

LESSICO MULTILINGUE DEI BENI CULTURALI

I Corpora LBC

Informatica Umanistica per il Lessico dei Beni Culturali

a cura di

**RICCARDO BILLERO, ANNICK FARINA,
MARÍA CARLOTA NICOLÁS MARTÍNEZ**



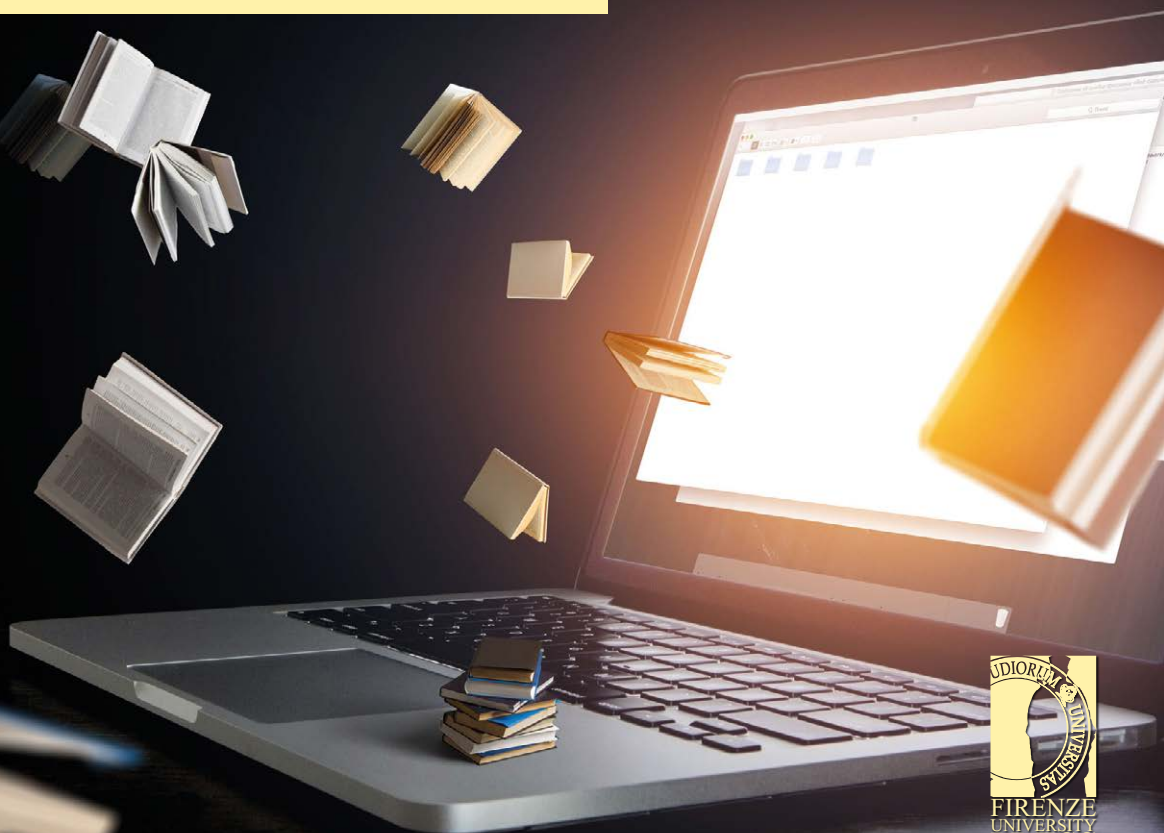
LESSICO MULTILINGUE DEI BENI CULTURALI

I Corpora LBC

Informatica Umanistica per il Lessico dei Beni Culturali

a cura di

**RICCARDO BILLERO, ANNICK FARINA,
MARÍA CARLOTA NICOLÁS MARTÍNEZ**



STRUMENTI PER LA DIDATTICA E LA RICERCA

ISSN 2704-6249 (PRINT) - ISSN 2704-5870 (ONLINE)

– 218 –

LESSICO MULTILINGUE DEI BENI CULTURALI

Director

Annick Farina, University of Florence, Italy

Christina Samson, University of Florence, Italy

Scientific Board

Sabrina Ballestracci, University of Florence, Italy

Marco Biffi, University of Florence, Italy

Elena Carpi, University of Pisa, Italy

Dave Coniam, The Education University of Hong Kong, China

Christina Dechamps, Nova University of Lisbon, Portugal

Isabella Gagliardi, University of Florence, Italy

Marcello Garzaniti, University of Florence, Italy

Paul Geyer, University of Bonn, Italy

Donata Levi, University of Udine, Italy

Valentina Pedone, University of Florence, Italy

Federica Rossi, Max Planck Institute - Florence Institute for Art History, Italy

Geoffrey Williams, Bretagne Sud University, France

Published book

Raus R., Cappelli G., Flinz C. (édité par), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel. Vol. II*, 2017

Zotti V., Pano Alamán A. (a cura di), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*, 2017

Flinz C., Carpi E., Farina A., (édité par), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel. Vol. I*, 2018

Farina A., Funari F. (a cura di), *Il passato nel presente: la lingua dei beni culturali*, 2020

Billero R., Farina A., Nicolás Martínez M.C. (a cura di), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, 2020

I Corpora LBC

Informatica Umanistica per il Lessico dei Beni Culturali

a cura di

Riccardo Billero, Annick Farina,
María Carlota Nicolás Martínez

FIRENZE UNIVERSITY PRESS

2020

I Corpora LBC : informatica Umanistica per il Lessico dei Beni Culturali / a cura di Riccardo Billero, Annick Farina, María Carlota Nicolás Martínez. – Firenze : Firenze University Press, 2020.
(Strumenti per la didattica e la ricerca ; 218)

<https://www.fupress.com/isbn/9788855182539>

ISSN 2704-6249 (print)

ISSN 2704-5870 (online)

ISBN 978-88-5518-252-2 (print)

ISBN 978-88-5518-253-9 (PDF)

ISBN 978-88-5518-255-3 (XML)

DOI 10.36253/978-88-5518-253-9

Graphic design: Alberto Pizarro Fernández, Lettera Meccanica SRLs

Front cover: Sergey Nivens © 123RF.COM

FUP Best Practice in Scholarly Publishing (DOI https://doi.org/10.36253/fup_best_practice)

All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Boards of the series. The works published are evaluated and approved by the Editorial Board of the publishing house, and must be compliant with the Peer review policy, the Open Access, Copyright and Licensing policy and the Publication Ethics and Complaint policy.

Firenze University Press Editorial Board

M. Garzaniti (Editor-in-Chief), M.E. Alberti, F. Arrigoni, M. Boddi, R. Casalbuoni, F. Ciampi, A. Dolfi, R. Ferrise, P. Guarnieri, A. Lambertini, R. Lanfredini, P. Lo Nostro, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, A. Novelli, A. Orlandi, A. Perulli, G. Pratesi, O. Roselli.

📖 The online digital edition is published in Open Access on www.fupress.com.

Content license: the present work is released under Creative Commons Attribution 4.0 International license (CC BY 4.0: <http://creativecommons.org/licenses/by/4.0/legalcode>). This license allows you to share any part of the work by any means and format, modify it for any purpose, including commercial, as long as appropriate credit is given to the author, any changes made to the work are indicated and a URL link is provided to the license.

Metadata license: all the metadata are released under the Public Domain Dedication license (CC0 1.0 Universal: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>).

© 2020 Author(s)

Published by Firenze University Press

Firenze University Press

Università degli Studi di Firenze

via Cittadella, 7, 50144 Firenze, Italy

www.fupress.com

This book is printed on acid-free paper

Printed in Italy

Indice

La banca dati LBC <i>Annick Farina, María Carlota Nicolás Martínez</i>	7
Implementazione di software per la gestione dei corpora LBC <i>Riccardo Billero</i>	19
Il <i>Corpus LBC-español</i> : basi, sviluppo e applicazioni <i>Elena Carpi, Ana Pano Aláman</i>	33
Un itinerario letterario nel corpus LBC in lingua inglese <i>Ilaria Natali</i>	43
Il corpus LBC tedesco: costruzione e possibili applicazioni <i>Sabrina Ballestracci, Claudia Buffagni, Carolina Flinz</i>	55
Il corpus LBC francese <i>Annick Farina, Lorella Sini</i>	77
Il corpus LBC russo <i>Valentina Rossi, Marcello Garzaniti, Natalia Zhukova</i>	101

La banca dati LBC

Annick Farina, María Carlota Nicolás Martínez

Introduzione

Il progetto specifico della creazione della Banca Dati LBC è stato eseguito dall'Unità di ricerca *Lessico multilingue dei Beni Culturali* che fa capo all'Università di Firenze. Uno dei principali scopi dell'Unità di Ricerca è sviluppare risorse e strumenti informatici di supporto per i membri del gruppo di ricerca, con l'intento di renderli fruibili anche ad altri ricercatori mediante la pubblicazione Open Access delle risorse. La Banca Dati permette di effettuare ricerche all'interno di corpora testuali nelle diverse lingue del progetto (per ora sei: francese, inglese, italiano, russo, spagnolo, tedesco) contenenti un numero sufficiente di testi per poter mettere a disposizione dati esaurienti. La struttura delle banche dati permette di consultare i corpora sulla base di diversi criteri: autore, periodo cronologico, genere, formato ecc., fornendo informazioni lessicali specifiche necessarie nello svolgimento dei propri lavori di lessicografia e di traduzione. Ora che abbiamo raggiunta la fine della prima fase di questo progetto, l'Unità di ricerca dispone di una piattaforma che consente la consultazione dei corpora e offre informazioni su di essi¹.

1. Cornice del progetto

Inizialmente il progetto, denominato *Informatica Umanistica e multilinguismo: sviluppo e divulgazione dei Beni Culturali toscani*, si poneva il seguente obiettivo:

¹ Consultabile su <<http://corpora.lessicobeniculturali.net/>>.

promuovere la realizzazione di studi, di ricerche e di altre attività incentrate sul lessico della lingua italiana nell'ambito dei Beni culturali della città di Firenze e della Toscana in relazione a varie lingue (cinese, francese, inglese, italiano, portoghese, russo, spagnolo, tedesco). Tale ricerca sarebbe stata condotta sulla base di corpora originati da testi letterari classici, filosofici, opere scientifiche e tecniche, libri di cucina, manuali tecnici, ecc. In tale direzione, l'Unità di ricerca intendeva, mettendo a frutto le conoscenze nell'ambito sia tecnologico sia umanistico, disporre di uno spazio digitale con vari strumenti utili a diffondere a livello internazionale la conoscenza del patrimonio artistico e culturale toscano (Farina 2016).

Le tipologie di utenti mirati sono:

- Linguisti, letterati, ricercatori in scienze umane e sociali, il cui lavoro necessita di ricerche all'interno di vari corpora i cui dati sono strutturati dal sistema in modo tale da ottenere informazioni sul lessico per autore, periodo cronologico, genere ecc.;
- Traduttori che hanno necessità di consultare banche dati testuali comparabili o parallele nelle lingue di competenza² e risorse lessicali specifiche basate su testi della stessa specialità/tipologia di quelli che devono tradurre;
- Specialisti del settore turistico, quali guide, agenzie turistiche, ristoratori ecc., che intendono presentare il patrimonio culturale a persone provenienti da varie nazioni, attraverso ad es. il legame tra una citazione e un luogo, permettendo di arricchire la scoperta turistica con informazioni che non sono immediatamente percepibili visitando un luogo;
- Turisti interessati ad approfondire la propria conoscenza del territorio e della cultura legata ad esso, che potrebbero consultare informazioni su vari luoghi e autori, su artisti che ne hanno parlato o che ci sono vissuti, realizzando la propria visita 'accompagnati' dai più noti turisti dei secoli passati, es. Unamuno, Stendhal, Lawrence ecc.

