche, semantiche o anche sintattiche) della stessa locuzione idiomatica: si vedano gli esempi di *farsi una ragione di qlcs* e *saperla lunga*. La prima vede l'effetto ironico-spregiativo limitato ad una certa variante pragmatica usata al futuro semplice (*me ne farò/ce ne faremo una ragione*<sup>46</sup>). Il valore dell'uso di *saperla lunga* invece può variare molto a seconda della variante semantica intesa dal parlante: nelle accezioni f. e c. elencate sotto (11) e (12) *saperla lunga* ha valore spregiativo, mentre l'accezione a. implica un giudizio positivo da parte del parlante.

Il criterio c., inoltre, è spesso legato ai punti d. e e.: fraseologismi come *scoprire l'acqua calda* o *avere la coda di paglia*, infatti (che non solo hanno valore spregiativo ma implicano anche, entrambi in modo diverso, un sentimento di superiorità da parte del parlante nei confronti dell'interlocutore o di terzi), di solito si usano in contesti sociali di tipo familiare, soprattutto per quanto riguarda l'uso *face-to-face*. L'apprendente andrebbe dunque avvisato del fatto che criticando ad esempio una persona sconosciuta di aver scoperto l'acqua calda, molto probabilmente rischierebbe di offenderla o di provocarla.

Le altre due rubriche presenti nella modalità base, *motivazione* e *contesto situazionale*, servono invece a spiegare rispettivamente con quale intenzione illocutiva e in che tipo di situazione comunicativa usare un determinato fraseologismo. Una locuzione come *attaccare un bottone a qlcu*, ad esempio, viene usata spesso con la motivazione di lamentarsi del fatto di essere coinvolti in una conversazione lunga e noiosa, cfr.: “*Durante la mattinata un ragazzo ci ha attaccato un bottone che non finiva più [...]*.” (WEBBIT). Tuttavia, di solito un fraseologismo può assolvere diverse funzioni illocutive: *attaccare un bottone*, infatti, si presta altrettanto bene a esprimere una scusa, cfr.: “*Mi scusi se le ho attaccato un bottone, vedo che lei sta partendo.*” (WEBBIT). Allo stesso modo, con *avere la coda di paglia* il parlante non deve esternare necessariamente un sospetto sul comportamento altrui, ma può anche cercare di difendersi da un tale sospetto, cfr.: “*e poi per quale diavole di motivo dovrei avere la coda di paglia scusa? ho solo espresso un mio pensiero...*” (WEBBIT).

A volte la funzione illocutiva è legata strettamente ad un determinato contesto situazionale, come nel caso di *non stare più nella pelle*, che solitamente si usa quando si vive o quando si è in attesa di un'esperienza felice e molto positiva, cfr.: “*Non stavo più nella pelle all'idea di andare a scuola, dove avrei conosciuto altri bambini, avrei letto e disegnato.*” (CORIS). *Dirne quattro a qlcu* si usa invece spesso in situazioni in cui l'agente è molto arrabbiato col paziente (per es. per aver subito un torto) e sente la forte volontà di rimproverarlo e sfogarsi con lui, cfr.: “*Ancora! Adesso gliene dico quattro!*” (WEBBIT).

46 Cfr. “partendo dal presupposto che non sentirò la sua mancanza... che scelta stupida che ha fatto!!! [...]. a meno che Marano non le abbia fatto una promessa..non credo che la rivedremo molto presto in tv. bè ce ne faremo una ragione.” (PAISÀ).

#### 4.4.2. Modalità dettagliata

Nella modalità dettagliata l'analisi pragmatica è costituita dalla rubrica *atti linguistici*. Per quanto riguarda la classificazione degli atti linguistici, si adotta qui il modello di Pörings/Schmitz (2003: 163 pp.) che propongono la seguente tassonomia:

(16) **Atti costitutivi**

**Espressivi:** ringraziare, salutare, augurare, scusarsi, congratularsi, elogiare ecc.

**Dichiarativi:** battezzare, nominare, rilasciare, condannare ecc.

**Atti informativi**

**Assertivi:** descrivere, sostenere, annunciare, presumere ecc.

Richiesta d'informazione: domandare

**Atti obbligativi**

**Direttivi:** pregare, ordinare, consigliare, proporre ecc.

**Commissivi:** promettere, accordare, offrire, minacciare ecc.

(Pörings/Schmitz 2003: 166)<sup>47</sup>

Il modello segue un approccio cognitivo e si distingue dalla classificazione di John Searle per due fondamentali aspetti (cfr. *ibid.*): 1. L'introduzione della nuova categoria *richiesta d'informazione* che consiste nel solo atto linguistico *domandare* (non previsto da Searle); 2. L'introduzione delle tre categorie sovraordinate *atti costitutivi*, *atti informativi* e *atti obbligativi* comprendenti tipi di atti linguistici che, secondo gli autori, condividono rispettivamente almeno una caratteristica fondamentale<sup>48</sup>. Sulla base di tale modello si possono classificare gli atti linguistici di cui un parlante si serve tipicamente usando un determinato fraseologismo. Nel caso di *scoprire l'acqua calda*, ad esempio, l'atto linguistico più frequentemente usato è *criticare*, appartenente al gruppo degli atti espressivi (costitutivi) e realizzato quasi sempre in modo indiretto<sup>49</sup>. Un fraseologismo come *darci dentro* invece, si caratterizza prevalentemente per l'uso di atti assertivi (informativi)<sup>50</sup> e di atti direttivi (obbligativi)<sup>51</sup>, realizzati quasi sempre in modo diretto. Per un esempio d'analisi completo si veda il capitolo 4.4.3.

#### 4.4.3. Fraseoframe di farsi una ragione di qualcosa (*pragmatica*)

(17) ***farsi una ragione di qualcosa***

**consigli per l'uso:**

47 Il quadro è stato tradotto dagli autori dal tedesco in italiano.

48 Gli atti obbligativi hanno in comune la funzione di vincolare l'interlocutore o il parlante stesso a un'azione futura, gli atti informativi riguardano la comunicazione di informazioni o la richiesta di informazioni e gli atti costitutivi condividono la proprietà di costituire forme di realtà sociale (cfr. Pörings/Schmitz 2003: 166 pp.).

49 Cfr.: "Si continua a dire che Stoner era costretto a pelare o frenare prima per non tamponare Rossi? Bene signori, avete **scoperto l'acqua calda**. Normalmente i piloti (di auto, moto, ed anche i fantini) la prima cosa che non devono fare è tamponare chi hanno davanti anche se è più lento." (PAISÀ)

50 Cfr.: "ho iniziato da piccolo a darci dentro con la musica [...] complice l'entusiasmo dei miei genitori [...]." (WEBBIT).

51 Cfr.: "Il tempo scorre inesorabile e il countdown segna -3: avete ancora 3 giorni per votare i candidati alla fase finale dei SoundsBlog Awards 2009, quindi **dateci dentro!**" (PAISÀ).

- frequenza: alta
- registro: neutro
- si consideri il possibile uso ironico, cfr.: “*Detto questo, se vuoi smettere di rispondere ai miei post, pazienza: me ne farò una ragione.*” (PAISÀ); l’effetto ironico sta nel fatto che di solito *farsi una ragione di qlcs* viene usato per riferirsi a fatti sgradevoli o addirittura tragici, mentre nell’esempio citato il parlante lo usa intendendo l’esatto contrario: e cioè che il presunto fatto negativo, in questo caso la minaccia dell’interlocutore di non rispondere più ai messaggi del parlante, lascia quest’ultimo completamente indifferente (si tratta quindi di una variante efficace anche a livello retorico, utilizzabile per replicare a esternazioni volte a indispettire); la variante viene usata quasi sempre al futuro semplice.
- *farsi una ragione* si differenzia dal fraseologismo *mettersi il cuore in pace* (appartenente allo stesso campo semantico della *rassegnazione*) per almeno due aspetti importanti: 1. *f. una ragione* si usa prevalentemente in contesti sociali seri e in merito a fatti dolorosi o quantomeno difficili da accettare (eccezion fatta per la variante ironica), mentre *m. il cuore in pace* non può essere usato per designare il superamento di eventi tragici come, ad esempio, la perdita di una persona amata; 2. *m. il cuore in pace* riguarda più la dimensione emotiva del superamento di un fatto negativo, mentre *f. una ragione* concerne sia il livello emotivo, sia quello cognitivo-razionale, che di solito consiste nel realizzare l’irreversibilità del fatto accaduto o nel comprenderne i motivi.

#### motivazione:

- a. dire che il paziente/agente<sup>52</sup> sa ormai accettare un certo fatto negativo, cfr.: “*Ormai anche mia figlia Elena, che ha 6 anni, si è abituata a vedermi partire e se n’è fatta una ragione.*” (CORIS)
- b. evidenziare che nella vita certi fatti inevitabili vanno accettati, cfr.: “*Con lei sono stato bene e avrei preferito che fosse durata per sempre. Ma ogni tanto va così e bisogna farsene una ragione.*” (WEBBIT)
- c. invitare il paziente/agente a rassegnarsi a un certo fatto negativo accettandolo razionalmente, cfr.: “*la Raffa quando qualcuno le sta antipatico glielo dimostra... quindi cara Simo... fattene una ragione... non piaci alla Raffa!!!*” (PAISÀ); (anche) invitare il paziente/agente a riconoscere l’evidenza di un fatto, cfr.: “*Ragazzi, fatevene una ragione: Kobe è un fenomeno!*” (WEBBIT)
- d. dire che al paziente/agente non resta che rassegnarsi a un certo fatto sgradevole, cfr.: “*Ora, con le Pagine Utili, stiamo rompendo un altro monopolio, quello delle Pagine Gialle, e a qualcuno dà fastidio, molto fastidio. Se ne facciamo una ragione: c’è chi è bravo a rompere le scatole, e chi a rompere i monopoli.*” (CORIS)
- e. dire di essere incapace di accettare un certo fatto tragico soffrendone ancora troppo, cfr.: “*La mia mamma io l’ho perduta di recente e non so farmene ancora una ragione.*” (La Repubblica), oppure evidenziare la sofferenza del paziente, cfr.: “*Due tragedie assurde delle quali i parenti e gli amici non riescono a farsi una ragione.*” (La Repubblica)
- f. (uso ironico) far capire di non farsi turbare facilmente da certi fatti negativi o da certe esternazioni volte a indispettire, cfr.: “*non credo che la rivedremo molto presto in tv. bè ce ne faremo una ragione.*” (PAISÀ); “*Detto questo, se vuoi smettere di rispondere ai miei post, pazienza: me ne farò una ragione.*” (PAISÀ)

#### contesto situazionale:

- usato di solito in contesti di eventi negativi o tragici irreversibili che si è obbligati ad accettare

52 Il soggetto di *farsi una ragione di qlcs* è interpretabile sia come paziente, sia come agente, a seconda che prevalga il condizionamento psichico-emotivo causato dal fatto negativo (paziente) o lo sforzo attivo per superare tale condizionamento (agente).



- l'uso ironico è legato a contesti in cui il parlante si vede confrontato con fatti o esternazioni altrui, di cui l'altro pensa (o spera) che siano difficili da accettare per il parlante colpendolo emotivamente, mentre è vero il contrario

#### atti linguistici:

##### costitutivi

- **espressivo [indiretto]**: lamentarsi: “*spero ogni giorno che lui torni da me.. [...] non so proprio [sic!] come andare avanti... vorrei ammazzarmi! non riesco a farmene una ragione...*” (CORIS)

##### informativi

- **assertivo [diretto]**: dire: “*Lukas soffre molto per il distacco da Sandra ma se ne fa una ragione, [...].*” (PAISÀ)
- **richiesta d'informazione [diretto]**: domandare: “*«È riuscito a farsene una ragione? Voglio dire, dopo un po'?»*” (CORIS)

##### obbligativi

- **direttivo [diretto]**: esortare/invitare a fare qlcs: “*la Raffa quando qualcuno le sta antipatico glielo dimostra... quindi cara Simo... fattene una ragione.... non piaci alla Raffa!!!*” (PAISÀ)

## 5. Conclusione

In questo articolo abbiamo cercato di dare una panoramica dei lavori della fraseologia italiana basata sui principi della Grammatica delle Costruzioni. Questa applicazione fa parte di un progetto linguistico-didattico più vasto: la banca dati *Piattaforma per chi impara italiano*. Essa è ancora in fase di sviluppo, ma già disponibile gratuitamente in rete e aperta a tutti coloro che non riescono a soddisfare il loro interesse per la lingua italiana con i comuni dizionari gratuiti online o le grammatiche italiane in versione cartacea.

I vantaggi di un progetto come la *Piattaforma* sono evidenti:

- orientamento didattico verso gli apprendenti L2;
- elaborazione di informazioni linguistiche secondo i bisogni degli utenti;
- riferimento all'uso linguistico autentico (corpora);
- inserimento dei principali dizionari di italiano e delle più importanti grammatiche (anche quelle per l'insegnamento L2);
- inclusione di video per illustrare l'uso di un segnale discorsivo o di un fraseologismo;
- applicazione del materiale nei seminari e corsi d'italiano per germanofoni;
- sostenibilità tramite la realizzazione digitale.

Per quanto riguarda i fraseologismi e i segnali discorsivi, gli utenti della piattaforma hanno a disposizione una vasta e affidabile gamma d'informazioni sulle loro caratteristiche pragmatico-semantiche. Perché chi vuole padroneggiare una lingua deve conoscere a fondo e saper usare questi fenomeni in modo naturale e sicuro come i madrelingua. La piattaforma offre quanto più c'è da sapere per favorire non solo la comprensione, ma anche l'utilizzo attivo delle locuzioni idiomatiche e dei segnali discorsivi.