Frutto del progetto dovevano essere strumenti digitali necessari alla gestione delle fonti primarie.

Dal 2016 l'Unità di Ricerca si è progressivamente concentrata sulla creazione di una banca dati formata da corpora in diverse lingue con testi storico-artistici e/o letterari. Questo progetto è stato possibile anche grazie alla collaborazione con un ricercatore formato in Digital Humanities con il quale si è realizzato uno scambio di *know how* tra due discipline diverse (la linguistica e l'informatica), creando nel corso della collaborazione gli strumenti necessari a linguisti, letterati e ricercatori, progettando e garantendo al tempo stesso la possibile espansione del progetto verso le altre tipologie di utenti previste.

Il progetto nella sua prima fase ha potuto disporre di alcune importanti collaborazioni con università e/o enti su base internazionale, tra le quali il Laboratoi-

² A proposito della futura implementazione di corpora paralleli per il nostro progetto, cfr. Zotti 2017.

re d'Informatique dell'Università di Tours (Francia), il Laboratorio UMR 5206 "Triangle" del francese Centre national de la recherche scientifique, il gruppo LiCoRN della Université de Bretagne Sud, CHIMERA Romance Corpora and Linguistics Studies (Universidad Autónoma de Madrid).

2. La banca dati LBC

La prima fase del progetto è consistita nella creazione della banca dati delle diverse lingue con testi storico-artistici e/o letterari su Firenze e la Toscana.

Gli obiettivi programmati in questa fase di lavoro sui corpora sono stati i seguenti³:

- creare corpora testuali nelle varie lingue del progetto con la massima omogeneità possibile per permettere la loro comparabilità;
- predisporre gli strumenti necessari per la strutturazione dei dati riguardanti la gestione bibliografica delle fonti prevedendo il collegamento fra testi inseriti in più corpora (testi tradotti);
- definire le modalità di ricerca delle fonti (digitali e cartacee);
- definire il tipo di conversione OCR dei testi secondo il loro formato;
- adeguare il formato dei testi al loro uso all'interno dei corpora anticipando i futuri tipi di uso sia documentario sia lessicografico;
- scegliere un software di gestione con funzioni adeguate all'utenza e agli sviluppi mirati (fra cui creazione di lemmari, concordanze e liste di frequenze) e strutturato per agevolarne la gestione a vari livelli, in particolare per quanto riguarda l'estrazione di informazioni lessicografiche nelle diverse lingue.
- programmare uno spazio digitale in cui alloggiare i corpora e i metadati con l'idea di diffondere strumenti utili alla ricerca a livello internazionale, mettendo a frutto le attuali conoscenze in ambito sia tecnologico sia umanistico;
- diffondere questi strumenti tramite il sito dell'Unità di ricerca⁴ già disponibile in Open Access.

Partendo da questi obiettivi si è progettato lo svolgimento delle seguenti azioni:

- concordare tra tutti i ricercatori delle varie lingue le tipologie di testi da inserire nel corpus, prevedendo anche la realizzazione di corpora paralleli fra le varie lingue del progetto e riservando, dunque, una parte cospicua ad opere di cui esistono traduzioni nelle varie lingue del progetto prestando attenzione alla loro qualità⁵;

³ Cfr. Billero, Nicolás 2017 per una descrizione più dettagliata.

⁴ Nel 2015 l'Unità firmò un contratto di edizione con la Firenze University Press per la pubblicazione del Portale digitale "Lessico per i Beni Culturali", che è consultabile on-line <<http://www.lessicobeniculturali.net>> nelle versioni delle diverse lingue previste dal progetto (cinese, francese, inglese, italiano, portoghese, russo, spagnolo, tedesco). La pubblicazione è stata implementata dal Partner tecnologico: Progettinrete S.r.l.

⁵ Per il processo di verifica di qualità semi-automatizzato cfr. Farina, Billero 2018.

- definire la macrostruttura della banca dati e delle subcategorie (tipologie testuali, traduzione/lingua originale ecc.);
- determinare le specifiche per l'etichettatura dei testi digitalizzati;
- coordinare e validare la digitalizzazione del materiale testuale;
- realizzare un software on line, la cui interfaccia sia accessibile inizialmente ai soli membri dell'unità di ricerca e che consenta la consultazione dei suddetti corpora, in ogni lingua o in diverse lingue contemporaneamente, sulla base dei risultati già ottenuti dalla Umanistica Digitale;
- svolgere incontri formativi destinati ai membri del gruppo LBC e ai tirocinanti al fine di fornire le competenze necessarie per creare i corpora e per utilizzare il software di gestione implementato;
- svolgere Workshop e riunioni di lavoro periodici per verificare l'andamento dei lavori, individuare problemi e per scambi di *know-how* anche con specialisti invitati esterni al gruppo;
- presentare e pubblicare i risultati ottenuti nelle varie fasi del progetto in conferenze e/o riviste specialistiche nazionali e internazionali dando modo al tempo stesso sia di ottenere dalla comunità scientifica riscontri in merito, sia di comunicare ad essa eventuali innovazioni derivanti dal progetto stesso.

3. I corpora LBC: fasi specifiche del lavoro

3.1 Creazione delle bibliografie di testi da inserire nei corpora

È stata realizzata una bibliografia di testi di interesse (corrispondenti per tematica e genere a quelli dell'intero progetto) relativi ad ogni lingua con due criteri di priorità per i testi in lingua originale: autorità riconosciuta del testo/ autore nella cultura di appartenenza e diffusione⁶; facilità di conversione in formato Word, evitando testi di difficile digitalizzazione. Questi criteri si sono poi concretizzati in ulteriori passaggi: è stato redatto un elenco di testi fondamentali in italiano e in altre lingue, primordiali per la conoscenza a livello internazionale del patrimonio artistico-culturale toscano che sarebbe servito come punto di partenza in tutte le lingue per la ricerca di possibili traduzioni. Siccome la lingua di partenza è l'italiano, si è concordato prima di tutto una lista di testi di base di Storia dell'Arte essenziali e riferiti alla Toscana, concretamente: *Le vite* del Vasari, i libri di architettura di Alberti, Palladio, Sellio, alcuni scritti di Machiavelli e di Leonardo. Oltre a questi, ogni lingua ha considerato libri di viaggio di nota fama, come i viaggi in Toscana di Stendhal e Ruskin, e libri d'arte come il Burckhardt.

⁶ «I criteri di selezione dei testi sono stati: la rilevanza storico-culturale dell'opera dell'ambito specifico di studio (ad es. testi di Vitruvio o Leonardo); la diffusione internazionale di un'opera relazionata con l'ambito di studio (es. libri di Vasari); il prestigio dato a livello internazionale al patrimonio italiano da parte di un'opera (es. testi di Stendhal o Ruskin); la specificità dell'argomento in rapporto alla storia dell'arte italiana, e in particolare della Toscana (es. Burckhardt)» (Billero, Nicolás 2017: 208).

La scelta delle diverse tipologie di testo è stata concordata fra i vari gruppi di lavoro, aprendo un dibattito legato alla decisione dei testi da introdurre ma anche alla classificazione delle tipologie testuali. Un punto, per esempio, che ha richiesto una lunga riflessione riguarda la categorizzazione delle *Vite* di Vasari come testo tecnico, poiché si tratta di un testo caratterizzato dalla presenza di stili diversi e dunque attribuibile a generi testuali diversi⁷. La decisione di classificarlo come testo tecnico si basa sulla constatazione che le *Vite* è un testo di riferimento non solo per gli storici dell'arte ma anche per chi si interessa delle tecniche artistiche del periodo rinascimentale.

Anche la classificazione di testi epistolari nella categoria "Letterario-Biografico" è stata oggetto di discussione: le prime corrispondenze introdotte nei corpora sono state quelle di viaggiatori del Grand Tour francese in uno stile molto simile a quello dei quaderni di viaggio, si è quindi pensato di inserirli nella stessa categoria. Diversamente, nelle corrispondenze di artisti come quella di Michelangelo, stile e contenuto sono molto diversi. Si è quindi pensato di creare una categoria apposita, "Documentario", contenente la sottocategoria epistolare. Questa possibilità sarà valutata per l'estensione futura della banca dati testuali.

Non abbiamo comunque definito nella prima fase la priorità e proporzione da dare ad ogni tipologia nei diversi corpora: i criteri sopra descritti rendevano la scelta dipendente dalle specificità di ogni lingua/cultura e del tipo di risorse facilmente reperibili nei vari Paesi. Infatti, se l'Italia è stata meta di numerosi viaggiatori del Grand Tour, il percorso è diverso in base alla nazionalità dei viaggiatori e non comprende per forza la Toscana, inoltre i loro resoconti di viaggio non sono stati pubblicati nella stessa proporzione. L'interesse per l'arte rinascimentale italiana non si è sviluppato negli stessi periodi nelle varie nazioni e non ha dato luogo alla stessa quantità di opere critiche. Da queste osservazioni deriva una eterogeneità fra corpora che vorremmo limitare negli sviluppi futuri del progetto. Infatti l'analisi della distribuzione delle tipologie di testi scelti in ogni corpus e dei secoli rappresentati alla fine di questa prima fase di costituzione dei corpora potrà permettere una più ampia omogeneizzazione in futuro, consentendo lavori di comparazione dei testi. Nella prima fase la priorità data

⁷ «L'opera nel suo complesso comprende, oltre alla descrizione delle vite degli artisti, un apparato testuale di cornice costituito dalla Dedicazione a Cosimo I de' Medici e dal Proemio, definibili come microtesti appartenenti a generi testuali differenti da quello biografico, con funzione appellativa e stile argomentativo. Anche le stesse singole *Vite* non sono riconducibili esclusivamente al genere della biografia, ma presentano in linea di massima tre nuclei tematici, ognuno caratterizzato dalla predominanza di uno stile peculiare: nella parte iniziale, in cui si descrivono l'infanzia e la vita dell'artista, predomina il carattere biografico e lo stile è descrittivo-narrativo [...]. Nella parte centrale si ha la descrizione dei viaggi e delle opere dell'artista fino alla sua morte. In essa predomina il carattere artistico-storiografico e lo stile è descrittivo-esplicativo [...] Nella parte finale si annoverano i discepoli e sono narrati aneddoti sulla vita dell'artista: qui emerge nuovamente il tratto biografico, arricchito però di elementi narrativi come dialoghi e poetici come parti in versi, spesso citate da altri autori» (Ballestracci 2018: 22-23).

all’inserimento di testi di riferimento della propria lingua e/o di facile conversione ha permesso di ottenere una base di testi consistente e sufficiente per ricerche in un’unica lingua⁸.

3.2 Creazione degli strumenti necessari alla gestione delle fonti

3.2.1 Condivisione delle cartelle

Per consentire la corretta gestione del flusso di lavoro⁹ si sono stabilite delle fasi, che sono state seguite in ogni lingua, facilitate dalla creazione di cartelle condivise dai responsabili del lavoro di ogni lingua e dell’informatico. In un primo momento sono state create in Dropbox e successivamente in un Cloud (Figura 1). Questo sistema ha facilitato enormemente il controllo e il processo del lavoro. Nelle cartelle venivano di volta in volta inseriti sia i risultati sia le informazioni sui compiti da svolgere/svolti in ogni tappa del lavoro.

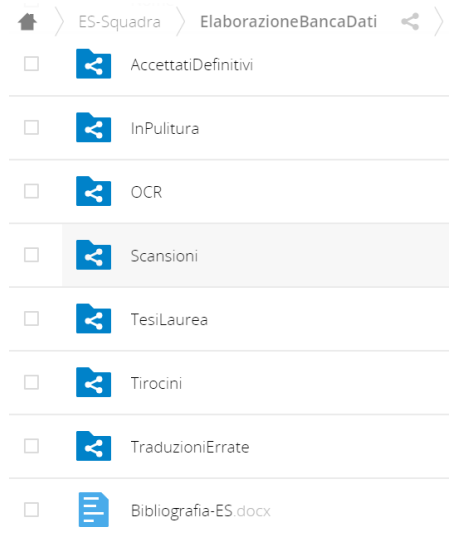


Figura 1. Cartelle Cloud del corpus spagnolo.