In conclusione la *Piattaforma per chi impara italiano* è la base metodologica e l'interfaccia per il progetto internazionale FRAME (Fraseologia Multilingue Elettronica), realizzato in cooperazione con diverse università in Germania, Belgio, Italia e Spagna, che, sotto la responsabilità di Paola Cotta Ramusino, Fabio Mollica (Università degli Studi di Milano) e Elmar Schafroth (Heinrich-Heine-Universität Düsseldorf), si concentra sulla fraseologia di sette lingue in ottica contrastiva seguendo il modello descrittivo della Grammatica delle Costruzioni (cfr. Benigni/Cotta Ramusino/Mollica/Schafroth 2015).

## BIBLIOGRAFIA

- Adamo, G., Della Valle, V. (2005), *Introduzione*, in Id. (2006), *Parole nuove, Un dizionario di neologismi dai giornali*, Roma, Enciclopedia italiana, pp. V-XII.
- Aprile, M. (2008), *Dalle parole ai dizionari*, Bologna, Il Mulino.
- Baranov, A. N., Dobrovol'skij, D. (1996), *Idiomatičnost' i idomy*, in "Voprosy jazykoznanija", 5, pp. 51-64.
- Bárdosi, V., Ettinger, S., Stölting, C. (2003), *Redewendungen Französisch/Deutsch. Thematisches Wörter- und Übungsbuch*, 3. Auflage, Tübingen, Francke.
- Beccaria, G. L. (2002), *Italiano: antico e nuovo*, 2a ed., Milano, Garzanti.
- Benigni, V., Cotta Ramusino, P.M., Mollica, F., Schafroth, E. (2015), *How to apply CxG to phraseology: a multilingual research project*, in "Journal of Social Sciences", 11/3, pp. 275-288, from <http://thescipub.com/PDF/jssp.2015.275.288.pdf> (28.11.2016).
- Blumenthal, P., Rovere, G. (2016), *Wörterbuch der italienischen Sprache: Deutsch-Italienisch, Italienisch-Deutsch*, 2. Auflage, Nürnberg, Acolada, from <http://www.unilex-online.de/pages/search.aspx> (28.11.2016).
- Burger, H. (2015), *Phraseologie. Eine Einführung am Beispiel des Deutschen*, 5. Auflage, Berlin, Schmidt.
- Burger, H. et al. (2007), *Objektbereich, Terminologie und Forschungsschwerpunkte*, in H. Burger et al. (Hrsg.), 2 Bd., Berlin/New York, de Gruyter, pp. 1-10.
- Burger, Harald et al. (Hrsg.) (2007), *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*, 2 Bd. Berlin/New York, de Gruyter.
- Cacia, D., Papa, E., Verdiani, S. (2013), *Dal mondo alle parole. Definizioni spontanee e dizionari d'apprendimento*, Roma, Società Editrice Romana.
- Casadei, F. (1995a), *Per una definizione di 'espressione idiomatica' e una tipologia dell'idiomatico*, in "Italiano. Lingua e Stile", XXX, pp. 335-358.
- Casadei, F. (1995b), *Flessibilità lessico-sintattica e produttività semantica delle espressioni idiomatiche: un'indagine sull'italiano parlato*, in F. Casadei, G. Fiorentino, V. Samek-Ludovici (a. c. di), *L'italiano che parliamo*, Sant'Arcangelo di Romagna, Fare Editore, pp. 11-33.
- Corda, A., Marellò, C. (2004), *Lessico. Insegnarlo e impararlo*, Perugia, Guerra.
- Cresti, E., Panunzi, A. (2013), *Introduzione ai corpora dell'italiano*, Bologna, Il Mulino.
- Crocco, C. (2015), *Corpora e testi di italiano contemporaneo*, in M. Iliescu, E. Roegiest (eds.), *Manuel des anthologies, corpus et textes*, Berlino/Boston, de Gruyter, pp. 509-534.
- Croft, W. (2001), *Radical construction grammar: syntactic theory in typological perspective*, Oxford, Oxford University Press.
- Dardano, M. (1974), *Il linguaggio dei giornali italiani*, 2a ed., Roma, Laterza.
- Dardano, M. (2012), *Traslati, eufemismi e tabù della quotidianità*, in U. Reutner, E. Schafroth (eds.), *Political Correctness. Aspetti politici, social, letterati e mediatici della censura linguistica*, Frankfurt a. M., Lang, pp. 301-317.
- Della Valle, V., Patota, G. (2016), *Lezioni di lessicografia. Storie e cronache di vocabolari*, Roma, Carocci.
- De Mauro, T. (1994), *Premessa: Il LIP*, in T. De Mauro (a. c. di), pp. XI-XXVI.
- De Mauro, T. (2000), *Il dizionario della lingua italiana*, Torino, Paravia.
- De Mauro, T. (2005), *La fabbrica delle parole. Il lessico e problemi di lessicologia*, Novara, Istituto Geografico De Agostini.
- De Mauro, T. (a. c. di) (1994), *Come parlano gli italiani*, Firenze, la Nuova Italia.
- Dobrovol'skij, D. (2011), *Phraseologie und Konstruktionsgrammatik*, in A. Lasch, A. Ziem (Hrsg.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, Tübingen, Stauffenburg, pp. 111-130.
- Ettinger, S. (2007), *Phraseme im Fremdsprachenunterricht*, in H. Burger et al. (Hrsg.), pp. 893-908.
- Ettinger, S. (2009a): *Haben die Männer am Grill die Hosen an? Phraseographie und Sprachwirklichkeit*, in C. Mellado Blanco (Hrsg.), *Theorie und Praxis der idiomatischen Wörterbücher*, Tübingen, Niemeyer, pp. 45-64.

- Ettinger, S. (2009b), *Phraseologie und Phraseodidaktik*, from <http://www.ettinger-phraseologie.de> (28.11.2016).
- Ettinger, S. (2013), *Aktiver Phrasengebrauch und/oder passive Phrasemkenntnisse im Fremdsprachenunterricht. Einige phraseodidaktische Überlegungen*, in I. González Rey (Hrsg.), *Phraseodidactic Studies on German as a Foreign Language. Phraseodidaktische Studien zu Deutsch als Fremdsprache*, Hamburg, Dr. Kovač, pp. 11-30, from [http://www.ettinger-phraseologie.de/media/pdf/Beitrag\\_Ettinger.pdf](http://www.ettinger-phraseologie.de/media/pdf/Beitrag_Ettinger.pdf) (28.11.2016).
- Ettinger, S., Nunes, M. (2006), *Portugiesische Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene*, Hamburg, Buske.
- Fenati, B., Rovere, G., Schemann, H. (con la collaborazione di L. Giacoma) (2009), *Dizionario Idiomatico Tedesco-Italiano*, Bologna, Zanichelli.
- Fillmore, C.J. (1985), *Frames and the semantics of understanding*, in "Quaderni di Semantica", 6, pp. 222-254.
- Fillmore, C.J., Kay, P., O'Connor, M.C. (1988), *Regularity and idiomaticity in grammatical constructions: the case of let alone*, in "Language", 64, pp. 501-538.
- Fleischer, W. (1997), *Phraseologie der deutschen Gegenwartssprache*, 2. Auflage, Tübingen, Niemeyer.
- FRAME = Cotta Ramusino, P., Mollica, F., Schafroth, E. (a c. di) (2014-): *Fraseologia multilingue elettronica*, from [www.fraseologia.it](http://www.fraseologia.it) (28.11.2016).
- Fried, M., Östman, J.-O. (2004), *Construction Grammar: A thumbnail sketch*, in: M. Fried, J.-O. Östman, (eds.), *Construction Grammar in a Cross-Language Perspective*, Amsterdam/Philadelphia, Benjamins, pp. 11-86.
- Giacomini, L., Rovere, G. (2009), *Italianische Lernerwörterbücher aus soziolinguistischer Sicht*, in "Lexicographica", 25, pp. 21-45.
- Giacoma, L. (2012), *Fraseologia e fraseografia bilingue. Riflessioni teoriche e applicazioni pratiche nel confronto Tedesco-Italiano*, Francoforte, Lang.
- Götz, D., Lorenz, G. (2002), *Englische Idioms von A-Z. Mit Erklärungen, Beispielen aus dem Sprachgebrauch und Übersetzungen*, Ismaning, Hueber.
- Goldberg, A.E. (1995), *Constructions: a construction grammar approach to argument structure*, London/Chicago, University of Chicago Press.
- Goldberg, A.E. (2006), *Constructions at work: the nature of generalization in language*, Oxford, Oxford University Press.
- GRADIT = De Mauro, T. (2000), *Grande dizionario italiano dell'uso*, 6 voll., Vol. 7, *Nuove parole I* (2003), Vol. 8, *Nuove parole II* (2007), Torino, UTET (1 CD-ROM 2000, 1 CD-ROM 2003).
- Granger, S., Meunier, F. (eds.) (2008), *Phraseology: an interdisciplinary perspective*, Amsterdam/Philadelphia, Benjamins.
- Gries, S.T. (2008), *Phraseology and linguistic theory: a brief survey*, in S. Granger, F. Meunier (eds.), pp. 3-25.
- Gross, M. (1988), *Les limites de la phrase figée*, in "Langages", 90, pp. 7-22.
- Hallsteinsdóttir, E., Šajánková, M., Quasthoff, U. (2006), *Phraseologisches Optimum für Deutsch als Fremdsprache. Ein Vorschlag auf der Basis von Frequenz- und Geläufigkeitsuntersuchungen*, in "Linguistik online" 27,2, from [http://www.linguistik-online.de/27\\_06/hallsteinsdottir\\_et\\_al.html](http://www.linguistik-online.de/27_06/hallsteinsdottir_et_al.html) (28.11.2016).
- Hallsteinsdóttir, E., Winzer-Kiontke, B., Laskowski, M. (2011), *Phraseodidaktik/Phraseodidactics*, in "Linguistik online", 47,3, from [http://www.linguistik-online.de/47\\_11/](http://www.linguistik-online.de/47_11/) (28.11.2016).
- Hausmann, F.J. (1993), *Ist der deutsche Wortschatz lernbar?*, in "Informationen Deutsch als Fremdsprache", 20, pp. 471-485.
- Hessky, R., Ettinger, S. (1997), *Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene*, Tübingen, Narr.
- Hoffmann, T., Trousdale, G. (eds.) (2013), *The Oxford handbook of construction grammar*, Oxford, Oxford University Press.
- Imperiale, R. (2016), *Fraseoframe saperla lunga*, from <http://lp-italienisch.de/wp-content/uploads/2015/10/Muster-PhraseoFrame-saperla-lunga.pdf> (28.11.2016).

- Lehmann, C. (2016), *Semantische Rollen*, from [www.christianlehmann.eu/ling/lg\\_system/sem/index.html?http://www.christianlehmann.eu/ling/lg\\_system/sem/semant\\_rolle.php](http://www.christianlehmann.eu/ling/lg_system/sem/index.html?http://www.christianlehmann.eu/ling/lg_system/sem/semant_rolle.php) (30.09.2016).
- Lemnitzer, L., Zinsmeister, H. (2015), *Korpuslinguistik. Eine Einführung*. 3. Auflage, Tübingen, Narr.
- Lo Cascio, V. (a c. di) (2013), *Dizionario combinatorio italiano*, Amsterdam/Philadelphia, Benjamins.
- Loporcaro, M. (2006), *Cattive notizie. La retorica senza lumi dei mass media italiani*, Milano, Feltrinelli.
- Lurati, O. (2001), *Dizionario dei modi di dire*, Milano, Garzanti.
- Makkai, A. (1972), *Idiom structure in English*, Den Haag/Paris, Mouton.
- Marello, C. (1996), *Le parole dell'italiano. Lessico e dizionari*, Bologna, Zanichelli.
- Mel'čuk, I. (1995), *Phrasemes in Language and Phraseology in Linguistics*, in M. Everaert et al. (eds.), *Idioms. Structural and Psychological Perspectives*, Hillsdale, NJ & Hove, UK, Erlbaum, pp. 167-232.
- Mel'čuk, I. (2008), *Phraséologie dans la langue et dans le dictionnaire*, in "Repères & Applications", VI, pp. 1-13.
- Mel'čuk, I. (2012), *Phraseology in the language, in the dictionary, and in the computer*, in "Yearbook of Phraseology", 3, pp. 31-56.
- Mel'čuk, I. (2013), *Tout ce que nous voulions savoir sur les phrasèmes, mais ...*, in "Cahiers de lexicologie", 102, pp. 129-149.
- Mel'čuk, I., Polguère, A. (2007), *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, Bruxelles, De Boeck & Larcier.
- Meunier, F., Granger, S. (eds.) (2008), *Phraseology in foreign language learning and teaching*, Amsterdam/Philadelphia, Benjamins.
- Mollica, F. (2015), *Die Rolle der Kontrastivität in der Phraseodidaktik: Eine kognitive und konstruktionsgrammatische Perspektive*, in C. Di Meola, D. Puato (Hrsg.), *Deutsch kontrastiv aus italienischer Sicht: Phraseologie, Temporalität und mehr*, Frankfurt a. M., Lang, pp. 13-35.
- Mollica, F., Schafroth, E. (in corso di stampa), *Der Ausdruck der Intensivierung in komparativen Phrasem-Konstruktionen im Deutschen und im Italienischen*, in K. Steyer, Kathrin (Hrsg.), *Sprachliche Verfestigung. Chunks, Muster, Phrasem-Konstruktionen*, Mannheim, Institut für deutsche Sprache.
- Nuccorini, S. (2007a), *Italian phraseology*, in H. Burger et al. (Hrsg.), pp. 691-703
- Nuccorini, St. (2007b), *Note su alcune fraseologie nei dizionari pedagogici più recenti*, in N. Minerva (a c. di), *Lessicologia e lessicografia negli insegnamenti linguistici 3*, Bologna, Clueb, pp. 35-151.
- Nunberg, G., Sag, I.A., Wasow, T. (1994), *Idioms*, in "Language", 70, pp. 491-538.
- Pittàno, G. (2009), *Dizionario dei modi di dire*, Bologna, Zanichelli.
- Pörings, R., Schmitz, U. (Hrsg.) (2003), *Sprache und Sprachwissenschaft. Eine kognitiv orientierte Einführung*, 2. Auflage, Tübingen, Narr.
- Polenz, P. von (2008), *Deutsche Satzsemantik*, 3. Auflage, Berlin/New York, de Gruyter.
- Polguère, A. (2003), *Collocations et fonctions lexicales: pour un modèle d'apprentissage*, in "Revue Française de Linguistique Appliquée", pp. 117-133.
- Quartu, M., Rossi, E. (2012), *Dizionario dei modi di dire della lingua italiana*, Milano, Hoepli.
- Rovere, G. (2003), *Phraseme in zweisprachigen Wörterbüchern mit Italienisch und Deutsch*, in "Lexicographica", 19, pp. 119-139.
- Rovere, G. (2009), *Komplexe Phraseme im Sprachenpaar Italienisch-Deutsch. Anmerkungen zu ihrer lexikographischen Darstellung*, in W. Heinrich, C. Heiss (Hrsg.), *Fachsprache, elektronische Wörterbücher, multimediale Datenbanken. Empirische Forschungsansätze der Sprach- und Übersetzungswissenschaft. Festschrift für Marcello Soffritti zum 60. Geburtstag*, Monaco, Iudicium, pp. 117-135.
- Schafroth, E. (2009), *Wörterbücher des Italienischen im Vergleich. Zur aktuellen Situation der italienischen Lexikographie*, in "Italienisch", 61, pp. 72-93.
- Schafroth, E. (2011), *Caratteristiche fondamentali di un learner's dictionary italiano. Italiano LinguaDue*, 3, from [riviste.unimi.it/index.php](http://riviste.unimi.it/index.php) (28.11.2016).
- Schafroth, E. (2013a), *Das pragmatische Potential von Phrasemen – illustriert am Deutschen und Italienischen*, in S. Cantarini (Hrsg.), *Wortschatz, Wortschätze im Vergleich und Wörterbücher: Methoden, Instrumente und neue Perspektiven*, Frankfurt a. M., Lang, pp. 185-208.