Per i testi in versione cartacea si è proceduto alla scannerizzazione, inserendo la versione in PDF nella cartella *Scansioni*. Finito questo lavoro, si sono trasferiti i testi scansionati nella cartella OCR insieme ai testi già disponibili.

⁸ Di fatto, i membri del gruppo di lingue diverse hanno già cominciato studi comparativi fra corpora o sotto-corpora LBC che potranno essere utili in fase di bilanciamento dei corpora. Citiamo ad esempio Farina, Flinz 2020 per la comparazione fra francese e tedesco.

⁹ Cfr. Billero 2020 per una descrizione più dettagliata del flusso di lavoro.

li in versione PDF (principalmente testi liberi da diritti d'autore disponibili nelle biblioteche in rete). Da questa cartella l'informatico ha proceduto alla trasformazione dei PDF in formato Word.

I testi in Word sono stati messi nella cartella *InPulitura* e sono stati distribuiti fra i membri del gruppo o consegnati agli studenti che effettuavano i tirocini per l'unità di ricerca. La scrupolosa verifica di un testo in una lingua studiata dagli studenti comparandolo con la versione originale PDF, eliminando gli errori creati nella conversione tramite OCR ci è parsa molto formativa, e ha costituito perciò la parte principale dei tirocini svolti all'interno del gruppo. Tutti i tirocini comprendevano comunque l'acquisizione di *savoir-faire* che riguardava l'insieme del processo legato alla creazione delle banche dati. Nella cartella *Tirocini*, il responsabile del tirocinio raccoglieva i testi 'puliti' che necessitavano di un secondo controllo prima di passare alla cartella *AccettatiDefinitivi*, contenente i testi revisionati dai membri del gruppo.

I testi sono stati spogliati di tutto ciò che non era testo, di tutti gli elementi paratestuali e delle formattazioni dovute all'edizione (intestazioni e piè di pagina, numeri di pagina, illustrazioni, bibliografie ecc.). Si sono mantenute a parte le note originali nel testo distinguendo le note degli autori da quelle di traduttori o editori con etichette (nda/ndt/nde), ma sono state eliminate in fase di introduzione dei testi nel software di gestione dei corpora per evitare errori di datazione e di paternità.

3.2.2 Creazione della macrostruttura dei corpora

Nel disegno del corpus si sono seguite regole tradizionali rispettando criteri condivisi da tutti i membri del gruppo per la gestione dei metadati quali:

- lingua: è inserita sia la lingua del testo sia la lingua di origine per i testi in traduzione;
- informazione cronologica: differenziazione tra data di redazione dei testi (laddove possibile) e data di edizione; per i testi tradotti sono state inserite le stesse informazioni sia sul testo di origine che sul testo tradotto¹⁰. In caso di datazioni poco chiare è stata aggiunta l'etichetta "Ipotesi". Per le pubblicazioni online viene indicata la data di consultazione;
- tipologia di testo: tutti i testi hanno come argomento il patrimonio artistico e il suo lessico, e in particolare un'ampia visione di Firenze e della Toscana descritta da diversi punti di vista. Sono state individuate quattro macrocategorie (Divulgativo, Tecnico, Dizionario e Letterario) e le loro relative sottoca-

¹⁰ I testi contenuti vanno dal Rinascimento ai giorni nostri. Sebbene siano presenti entrambe le datazioni, l'anno di pubblicazione è secondario rispetto a quello di redazione. Quest'ultimo, infatti, è il dato di maggiore interesse per l'estrazione di informazioni, poiché rappresentativo delle caratteristiche linguistiche del periodo considerato; infatti, i testi sono stati inseriti nella banca dati rimanendo fedeli all'edizione usata, senza produrre alcun tipo di modernizzazione o di correzione ortografica.

tegorie (Divulgativo: Blog, Guida, Ricettario, Rivista; Tecnico: Architettura, Arte, Edilizia, Enogastronomia, Storia; Letterario: Biografico, Fiction, Saggistica; Dizionario: Monolingue, Bilingue/plurilingue). Si è tenuto conto per individuare queste categorie della destinazione principale dell'opera e del tipo di lettore a cui è rivolta, dati che condizionano il tipo di lingua usata e il suo livello di specializzazione;

- autore: sono indicati cognome e nome e l'indicazione "sa" (senza autore) quando inesistente;
- nome del traduttore: quando reperibile, si indica il nome del traduttore;
- titolo e sottotitolo: si è scelto l'introduzione sia di testi interi sia di frammenti che corrispondono ad un'unità testuale perché provvisti di titoli, quali capitolo di libro, lettera completa, articolo di rivista ecc. Tale scelta è stata effettuata poiché in molti casi l'intero libro non coincideva con gli interessi del progetto ma anche per facilitare la futura realizzazione di versioni in parallelo di testi tradotti¹¹. Per i testi tradotti sono stati inseriti sia titoli originali sia titoli tradotti;
- informazione editoriale: si indica nome editore e luogo di edizione;
- delimitazione geografica: per testi che hanno come oggetto una città o regione definita si è inserito il nome della città o regione. Questa indicazione è presente principalmente per i libri di viaggio e per le corrispondenze.

Con i testi definitivi pronti si è proceduto alla schedatura del testo e alla creazione di metadati condivisi da tutte le lingue.

Per questa fase di compilazione dei metadati si è usato una struttura informatica creata su Excel per poter estrarre successivamente l'informazione (Figura 2). La prima informazione generata automaticamente usando questi dati è il nome di ogni file di testo da cui si possono ricavare in modo immediato le principali informazioni sul testo¹².

¹¹ Abbiamo notato che molte traduzioni dei testi ritenuti da noi essenziali per la conoscenza del patrimonio toscano quali ad esempio *Le Vite* di Vasari o la corrispondenza di Michelangelo non hanno sempre traduzioni integrali nelle varie lingue. La divisione di queste opere nei singoli frammenti permette di mettere in parallelo testi originali e traduzioni anche nel caso di traduzioni parziali.

¹² Il nome del file fr-FR_LET_bio_Bros_LetFam_LetXXIV_1739r_1858p, ad esempio, corrisponde all'informazione: Lingua Originale: Francese (Francia), Categoria e sottocategoria: Letterario-Biografico; Autore: Brocchi, Charles de; Titolo: Lettres familières écrites d'Italie à quelques amis, en 1739 et 1740, avec une étude littéraire et des notes par Hippolyte Babou; Frammento: Lettre XXIV À M. De Quentin. Mémoire sur Florence; Anno di redazione: 1739; Info anno di redazione: Redazione; Anno di pubblicazione: 1858; Info anno di pubblicazione: Pubblicazione. Il nome del file (testo in traduzione) it_TEC_art_Vas_VitGiu_Gio_1568r_1966p_fr-FR_1839t_1839p corrisponde invece a: Lingua originale: Italiano; Lingua di traduzione: Francese (Francia); Categoria e sottocategoria: Tecnico-Arte; Autore: Vasari, Giorgio; Titolo: Vite de' più eccellenti pittori scultori e architettori; Frammento: Vita di Giotto, pittore, scultore e architetto fiorentino; Anno di redazione: 1568; Info anno di redazione: Redazione; Anno di pubblicazione: 1966; Info anno di pubblicazione: Pubblicazione; Titolo traduzione: Vies des peintres, sculpteurs et architectes;

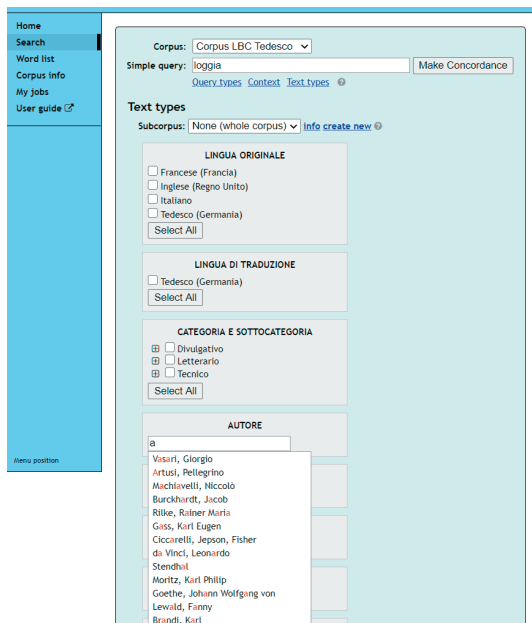


Figura 3. Ricerca con text type su corpus tedesco.

Corpus LBC Francese		General info		Lexicon sizes	
Counts		Corpus description	Document	word	99,715
Tokens	3,818,747	Language	French	tag	34
Words	3,164,995	Encoding	UTF-8	lemma	25,673
Paragraphs	37,464	Compiled	11/12/2019 11:04:35	lc	88,549
Documents	252	Tagset	Description	lemma_lc	25,403

Structures and attributes				
doc 252				
Subcorpora statistics				
Subcorpus	Tokens	Words	%	
DIVBlog	32,613	- 27,029	0.85	
DIVGuida	334,433	- 277,179	8.75	
DIVRivista	50,604	- 41,940	1.32	
DIZMono	850,167	- 704,622	22.26	
LETBio	884,995	- 733,487	23.17	
LETFic	632,831	- 524,493	16.57	
LETSag	710,938	- 589,228	18.61	
LinguaOriginale	3,515,483	- 2,913,648	92.05	
LinguaTraduzione	303,262	- 251,344	7.94	
TECArc	29,308	- 24,290	0.76	
TECArt	292,468	- 242,398	7.65	

Figura 4. Informazione dettagliata sul corpus francese disponibile su “Corpus info”.

4. Conclusioni

La realizzazione di questa prima fase dei nostri corpora è da ritenere soddisfacente in quanto ha creato le basi necessarie per i primi lavori e per le ricerche del nostro gruppo (Carpi 2016; Ballestracci 2018; Lanini, Nicolás 2018; Farina, Frinz 2020). Sono stati già realizzati i primi lemmari di ogni lingua corredati di concordanze estratte dai corpora che verranno pubblicate sulla piattaforma e potranno essere usate per l'elaborazione di futuri dizionari.

Questo primo lavoro realizzato da ogni gruppo aveva lo scopo di validare i corpora nella prospettiva di identificare problemi che emergono solo dal loro uso effettivo.

Nel futuro si pensa di ampliare sia il numero di lingue (mancano ancora i corpora di cinese, portoghese e turco) sia quello dei testi con l'idea di omogeneizzazione già descritta, per cercare di rendere i corpora quanto più possibile comparabili fra loro.

Bibliografia

- Ballestracci S. 2018, *Tra norma e variazione: allora, ora e poi in Vite de' più eccellenti pittori scultori e architettori di Giorgio Vasari. Un'analisi in ottica contrastiva con il tedesco*, in Meli M. (ed.), *Le norme stabilite e infrante: saggi italo-tedeschi in prospettiva linguistica, letteraria e interculturale*, FUP, Firenze: 21-40.
- Billero R., Nicolás Martínez M.C. 2017, *Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 4(2): 203-216.
- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Carpi E. 2016, *El español de la pintura y los recursos lexicográficos y terminológicos: cómo traducir al español tempera y guazzo*, «Cuadernos Aispi», 6: 111-126.
- Farina A., Billero R. 2018, *Comparaison de corpus de langue «naturelle» et de langue «de traduction»: les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues* in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 108-116.
- Farina A., Flinz C. 2020, *Analisi linguistica comparativa dei corpora LBC. La visione del patrimonio fiorentino francese e tedesco: l'esempio del Duomo*, in Farina A., Funari F. (eds.), *Past in Present / Le passé dans le présent / Il passato nel presente*, FUP, Firenze: 75-98.
- Farina A. 2016, *Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique*, «Publif@rum», 24, <http://www.farum.it/publifarum/ezone_articles.php?art_id=335>.
- Lanini L., Nicolás Martínez M.C. 2018, *Verso un dizionario corpus-based del lessico dei beni culturali: procedure di estrazione del lemmario* in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 411-418.
- Zotti V. 2017, *L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle Vite di Vasari*, in Zotti V., Pano Alamán A. (eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*, FUP, Firenze: 105-134.

Implementazione di software per la gestione dei corpora LBC

Riccardo Billero

Introduzione

L'importanza del lessico dei beni culturali italiani va ben oltre i confini delle comunità linguistiche ed è oggetto di interesse per studiosi afferenti a diverse discipline (come storia dell'arte, letteratura, linguistica), per i professionisti (traduttori, guide turistiche, organizzatori di eventi) e per gli studenti di tali discipline. Attraverso l'utilizzo delle moderne tecnologie messe a disposizione dall'Informatica Umanistica, l'Unità di Ricerca "Lessico multilingue dei Beni Culturali" (LBC) si pone l'obiettivo di creare strumenti utili per i soggetti summenzionati, concentrandosi sull'arte di Firenze e della Toscana. Tra tali strumenti rivestono particolare importanza i sei corpora LBC, liberamente disponibili online, la cui creazione è descritta nel seguito.

1. Informatica Umanistica e Lessico dei Beni Culturali

L'Informatica Umanistica è un campo di studio in continua evoluzione, di cui esistono molteplici definizioni, alcune più orientate verso le tecnologie e le metodologie di indagine, altre verso l'innovazione dei contenuti e dei programmi di ricerca all'interno degli studi umanistici, mediante l'ausilio delle nuove tecnologie (Numerico, Vespignani 2003: 13-14). Per Svensson (2010) si tratta fondamentalmente dell'intersezione tra le scienze umanistiche e le tecnologie dell'informazione; ad oggi «si privilegia il concetto di *digital humanities* per riferirsi a una realtà più vasta che riguarda non solo le metodologie di costituzione

Riccardo Billero, University of Florence, Italy, riccardo.billero@unifi.it, 0000-0002-6099-222X

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Riccardo Billero, Annick Farina, Maria Carlota Nicolás Martínez (edited by), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, © 2020 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5870 (online), ISBN 978-88-5518-253-9 (PDF), DOI 10.36253/978-88-5518-253-9

di corpora e di annotazione di testi, nonché l'elaborazione di programmi atti a leggere i dati raccolti in essi, ma anche la realizzazione di applicazioni pensate per poter funzionare su diversi dispositivi elettronici, utili sia alla visualizzazione, fruizione e condivisione di enormi quantità di dati, che all'interpretazione semantica degli stessi da parte dei computer, per menzionare solo alcune possibilità» (Zotti, Pano Alamán 2017: 8).

Qualunque sia la definizione che preferiamo adottare, possiamo comunque affermare che l'incontro tra l'informatica e le scienze umane offre alle nuove generazioni di ricercatori, docenti e studenti un modo nuovo di avvicinarsi alla cultura, mediante l'uso degli strumenti messi oggi a disposizione delle tecnologie contemporanee.

In questa ottica intendiamo qui descrivere il contributo che il progetto Lessico dei Beni Culturali ha ricevuto dall'Informatica Umanistica grazie al lavoro dell'autore.

Il progetto Lessico multilingue dei Beni Culturali (in breve LBC) è attivo da diversi anni presso l'omonima Unità di Ricerca con sede nel Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia (FORLILPSI) dell'Università degli Studi di Firenze (Billero, Nicolás Martínez 2017; Billero 2020) e si pone gli obiettivi di:

- a. creare una piattaforma web di riferimento utile allo studio del lessico dei beni culturali;
- b. creare dei corpora di riferimento per ognuna delle lingue coinvolte nel progetto¹;
- c. fornire l'accesso a tali corpora e degli strumenti utili per la ricerca al loro interno, al fine di abilitare e promuovere studi in settori quali linguistica, lessicografia o ambiti socioculturali.
- d. creare un dizionario multilingue dei termini artistici che utilizzi i corpora come riferimento.

Attualmente, le risorse principali del progetto LBC sono sei corpora contenenti una raccolta di testi in varie lingue (come di seguito descritto) e sei lemmari relativi al linguaggio dell'arte, estratti dai corpora per ciascuna lingua e successivamente elaborati dai membri dei gruppi linguistici coinvolti nel progetto. Nel seguito saranno descritti in particolare i corpora LBC ed il relativo processo di creazione.

2. I corpora LBC

I corpora LBC sono una delle principali risorse dell'Unità di Ricerca "Lessico multilingue dei Beni Culturali". Fin dall'inizio dei lavori (nel 2016), è stato previsto che i corpora LBC sarebbero stati utilizzati non solo dai membri dell'Unità

¹ Al momento le lingue coinvolte nel progetto sono: cinese, francese, inglese, italiano, portoghese, russo, spagnolo, tedesco, turco.

di Ricerca LBC² nella fase di implementazione del dizionario, ma anche da altri utenti, quali ad es. traduttori e specialisti in storia dell'arte italiana.

In considerazione di ciò, i corpora LBC sono stati progettati per avere un ampio uso specialistico, con i seguenti requisiti di base:

- a. disporre di materiale linguistico utile per creare e documentare future voci del dizionario multilingue LBC;
- b. disporre di materiale linguistico utile per svolgere studi linguistici, letterali o culturali;
- c. includere testi che sensibilizzino il grande pubblico al patrimonio culturale di Firenze e della Toscana.

Per la creazione dei corpora LBC, abbiamo considerato testi scritti dal Rinascimento ai giorni nostri ed aventi per soggetto il patrimonio culturale e, in particolare, una visione di Firenze o della Toscana descritta da diversi punti di vista, sia oggettivi che personali. L'ambito geografico cambierà in futuro, per coprire l'intera Italia e altre culture. Tra i testi inseriti all'interno dei nostri corpora, possiamo enumerare:

- a. biografie, es. Cellini, *La vita* o Dumas, *Une année à Florence*;
- b. testi di arte, es. Vasari, *Le Vite* o Alarcón, *Vidas de los pintores y estatuarios*;
- c. romanzi, es. Forster, *A Room with a View*;
- d. saggi, es. Feuillet, *L'art italien* o Burckhardt, *Geschichte der Renaissance*;
- e. guide turistiche, es. Routard, *GeoGuide*;
- f. dizionari specialistici, es. Viollet Le Duc, *Dictionnaire de l'architecture*.

Per ognuno dei testi presenti nei corpora sono memorizzati sia l'anno di redazione che l'anno di pubblicazione, con quest'ultimo considerato secondario rispetto al primo. In effetti, l'anno di redazione sarà il dato di fondamentale importanza per l'estrazione delle informazioni, poiché è in qualche modo rappresentativo delle caratteristiche linguistiche del periodo in esame; infatti i testi sono stati inseriti in tutti i corpora rimanendo fedeli all'edizione utilizzata, senza produrre alcun tipo di modernizzazione.

Va sottolineato che all'interno dei nostri corpora sono stati inclusi sia libri interi che frammenti 'auto-conclusivi' (come un capitolo di un libro, un articolo di una rivista); questa scelta è stata fatta perché in molti casi i contenuti dei libri nella loro interezza non sempre coincidevano con gli interessi del progetto.

È stato svolto un lavoro preliminare per determinare le categorie da utilizzare per catalogare i testi, a seconda dei generi testuali inclusi nei corpora. Innanzitutto abbiamo differenziato i testi *tecnici* da quelli *informativi*, soprattutto in base al target di riferimento, ovvero se il destinatario è o meno uno specialista. La terza categoria individuata è quella dei *dizionari*, che sono da considerarsi testi tecnici nel nostro campo di interesse. Infine, è stato inserito il genere *lette-*

² Esempi di utilizzo possono essere trovati in Carpi 2017; Farina, Billero 2018; Billero, Carpi 2018; Lanini, Nicolás 2018; Garzaniti 2020; Farina, Frinz 2020.

rario in quanto si rivela utile per fornire una visione di Firenze e della Toscana, esprimendo come il suo patrimonio è percepito da un punto di vista esterno. In dettaglio, sono state individuate le seguenti categorie e sottocategorie:

1. Divulgativo – Blog
2. Divulgativo – Guida
3. Divulgativo – Ricettario
4. Divulgativo – Rivista
5. Tecnico – Architettura
6. Tecnico – Arte
7. Tecnico – Edilizia
8. Tecnico – Enogastronomia
9. Tecnico – Storia
10. Letterario – Biografico
11. Letterario – Fiction
12. Letterario – Saggistica
13. Dizionario – Monolingue
14. Dizionario – Multilingue

La suddetta classificazione si è rivelata finora efficace, tuttavia occorre sottolineare che è ancora provvisoria, come notano gli stessi membri dell'Unità di Ricerca, e potrebbe essere soggetta a revisioni nelle successive fasi di lavoro.

3. Il flusso di lavoro per la creazione dei corpora

È stato creato un flusso di lavoro adeguato alla generazione dei corpora LBC che ha permesso di eseguire tutti i passaggi necessari ad ottenere corpora consultabili online partendo da un semplice elenco di testi da utilizzare.

Le responsabilità per l'esecuzione di tale flusso di lavoro sono suddivise nel modo seguente:

1. Il processo di selezione, digitalizzazione e raccolta dei testi (con relativi metadati) da inserire nei corpora è curato dai responsabili dei gruppi linguistici;
2. La gestione IT dell'intero progetto LBC era invece sotto la nostra responsabilità.

Le due aree di responsabilità sono essenzialmente indipendenti l'una dall'altra, o più correttamente vengono svolte in momenti diversi. Naturalmente, va notato che la seconda richiede i risultati della prima come dati in input. Le due sezioni seguenti illustreranno le due diverse aree delineate sopra, concentrandosi in particolare sugli aspetti tecnici e informatici.

4. La raccolta dei testi

Il flusso di lavoro relativo al processo di selezione, digitalizzazione e raccolta dei testi (con relativi metadati) da inserire nei corpora è stato condotto dai diversi gruppi linguistici, uno per ogni lingua inclusa nel progetto (Farina 2016).

In primo luogo, ogni gruppo ha compilato una bibliografia di testi (scritti nella relativa lingua di riferimento) aventi come tema il patrimonio culturale di Firenze e della Toscana. In particolare occorre osservare che il gruppo di lavoro di lingua italiana ha condotto una ricerca al fine di creare una bibliografia dei testi più importanti che dovevano essere inclusi nei corpora, i cosiddetti ‘testi fondatori’; tali testi sono considerati rilevanti o per la loro terminologia artistica (l’arte italiana è un punto di riferimento per la storia dell’arte) o perché sono classici dell’arte italiana (come Vasari o Michelangelo) che hanno come oggetto Firenze.

Successivamente, ogni gruppo ha individuato le traduzioni di questi testi e ha cercato di identificare altri testi, nella propria lingua, aventi come soggetto l’arte a Firenze o in Toscana, cioè libri di viaggiatori di passaggio a Firenze (es. *Geschichte der Renaissance* di Burckhardt).

Al fine di consentire l’accesso ai file a tutti i membri dell’Unità di Ricerca, si è deciso di utilizzare una soluzione di archiviazione dati su cloud, basata dapprima su cartelle Dropbox, una per ogni lingua coinvolta nel progetto. I responsabili dei diversi gruppi linguistici hanno avuto accesso ai propri dati in modalità di lettura e scrittura, accedendo invece ai dati delle altre lingue in modalità di sola lettura.