- Schafroth, E. (2013b), *Fraseoframe Che (cosa, che cosa) me lo chiedi a fare?*, from [http://www.romanistik.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische\\_Fakultaet/Romanistik/Dateien\\_Romanistik/Forschung\\_und\\_Projekte/FraseoFrame\\_a\\_fare\\_online\\_it.pdf](http://www.romanistik.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Romanistik/Dateien_Romanistik/Forschung_und_Projekte/FraseoFrame_a_fare_online_it.pdf) (06.10.2016).
- Schafroth, E. (2014), *Eine Sache des Verstehens: Phraseme als Konstruktionen und ihre Beschreibung in der Lexikographie Französisch/Deutsch*, in M.J. Domínguez Vázquez, F. Mollica, M. Nied Curcio, (Hrsg.), *Zweisprachige Lexikographie zwischen Translation und Didaktik*, Berlin/New York, de Gruyter, pp. 83-111.
- Schafroth, E. (dir.) (2014-): *Piattaforma per chi impara l'italiano/Lernerplattform Italienisch*, Istituto di romanistica, università di Düsseldorf, from <http://lp-italienisch.de> (28.11.2016).
- Schafroth, E. (2015), *Italian phrasemes as constructions: how to understand and use them*, in "Journal of Social Sciences", 11/3, pp. 317-337, from <http://thescipub.com/PDF/jssp.2015.317.337.pdf> (28.11.2016).
- Serianni, L., Trifone, M. (a c. di) (2014): (Il) *Devoto-Oli 2014. Vocabolario della lingua italiana*, Firenze, Le Monnier.
- Sinclair, J. (1991), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Sornicola, R. (1994), *Quattro dimensioni nello studio del parlato*, in T. De Mauro (a c. di), pp. 111-130.
- Stathi, K. (2006), *Korpusbasierte Analyse der Semantik von Idiomen*, in "Linguistik online", 27/2, 73-89, from [http://www.linguistik-online.de/27\\_06/stathi.html](http://www.linguistik-online.de/27_06/stathi.html) (28.11.2016).
- Stathi, K., Hümmel, C. (2006), *Polysemy and vagueness in idioms: a corpus-based analysis of meaning*, in "International Journal of Lexicography", 19, pp. 361-377.
- Treccani. *Dizionario della lingua italiana* (2013), Roma, Istituto della Enciclopedia Italiana.
- Turrini, G., Alberti, C., Santullo, M.L., Zanchi, G. (a c. di) (1995), *Capire l'antifona. Dizionario dei modi di dire con esempi d'autore*, Bologna, Zanichelli.
- Ziem, A. (2015), *Nullinstanziierungen im gesprochenen Deutsch: Kohärenz durch grammatische Konstruktionen?*, in J. Bücker, S. Günthner, W. Imo (Hrsg.), *Konstruktionsgrammatik V. Konstruktionen im Spannungsfeld von sequentiellen Mustern, kommunikativen Gattungen und Textsorten*, Tübingen, Stauffenburg, pp. 45-80.
- Ziem, A., Lasch, A. (2013), *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze*, Berlin/Boston, de Gruyter.

## B. Corpora

- BADIP (Banca Dati dell'Italiano Parlato), from [badip.uni-graz.at/it/](http://badip.uni-graz.at/it/) (28.11.2016).
- Corpus CORIS, from [corpora.dslo.unibo.it/TCORIS/](http://corpora.dslo.unibo.it/TCORIS/) (28.11.2016).
- Cosmas II = Institut für Deutsche Sprache (2016), Corpus Search, Management and Analysis System, from <http://www.ids-mannheim.de/cosmas2/> (28.11.2016).
- itTenTen (2010-), from <https://www.sketchengine.co.uk/ittenten-corpus> (28.11.2016)
- PAISÀ, from <http://www.corpusitaliano.it/> (28.11.2016).
- "La Repubblica" Corpus, from [dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica](http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica) (28.11.2016).
- WEBBIT, from [clic.cimec.unitn.it/marco/webbit/](http://clic.cimec.unitn.it/marco/webbit/) (28.11.2016).

**RICCARDO IMPERIALE** • He is part of the Academic staff at Heinrich-Heine-Universität, Düsseldorf, Germany, where he actively cooperates on the research group's projects.

**E-MAIL** • [Riccardo.Imperiale@hhu.de](mailto:Riccardo.Imperiale@hhu.de)

**ELMAR SCHAUFROTH** • He is Professor and chair in Romance Linguistics (Italian and French) at Heinrich-Heine-Universität, Düsseldorf, Germany. Among his research interests there are Italian, French and German lexicology and lexicography, especially learners' lexicography and phraseology; gender; language and music; discourse analyses; language variation. He is co-editor of *Studia*

---

*Romanica et Linguistica, Grundlagen der Romanistik, and Kontrastive Linguistik/Linguistica contrastive.* His latest work: *Französische Lexikographie*, de Gruyter, Berlin/Boston 2014.

**E-MAIL** • [schafroth@phil.hhu.de](mailto:schafroth@phil.hhu.de)

# PaSSAGGI

---





# LA TRADUCCIÓN DEL TEXTO POÉTICO

Los perros románticos de Roberto Bolaño

---

Giorgia ESPOSITO

**ABSTRACT** • Italian poetry took its first steps during the 13th century, when the poets of the “Scuola Siciliana” started translating troubadoursque poems. Notwithstanding, along the centuries, many have branded poetic translation as an impossible task. The emergence of free verse poems, typical of contemporary poetry, reopens the debate on which have to be considered the characteristic elements of the genre. If rhyme and metre are not the poetry's dominant constituents anymore, thence, the translation problems will resemble those of the prose texts. Still, terminological rigour and narrative rhythm represent the genre main cores. Poets and translators' activities, therefore, can be considered akin, to the extent that both face language, looking for the suitable expression and purchasing perfection, yet knowing that their task, though necessary, will still be utopian.

**KEYWORDS** • Bolaño, Faithfulness, Free Verse, The Romantic Dogs, Untranslatability

## 1. El concepto de fidelidad en la traducción del texto poético

Entre los grandes interrogativos que no se ha conseguido dirimir nunca, destaca la dicotomía entre una traducción fiel al espíritu y otra fiel a la letra del texto.<sup>1</sup> Estas dos manifestaciones de fidelidad al texto fuente<sup>2</sup> siguen pareciendo irreconciliables, aun en la segunda mitad del siglo XX, cuando la teoría de la traducción o traductología<sup>3</sup> ya se ha configurado como disciplina

---

<sup>1</sup> Cicerón es el primero en sostener que prefiere traducir: «nec conuerti ut interpres, sed ut orator, sententiis isdem et earum formis tamquam figuris, uerbis ad nostram consuetudinem aptis. In quibus non uerbum pro uerbo necesse habui reddere, sed genus omne uerborum uimque seruaui. Non enim ea me annumerare lectori putauit oportere, sed tamquam appendere.» Cicero, *De optimo genere oratorum*, V, 14, in Cicerón (1964), *L'orateur*, Paris, Les Belles Lettres, p. 114. Si bien, para los Romanos, el concepto de traducción coincidiese con el de *imitatio*, es Cicerón el que da comienzo a la dicotomía entre letra y espíritu en traducción.

<sup>2</sup> Recurro a la locución «texto fuente», en detrimento de otras, cuyo uso tal vez sea más extendido (texto de partida, prototexto, o ST, sigla del anglicismo *source text*), porque estoy de acuerdo con Umberto Eco, cfr. Eco U. (2003), *Dire quasi la stessa cosa*, Milano, Bompiani, p. 17, cuando afirma que la locución *testo fonte* permite algunas inferencias metafóricas, representativas de la relación entre texto original y texto traducido.

<sup>3</sup> Entre las numerosas denominaciones en uso para referirse a la disciplina que se ocupa de estudiar la traducción, Hurtado Albir (2001) señala que, en español, «la denominación más extendida para referirse a la disciplina que se ocupa de estudiar la traducción parece ser la de *Teoría de la traducción* o *Traductología*, [sin embargo] ésa no es la única denominación existente; otras denominaciones coexisten: *Lingüística aplicada a la traducción*, *Translémica*, *Translatología*, *Ciencia de la traducción*, *Estudios sobre la traducción* y *Estudios de la traducción*.» Hurtado Albir A. (2001), *Traducción y Traductología. Introducción a la Traductología*, Madrid, Cátedra, p. 133. Al tener que elegir, Hurtado opta por «la denominación *Traductología*, concibiéndola como la disciplina, con entidad propia, encargada de analizar

autónoma, relacionada, pero no subordinada, a la lingüística, a la filosofía y a la crítica literaria. Dicha divergencia se acentúa a la hora de hablar de traducción de textos poéticos; aquí también, las raíces del debate acerca de la supuesta imposibilidad de una fidelidad al espíritu y a la letra a la vez, de la que derivaría la intraducibilidad del texto poético, cuentan con antecesores ilustres. El primer gran detractor de la traducción poética es el poeta Dante Alighieri, que, en el *Convivio*, afirma: “E però sappia ciascuno che nulla cosa per legame musaico armonizzata si può de la sua loquela in altra trasmutare senza rompere tutta sua dolcezza e armonia” (Alighieri 1965: 21). Dante rechaza, en sede teórica, la traducibilidad de la poesía, poniendo el acento sobre la imposibilidad de *trasmutare*, sin corromper su gracia y armonía, el *legame musaico*, es decir específicamente musical y rítmico.<sup>4</sup> Llega a las mismas conclusiones Roman Jakobson (1959), en su ensayo *On linguistic aspects of translation*, donde afirma que:

In poetry, verbal equations become a constructive principle of the text. Syntactic and morphological categories, roots, and affixes, phonemes and their components (distinctive features) – in short, any constituents of the verbal code – are confronted, juxtaposed, brought into contiguous relation according to the principle of similarity and contrast and carry their own autonomous signification. Phonemic similarity is sensed as semantic relationship. (Jakobson 1959: 238)

Jakobson defiende la intraducibilidad del texto poético, a causa del valor semántico de la similitud entre los fonemas, que se reconduce a la dicotomía inconciliable entre espíritu (*genum*) y letra (*verbum*), puesto que, en el texto poético, el sentido y la forma se encuentran en estrecha relación. Al transmutar un texto a otro idioma, es imposible mantener esta armonización entre significado y significante, de lo que derivaría la intraducibilidad del texto poético, del que Jakobson solo admite una trasposición creativa, pero nunca una traducción: «poetry by definition is untranslatable. Only creative transposition is possible».<sup>5</sup> Según el lingüista, es la estructura misma del lenguaje la que invalida la traducibilidad de la poesía. Sin embargo, y a pesar de los muchos e ilustres detractores, se sigue traduciendo todo tipo de textos, sin exceptuar los poéticos.<sup>6</sup>

---

la traducción (oral, escrita, audiovisual) y que asume, pues, el conjunto de estudios en torno a ella.» *Ivi*, p. 135.

<sup>4</sup> Sin embargo, la poesía en lengua italiana nace a través de las traducciones de la poesía trovadoresca, que los poetas de la Scuola Siciliana llevaron a cabo, a partir del siglo XIII. Gianfranco Folena señala que «Il primo poeta-traduttore della letteratura italiana è anche probabilmente il primo poeta nostro in senso assoluto, Iacopo da Lentini, l'instauratore, come pare sempre più evidente, della poesia siciliana.» Folena G. (1991), *Volgarizzare e tradurre*, Torino, Einaudi, p. 25. Octavio Paz (1990) también ve, en las traducciones, el origen de la poesía europea: «Los grandes períodos creadores de la poesía de Occidente, desde su origen en Provenza hasta nuestros días, han sido precedidos o acompañados por entrecruzamientos entre diferentes tradiciones poéticas. Esos entrecruzamientos a veces adoptan la forma de la imitación y otras de la traducción. Desde este punto de vista la historia de la poesía europea podría verse como la historia de las conjunciones de las diversas tradiciones que componen lo que se llama literatura de Occidente». Paz O. (1990), *Traducción: literatura y literalidad*, Barcelona, Tusquets Editores, p. 24.