Dropbox è stato inizialmente scelto per i suoi vantaggi, come la facilità della configurazione iniziale (da parte del responsabile IT) e la facilità d’uso per tutti i membri del gruppo. In un secondo momento, sia perché questa soluzione si è rivelata troppo limitata in termini di spazio disponibile per le esigenze del gruppo, sia per gestire i nostri documenti ‘privatamente’, è stato noleggiato un VPS (Virtual Private Server) dedicato al progetto e opportunamente configurato con NextCloud³, che ha sostituito in breve tempo Dropbox, seppur utilizzando lo stesso criterio di gestione dei file.

I responsabili dei vari gruppi linguistici sono stati incaricati di raccogliere – all’interno della propria cartella cloud – i file contenenti i vari testi utilizzati per costruire il corpus.

Per i testi da inserire nel corpus viene utilizzato Microsoft Word; questo software è stato scelto in considerazione della necessità dei membri dell’Unità di Ricerca di mantenere alcuni degli elementi di formattazione di base come grassetto, corsivo e note a piè di pagina. Word è stato preferito anche ad altre soluzioni (come LibreOffice Writer) a causa dell’esperienza dei membri dei gruppi

³ Per quanto riguarda la scelta del software, sono stati preferiti strumenti *open source*, per vari motivi: il codice open source è di qualità superiore e rimane affidabile nel tempo; la trasparenza dell’open source e il controllo fornito da tutti i contributori garantisce una maggiore sicurezza; l’open source ha ampie comunità di utenti che consentono un miglioramento continuo delle prestazioni e un’ampia fonte di risposte a tutte le preoccupazioni; le piattaforme open source consentono una maggiore interoperabilità, contro la rigidità di un’unica soluzione che non può soddisfare tutte le esigenze. Tuttavia, tali considerazioni non erano necessarie per quanto riguarda Microsoft Word ed Excel, gli altri strumenti usati dai componenti dei gruppi linguistici, essendo strumenti di provata efficienza e già a disposizione dei ricercatori coinvolti nel progetto.

con Word e perché Dropbox (la soluzione cloud utilizzata inizialmente) ha una versione online del software. Inoltre, dato che il formato *.docx* (il formato di file utilizzato da Word 2007 in poi) è uno standard XML, è stato possibile implementare un software – descritto di seguito – per l'estrazione automatica dei testi dai file e il loro inserimento nei corpora.

Prima di inserire ogni file di testo, i responsabili dei vari gruppi linguistici devono salvare i relativi metadati compilando una riga in un file Microsoft Excel⁴, contenuto nella cartella cloud della propria lingua; in questo modo ottengono il nome che deve essere assegnato al file Word (stabilito algebricamente dal suddetto file Excel).

L'utilizzo di un nome di file ottenuto algebricamente si è reso necessario per facilitare il lavoro dei membri dell'Unità di Ricerca LBC nella fase di consultazione dei corpora; questo nome deve essere 'intelligibile', cioè costituito da una serie di elementi che facilitano la ricerca del testo di interesse, distinguendolo facilmente dagli altri presenti nei corpora. In particolare, alcuni dei metadati inclusi nel foglio Excel contribuiscono a creare il nome univoco assegnato al file contenente il testo; questo nome viene stabilito utilizzando una apposita formula Excel e associando delle abbreviazioni per i vari metadati.

Le informazioni utilizzate per la creazione dei nomi dei file sono:

- a. lingua originale del testo; ogni lingua è associata al proprio codice, secondo la codifica ISO 639-2 Alpha-2, che consente di specificare un paese o una regione⁵. Ad esempio, un libro in inglese britannico sarà identificato dal codice *en-GB*.
- b. Categoria e sottocategoria del testo come sopra descritto. Ad esempio, un libro classificato come Letterario – Saggistica sarà identificato dall'acronimo: *LET_sag*.
- c. Autore del testo. La cartella di lavoro di Excel include un foglio di lavoro degli autori che deve essere compilato dai responsabili dei gruppi linguistici con il nome dell'autore e un'abbreviazione univoca di 3-4 lettere che contribuirà a creare l'identificativo del file di testo utilizzato nei corpora. Ad esempio, tutti i libri scritti da Ruskin avrebbero *Rusk* nel loro identificativo. All'interno del foglio di lavoro dei metadati, le celle della colonna Autore mostrano una casella a discesa riempita automaticamente con i nomi degli autori presenti nel suddetto foglio di lavoro degli autori; in questo modo è possibile ridurre al minimo gli errori di battitura o errate attribuzioni degli acronimi.
- d. Titolo del libro (o rivista o miscellanea). Come nel caso degli autori, all'interno della cartella di lavoro di Excel è presente un foglio di lavoro *Titoli* che

⁴ La scelta di Microsoft Excel per l'archiviazione dei dati è stata fatta per gli stessi motivi per cui è stato scelto Word, ovvero la sua pervasività, la sua semplicità d'uso e il fatto che il formato *.xlsx* (il formato di file utilizzato da Excel 2007 in poi) è uno standard XML.

⁵ In particolare: per il francese si può scegliere tra la variante della Francia e quella del Canada; per l'inglese tra le varianti del Regno Unito e degli USA; per il portoghese tra Portogallo e Brasile; per lo spagnolo tra Spagna e America Latina; per il tedesco tra Germania, Austria o Svizzera.

deve essere compilato dai responsabili dei gruppi linguistici, che inseriscono il titolo del libro e un'abbreviazione univoca di 3-4 lettere che contribuirà a creare l'identificativo del file di testo utilizzato nei corpora. Ad esempio, il libro *Mornings in Florence*, sarà associato all'abbreviazione *Morn*. All'interno del foglio dei metadati, le celle della colonna *Titolo* visualizzano una casella a discesa riempita automaticamente con i titoli presenti nel suddetto foglio di lavoro *Titoli*, come nel caso degli autori.

- e. Titolo del frammento (o articolo su rivista o miscellanea). Come nel caso degli autori e dei titoli, all'interno della cartella di lavoro di Excel è presente un foglio di lavoro *Frammenti* che deve essere compilato dai responsabili dei gruppi linguistici, che inseriscono il nome del frammento e un'abbreviazione univoca di 3-4 lettere che contribuirà a creare l'identificativo del file di testo utilizzato nei corpora. Se il testo considerato è stato preso nella sua interezza, deve essere indicato utilizzando l'abbreviazione *integ*. Analogamente ai due casi precedenti, all'interno del foglio di lavoro dei metadati, la colonna *Frammenti* mostra una casella a discesa riempita automaticamente con i titoli presenti nel suddetto foglio di lavoro *Frammenti*, come nel caso degli autori e dei titoli.
- f. Anno di redazione del testo. Ad esempio, un libro scritto nel 1875 sarà identificato dall'acronimo *1875r*, dove *r* significa redazione. Per la data di redazione è possibile indicare l'anno preciso (se noto), oppure un'ipotesi dell'anno o del secolo di riferimento.
- g. Anno di pubblicazione del testo. Ad esempio, un'edizione del 1881 di un libro utilizzato sarà identificata dall'acronimo *1881p*, dove *p* sta per pubblicato. Per la data di pubblicazione è possibile specificare l'anno preciso (se noto), oppure un'ipotesi dell'anno o del secolo di riferimento.

Pertanto, il libro *Mornings in Florence*, scritto in inglese nel 1875 e pubblicato nell'edizione del 1881, classificato come Letterario – Saggistica, sarà identificato dal nome di file seguente:

en-GB_LET_sag_Rusk_Morn_integ_1875r_1881p

dove:

en-GB indica la lingua inglese nella variante del Regno Unito;

LET_sag indica la categoria;

Rusk è un'abbreviazione per l'autore Ruskin;

Morn è una abbreviazione del titolo del testo;

integ indica che il testo è stato considerato nella sua interezza;

1875r indica che il testo fu redatto nel 1875;

1881p indica che il testo fu pubblicato nel 1881.

Per rendere i testi riconoscibili rispetto a quelli in lingua originale vengono giustapposti alcuni elementi al nome che il testo ha nella versione in lingua originale. A tal fine i responsabili linguistici devono inserire – nella stessa cartel-

la di lavoro Excel fin qui considerata – informazioni sulla versione tradotta del testo; in particolare, i dati essenziali per la creazione del nome sono i seguenti:

- a. lingua di traduzione; come per la lingua originale, è possibile specificare una lingua relativa a un paese o una regione;
- b. anno di traduzione (come l'anno di redazione del testo in lingua originale);
- c. anno di pubblicazione (come quello del testo in lingua originale).

Ad esempio, il caso precedente del testo inglese *Mornings in Florence*, scritto nel 1875 e considerato nell'edizione del 1881, classificato come Letterario – Saggistica e identificato nell'edizione inglese dal nome del file già visto nell'esempio:

en-GB_LET_sag_Rusk_Morn_integ_1875r_1881p

è identificato nella traduzione spagnola dal nome:

en-GB_LET_sag_Rusk_Morn_integ_1875r_1881p_es-ES_1910t_1910p

dove la parte giustapposta include:

es-ES a indicare la lingua spagnola (nella sua variante di Spagna);

1910t a indicare che il testo fu tradotto (*t*) nel 1910;

1910p a indicare che il testo è considerato nella versione pubblicata (*p*) nel 1910.

Oltre ai metadati richiesti per creare il nome del file, è anche possibile inserire le seguenti informazioni nel foglio di lavoro di Excel, quali metadati aggiuntivi per il testo:

- supporto per il quale il testo è stato originariamente creato (carta, CD/DVD, sito web);
- nome e cognome del curatore (se miscellanea);
- numero di volume della rivista o mese (se presente);
- indirizzo web (se presente) da cui proviene il libro o la pagina consultata;
- casa editrice;
- luogo di pubblicazione.

Se il testo è tradotto si possono inserire anche i seguenti metadati:

- titolo del libro o della rivista o miscellanea tradotta;
- titolo del frammento (o dell'articolo se è incluso in una rivista o miscellanea);
- luogo di pubblicazione;
- nome e cognome del traduttore;
- indirizzo web (se presente);
- casa editrice;
- commenti vari.

Come risultato di questo flusso di lavoro, all'interno della nostra installazione di NextCloud saranno presenti diverse cartelle, una per ogni lingua coinvolta nel progetto, contenenti tutti i testi da inserire nei corpora; tale operazione è gestita dalla seconda parte del flusso di lavoro.

5. La creazione dei corpora

La seconda parte del flusso di lavoro, riguardante tutti i processi digitali necessari per raccogliere i testi (con i relativi metadati) e inserirli all'interno di corpora strutturati per l'utilizzo da parte del software di gestione dei corpora, era invece di nostra responsabilità, come sopra indicato. In particolare, a tal fine, sono stati realizzati due appositi software (o, più correttamente, script), utilizzando i linguaggi di programmazione Bash e Python, automatizzando così l'intero flusso di lavoro.

Come accennato in precedenza, dopo una fase iniziale durante la quale Dropbox è stato utilizzato come soluzione cloud, si è deciso di utilizzare NextCloud sul VPS proprietario. Abbiamo noleggiato una macchina virtuale fornita da un servizio online, su cui era installata la distribuzione Debian GNU/Linux; tale VPS non solo ha permesso la creazione di un nostro cloud, ma ha anche consentito l'esecuzione del suddetto script e la possibilità di rendere disponibili i vari corpora tramite un servizio web.

La seconda parte del flusso di lavoro viene coordinata da uno script Bash⁶, il quale richiama altri script o software in un ordine ben determinato, ognuno responsabile dell'esecuzione di uno dei tre passaggi che trasformano una raccolta di testi in formato Microsoft Word (e i loro metadati in formato Microsoft Excel) in un insieme di corpora, ovvero:

1. riconciliare ogni testo con i suoi metadati e raccogliarlo in un unico documento con tutti gli altri testi;
2. tokenizzare e lemmatizzare tutti i testi;
3. compilare il corpus.

Il primo passaggio viene eseguito da un apposito script Python⁷, il quale legge il foglio Excel contenente l'elenco dei testi da inserire nel corpus e dei relativi metadati e per ogni riga di informazione in esso presente individua i vari metadati e il nome assegnato al relativo file Word (generato automaticamente da una apposita formula, come sopra indicato); con queste informazioni, lo script può quindi cercare nella cartella cloud il file Word e aprirlo.

A partire dal 2007, ogni documento Microsoft Word viene archiviato utilizzando il formato Office Open XML (standard internazionale ISO/IEC DIS 29500). Questo tipo di documento è costituito da un file compresso in cui sono presenti vari documenti XML che specificano i molteplici elementi del documen-

⁶ Bash è una shell e un processore di comandi per i sistemi operativi GNU/Linux che è ampiamente utilizzata come shell di login predefinita per la maggior parte delle distribuzioni Linux e può leggere ed eseguire comandi da un file, il cosiddetto script di shell.

⁷ Alla base della scelta di utilizzare Python come linguaggio di programmazione ci sono numerosi motivi: si tratta di un linguaggio completamente gratuito, utilizzabile su varie piattaforme (Linux, Windows ecc.), già installato di default in molte distribuzioni Linux (inclusa Debian), facile da usare, dotato di librerie per gestire agevolmente documenti in formato Microsoft Word ed Excel e più in generale in formato testo e XML, con risultati altamente performanti.

to Word, come il testo in esso contenuto, la sua formattazione, gli oggetti inseriti, ecc. Grazie alla conoscenza di questo standard, lo script Python apre il documento Word, ne legge il testo e lo copia come contenuto in un opportuno nodo di un documento XML, senza alcun elemento di formattazione; invece gli attributi di tale nodo vengono impostati utilizzando tutti i relativi metadati contenuti nel file Excel. Come risultato del processo, lo script Python genera in output – per ogni lingua – un documento XML contenente tutti i testi con i relativi metadati; quel documento verrà quindi utilizzato come input per il passo seguente.

Tuttavia, poiché la lettura di file Word ed Excel richiede una certa quantità di tempo per elaborare il contenuto che diventa eccessivo quando centinaia o migliaia di testi vengono inseriti nel corpus, è stata implementata una cache per accelerare la ricompilazione del corpus. Tale cache si basa sul principio che una volta che i testi e i relativi metadati vengono memorizzati, saranno modificati molto raramente e sostanzialmente solo per correggere errori materiali riscontrati successivamente all’inserimento del testo e dei relativi metadati. In questo modo diventa più veloce controllare – ogni volta che il corpus viene ricompilato – se sono state apportate modifiche; l’introduzione della cache ci ha permesso di ridurre il tempo di ricompilazione di un corpus da poche ore a circa dieci minuti, per un corpus di un milione di parole. Nonostante aggiunga complessità al flusso di lavoro, l’utilizzo della cache ne riduce i tempi di esecuzione in base al principio che nuovi testi e relativi metadati possono essere continuamente inseriti, mentre raramente accade che i testi esistenti vengano modificati (Billero 2020).

Il secondo passaggio dello script Bash consente la tokenizzazione e quindi la lemmatizzazione del testo. La procedura di tokenizzazione trasforma il file XML ottenuto come output del passaggio precedente nel cosiddetto file verticale, ovvero un file strutturato in modo che su ogni riga sia presente un solo token (una singola parola o un unico simbolo grammaticale). Questa modalità di memorizzazione dei dati è funzionale alla procedura di lemmatizzazione dove il cosiddetto Part-Of-Speech (o più brevemente POS) e il lemma vengono memorizzati all’interno del file verticale accanto a ciascun token.

Ad esempio, la frase «This book is easy to read» verrà memorizzata come segue:

Word	POS	Lemma
This	DT	this
book	NN	book
is	VBZ	be
easy	JJ	easy
to	TO	to
read	VB	read
.	SENT	.

Per questa seconda fase, abbiamo scelto di utilizzare TreeTagger, un tagger per parti del discorso indipendente dalla lingua (Schmid 1994, 1995), uno dei software più importanti della sua categoria. Pertanto, lo script Bash richiama

TreeTagger, utilizzando come input il file XML ottenuto come risultato del precedente passaggio contenente tutti i testi con i relativi metadati.

Infine, il terzo passaggio consiste nella compilazione del corpus, ovvero nel memorizzarlo in un formato facilmente accessibile agli utenti finali. Dato il numero diversificato di utenti, la consultazione delle informazioni dovrebbe essere quanto più ampia possibile; pertanto, dovrebbe essere utilizzato un buon software di analisi e gestione dei corpora. Esistono due diverse tipologie di tale software: quelli che possono essere utilizzati solo su un computer desktop e quelli disponibili come servizio web. Per soddisfare gli scopi del progetto, abbiamo scelto un software che potesse essere utilizzato come servizio web.

Una ricerca dello stato dell'arte ha identificato l'esistenza di tre principali software all'interno della famiglia sopra descritta: CWB (IMS Open Corpus WorkBench), INL BlackLab e SketchEngine. Va notato che mentre i primi due software sono disponibili come open source, SketchEngine è solitamente disponibile solo a pagamento, ma attualmente è gratuito per i membri degli istituti di ricerca in Europa. Tuttavia, è stata rilasciata anche una versione gratuita ridotta, chiamata NoSketchEngine, priva delle caratteristiche più interessanti per lo studio lessicografico incluse nella versione commerciale, come la funzione Word Sketch e il *thesaurus*.

Tutti e tre i software sopra menzionati hanno una serie di funzionalità di base che sono di fondamentale importanza nella ricerca di corrispondenze, tra cui: la possibilità di effettuare ricerche semplici, ricerche utilizzando espressioni regolari o Context Query Language⁸, oltre alla possibilità di utilizzare lemmatizzatori come TreeTagger, Freeling, RfTagger. Inoltre, tutti e tre i software dispongono di una speciale API (Application Programming Interface) che consente loro di eseguire operazioni come l'interrogazione o l'inserimento di dati tramite software di terze parti appositamente implementati per le proprie esigenze.

Mentre CWB e SketchEngine dispongono attualmente di un'eccellente documentazione e di una solida comunità di riferimento, lo stesso non si può dire per BlackLab, che mostra ancora debolezze in questi aspetti.

SketchEngine offre anche la funzionalità di Word Sketch, utile per la creazione di elenchi di parole di termini specifici delle diverse arti e che può essere utilizzata per mettere in relazione parole che si riferiscono allo stesso oggetto artistico, per delimitare campi di conoscenza o di riferimento come monumenti, o distinguere tra oggetti e persone e disambiguare i nomi propri che si riferiscono a queste due categorie. Tuttavia, questa funzione non è disponibile nella versione gratuita, ovvero NoSketchEngine.

Alla fine, abbiamo scelto per la nostra ricerca il software di gestione dei corpora SketchEngine. Abbiamo anche provveduto ad effettuare un'installazione di NoSketchEngine sul nostro VPS rendendo così i nostri corpora disponibili

⁸ Context Query Language (CQL in breve) è un linguaggio di query originariamente creato per CWB e successivamente implementato negli altri software menzionati.

online⁹ agli utenti finali. Attualmente, i corpora contengono otto milione di parole, divise per lingua, come indicato nella tabella seguente.

Tabella 1. Numero attuale di parole per lingua.

Lingua	Parole
Francese	3.165.000
Russo	1.900.000
Inglese	1.036.000
Spagnolo	1.020.000
Tedesco	1.018.000
Italiano	1.009.000

L'utilizzo del software di gestione dei corpora NoSketchEngine consente agli utenti di eseguire vari tipi di ricerche. Un primo tipo di ricerca è la ricerca per parola, in cui gli utenti cercano una parola scritta nel modo in cui si aspettano di trovarla nei corpora; altrimenti è anche possibile cercare il lemma, dove NoSketchEngine fornirà i risultati come un elenco di occorrenze di quel lemma nelle varie forme flesse presenti all'interno del corpus. È anche possibile utilizzare il linguaggio di query CQL per eseguire ricerche più selettive, nonché la possibilità di specificare un particolare POS (Part-of-Speech) da cercare. Ad esempio, l'utente finale può cercare aggettivi vicino alla parola opera o cercare verbi dopo il nome di un pittore, ecc.

Per quanto riguarda i filtri di ricerca, è possibile utilizzare molti dei metadati per limitare l'ambito delle ricerche, dall'intero corpus a una sezione; ad esempio, gli utenti finali possono filtrare per lingua, autore, titolo dell'opera, categoria, sottocategoria e così via.

Di conseguenza, possiamo affermare che NoSketchEngine è un potente strumento per la ricerca all'interno dei nostri corpora utilizzando ricerche generiche o dettagliate e che soddisfa i requisiti del progetto LBC e le diverse esigenze degli utenti specificate in precedenza.

Conclusioni

La realizzazione di questa prima fase dei nostri corpora è da ritenere soddisfacente in quanto ha creato le basi necessarie per i primi lavori e per le ricerche del nostro gruppo (Carpi 2017; Farina, Billero 2018; Billero, Carpi 2018; Lanini, Nicolás 2018; Farina, Frinz 2020; Garzaniti 2020), dimostrando il fondamentale apporto dell'Informatica Umanistica al progetto Lessico dei Beni Culturali.

Senza dubbio quanto realizzato finora si tratta solo di un primo passo verso il raggiungimento di altri prodotti utili alla ricerca nell'ambito della lessicogra-

⁹ Consultabile su <<http://corpora.lessicobeniculturali.net/>>.

fia dell'arte; ad esempio, in futuro dovrebbe essere introdotta la possibilità di etichettare i testi a un livello molto più avanzato rispetto a quanto fatto finora, utilizzando standard universalmente riconosciuti. In particolare, poiché i nostri corpora sono strettamente legati ai nomi propri di persona e luogo, è ipotizzabile che ISNI possa essere utilizzato applicando un'etichetta a ciascuno dei nostri nomi per raccogliere i nomi propri di tutte le varianti in tutte le lingue presenti nei corpora.

Bibliografia

- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Billero R., Carpi E. 2018, *Corpora e terminologia artistica: il caso del corpus spagnolo LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 5(1): 85-91.
- Billero R., Nicolás Martínez M.C. 2017, *Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 4(2): 203-216.
- Carpi E. 2017, *El lenguaje para fines artísticos: traducciones de tondo al español*, in Curado A. (ed.), *EPiC Series in Language and Linguistics*, 3, *LSP in Multi-disciplinary contexts of Teaching and Research. Papers from the 16th International AELFE Conference*: 79-84, <<https://doi.org/10.29007/wx3m>>.
- CWB, <<http://cwb.sourceforge.net/>>, (10/2020).
- Farina A. 2016, *Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique*, «Publif@rum», 24, <http://www.farum.it/publifarum/ezone_articles.php?art_id=335>.
- Farina A., Billero R. 2018, *Comparaison de corpus de langue «naturelle» et de langue «de traduction»: les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues* in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 108-116.
- Farina A., Flinz C. 2020, *Analisi linguistica comparativa dei corpora LBC. La visione del patrimonio fiorentino francese e tedesco: l'esempio del Duomo*, in Farina A., Funari F. (eds), *Past in Present / Le passé dans le présent / Il passato nel presente*, FUP, Firenze: 75-98.
- Garzaniti M. 2020, *Il termino russo friag e le sue radici nelle relazioni culturali e artistiche fra la Russia e l'Italia*, in Pano Alamán A., Zotti V. (eds.), *The language of art and culture heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 104-119.
- ISNI, <<http://www.isni.org/>>, (10/2020).
- Lanini L., Nicolás Martínez M.C. 2018, *Verso un dizionario corpus-based del lessico dei beni culturali: procedure di estrazione del lemmario* in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 411-418.
- NextCloud, <<https://www.nextcloud.com/>>, (10/2020).
- Numerico T., Vespignani A. 2003, *Informatica per le scienze umanistiche*, il Mulino, Bologna.