<sup>5</sup> Jakobson R. (1959), «*On linguistic aspects of translation*», *On Translation*, New York, Oxford University Press, p. 238.

<sup>6</sup> Umberto Eco señala que «dalla prima metà dello scorso secolo in avanti sono state elaborate teorie della struttura di una lingua, o della dinamica dei linguaggi, che ponevano l'accento sul fenomeno della radicale impossibilità della traduzione; sfida non da poco per gli stessi teorici che, pur elaborando queste teorie, si rendevano conto che di fatto, e da millenni, la gente traduce.» Eco U. (2003), *op. cit.*, p. 17.

Octavio Paz ve, en el género poético, «un más allá del lenguaje»<sup>7</sup> y, en la traducción y la creación poética, «operaciones gemelas.»<sup>8</sup> Si la lectura es una mirada sobre el mundo a través de ojos ajenos, y si la traducción consiste en captar esa mirada y devolverla a otro idioma, diciendo «casi la misma cosa»,<sup>9</sup> la traducción de la poesía – que tal vez sea la mirada más profunda que la literatura admita – acaso sea una forma de poesía, en el momento en que capta esa profundidad y la hace inteligible en otro idioma. Entender el sentido de las palabras es una operación propia de la lectura, y máxime de la lectura de un texto poético; la traducción solo añade el movimiento, ese conducir,<sup>10</sup> de una a otra lengua, el desarrollo de un pensamiento lingüístico. En el proceso interlingüístico, a la ecuación entre poesía y traducción, se suma el movimiento: la traducción es poesía en movimiento entre dos sistemas lingüísticos. Para Octavio Paz la traducción poética «es una operación análoga a la creación poética, sólo que se despliega en sentido inverso».<sup>11</sup>

El punto de partida del traductor no es el lenguaje en movimiento, materia prima del poeta, sino el lenguaje fijo del poema. Lenguaje congelado y, no obstante, perfectamente vivo. Su operación es inversa a la del poeta: no se trata de construir con signos móviles un texto inamovible, sino desmontar los elementos de ese texto, poner de nuevo en circulación los signos y devolverlos al lenguaje. (Paz 1990: 22)

Si la labor del poeta consiste en moldear el lenguaje en movimiento, el traductor tiene que trabajar con el lenguaje fijo del texto fuente. Sin embargo, el papel del poeta es diferente al papel del traductor, porque, como señala Walter Benjamin (1923), la intención hacia el lenguaje de cada uno es diferente: «la intención de un autor es natural, primitiva e intuitiva, la del traductor es derivada, ideológica y definitiva».<sup>12</sup>

Precisamente por ser la traducción una forma peculiar, la función del traductor tiene también un carácter peculiar, que permite distinguirla exactamente de la del escritor. Esta función consiste en encontrar en la lengua a la que se traduce una actitud que pueda despertar en dicha lengua un eco del original. Esta es una característica de la traducción que marca su completa divergencia respecto a la obra literaria, porque su actitud nunca pasa al lenguaje como tal, o sea a su totalidad, sino que se dirige sólo de manera inmediata a determinadas relaciones lingüísticas. Porque la traducción, al contrario de la creación literaria, no considera como quien dice el fondo de la selva idiomática, sino que la mira desde afuera, mejor dicho, desde en frente y sin penetrar en ella hace entrar al original en cada uno de los lugares en que eventualmente el eco puede dar, en el propio idioma, el reflejo de una obra escrita en una lengua extranjera. (Benjamin 1923: 136)

<sup>7</sup> Paz O. (1990), *op. cit.*, p. 22.

<sup>8</sup> *Ivi*, p. 23.

<sup>9</sup> «Per bene che vada, traducendo si dice *quasi* la stessa cosa. Il problema del *quasi* diventa ovviamente centrale nella traduzione poetica, sino al limite della ricreazione così geniale che dal quasi si passa a una cosa assolutamente *altra*, un'altra cosa, che con l'originale ha solo un debito, vorrei dire, morale. Pero è interessante vedere dove talora il traduttore, sapendo che può dire solo un *quasi*, va a cercare il *nucleo della cosa* che vuole rendere (sia pure quasi) a ogni costo.» Eco U. (2003), *op. cit.*, p. 277.

<sup>10</sup> Los términos utilizados para referirse a la traducción «si rifanno tutti alla stessa metafora: l'idea cioè di *far passare*, di facilitare il passaggio da una lingua all'altra, di *trasportare* in un'altra lingua il significato di un determinato idioma, idea che si ritrova a partire dal latino *tra-duco* o *transfero* [...] fino all'italiano *tradurre*, al francese *traduire*, al tedesco *übersetzen*, al russo *perevodit'*». Mounin G. (1965), *Teoria e storia della traduzione*, Torino, Einaudi, p. 19. (Traducción al italiano de Stefania Morganti).

<sup>11</sup> Paz O. (1990), *op. cit.*, p. 20.

<sup>12</sup> Benjamin W. (1923), *Angelus novus*, Barcelona, Edhasa, 1971, p. 137. (Traducción al español de Héctor Álvarez Murena).

Para que un traductor de poesía consiga ver, dentro de 'la selva idiomática', la obra poética, y logre reproducir su 'eco', es preciso que cuente con una sensibilidad poética.<sup>13</sup> Paralelamente, al poeta que quiera traducir, se le exige una competencia traductora.<sup>14</sup>

## 2. La traducción de *Los perros románticos* de Roberto Bolaño

Roberto Bolaño,<sup>15</sup> conocido por el gran público por sus novelas,<sup>16</sup> empezó su carrera literaria como poeta vanguardista.<sup>17</sup> Los poemas que forman parte de *Los perros románticos*<sup>18</sup> se configuran, por lo tanto, como textos libres de restricciones métricas, y en los que el ritmo de la narración no se aleja del de la prosa del autor. El aspecto formal de los poemas no ha representado, pues, un problema primario a la hora de traducir. Sin embargo, he intentado respetar, en la medida de lo posible, la organización sintáctica del texto fuente, proceso favorecido por la cercanía entre la lengua española y la italiana, y he procurado, especialmente, no cambiar la colocación de las palabras en posición final de verso. He evitado recurrir a sinónimos que hubieran debilitado, o desviado, la semántica del texto fuente, aún a costa de renunciar a la correspondencia silábica con el original, porque he considerado el contenido semántico más relevante que el metro. Los problemas traductológicos mayores no han sido de tipo lingüístico-formal, sino en lo que

<sup>13</sup> En el prefacio al *Libro secondo della Eneide*, Giacomo Leopardi (1817) afirma: «senza esser poeta non si può tradurre un vero poeta.» Virgilio (1817), *Eneide. Libro secondo*, Milano, Pirotta, traducción al italiano de Giacomo Leopardi. Franco Buffoni (1989) vuelve a proponer la frase de Leopardi, para ampliar su significado: «Occorre essere poeti soprattutto per tradurre non traducendo.» Buffoni F., «*Leopardi in lingua inglese come paradigma della simbolicità del compito di un poeta traduttore*», in Buffoni F. (a cura di) (1989), *La traduzione del testo poetico*, Milano, Guerini e Associati, p. 107.

<sup>14</sup> «En lo que se refiere a la competencia traductora», señala Hurtado Albir (2001), «no existe una tradición investigativa comparable a la que se ha dado en torno a la competencia comunicativa, y la reflexión es todavía incipiente.» Hurtado Albir (2001), *op. cit.*, p. 382.

<sup>15</sup> Roberto Bolaño Ávalos (Santiago de Chile, 1953 - Barcelona, 2003).

<sup>16</sup> Hasta 1998, fecha de publicación de *Los detectives salvajes* (Barcelona, Anagrama), novela que le otorga el “Premio Herralde de Novela”, en 1998, y el “Premio Internacional de Novela Rómulo Gallegos”, en 1999, Bolaño es un escritor casi desconocido, que se gana la vida participando en concursos literarios de provincias.

<sup>17</sup> En la primera mitad de los años Setenta, Bolaño funda, juntos con su amigo, el poeta Mario Santiago Papasquiaro (Ulises Lima en *Los detectives salvajes*), el “movimiento infrarrealista”, una vanguardia más literaria que real. Matías Ayala Munita señala que el infrarrealismo «se volvió real en la medida en que fue ficcionalizado [bajo el nombre de “real visceralismo” o “realismo visceral”] en *Los detectives salvajes*.» Ayala M., «*Notas sobre la poesía de Roberto Bolaño*», in Paz Soldán E., Faverón Patriau G. (a cargo de) (2008), *Bolaño salvaje*, Barcelona, Candaya, p. 93.

<sup>18</sup> *Los perros románticos* es el primer libro de poemas publicado en España por Roberto Bolaño. La primera edición, Zarautz, Fundación Social y Cultural Kutxa, 1993, que le otorga el “Premio Literario Ciudad de Irún 1994”, contiene cuarenta y cinco poemas, escritos entre 1977, año en que Bolaño llega a Europa, a Barcelona, y 1990. Los poemas están subdivididos en cinco capítulos temáticos: Poetas, Detectives, Amores, Hospitales y Crepúsculos. Solo veintitrés poemas confluyen, juntos con veinte textos inéditos, en las dos ediciones sucesivas: Barcelona, Lumen, 2000 y Barcelona, Acantilado, 2006. A partir de la segunda edición, *Los perros románticos*, ya sin división en capítulos temáticos y acompañado del prefacio del poeta y crítico literario Pere Gimferrer, contiene los poemas escritos entre 1980, fecha en que Bolaño deja Barcelona para mudarse a Gerona, y 1998, año de publicación de *Los detectives salvajes* (Barcelona, Anagrama).

concierna a los elementos culturales específicos de la realidad latinoamericana y al diálogo incesante que Bolaño establece con el resto de su obra.

A pesar de la invitación de Benjamin a ignorar, en los procesos de creación y de traducción, al receptor de una obra de arte,<sup>19</sup> en el caso de Bolaño, había que tener en cuenta por lo menos dos factores: en primer lugar, la posible familiaridad del lector con el resto de la producción del autor y, en segundo lugar, la comparación con el texto original, facilitada por la edición bilingüe y la cercanía entre los dos idiomas. Por esta razón he optado por una fidelidad a la letra, casi servil, allí donde un sinónimo tal vez hubiera hecho el texto más fluido, mas corriendo el riesgo de perder la confianza del lector. Franco Buffoni advierte: «ciò che serve a un lettore italiano di poesia francese o spagnola [...] è un buon glossario a fronte o una fedelissima traduzione di servizio».<sup>20</sup> La familiaridad con el resto de la producción del autor ha sido necesaria para no incurrir en el error de corregir, en nombre de la gramática italiana, solecismos que tienen un significado claro para el lector experto, como en el caso del poema *La francesa* (v. 52):

No me quiero morir, susurraba  
mientras se corría  
En la perspicaz oscuridad del  
dormitorio,  
Y yo no sabía qué decir,  
En verdad no sabía qué decir,  
Salvo acariciarla y sostenerla  
mientras se movía  
Arriba y abajo como la vida,  
Arriba y abajo como las  
**poetas**<sup>21</sup> de Francia  
Inocentes y castigadas

Non voglio morire, sussurrava  
mentre veniva  
Nel buio penetrante della  
camera,  
E io non sapevo cosa dire,  
Non sapevo davvero cosa dire,  
Salvo accarezzarla e sostenerla  
mentre si muoveva  
Su e giù come la vita,  
Su e giù come le **poete** di  
Francia  
Innocenti e castigate

Los solecismos voluntarios se han de mantener en la traducción,<sup>22</sup> sobre todo si representan una marca del autor, como en el caso de “poeta” para el género femenino.<sup>23</sup> En efecto, para Bolaño, el término “poetisa” tiene una connotación peyorativa:<sup>24</sup>

<sup>19</sup> «Cuando nos hallamos en presencia de una obra de arte o de una forma artística nunca advertimos que se haya tenido en cuenta al destinatario para facilitarle la interpretación. No se trata sólo de que la referencia a un público determinado o a sus representantes contribuya a desorientar, sino de que incluso el concepto de un destinatario «ideal» es nocivo para todas las explicaciones teóricas sobre el arte, porque éstas han de limitarse a suponer principalmente la existencia y la naturaleza del ser humano. De tal suerte, el arte propiamente dicho presupone el carácter físico y espiritual del hombre; pero no existe ninguna obra de arte que trate de atraer su atención, porque ningún poema está dedicado al lector, ningún cuadro a quien lo contempla, ni sinfonía alguna a quienes la escuchan.» Benjamin W. (1923), *op. cit.*, p. 127.

<sup>20</sup> Buffoni F., *op. cit.*, in Buffoni F. (1989), *op. cit.*, p. 112.

<sup>21</sup> A fin de facilitar la comparación entre texto fuente y traducción, se señalan, con letras en negrita, los elementos textuales analizados en clave traductológica.

<sup>22</sup> «Non importa se il purista si irrita [...]: volute o non volute, erano nell'originale. Non importa nemmeno se l'eventuale solecismo verrà magari attribuito al traduttore.» Osimo B. (2004), *Manuale del traduttore*, Milano, Hoepli, p. 88.

<sup>23</sup> El uso de “poeta”, para referirse a ambos géneros, está cada vez más extendido, y también el DRAE lo registra (<http://dle.rae.es/?id=TUUD7eg>). Sin embargo, en una visión diacrónica, la elección de Bolaño ha de considerarse un solecismo voluntario. También la traductora de la novela del mismo autor, *Los detectives salvajes*, recurre al término “poeta” en contraposición a “poetessa”. Cfr., *ivi*, nota 27.

<sup>24</sup> También en *Amuleto*, la novela que cuenta la historia de Auxilio Lacouture, «madre de la poesía mexicana» (Bolaño R., 1999, *Amuleto*, Barcelona, Anagrama, p. 11), la protagonista no se define nunca “poetisa”, sino siempre “poeta”. Cfr., *ivi*, pp.: 35, 101, 103, 114, 117.

- ¿Hay muchas poetisas?  
 – Decirles poetisas queda un poco gacho – dijo Pancho.  
 – Se les dice poetas – dijo Barrios.  
 (Bolaño R., 1998, *Los detectives salvajes*, Barcelona, Anagrama, p. 31)<sup>25</sup>

La traducción de las figuras retóricas puede que implique alejarse de la letra del texto, para transmitir el espíritu y la imagen mental que dichas figuras evocan en el lector. En el poema *Los pasos de Parra* (vv.: 6, 19, 31, 42), el autor se refiere a Chile, recurriendo a la metonimia “el Flandes indiano”, cuyo equivalente en italiano sería “le Fiandre americane”:<sup>26</sup>

|  |  |
|--|--|
| Oh el Flandes indiano                          | Oh le Fiandre americane                        |
| El interminable pasillo de nuestro descontento | Lo sconfinato corridoio del nostro malcontento |
| En donde todo lo hecho                         | Dove ogni cosa fatta                           |
| parece deshecho                                | sembra disfatta                                |
| El país de Zurita                              | Il paese di Zurita <sup>27</sup>               |
| y de las cordilleras fritas                    | e delle cordigliere fritte                     |
| El país de la eterna juventud                  | Il paese dell'eterna giovinezza                |

Las expresiones idiomáticas constituyen un problema traductológico de primera importancia: al estar, en muchos casos, arraigadas en los referentes culturales propios de la lengua fuente, no siempre cuentan con un equivalente en otra cultura y otro idioma. A falta de un homólogo, el traductor deberá encontrar una expresión análoga, que contenga, en la medida de lo posible, alguna continuidad lógica con la locución del texto fuente. Lo mismo se aplica a los casos en que el autor distorsiona adrede una expresión idiomática, cambiando algún elemento, con el propósito de desorientar o producir un efecto cómico, como en el caso del poema *Atole* (v. 3), en que Bolaño hace referencia a la expresión idiomática: “dar atole con el dedo”,<sup>28</sup> pero sustituye el verbo “dar” por “tomar”:

Vi a Mario Santiago y Orlando Guillén  
 los poetas perdidos de México

Ho visto Mario Santiago e  
 Orlando Guillén

<sup>25</sup> ««Ci sono molte poetesse?»./ «La parola poetessa è un po' brutta» ha detto Pancho./ «Si dice una poeta» ha detto Barrios.» Bolaño R. (2014), *I detective selvaggi*, Milano, Adelphi, p. 22 (Traducción al italiano de Ilide Carmignani).

<sup>26</sup> Hacen falta dos apuntes, con respecto al uso de los dos términos en el español de Latinoamérica: “indiano” se usa para referirse a los españoles que hayan vivido en América, el “Flandes” es una metonimia para indicar las colonias españolas, aportadoras de riqueza a la Corona (el Flandes europeo y el americano), de la que deriva la expresión idiomática: “clavar una pica en Flandes” (obtener éxito en el extranjero). Por esto, he optado por traducir “indiano” con “americano”, puesto que, para el lector italiano, el término “indiano” remite más al subcontinente asiático que al americano.

<sup>27</sup> Raúl Zurita Canessa (Santiago de Chile, 1950) es un poeta chileno, ganador del “Premio Nacional de Literatura de Chile”, en el año 2000. Bolaño dice de él: «Zurita crea una obra magnífica, que descuella entre los de su generación y que marca un punto de no retorno con la poética de la generación precedente, pero su escatología, su mesianismo, son también los puntales de un mausoleo o de una pira funeraria hacia la que se encaminaron, en los años ochenta, casi todos los poetas chilenos.» Bolaño R. (2004), *«La poesía chilena y la intemperie»*, *Entre Paréntesis*, Barcelona, Anagrama, pp. 88, 89.

<sup>28</sup> El atole es una bebida caliente, a base de harina de maíz y agua, típica de América Central; la locución a la que remite, en cambio, significa: engañar, defraudar. Hacía falta encontrar una expresión en italiano que contuviese, por un lado, la idea del engaño y, por el otro, la referencia al acto de beber.

tomando atole con el dedo

i poeti perduti del Messico  
bere dalla coppa dell'inganno

A causa del anisomorfismo de las lenguas naturales, hay términos que son intraducibles o que solo pueden traducirse a través de una perífrasis. Si, para los textos en prosa, el recurso a una perífrasis puede ser una solución, lo mismo no puede aplicarse al género poético. En estos casos, es preciso buscar un equivalente que, aunque no consituya una traducción fiel del original, produzca un efecto análogo y no desgaje, tal como haría una perífrasis, el ritmo del texto. El verbo “mantear”, que aparece en el poema *Palingenesia* (v. 10), no tiene un equivalente en italiano, por lo que ha sido necesario poner en práctica las estrategias de compensación antes mencionadas:<sup>29</sup>

Los bárbaros avanzan, susurró  
melodiosamente, una masa  
disforme,  
Grávida de aullidos y  
juramentos, una larga  
noche **manteada**  
Para iluminar el matrimonio de  
los músculos y la grasa.

I barbari avanzano, sussurrò  
dolcemente, una massa  
informe,  
Gravida di ululati e bestemmie,  
una lunga notte **sbrigliata**  
Per illuminare il matrimonio  
fra i muscoli e il grasso.

Ahí donde el texto fuente presente rimas internas, que no se pueden mantener sin traicionar la semántica, el traductor puede adoptar estrategias de compensación. Una opción viable consiste en desplazar el elemento de la rima a otro término del mismo verso, como en el caso de *Sucio, mal vestido* (vv. 19-20):

Sumulistas y teólogos,  
adivinadores  
y **salteadores** de caminos  
emergieron  
como realidades acuáticas en  
medio de una realidad  
metálica.

Logici e teologi, **indovini**  
e pirati di **cammini**<sup>30</sup> emersero  
come realtà acquatiche in  
mezzo a una realtà metallica.

Si no es posible cambiar los elementos que producen la rima, otra estrategia de compensación, para preservar el fluir rítmico del texto, reside en convertir la rima interna en una asonancia. En el poema *Junto al acantilado* (vv. 6, 9), no se podía desplazar el elemento de la rima, puesto que los dos términos revestían una carga semántica esencial en la economía del verso,

<sup>29</sup> “Mantear” significa «Lanzar al aire entre varias personas, con una manta cogida por las orillas, a otra, que al caer sobre la manta vuelve a ser lanzada repetidas veces hacia arriba» (<http://dle.rae.es/?id=OGbLbD8>). En italiano no existe un término para designar este proceso, por lo que he buscado un equivalente que contuviese la imagen mental que el acto sugiere: alborozo, movimiento repentino, libertad. “Una larga noche manteada” se ha convertido, pues, en “una lunga notte sbrigliata”. En la versión en inglés, la traductora ha optado por el verbo “to toss-up” (tirar una moneda, echar a cara o cruz). Bolaño R. (2008), «*Palingenesia*», *The Romantic Dogs*, New York, New Directions, p. 71. (Traducción al inglés de Laura Healy).

<sup>30</sup> Desde el punto de vista lingüístico, la solución «pirati di cammini» tal vez exceda en sugestión al original, puesto que «salteadores de caminos» es prácticamente una colocación. «Briganti di cammini» sería una traducción más literal, sin embargo la palabra «pirati» también remite a «salteadores» y, además, engarza muy bien con el verso siguiente («realidades acuáticas»).

así que, siguiendo un discurso analógico, “las pesadillas” se han convertido en “le paure”, y “las astillas” en “le macerie”:

Y una mujer, la proyección  
holográfica de una mujer,  
salía a la terraza  
a contemplar las **pesadillas** o  
las **astillas**.  
Nadie entendía nada.  
Todo fallaba: el sonido, la  
percepción de la imagen.  
Pesadillas o astillas  
empotradas en el cielo  
a las nueve de la noche.

E una donna, la proiezione  
olografica di una donna,  
usciva sul terrazzo  
a contemplare le **paure** o le  
macerie.  
Nessuno capiva niente.  
Tutto veniva meno: il suono, la  
percezione dell'immagine.  
Le paure o le macerie  
incassate nel cielo  
alle nove di sera.

### 3. Conclusión

Si bien es verdad que no todos los textos poéticos tienen como elemento dominante la rima o el metro y que, por consiguiente, los problemas traductológicos que suponen no se alejan de los de la prosa, también es cierto que el lenguaje poético, a pesar de la versificación libre, típica de la poesía contemporánea, suele caracterizarse por una rigurosidad terminológica, que requiere el máximo tesón por parte del traductor. Preservar las rimas internas y las asonancias (cfr. *supra*, p. 13), así como encontrar una solución para los términos intraducibles, sin recurrir a una perífrasis (cfr. *supra*, p. 12), son cuestiones que, en un texto en prosa, se pueden pasar por alto, sin que esto conlleve, necesariamente, pérdidas graves. Al contrario, el género poético admite un grado de aproximación mucho menor por parte del traductor, puesto que forma y contenido se entrelazan para conferir significado al texto.

La creación poética, así como la traducción, consiste en seguir el desarrollo del pensamiento lingüístico, hasta hallar su origen, para traducirlo en palabras. Si, en la primera, la respuesta lingüística puede darse a través de una fulguración instantánea y de una iluminación, en la segunda, el resultado es el fruto de elecciones razonadas: al traductor, a diferencia del poeta, se le exige lucidez hasta el final de su labor, puesto que persigue el origen de un pensamiento lingüístico ajeno. Creación poética y traducción son operaciones análogas, que difieren, básicamente, por la existencia, en la segunda, del *otro* como sujeto lingüístico. En cuanto a fines, la práctica de la una puede fortalecer la otra: el estudio de la poesía y la práctica de la traducción poética pueden ayudar a traducir mejor, incluso los textos que no son poéticos. La puntualidad del lenguaje poético, de hecho, impone al traductor un esmero, del cual otros tipos de textos pueden prescindir; de ahí, la importancia que la práctica de la traducción poética cobra, en cuanto ejercicio finalizado a afinar el ahínco del traductor.

### BIBLIOGRAFÍA

- Alighieri D. (1965), *Convivio*, Alpignano, Tallone Editore.  
Benjamin W. (1923), *Angelus Novus*, Barcelona, Edhasa, 1971. (Traducción al español de Héctor Álvarez Murena).  
Bolaño R. (1998), *Los detectives salvajes*, Barcelona, Anagrama.  
Bolaño R. (1999), *Amuleto*, Barcelona, Anagrama.  
Bolaño R. (2004), *Entre paréntesis*, Barcelona, Anagrama.



- 
- Bolaño R. (2006), *Los perros románticos*, Barcelona, Acantilado, III edición. [I edición: 1993, II edición: 2000]
- Bolaño R. (2008), *The Romantic Dogs*, New York, New Directions. (Traducción al inglés de Laura Healy).
- Bolaño R. (2014), *I detective selvaggi*, Milano, Adelphi. (Traducción al italiano de Ilide Carmignani).
- Buffoni F. (a cargo de) (1989), *La traduzione del testo poetico*, Milano, Guerini e Associati.
- Cicero (1964), *L'orateur*, Paris, Les Belles Lettres. (Traducción al francés de Albert Yon).
- Eco U. (2003), *Dire quasi la stessa cosa*, Milano, Bompiani.
- Folena G. (1991), *Volgarizzare e tradurre*, Torino, Einaudi.
- Hurtado Albir A. (2001), *Traducción y Traductología. Introducción a la Traductología*, Madrid, Cátedra.
- Jakobson R. (1959), «*On linguistic aspects of translation*», *On Translation*, New York, Oxford University Press.
- Mounin G. (1965), *Teoria e storia della traduzione*, Torino, Einaudi. (Traducción al italiano de Stefania Morganti).
- Ortega y Gasset J. (1937), «*Miseria y esplendor de la traducción*», en Ortega y Gasset J. (1957), *Obras completas*, vol. 5, Madrid, Alianza Editorial: Revista de Occidente.
- Osimo B. (2004), *Manuale del traduttore*, Milano, Hoepli.
- Paz O. (1990), *Traducción: literatura y literalidad*, Barcelona, Tusquets Editores, III edición. [I edición: 1971]
- Paz Soldán E., Faverón Patriau G. (a cargo de) (2008), *Bolaño salvaje*, Barcelona, Candaya.
- Schulte R., Biguenet J. (a cargo de) (1992) *Theories of Translation: An Anthology of Essays from Dryden to Derrida*, Chicago, University of Chicago Press.
- Virgilio (1817), *Eneide. Libro secondo*, Milano, Pirotta. (Traducción al italiano de Giacomo Leopardi).

**GIORGIA ESPOSITO** • is a Ph.D. student in Digital Humanities (curriculum: Foreign languages and literatures, linguistics and onomastics) at the University of Torino, where she graduated in March 2016 with a thesis entitled “Tradurre l’antipoesia. *Los perros románticos* di Roberto Bolaño.”

**E-MAIL** • giorgia.esposito42@gmail.com



SeGNALI

---





*L'italiano della musica nel mondo*

A cura di Ilaria BONOMI e Vittorio COLETTI

Firenze, Accademia della Crusca-goWare, 2015-2016, 203 pp.