- Schmid H. 1995, *Improvements in Part-of-Speech Tagging with an Application to German* in *Proceedings of the ACL SIGDAT-Workshop*, Dublin.
- Schmid H. 1994, *Probabilistic Part-of-Speech Tagging Using Decision Trees* in *Proceedings of International Conference on New Methods in Language Processing*, Manchester.
- SketchEngine, <<https://www.sketchengine.eu/>>, (10/2020).
- Svensson P. 2010, *Landscape of Digital Humanities*, «Digital Humanities Quarterly», IV (1), <<http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>> (11/2020).
- Zotti, V., Pano Alamán A. 2017, *Introduzione*, in Zotti V., Pano Alamán A. (eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*, FUP, Firenze: 7-15.

Il *Corpus LBC español*: basi, sviluppo e applicazioni

Elena Carpi, Ana Pano Alamán

Abstract (italiano): In questo contributo si presentano i criteri per la realizzazione, le fasi di sviluppo e le possibili applicazioni del Corpus LBC in spagnolo, creato nell'ambito del progetto Lessico multilingue dei Beni Culturali (LBC). In primo luogo, si commentano brevemente le caratteristiche principali del corpus e i criteri di selezione dei testi che contiene. In secondo luogo, vengono illustrati i possibili usi del corpus da parte di diversi tipi di utenti, in aree di interesse come il turismo, la traduzione specializzata e l'insegnamento dello spagnolo per scopi specifici.

Parole chiave: linguistica dei corpora, lessico, terminologia, arte, beni culturali.

Abstract (inglese): This brief essay presents the criteria for the realization, the development phases and the possible applications of the LBC Corpus in Spanish, created in the framework of the Lexicon of Cultural Heritage (LBC) project. First, the authors briefly comment the main features of the corpus and the selection criteria of the texts that are included in the corpus. Second, they discuss and illustrate the possible uses of the corpus by different types of users and in areas of interest such as tourism, specialized translation and the teaching of Spanish for specific purposes.

Keywords: corpus linguistics, lexicon, terminology, art, cultural heritage.

Introduzione

Questo contributo presenta le principali caratteristiche e le possibili applicazioni del *Corpus LBC Español*, un corpus monolingue specialistico¹ che si propone di offrire una rappresentazione adeguata del discorso specialistico sull'arte e sul patrimonio culturale fiorentino e italiano in un campione ampio e variegato di testi, con la finalità di effettuare ricerche automatizzate e analisi approfondite, quantitative e qualitative, in particolare, sul lessico artistico. Il corpus è stato ideato e realizzato

¹ I corpora specialistici sono rappresentativi di «una variedad lingüística específica o de algún tipo de sublenguaje o lengua especializada. [...] Se crean para el estudio de la lengua usada para fines específicos y de la terminología usada en sublenguajes» (Pérez Hernández, 2002). Le loro caratteristiche sono simili a quelle dei corpora di riferimento (quantità, qualità, semplicità e documentazione), benché normalmente siano rappresentativi dello specifico dominio di studio per il quale vengono creati.

Elena Carpi, University of Pisa, Italy, elena.carpi@unipi.it, 0000-0001-7503-6137

Ana Pano Alamán, University of Bologna, Italy, ana.pano@unibo.it, 0000-0003-3652-8694

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)


Riccardo Billero, Annick Farina, María Carlota Nicolás Martínez (edited by), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, © 2020 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5870 (online), ISBN 978-88-5518-253-9 (PDF), DOI 10.36253/978-88-5518-253-9

nell'ambito del progetto interuniversitario *Lessico multilingue dei Beni Culturali*², il cui obiettivo principale è la creazione di dizionari monolingui e bilingui correlati tra loro da collegamenti ipertestuali, basati sull'estrazione di dati contenuti nei corpora delle diverse lingue descritti in questo volume. Tali strumenti lessicografici si propongono di contribuire a migliorare l'attività di autori, redattori e traduttori che lavorano con testi di tematica artistica, e di selezionare i termini più appropriati nei diversi contesti di fruizione e in tipologie discorsive come, ad esempio, guide turistiche, pagine web di musei o cataloghi di mostre. Allo stesso modo, il progetto si propone di realizzare studi comparativi che permettano di specificare l'ambito semantico e di trovare le possibili equivalenze dei termini relativi all'arte e al patrimonio culturale toscano e italiano in diverse lingue (cinese, francese, inglese, portoghese, russo, spagnolo, tedesco). Da un lato, il progetto lavora sul lessico relativo ai beni culturali della città di Firenze e ai nomi propri relativi alla sua storia e al suo patrimonio artistico e architettonico; dall'altro, affronta i termini legati alla pittura, all'architettura, alla scultura e alle altre arti che caratterizzano il patrimonio artistico e culturale in un senso più applicato e focalizzato sulle tecniche, i materiali e gli strumenti utilizzati nella creazione, conservazione e diffusione di tale patrimonio. Al fine di raggiungere i suddetti obiettivi sono stati selezionati una serie di testi rappresentativi del patrimonio artistico fiorentino per la loro integrità, rilevanza nella materia e capacità di leggere termini ed espressioni antiche in chiave moderna. Le parole e le collocazioni tratte da questi testi saranno incorporate nella struttura del dizionario LBC, con particolare attenzione al loro rapporto con le connotazioni che hanno acquisito nel tempo.

1. Corpus LBC Spagnolo

Prima di costituire un corpus di testi rappresentativi del lessico dell'arte e del patrimonio fiorentino e italiano in spagnolo, è stato necessario considerare alcune questioni relative alla finalità del corpus, alla sua dimensione, rappresentatività, autenticità ed equilibrio. In primo luogo, si è riflettuto sulla distinzione stabilita nella linguistica dei corpora tra l'approccio *corpus-based* e *corpus-driven* (Tognini-Bonelli 2001), un aspetto connesso agli obiettivi che avrebbero guidato la progettazione del corpus e l'estrazione dei dati per la preparazione del lemmaario. Il primo tipo di corpus funge da metodo di indagine e aiuta a validare ipotesi preesistenti, mentre nel secondo, adottato in questo caso, l'insieme dei testi raccolti fa parte della ricerca e può dare origine a nuove conoscenze e teorie. D'altra parte, trattandosi di un corpus specialistico, si è ipotizzato che la sua dimensione sarebbe stata inferiore a quella di un corpus generale o di riferimento, senza rinunciare comunque ad ottenere un numero di parole di lessico artistico superiore al milione; nello specifico, attualmente il corpus (20.08.2019) è formato da 1.020.872 parole (1.162.849 token) e da un totale di 207 documenti (Figura 1).

² Informazioni sul progetto: <<http://www.lessicobeniculturali.net/contenuti/il-progetto/818>>.

Corpus LBC Spagnolo 

Corpus LBC Spagnolo



Counts		General info		Lexicon sizes	
Tokens	1,162,849	Corpus description	Document	word	59,196
Words	1,020,872	Language	Spanish	tag	70
Paragraphs	14,776	Encoding	UTF-8	lemma	38,986
Documents	207	Compiled	08/27/2018 17:11:02	lc 	53,835
		Tagset	Description	lemma_lc 	36,350

Figura 1. Dati quantitativi del *Corpus LBC español*.

Consideriamo il *Corpus LBC Español* rappresentativo del lessico dell'arte³ e del patrimonio artistico fiorentino in spagnolo grazie all'equilibrio che lo caratterizza tra la quantità di testi relativi al patrimonio artistico, storico, politico e culturale della città di Firenze, e di quelli che si occupano di oggetti, materiali e tecniche principalmente legati alla pittura, architettura e scultura. Il criterio di rappresentatività è soddisfatto anche grazie alla selezione di unità testuali che riguardano periodi storici diversi, così come di autori, temi, registri e culture differenti. Infine, va sottolineato che si tratta di un corpus semiaperto, poiché prevede l'inserimento futuro di nuovi testi (Rojo 2015: 373).

1.1. Selezione testuale

Il lessico dell'arte sorge nel Medioevo nell'ambito delle *ars*, attività artigianali con una propria terminologia tecnica (*vid.* Biffi 2010). Successivamente, a partire dal XVIII secolo, per riferirsi alla pittura, all'architettura e alla scultura nasce un nuovo concetto, quello di «belle arti», che allude nello stesso tempo alla tecnica e alle diverse tipologie di attività con valore estetico (Carchia, D'Angelo 2007: 17). La genesi e l'evoluzione del linguaggio a fini artistici spiega perché le parole legate all'arte e alla critica d'arte siano lontane dal costituire un insieme di termini ed espressioni univoche (De Mauro 1965: 37). In effetti, il lessico dell'arte si colloca a metà tra i linguaggi specialistici e il linguaggio co-

³ Questo corpus costituisce una risorsa senza precedenti nel panorama spagnolo, lingua nella quale sono presenti numerosi corpora di riferimento per lo studio della lingua sia in sincronia sia in diacronia, come il CORPES XXI (Corpus del Español del Siglo XXI), il CORDE (Corpus Diacrónico del Español) o il CREA (Corpus de Referencia del Español Actual, che integra altri corpus orali), della Real Academia de la Lengua (<<https://www.rae.es/recursos/banco-de-datos>>); citiamo inoltre, per lo studio delle varietà di spagnolo, il corpus PRESEEA (Proyecto para el estudio sociolingüístico del español de España y América, <<http://pre-seea.linguas.net/>>). Per quanto riguarda i corpora di specialità, si segnala lo IULA Corpus (Institut de Lingüística Aplicada) dell'Università Pompeu Fabra, che contiene testi scritti in diverse lingue, riguardanti economia, diritto, ambiente, medicina, informatica e scienze del linguaggio (<<https://www.upf.edu/web/iula/corpus>>), el Archivo Gramatical de la Lengua Española (AGLE, <<https://cvc.cervantes.es/lengua/agle/>>) e el Léxico de la Alimentación y el Arte culinarios tradicionales (<<https://people.unil.ch/rolfeberenz/corpus/>>).

mune, e presenta un alto grado di variazione a seconda del tipo di testo in cui si trova⁴. Infatti, come segnala Carpi (2015: 112):

El lenguaje para fines artísticos (LFA) integra varios subconjuntos de lenguajes de especialidad – los de la música, las artes visuales, la arquitectura, el cine, el teatro, la fotografía, etc. – cada uno dotado de un discurso propio y caracterizado por rasgos léxicos, estilísticos y retóricos.

Anche per questo, «[...] el carácter polisémico de las unidades léxicas» (Cabré 2006: 29) si manifesta con frequenza in questa lingua specialistica, benché, in un dato contesto, esse costituiscano unità di discorso specializzato.