(«La lingua italiana nel mondo», Nuova serie e-book)

ISBN 978-88-6797-423-8

---

Virgilio BERNARDONI

Il volume raccoglie gli interventi promossi dall'Accademia della Crusca in occasione della Settimana della lingua italiana nel mondo promossa dal Ministero degli Affari Esteri e dedicata nell'anno 2015 all'italiano in musica. Le argomentazioni degli autori – per lo più linguisti e storici della lingua italiana, più una sparuta rappresentanza di musicologi – s'incardinano sul concetto fondamentale che la musica sia stata – e continui tuttora a essere – uno dei fattori più «duraturi e potenti» (p. 7) di diffusione della lingua italiana all'estero. A farsene portatore è stato (ed è) quello che i curatori Ilaria Bonomi e Vittorio Coletti definiscono «cosmopolitismo [...] costitutivo dell'italiano cantato» (p. 8): ovvero, l'immagine via via sempre più pervasiva, tanto da diventare mitica, della 'cantabilità' quale marca specifica della musica italiana, attestatasi fra Seicento e Ottocento nell'opera e transitata nel corso del Novecento al genere della canzone d'intrattenimento e *popular*. I contributi, editi nella veste di libro elettronico, si distribuiscono in modo pressoché equo fra i suddetti ambiti: nell'ordine, sei insistono su aspetti connessi alla musica d'arte, e al melodramma in particolare, e quattro s'incentrano sulla canzone. Un'appendice offre in aggiunta i due interventi letti in occasione della presentazione della prima edizione nel gennaio 2016, nei quali una musicologa (Elisabetta Fava) e un esperto di musica *pop* (Felice Liperi) ripercorrono da ulteriori punti di

vista gli itinerari internazionali di opera e canzone italiana.

Le due sezioni sostanziali di *L'italiano della musica nel mondo* intersecano studi di orientamento storico-cronologico a studi d'interesse tipologico. Quelli storici centrati sull'opera s'inseriscono nella linea linguistica della librettologia (una prospettiva tutta italiana d'indagine sul teatro d'opera) e si connettono in linea più o meno diretta ai contributi fondativi di Gianfranco Folena (1983) e Daniela Goldin (1985) e a quelli più recenti della stessa Bonomi (1998) e di Luca Serianni (2002). Due saggi mettono al centro le politiche culturali connesse alla diffusione internazionale dell'opera seria. Stefano Saino (*Sulle prime rappresentazioni di melodrammi italiani all'estero*, pp. 31-50) ripercorre per sommi capi la geografia delle irradiazioni dell'opera italiana nelle corti europee, seguendone le trame dettate da relazioni parentali (come quella che fra il 1653 e il 1656 portò un musicista della cerchia dei Medici, Pietro Cesti, a Innsbruck alla corte dell'arciduca Ferdinando Carlo del Tirolo, imparentato con la potente famiglia fiorentina), da specifiche strategie di italianizzazione della cultura (come quelle promosse ripetutamente qua e là per l'Europa: a metà Seicento a Parigi dal cardinale Mazzarino col concorso del veneziano Francesco Cavalli; a metà Settecento a Stoccarda dal duca di Württemberg, Karl Eugen, con la chiamata a corte del napoletano Niccolò Jommelli e a San

Pietroburgo da Caterina II con una politica d'importazione del meglio dell'opera comica del tempo), da intricate relazioni dinastiche (come quelle che fecero di Vienna una capitale che nello spettacolo in musica parlò italiano per un lungo periodo). Vittorio Coletti (*L'italiano in Inghilterra nella musica del tedesco Händel*, pp. 51-66) ripercorre invece l'azione di un autore non italiano di origine, ma italianizzato per formazione, Georg Friedrich Händel, che è stato artefice del ventennio di maggior prestigio dell'opera italiana sulla piazza di Londra, in un ambiente culturale che sempre le fu ostile. La versatilità poliglotta di Händel è attestata: il tedesco era la sua lingua madre, l'italiano era la lingua alla quale si applicava nella professione musicale, il francese la lingua usata nella corrispondenza internazionale, l'inglese divenne la lingua della conversazione quotidiana da quando, nel 1710, trascorse i suoi giorni in territorio britannico. Non è un caso che egli abbia composto ben trentotto opere su libretti italiani (alcune delle quali poi raggiunsero anche la Germania, ma con i recitativi riscritti in tedesco) e tre soltanto in tedesco. È d'obbligo domandarsi che cosa capissero gli spettatori londinesi d'inizio Settecento di opere cantate in italiano. La traduzione inglese dei libretti, provvisti di prammatica di testo a fronte, benché in una versione letterale scarsamente attenta alla qualità retorica e iperletteraria dell'originale, fu la prassi corrente per aggirare il problema. Ancor più vi provvide, però, l'intuito drammaturgico del musicista, il quale propose al pubblico inglese un'opera italiana drasticamente scorciata nelle sue parti più ostiche (i recitativi) e invece integra, e possibilmente incrementata, nelle parti di pura espansione emotiva (le arie), contribuendo a rinforzare la convinzione comune fra i grammatici che l'italiano, lingua della musica per eccellenza, dovesse essere appreso dagli amanti di questa arte per meglio comprendere il significato dei versi cantati nelle arie.

Appare comunque evidente che il doppio registro dell'italianità circolante nel modo attraverso l'opera – quello della lingua e quello

dello stile musicale – si prestasse a un ventaglio di possibilità sul piano concreto dei modi della rappresentazione. La prassi dominante accompagnava con traduzione nella lingua locale spettacoli integralmente cantati in italiano, come accadde appunto nella Londra dei tempi di Händel o, nel medesimo periodo, ad Amburgo (un esempio è dato ancora dal libretto del *Viaggio a Reims* di Giacchino Rossini, qui analizzato da Eduardo Buroni, che fra l'altro rappresenta un caso clamoroso di italiano *nella* musica del mondo, poiché tic e caratteri del concerto internazionale di personaggi che vi agiscono in un contesto satirico – tedeschi, francesi, inglesi, polacchi, russi, oltre che italiani, ovviamente – sono espressi con i rispettivi stili musicali, ma nella lingua 'universale' del melodramma). All'opposto, si affiancavano casi di spettacoli presentati integralmente in traduzione (a ben vedere un corrispettivo dell'uso dei teatri italiani nell'Ottocento di volgere in italiano i libretti delle opere francesi e tedesche), così come forme miste di opere lasciate in italiano nelle arie, per non compromettere gli equilibri musicali, e riversate nella lingua locale nei recitativi, per favorire la comprensione dei nessi drammatici. Si davano anche casi singolari di spettacoli nello stile musicale italiano, però, concepiti fin da principio su testi in altra lingua, come gli oratori in inglese di Händel: un genere che ottenne il massimo gradimento del pubblico, associando alla musicalità italiana tanto apprezzata dagli inglesi un testo immediatamente comprensibile. E si arrivò perfino a spettacoli 'pensati', in italiano da autori operanti all'estero, ma mai rappresentati nella veste linguistica originaria, poiché immediatamente tradotti nella lingua locale, come nel caso delle opere buffe del veneziano Catterino Cavos date a Pietroburgo ai tempi di Caterina.

Il post-scriptum di Elisabetta Fava, che ragiona a vasto raggio sulle peculiarità in senso lato culturali dei contesti produttivi intercettati dall'opera italiana nelle sue peregrinazioni per il mondo nell'arco della propria storia, offre anche le motivazioni di fondo di questi

fenomeni. Da una parte, la sicurezza economica e l'agio organizzativo che gli artefici italiani trovavano all'estero furono fattori non secondari del fatto che fin dalle sue prime attestazioni fuori dai propri confini linguistici l'opera abbia parlato italiano. Le maggiori risorse e i mezzi produttivi più cospicui di cui gli operatori del settore poterono disporre nei teatri di molte corti europee, oltre alla possibilità di operare al riparo dallo stress policentrico del sistema italiano che a ogni nuova stagione li costringeva a peregrinazioni di città in città, hanno favorito il costituirsi di vere e proprie *enclaves* italiane di librettisti, compositori, cantanti e maestranze varie addette agli allestimenti di opere serie e comiche (nell'Ottocento, quella parigina del Théâtre Italien rivaleggiò con l'Opéra e l'Opéra-Comique nel cuore della metropoli europea dello spettacolo), tendenzialmente stabili per periodo prolungati. D'altra parte, la singolarità estetica di uno spettacolo interamente centrato sul canto garantì alle creazioni all'italiana l'attrattiva dell'unicum rispetto alle forme miste, parlate e cantate, del restante spettacolo in musica europeo (salvo le rare eccezioni del maggior teatro di Parigi, dall'epoca dell'Académie Royal de Musique di Luigi XIV all'Opéra, e dei drammi di Wagner).

Sul versante pertinente alla storia della lingua il volume indaga soprattutto due aspetti essenziali del percorso evolutivo dell'italiano dell'opera: la conquista nel medio Settecento di uno spazio specifico per la lingua del genere comico, vista nello specchio di una serie di testi destinati al pubblico germanofono (Paolo D'Achille, *Il Mozart comico prima di Da Ponte: aspetti linguistici dei libretti giocosi*, pp. 67-88) e la progressione dalla lingua del melodramma a quella dell'opera primo novecentesca (Pier Vincenzo Mengaldo, *Evoluzione linguistica dei libretti d'opera dal pieno Ottocento agli inizi del Nove*, pp. 113-121). Ne risulta conclamata la presenza pervasiva di una sorta di koiné librettistica, improntata alle forme proprie della lingua

letteraria, la cui volontà generica di tenere alto il tiro resiste anche alla lievità del comico. In questa sede mi limito a riassumere soltanto in pochissimi tratti la lunghissima durata del tono aulico del melodramma. Nel lessico l'uso del letterario *cor(e)* di gran lunga prevalente su *cuor(e)* nei libretti più antichi (nel libretto della *Finta giardiniera* mozartiana in trentadue occorrenze contro due) è ancora saldamente attestato nei libretti verdiani; le due forme si alternano ancora in un libretto innovativo sul piano linguistico quale il *Mefistofele* di Arrigo Boito e sono pressoché in parità in *Bohème*, *Tosca* e *Madama Butterfly* di Puccini; è all'altezza della *Fanciulla del West*, anno 1910, prima e per lungo tempo unica opera concepita da un team di autori italiani per essere tenuta a battesimo nel massimo teatro statunitense, il Metropolitan di New York, che *cuor(e)* diventa la forma esclusiva. Analogamente, nelle forme verbali dell'imperfetto quelle ridotte in *-ea*, *-ia* prevalgono nettamente e universalmente su quelle complete in *-eva*, *-iva*; perfino in un testo prosodico come *Fanciulla* la forma intera s'impone soltanto per la terza persona, mentre alla prima persona resiste quella ridotta.

Se la conquista di una lingua «perfettamente normale e prosastica» (Mengaldo, p. 121) contraddistingue l'italiano dei libretti soltanto nella fase estrema del melodramma d'inizio Novecento, altre dinamiche hanno investito invece la sfera della canzone. A prescindere dagli aspetti dialettali che hanno caratterizzato la presenza internazionale della canzone napoletana, promossa fin dagli albori della discografia da divi del melodramma come Enrico Caruso, il codice linguistico della canzone italiana maggiormente apprezzato all'estero s'identifica nella discorsività di fondo, resa 'poetica' da alcune forme linguistiche ricorrenti: i costrutti con inversione di termini, le rime bacciate facilissime e le assonanze, l'uso del verbo all'infinito; un caso, quest'ultimo, che è ampiamente attestato nel ritornello di quello che probabilmente rimane il maggior successo italiano nel mondo, *Nel blu dipinto di*

*blu*, con la catena di «volare... cantare... felice di stare...» che possono far sembrare il testo familiare anche ai non italofoeni che orecchiano un po' d'italiano (si veda per questo aspetto Lorenzo Coveri, *Italiano formato export. Dieci canzoni italiane per il mondo*, pp. 122-139, che riferisce del sondaggio condotto dalla Società Dante Alighieri in sessantasei paesi del mondo sulle cinque canzoni più famose in ogni continente).

I contributi sulla canzone aprono comunque una serie di prospettive che oltrepassano gli aspetti puramente linguistici. Gabriella Cartago (*Ius music*, pp. 140-150) affronta il fenomeno dell'inserimento dei cosiddetti 'nuovi italiani' nell'ambito della musica rap – un caso di italianità perentoriamente assunta come strumento identitario da parte di immigrati di seconda generazione – e ne mette in evidenza il carattere di lingua di protesta, in filiazione diretta dall'oralità e incline alla pratica del multilinguismo. Stefano Telve (*Italianismo "reale" e italianismo "percepito" nella musica leggera straniera: uno sguardo su Germania, Croazia e Olanda dagli anni Cinquanta a oggi*, pp. 151-167) offre un saggio di ampliamento di un suo precedente studio sul medesimo tema (cfr. Telve 2012), indagando soprattutto sull'immagine dell'Italia che scaturisce da una serie di canzoni non in lingua italiana, che però descrivono esperienze e aspetti della società italiana. Per cogliere i termini della questione è sufficiente citare qui il testo di una canzone in olandese del 1965, emblematicamente intitolata *Italiano* (le parti non in olandese nell'originale sono in corsivo): «*L'Italiano, l'Italiano* | è la lingua del sole, dei fiori e del cielo blu. | *L'Italiano, l'Italiano* | è la lingua del vino, dell'amore e dell'avventura. | *Non è e difficile parlarlo* [sic] *d'Italiano* [sic] | me lo ha detto anche quel ragazzo | *ma ditte bon giorno* | e io mi sono subito presa una cotta. || Continuava a parlare *d'amore* | e a baciarmi ha fatto *presto* | poi mi ha detto | *Io t'amo, pianissimo* | e io l'ho subito capito | che ci vuole a parlare l'italiano | canterai subito *spiegelbeeld* [titolo di una canzone famosa

negli anni Sessanta] | con il ritornello napoletano» (p. 161).

Alla fine, però, uno dei tratti più rilevanti del prestigio della musica italiana nel mondo e dell'influsso che la sua cultura e la sua lingua hanno esercitato su altre culture è determinato dall'elevato tasso di italianismi musicali esportati. Ilaria Bonomi (*Italianismi musicali nel mondo*, pp. 10-30), riprendendo in sintesi i suoi studi precedenti, enumera una serie di italianismi di diverse categorie semantiche – dalla teoria al linguaggio della partitura e alla tecnica esecutiva, dalle denominazioni di movimento alle forme musicali, dagli strumenti all'opera e al canto – e ne insegue le derivazioni dirette o indirette nel tedesco (che da Beethoven in avanti i compositori germanici di musica strumentale incominciarono ad affiancare con didascalie nella propria lingua, più a scopo di precisazione che di sostituzione) e in francese, inglese, spagnolo, catalano, rumeno, albanese, russo, ungherese, danese, islandese, fino a lingue extraeuropee come il giapponese e il cinese. La diffusione della terminologia musicale italiana disseminata grazie all'opera, infatti, si è rivelata più durevole dell'opera stessa e ha costituito la comunità linguistica più ampia di tutta la storia musicale. La sua suggestione se è ormai estesa oltre la specializzazione musicale ed è arrivata, per esempio, nel campo del *marketing* automobilistico, dove s'incontrano un *Concerto* fra i prodotti Honda e un florilegio di *Largo, Note e Fuga* fra quelli Nissan. Franco Fabbri (*Sui nomi delle musiche*, pp. 168-177) rileva che l'italiano ha fatto breccia perfino nella nomenclatura dei 'generi' della *popular music*, dagli anni '60 in avanti appannaggio pressoché esclusivo dell'anglo-americano, laboratorio della massiccia riformulazione semantica dalla quale è scaturito il lessico non tradotto delle tendenze musicali più recenti, al quale sono conformati i motori di ricerca della rete Internet. Anche in questo campo l'italiano oggi s'inserisce con termini nati in Italia per designare tendenze 'viventi' della nostra musica come *canzone*, appunto, *canzone d'autore*, *cantautore*.



Con ciò non si è dato conto di tutti i fili che s'intrecciano in questa pubblicazione efficace e utile, che tocca un vasto insieme di questioni, proponendone interessanti punti di osservazione. Per esempio, si è fatto soltanto qualche cenno indiretto alle implicazioni sociologiche sottese alle varie analisi, per cui nei diversi interventi appare chiaramente come nel tempo opera e canzone abbiano diffuso livelli linguistici rivolti ora alle *élites*

aristocratiche, ora alla borghesia, ora agli utenti dei media, ora alla sfera giovanile del *pop* e del *rock*. Comunque sia, proprio per l'ampiezza dei temi, affrontati con un'adeguata varietà di approcci, nell'insieme il volume si raccomanda non solo allo storico della lingua italiana, ma anche al musicologo, che ne può trarre illuminanti sollecitazioni argomentative e metodologiche.

### BIBLIOGRAFIA

- Bonomi, I. (1998), *Il docile idioma*, Roma, Bulzoni.  
 Folena, G. (1983), *L'italiano in Europa. Esperienze linguistiche del Settecento*, Torino, Einaudi.  
 Goldin, D. (1985), *La vera fenice. Librettisti e libretti tra Sette e Ottocento*, Torino, Einaudi.  
 Serianni, L. (2002), *Libretti verdiani e libretti pucciniani. Due modelli linguistici a confronto*, in L. Serianni, *Viaggiatori, musicisti, poeti. Saggi di storia della lingua italiana*, Milano, Garzanti, pp. 113-161.  
 Telve, S. (2012), *That's amore! La lingua italiana nella musica leggera straniera*, Bologna, Il Mulino.

**VIRGILIO BERNARDONI** • Virgilio Bernardoni is full professor in Musicology and Music History at University of Bergamo. His interests are focused in Italian opera, especially in 18th- and 19th-Century. His books and edited volumes concern items of *fin de siècle*. He is director of «Historiae Musicae Cultores» and «Centro studi Giacomo Puccini» series (Florence, Olschki) and president of the scientific board of Edizione Nazionale delle Opere di Giacomo Puccini.

**E-MAIL** • [virgilio.bernardoni@unibg.it](mailto:virgilio.bernardoni@unibg.it)



# LA FORMAZIONE INIZIALE DEGLI INSEGNANTI

Visita di ricerca comparativa di esperti giapponesi  
all'Università di Torino

---

*Akira KAWAMURA, Silvana MOSCA*

Nei giorni dal 21 al 25 novembre 2016, due professori giapponesi – Akira Kawamura, dell'Università del Kansai e Takao Kato, dell'Università Nanzan – hanno visitato diversi Dipartimenti dell'Università di Torino e alcune scuole, per una ricerca comparata sulla formazione iniziale degli insegnanti della scuola primaria e secondaria di primo grado nei due paesi. Si sono svolti incontri con docenti universitari, insegnanti esperti - tutor di tirocinio, studenti in formazione ed ex studenti attualmente neoinsegnanti; sono state condotte e documentate interviste e discussioni, visitati biblioteche e laboratori, incontrati i responsabili della formazione a vari livelli. Infine sono state concordate le fasi future della ricerca che si concluderà nel 2017.

La ricerca è finanziata dal governo giapponese (JSPS KAKENHI Grant Number JP25381282).

## **Il Progetto**

La riforma della formazione degli insegnanti incentrata sul tirocinio nell'epoca della globalizzazione. Uno studio comparativo Giappone – Italia.

Il presente studio si pone l'obiettivo di confrontare i sistemi di formazione incentrati sul tirocinio degli insegnanti di scuola primaria e secondaria di primo grado in Italia e in Giappone, dove attualmente è in atto una riforma di tali sistemi per sostenere la sfida

della globalizzazione. Il Giappone ambisce a elevare lo status dell'insegnante, pertanto si analizzerà il programma del tirocinio previsto all'interno del sistema di formazione degli insegnanti e si osserveranno i nuovi professionisti educati attraverso il percorso odierno per verificarne il profilo. In concreto, ci si propone di raggiungere le finalità del progetto attraverso: 1. interviste mirate, svolte in entrambi i paesi, che permetteranno di analizzare lo stato attuale dei programmi di formazione e i loro eventuali difetti o punti problematici; 2. analisi delle competenze in via di acquisizione degli insegnanti in formazione (*student teacher*) attraverso il lavoro sul campo, che permetterà di fare luce sulla struttura dei tirocini; 3. questionari rivolti a insegnanti in formazione freschi di tirocinio, per un inquadramento generale delle competenze acquisite. In Giappone e in molti altri paesi occidentali il tirocinio è incentrato sulla pratica, mentre in Italia si riscontra un equo bilanciamento tra pratica e teoria, mediante la partecipazione al tirocinio diretto e la stesura di un report finale. L'Italia appare dunque in controtendenza rispetto ad altre nazioni, ma è proprio questa caratteristica a indurci a riconsiderare il programma di formazione giapponese attualmente in vigore. I risultati delle interviste, dell'analisi del lavoro sul campo e dei questionari potranno offrire nuove e preziose conoscenze in termini di apporti politici e teorici come segue: 1. contribuiranno al dibattito mondiale sui

programmi di tirocinio nelle diverse società globali; 2. permetteranno di chiarire le tendenze della riforma e di riesaminare la politica di eccessiva enfasi dei risultati pratici in atto in Giappone; 3. suggeriranno nuove linee di condotta per il tirocinio nel sistema di formazione degli insegnanti a livello universitario.

### **Le fasi precedenti**

Il team giapponese ha compiuto quattro visite di studio a Torino dal 2012 al 2016, il team italiano ha compiuto una visita di studio in Giappone nel distretto di Osaka nel 2014. A seguito dello studio reciproco dei due modelli e delle osservazioni sul campo, è stato costruito un questionario per un'indagine quantitativa e di percezione dell'efficacia del tirocinio. A Torino il Questionario è stato somministrato a n° 122 studenti del TFA 2015 e a n° 187 studenti del quinto e ultimo anno 2016 del Corso di laurea magistrale a ciclo unico in Scienze della Formazione primaria.

### **La visita**

Si sono svolte interviste in profondità a testimoni privilegiati con un focus incentrato sulle attività e le metodologie che consentono di unire la teoria alla pratica, in particolare i Progetti di ricerca didattica e i Laboratori di formazione metodologica. La panoramica ha spaziato da contenuti di *Matematica argomentativa* alle attività di *Scienze sperimentali realizzate in laboratorio e nell'ambiente*, alla *Didattica delle lingue straniere con le ICT*, ai percorsi formativi della 5<sup>a</sup> annualità del nuovo corso di Laurea in Scienze della formazione primaria illustrati dai Tutor coordinatori e dai responsabili dell'USCOT (Ufficio Coordinamento Tirocini): *La narrazione di sé*, *Let's speak English*, *Didattica multimediale Classi 2.0*, *Storia, Geografia e cittadinanza*, *Dallo spazio del corpo allo*

*spazio geometrico*, *Inclusività*, *Costruzione dell'identità culturale*.

La visita, iniziata con l'accoglienza della professoressa Carla Marengo del Dipartimento di lingue e letterature straniere e culture moderne –nella saletta Graf della imponente Biblioteca del Rettorato, si è conclusa con la visita storico-scientifica al Dipartimento di Chimica e all'annesso museo, dove studiò Primo Levi e dove gli studenti e i loro docenti e formatori hanno mostrato una performance concreta con provette e reattivi per rilevare con semplici e sicure tecniche, idonee anche ai bambini della primaria, la presenza degli amidi nei comuni alimenti consigliati dai nutrizionisti. Inoltre i ricercatori hanno potuto intervistare due dirigenti scolastici e alcuni ex tirocinanti TFA attualmente insegnanti in scuole sperimentali della Rete di scuole AVIMES (Autovalutazione di Istituto per il Miglioramento dell'efficacia della scuola).

La visita ha anche consentito la presentazione degli ospiti ricercatori ad alcune autorità, nella cornice dell'inaugurazione del Festival dell'Educazione iniziato a Torino nella medesima settimana: tra gli altri, il Rettore dell'Università, il Direttore generale del Dipartimento istruzione del MIUR, l'Assessora comunale all'istruzione, il Direttore della Fondazione Agnelli.

### **Prospettive e prime riflessioni**

La delegazione ha definito con il team torinese gli sviluppi futuri della ricerca: applicazione, anche in Giappone, nel mese di dicembre, del medesimo questionario rivolto agli studenti tirocinanti dell'anno finale di corso; elaborazione dei dati, sistemazione della ricca documentazione raccolta, analisi e interpretazione dei dati qualitativi e quantitativi, stesura del Rapporto generale.

Come è noto, il sistema di formazione iniziale dei docenti in Italia –che persegue l'integrazione fra teoria e pratica- ha una

struttura diversa fra Scienze della formazione primaria, che ha concluso il suo primo quinquennio dopo la riforma del 2010, e il TFA, che costituisce formula transitoria per associare il Tirocinio formativo ai corsi di laurea disciplinari e che ancora non pare prospettare il superamento con l'inattuata delega conseguente alla legge 107/2015.

Il confronto con il sistema giapponese è stato, e certamente sarà, di particolare interesse per le nostre verifiche e per il reciproco arricchimento, come hanno affermato con soddisfazione anche i professori ospiti.

### Testo originale<sup>1</sup>

グローバル化下の教育実習を中核とした教員養成制度改革の日伊比較研究

本研究の目的は、グローバル化のもとで、現在、小学校教員養成改革を行っているわが国とイタリアの教育実習プログラムを中核とした教員養成システムを比較し、教職の高度化を目指すわが国の教員養成システムにおける教育実習プログラムのあり方について検討し、そのプログラムでどのような教員が養成されようとしているのかを考察することである。具体的には、①両国の教員養成制度と教育実習プログラムの改革の現状と課題についてヒアリング調査などを行うこと、②教育実習プログラムの実際を明らかにするために、フィールドワークを通して実習生が習得していくものについて詳細に明らかにすること、③教育実習を経験した学生に対する質問紙調査によって彼らが習得したもののより全体的な傾向を把握することを通して、前述の目的を達成する。

教育実習プログラムにおいて、わが国や多くの欧米諸国は実習経験に重点を置く実践的指導力重視であるのに対し、イタリアは、実習と実習レポートによる実践と理論の往還を重視しており、改革の方向性に差異がある。したがって、グローバル化のもとでの教育実習プログラム改革の方向性が異なるイタリアの教員養成システムを究明することにより、わが国のプログラムを相対化できる。

本研究は、ヒアリング調査、フィールドワーク、質問紙調査により、以下の点で理論的、政策的に貢献する有益な知見を提供することができる。①グローバル社会における教育実習プログラムに関する世界的議論に貢献する。②わが国の教職大学院における実践的指導力重視の政策に対する批判的検討を行い、現在の教員養成政策改革の方向性を明らかにできる。③大学院レベルの教員養成システムにおける教育実習のあり方を、政策的に提案することが可能になる。

AKIRA KAWAMURA • University of International Studies– Osaka. Principal Investigator of the project

E-MAIL • akira@kuins.ac.jp

<sup>1</sup> Traduzione dal giapponese di Anna Specchio.

**SILVANA MOSCA** •Former senior executive at MIUR and AVIMES Network – Project coordinator for the Italian team based in Turin (with the participation of Elisa Corino, Miranda Mosca, Cristina Onesti, Massimo Perotti, Elena Scalenghe)

**E-MAIL** •silvanamosca@virgilio.it



XL CONVEGNO INTERNAZIONALE  
AISTUGIA  
(Associazione Italiana STUdi GIApponesi)  
Torino, 22-24 settembre 2016

---

*Daniela MORO*

I giorni 22, 23 e 24 settembre 2016, presso il palazzo del Rettorato dell'Università degli Studi di Torino, si è tenuto il 40esimo convegno dedicato agli studi sul Giappone dell'Associazione Italiana per gli Studi Giapponesi (AISTUGIA). L'evento si è svolto con il contributo della Japan Foundation e il patrocinio dell'Ambasciata del Giappone e dell'Istituto Giapponese di Cultura di Roma, nell'ambito delle celebrazioni ufficiali del 150° anniversario delle relazioni Italia-Giappone.

Il convegno è stato organizzato grazie allo sforzo congiunto dei dipartimenti dell'Università di Torino che si occupano di studi giapponesi: il dipartimento di Lingue e Letterature Straniere e Culture Moderne e quello di Studi Umanistici, rappresentati rispettivamente da Gianluca Coci, principale responsabile e coordinatore dell'evento, e Matteo Cestari, i quali sono stati coadiuvati da Anna Specchio e Daniela Moro, con la collaborazione di Kobayashi Takae, Watanabe Asako e un team di studenti di entrambi i dipartimenti.

Il convegno si è aperto con i saluti inaugurali della prorettrice dell'Università di Torino Elisabetta Barberis, del vicedirettore alla didattica del dipartimento di Lingue e Letterature Straniere e Culture Moderne Matteo Milani, del segretario generale dell'AISTUGIA Matilde Mastrangelo, dell'ambasciatore del Giappone in Italia Umemoto Kazuyoshi e della direttrice dell'Istituto Giapponese di Cultura di Roma Takasu Naomi.

Dopo i saluti inaugurali, il celebre scrittore Ikezawa Natsukiha tenuto la conferenza di apertura in giapponese dal titolo: "La letteratura giapponese e la natura dei miei connazionali – Riflessioni durante la compilazione di un'antologia della letteratura". L'intervento è partito dal personale dubbio dello scrittore sul senso di vivere in un Paese continuamente tormentato dai fenomeni naturali, per poi entrare nel vivo della storia della letteratura giapponese. Partendo da un'analisi della peculiare forma del *bungakuzenshū* (antologia della letteratura), Ikezawa ha illustrato le modalità utilizzate nella scelta di opere, traduttori e autori per l'antologia da lui curata e ha effettuato una panoramica delle problematiche legate alla rilettura delle opere antiche in chiave moderna. Dal *Kojiki* agli scrittori emergenti contemporanei, passando per la visione della letteratura premoderna di Mishima Yukio, Ikezawa ha dimostrato con alcuni esempi concreti il lavoro di curatore di un'antologia della letteratura. Infine ha elencato alcuni aspetti comuni delle opere letterarie giapponesi del passato, paragonandole ad esempio alla letteratura della Grecia antica. Gli spunti di riflessione hanno preso in esame in particolare l'amore carnale e lo stretto legame tra politica e potere spirituale, puntando l'attenzione sul fatto che spesso le opere del periodo premoderno sono state interpretate in maniera errata nel corso della storia.

Sakai Naoki, ordinario di storia intellettuale e letterature comparate alla

Cornell University, ha tenuto una relazione sul concetto di *area studies*, studi umanistici legati a una specifica area geografica e culturale. Il suo intervento si intitolava: “The Ends of Area Studies – On the Question of Theory and Anthropological Difference”. Il professor Sakai ha esordito riconoscendo l’origine degli studi di area nella distinzione fra *humanitas* e *anthropos*, e, tracciando i punti salienti della storia del pensiero degli ultimi due secoli, ha spiegato come il mondo moderno internazionale sia stato caratterizzato dalla consacrazione dello stato-nazione da una parte e dall’ordine coloniale e internazionale del mondo dall’altra. Ha poi illustrato come con la globalizzazione, assieme all’indebolimento degli stati-nazione, si stia verificando il collasso dell’ordine coloniale e imperiale, con conseguente crollo della struttura della differenza antropologica tra *humanitas* e *anthropos*. Partendo da queste premesse, Sakai ha posto una serie di interrogativi, tra cui il ruolo costitutivo che hanno giocato gli studi di area nello sviluppo delle discipline umanistiche e sulle ragioni per cui gli studi di area continuano ad auto-legittimarsi, chiedendosi infine per quanto ancora e se continueranno a esistere in futuro.

A seguire si è tenuto un panel di giuristi moderato da Michele Graziadei dell’Università di Torino, che da anni si interessa di diritto comparato. Si è trattato di un confronto fra vari esperti, italiani e giapponesi, sul tema degli stereotipi relativi alla cultura giapponese nel contesto del diritto internazionale. Ognuno degli studiosi ha discusso la validità euristica dell’idea di unicità del Giappone nei vari aspetti del diritto: Giorgio Colombo ha discusso dei meccanismi di risoluzione delle controversie, Marco Giorgi ha affrontato le regole societarie, Andrea Ortolani ha analizzato le particolarità giapponesi nei rapporti contrattuali di durata; Michela Riminucci ha parlato dell’influenza del sistema giapponese dell’impiego sull’evoluzione del diritto del lavoro; Sakuramoto Masaki ha presentato criticamente le continuità e le

differenze nell’istituto del fallimento tra periodo Edo (1600-1868) e Giappone contemporaneo.

Nel pomeriggio e nei giorni successivi si sono svolte per lo più presentazioni singole, fatta eccezione per un secondo panel, inerente alla critica letteraria di periodo Meiji (1868-1912) e Taishō (1912-1926). Moderati da Ikuko Sagiyama, docente di Lingua e Letteratura Giapponese presso l’Università di Firenze, gli interventi di tre docenti di Ca’ Foscari, Bonaventura Ruperti, Pierantonio Zanotti e Luisa Bienati, hanno rispettivamente affrontato un’analisi dei saggi teorici relativi alla letteratura a opera di Kitamura Tōkoku e Izumi Kyōka, il contributo di Takamura Kōtarō al dibattito sulle nuove tendenze artistiche e letterarie del tempo e l’influenza del cinema sull’estetica e la letteratura di Tanikazi Jun’ichirō.

Gli interventi singoli hanno affrontato una grande varietà di temi, principalmente in ambito storico-politico e linguistico-culturale. Per quanto riguarda gli studi storici, si sono affrontati argomenti inerenti ai samurai, alle ambascierie tra Giappone e Italia (la prima delle quali è stata celebrata quest’anno, a sancire i 150 anni delle prime relazioni ufficiali tra i due paesi), alla propaganda durante la seconda guerra mondiale, fino alle ultime tendenze politiche ed economiche del Giappone sotto il governo Abe. Sugli aspetti culturali, si è potuto riscontrare una grande varietà tematica, a toccare ambiti come filosofie e religioni, cinema, letteratura (Tawada Yōko, Orihara Ichi, Hayashi Mariko), moda, *pachinko* e ultime tendenze del *cross dressing*. Anche la linguistica e la glottodidattica hanno ricevuto un’attenzione particolare, soprattutto nel campo della traduzione di termini specifici, nell’apprendimento tramite audiovisivi e nell’uso di strumenti digitali per trasformare la voce in testo.

In conclusione dei lavori, l’assemblea dei soci ha optato per Venezia come città che ospiterà il prossimo convegno, a settembre 2017. Il convegno è terminato simbolicamente



con il passaggio delle consegne a Bonaventura Ruperti, come rappresentante dell'Università Ca' Foscari di Venezia. Ulteriori informazioni sono reperibili sul sito: [www.aistugia.it](http://www.aistugia.it)

**DANIELA MORO** • Daniela Moro is lecturer of Japanese Language and Culture at the Department of Foreign languages and literatures and Modern cultures of the University of Turin.

**E-MAIL** • [daniela.moro@unito.it](mailto:daniela.moro@unito.it)





## VI CONGRESSO ITALIANO DI SLAVISTICA Torino, 28-30 settembre 2016:

*Roberta SALA*

Uno studio interdisciplinare dell'identità slava a contatto con l'Europa

Lo scorso settembre l'Università di Torino ha ospitato il Sesto Congresso Italiano di Slavistica, un evento di grande rilevanza accademica organizzato dall'Associazione Italiana degli Slavisti (AIS). Alle tre giornate di studio, introdotte dal saluto del presidente dell'AIS Giorgio Ziffer, hanno preso parte importanti figure della slavistica italiana legate all'area russa (cui la maggior parte dei contributi fa riferimento), serbo-croata, slovena e polacca. L'interdisciplinarietà, a tal proposito, costituisce certamente uno degli aspetti più significativi del convegno, volto alla ridefinizione in chiave contemporanea di un volto originale della cultura esteuropica. Nell'accostamento di discipline quali la linguistica, la letteratura e la filologia, infatti, è emersa l'importanza di uno studio eterogeneo e trasversale al fine di stabilire i tratti principali del carattere slavo, nonché le varie tappe della sua formazione.

A questo scopo, il filo conduttore dei vari contributi è legato alla definizione di un punto di vista esterno, volto a individuare le peculiarità del mondo slavo in relazione agli scambi linguistici e culturali intrattenuti con il resto d'Europa dal periodo antico fino ai giorni nostri. L'accento è stato posto, oltre che sulle contaminazioni derivanti da tale scambio, sulla percezione dei modelli "orientali" nella cultura europea occidentale e viceversa, al fine di

tracciare una linea di continuità coerente e multiforme nello sviluppo dell'identità slava.

Anche l'intreccio tra le varie discipline coinvolte, suddivise in sezioni tematiche nel corso delle varie giornate (storico-letteraria, storico-culturale, e linguistica), ha messo in luce l'importanza di un'analisi specifica e particolareggiata al fine di comporre un quadro d'insieme completo. Grazie al carattere diacronico delle discussioni, inoltre, si è evidenziata la volontà di proporre uno sguardo contemporaneo sulle dinamiche in corso nell'area slava, pur tenendo conto della profondità storica a esse sottesa.

A questo proposito, risulta significativo il dibattito sorto in seguito alla sessione mattutina di venerdì 30 dicembre in aula Principi d'Acaja (Sezione storico-letteraria e storico-culturale). Gli interventi di Barbara Ronchetti (Roma, La Sapienza) e Alessandro Achilli (Università degli Studi di Milano), infatti, hanno avviato una discussione molto produttiva a proposito dello studio dei fenomeni letterari nella Russia dei nostri giorni. L'accento è stato posto sulla definizione del postmodernismo (etichettato anche come "neo-modernismo" o "tardo modernismo") in base all'interscambio con i movimenti modernisti e a uno studio interculturale in grado di fornire un punto di vista dislocato. Ci si è ampiamente interrogati, inoltre, sul significato della letteratura contemporanea nei territori dell'ex Unione Sovietica, in relazione allo spirito di

clandestinità e di protesta tipico degli ultimi decenni del secolo scorso.

D'altra parte, i numerosi contributi in ambito letterario hanno rivelato un forte interesse per l'indagine dei fenomeni socio-culturali in atto nel tardo Ottocento e nei primi decenni del secolo scorso, al fine di fissare basi salde per la comprensione delle dinamiche odierne. Se, da un lato, ci si è focalizzati sulle radici dei movimenti avanguardisti in relazione alle correnti letterarie sorte contemporaneamente in Europa, dall'altro si è resa manifesta la volontà di individuare nell'opera di singole personalità artistiche i tratti tipici di un'epoca e le modalità di contaminazione esterna. Così, attraverso l'analisi delle influenze europee nella produzione di diverse figure appartenenti al mondo slavo, da Merežkovskij all'Achmatova, da Andreev a Grossman, dalla Hłakowiczówna a Slowacki, da Nenadović a Kundera, si è tentato di delineare i tratti, seppur discontinui, del quadro culturale su cui sorgono le basi della contemporaneità.

Risulta significativa, da questo punto di vista, l'introduzione nel dibattito di discipline mirate a fornire una prospettiva ampia e completa sui fenomeni artistici considerati. Negli interventi di Daniela Steila (Università di Torino) e Nadia Caprioglio (Università di Torino), per esempio, appare evidente la necessità di una visione filosofica per la definizione degli scambi interculturali tra l'area slava e quella europea. Allo stesso tempo, l'analisi della collaborazione

cinematografica italo-sovietica a opera di Claudia Olivieri (Università di Catania) mostra la volontà di estendere il dibattito culturale ad ambiti di indagine particolarmente rilevanti in ottica contemporanea. Il confronto tra la sfera slava e quella specificatamente italiana, d'altra parte, si è rivelato piuttosto produttivo nelle varie sessioni del convegno, sia da un punto di vista letterario (si pensi, per esempio, al contributo di Marco Sabbatini, dell'Università di Macerata, volto a individuare il legame dell'opera di Anna Achmatova con la cultura italiana) che prettamente culturologico (facciamo riferimento all'intervento di Alessandra Visinoni, dell'Università di Bergamo, finalizzato all'analisi dei rapporti della città di Bergamo con la Russia).

Anche nel campo della linguistica, il dibattito ha mostrato un forte interesse verso il concetto di equivalenza nella costituzione del corpus parallelo italo-russo NKRJa, illustrato da Anna Paola Bonola (Milano, Università Cattolica del Sacro Cuore), Francesca Biagini (Università di Bologna) e Valentina Nosedà (Milano, Università Cattolica del Sacro Cuore). Più in generale, i vari interventi relativi a quest'area disciplinare hanno messo in luce la rilevanza dei database linguistici per il perfezionamento di discipline quali la traduzione e la didattica delle lingue straniere. Allo stesso tempo, la forte presenza di contributi di carattere filologico ha ribadito la rilevanza di un terreno comune a tutta l'area slava, rappresentato, innanzitutto, delle radici linguistiche.

*RiCOGNIZIONI*  
Rivista di lingue, letterature e culture moderne

<http://www.ojs.unito.it/index.php/ricognizioni/index>  
[ricognizioni.lingue@unito.it](mailto:ricognizioni.lingue@unito.it)

© 2014  
Dipartimento di Lingue e Letterature straniere e Culture Moderne  
Università di Torino  
<http://www.dipartimentolingue.unito.it/>