Sulla base di quanto detto, e tenendo conto della natura eterogenea del lessico artistico e della varietà di tipologie discorsive in cui si manifesta, prima della selezione dei testi che avrebbero costituito il *Corpus LBC Español* il gruppo di ricerca di lingua spagnola del Progetto LBC, coordinandosi con quelli delle altre lingue, ha stabilito due grandi blocchi tematici. In un primo momento sono stati raccolti testi relativi alla città di Firenze, considerata la città d'arte per eccellenza, opere originali scritte in spagnolo o tradotte dall'italiano, dal francese e dall'inglese, pubblicate tra il XVI e il XX secolo. Questi testi sono stati selezionati per la loro rilevanza culturale e interculturale: si tratta infatti di opere letterario-biografiche e di viaggio, che consentono di comprendere i diversi punti di vista legati all'arte italiana e di ampliare così la scelta terminologica (Billero, Nicolás Martínez 2017). Questa prima serie di testi comprende principalmente scritti su Firenze e la Toscana, anche se, nelle fasi successive del lavoro, si sono aggiunte opere relative all'Italia nel suo complesso. L'obiettivo di ottenere la maggior varietà testuale possibile giustifica la decisione di includere opere di autori molto diversi tra di loro: il corpus comprende scrittori spagnoli molto conosciuti come Pedro Antonio de Alarcón (*De Madrid a Nápoles*, 1861), Leandro Fernández de Moratín (*Viaje a Italia*, 1867) o Vicente Blasco Ibáñez (*En el país del arte*, 1896) e altri meno noti, quali Gregorio Iribas y Sánchez (*Viaje por Italia y Suiza*, 1897) e José de Lasa (*De Madrid al Vesubio. Viaje por Italia*, 1873), che offrono uno sguardo concreto sull'arte italiana e utilizzano una terminologia propria. Questo primo gruppo include anche le opere tradotte di scrittori come François-René de Chateaubriand (*Viaje a Italia*, 1831, trad. Mariano de Rementería) e John Ruskin (*Las mañanas de Florencia*, 1901, trad. Carmen de Burgos), che propongono una ricca e dettagliata descrizione del patrimonio artistico fiorentino e italiano, e utilizzano un lessico che combina in modo equilibrato divulgazione e specializzazione.

⁴ Nonostante la ricchezza di questa lingua specialistica e delle tipologie testuali in cui si manifesta, le analisi sistematiche sulla lingua a fini artistici in spagnolo scarseggiano (Carpi 2017). Tuttavia, vale la pena citare alcuni studi relativi alla comunicazione della critica d'arte (García Rodríguez 1999) o ai termini del disegno e della pittura (Montijano García 2010; Carpi 2015). All'interno di questa tipologia di ricerca è inoltre possibile inserire studi sul discorso del turismo in spagnolo (Calvi e Mapelli 2011), in cui il lessico dell'arte è solitamente molto presente.

In un secondo momento, sono stati raccolti testi tecnici riguardanti l'architettura e l'arte (principalmente pittura e scultura), privilegiando trattati critici e manuali in quanto fonte essenziale di terminologia specialistica. Come nel caso precedente, sono state incluse sia opere originali spagnole, come il *Tratado de arquitectura* di Alonso de Vandelvira (1977 [1591]), che contiene sia termini architettonici sia quelli che caratterizzano l'arte del taglio delle pietre da costruzione, e i *Diálogos de la pintura* di Vicente Carducho (1633), sia le opere tradotte di Leon Battista Alberti (*Los diez libros de arquitectura*, 1797 [1485]), Andrea Palladio (*Los quatro libros de arquitectura*, 1797 [1570], tradotti da J. Francisco Ortiz y Sanz) e Leonardo da Vinci (*El Tratado de la Pintura*, 1784 [1651], tradotto da D. Antonio Rejón de Silva). Il corpus comprende anche la traduzione di classici riguardanti il periodo rinascimentale a Firenze, come *Las Obras históricas de Nicolás de Maquiavelo* (1892, tradotto da Luis Navarro), e altri saggi in spagnolo dedicati ad artisti italiani, per esempio *Las vidas de los pintores y estatuarios eminentes españoles* (1744) di Palomino de Castro e *El artista en Italia* di José Galofre (1851), che comprendono numerosi antroponimi e toponimi culturalmente rilevanti per la storia dell'arte italiana.

Una menzione speciale merita l'opera *Le vite de' più eccellenti pittori, scultori, e architettori* di Giorgio Vasari, le cui edizioni, Torrentina (1550) e Giuntina (1568), occupano un posto centrale riguardo l'arte rinascimentale italiana e la sua storia. Il testo del Vasari è una fonte essenziale del corpus poiché:

le due anime delle lingue delle arti, umanistica e tecnica, convivono in Vasari, che di fronte all'opera narra la percezione visiva dell'osservatore, ma si appoggia anche alla sua competenza pratica, come emerge a livello sia lessicale sia sintattico (Biffi, 2010).

Vasari introduce nella sua opera termini che collegano il lessico tecnico e alcuni concetti che anticipano il linguaggio della critica sopra menzionato, come, rispettivamente, *disegno*, e più precisamente, le «tecniche di disegno» e *maniera*, spesso tradotto come «stile» di un artista o di un'epoca (Pano Alamán 2017). Nello specifico, il *Corpus LBC español* raccoglie le traduzioni con la maggiore diffusione editoriale, alla base di antologie o selezioni di testi: quelle del 1960, eseguite da Julio E. Payrató sulla base dell'edizione del 1568; la versione del 2011, di María Teresa Méndez Baiges e Juan María Montijano García, e la versione del 2012 di Helena Aguilà, quest'ultima basata sulla prima edizione delle *Vite* del 1550. Per il momento, questi testi sono anche la base del corpus parallelo che stiamo preparando e annotando nell'ambito del progetto LBC.

1.2. Fasi di raccolta e archiviazione dei testi

La gestione dei contenuti del corpus è stata guidata da criteri di razionalità ed efficienza (Billero, Carpi 2018), con l'archiviazione sulla piattaforma NextCloud dei testi raccolti dopo una ricerca in cataloghi e banche dati online (Billero 2020). La cartella conteneva un foglio di calcolo Excel che ha permes-

so di generare i nomi dei file di testo presenti nel corpus, facilitando il lavoro in fase di consultazione delle concordanze; il nome del file è costituito da una serie univoca di descrittori e metadati relativi ai testi, quali lingua, autore, titolo, anno di pubblicazione, ecc. (*vid.* Billero, Nicolás Martínez 2017), che consentono anche di interrogare il corpus in modo intuitivo. I testi sono stati memorizzati in Word, programma scelto per il suo uso diffuso, per l'interoperabilità che presenta all'interno del progetto LBC e per la necessità di mantenere gli elementi di base del formato, come grassetto, corsivo, note a piè di pagina, ecc. D'altra parte, il formato *.docx* è uno standard XML, elemento che ha permesso di creare un software speciale in grado di rielaborare automaticamente tutti i file e generarne di nuovi con lo stesso formato, nome e metadati. Per poter effettuare ricerche, i testi sono stati caricati automaticamente in SketchEngine.

2. Uso del *Corpus LBC español*

Questa sezione presenta sinteticamente esempi di consultazione del corpus, al fine di mostrare il suo interesse per guide, traduttori, accademici e insegnanti di spagnolo per scopi specifici (*español para fines específicos*, EFE). Le interrogazioni effettuate con il programma SketchEngine possono essere molto utili: in particolare, quella effettuata utilizzando la funzione *context* permette di cercare una determinata parola e le sue corrispondenze con altre parole appartenenti a varie categorie grammaticali, e rende disponibili lessici combinati validati dall'uso.

Come sottolinea Cabré, la terminologia rappresenta una «*actividad de recopilación e ilustración de las fórmulas detectadas en el discurso especializado*» (2004: 98); la consultazione di corpora specializzati, quindi, consente di identificare tali formule e di ottenere elenchi di unità di conoscenza con diversi gradi di lessicalizzazione. Nell'ambito della redazione di testi per il turismo, come per esempio una guida al patrimonio artistico-culturale di Firenze, è importante considerare sia le frequenti collocazioni in cui un sostantivo è associato a un toponimo o antroponimo, sia quei termini scritti in volgare toscano o nell'italiano attuale -varietà diacroniche e diatopiche- relative a personaggi e luoghi della Firenze della metà del Cinquecento. Di fronte alla possibilità di adattare o tradurre alcuni riferimenti geografici o di lasciare gran parte dei nomi in italiano senza essere in grado di valutare se tali riferimenti siano noti nella cultura di arrivo, i professionisti del settore turistico possono trovare nel corpus LBC un valido strumento per stabilire qual è la forma più comune e approfondire la conoscenza enciclopedica del termine. Ad esempio, l'antroponimo «*Médicis*» presenta 508 occorrenze nel corpus rispetto a 7 per la forma italiana «*Medici*» e 14 per la forma ibrida «*Medicis*». Sebbene queste due forme siano validate dal loro uso nei testi, la maggiore frequenza della prima, adattata allo spagnolo, indica che è la più appropriata. L'utilizzo della funzione *Word Sketch* di SketchEngine, invece, consente di estendere l'indagine iniziale: oltre alle occorrenze del cognome generalmente associate ai nomi dei familiari, è possibile recuperare collocazioni relative ad opere d'arte, come «*Venus de Médicis*»:

doc#33	obra capital del Palacio Pitti, como la Venus de Médicis lo es de la Galería degli Uffizi. </p><p> El maravilloso
doc#33	reemplazó en la Tribuna de Uffizi á la Venus de Médicis , cuando los franceses se apoderaron de esta y se
doc#33	puede compararse á la de Cleomenes. La Venus de Médicis ha vuelto á Italia, y nos espera en la tribuna de
doc#33	de aquella galería, fruto del amor de los Médicis á las bellas artes: amor que heredaron, como una
doc#33	peregrinos á esta ciudad, es la Venus de Médicis , esculpida en Atenas, cerca de dos siglos antes
doc#33	artísticos de la antigüedad. </p><p> La Venus de Médicis ,—llamada así porque la adquirió Florencia en
doc#33	desapasionadamente que son inferiores á la de Médicis : no así los franceses, que tratan de hacer creer
doc#33	. </p><p> Conviniedo, pues, en que la Venus de Médicis es la mas bella obra del arte, comprended la
doc#33	. </p><p> De dos maneras hay que considerar á la Venus de Médicis : como mujer y como escultura, ó sea como modelo y
doc#33	de nuestros ojos. </p><p> Ahora bien, la Venus de Médicis es el modelo abstracto de la hermosura femenina
doc#33	fijaría al crear á nuestra madre Eva. </p><p> La Venus de Médicis causa el mismo efecto que el cielo, que la música
doc#33	. Considerada como estatua, la Venus de Médicis es todo lo que hemos dicho; ¡todo eso... fingido
doc#33	, por el contraste que hacia con la Venus de Médicis , fue una Venus de Ticiano, toda desnuda,

Figura 2. Concordanze di «Venus de Médicis».

< .previous del arte en general; la primera maravilla del mundo, al decir de la mayoría de los críticos; la joya de Florencia; la que por sí sola atraería innumerables peregrinos á esta ciudad, es la Venus de **Médicis** , esculpida en Atenas, cerca de dos siglos antes de la venida de Jesucristo, por el célebre Cleomenes, hijo del gramático Apollodoro, y encontrada hace 300 años en Tivoli, cerca de Roma, en la villa Adriana, donde estaba sepultada bajo escombros seculares, como tantos otros prodigios artísticos de la antigüedad. </p><p> La Venus de **Médicis**, —llamada así porque la adquirió Florencia en tiempo de un gran duque de esta familia,—es (dicen los florentinos) á las demás Venus, lo que Venus era á las demás diosas. También puede < .next >

Figura 3. «Venus de Médicis» in contesto, in Alarcón, P. A. de (1861), *De Madrid a Nápoles*. Riferimento dell'archivio nel corpus: Alarc_MadNap_Tosc95_1861r_1861p.

Allo stesso modo, la ricerca sul toponimo «Firenze» permette di estrarre nomi di edifici religiosi (*Santa Croce, Santo Spirito, Santa María del Fiore*), luoghi (*plaza San Giovanni*) e talvolta di artisti poco conosciuti (Rocco Zoppo o Taddeo Gaddi), che possono fornire indizi sulla formula linguistica più rilevante e arricchire il contenuto di un testo di viaggio sul patrimonio della città (*vid.* Pano Alamán 2019). In questo senso, consultare il corpus può risultare utile anche nell'insegnamento dello spagnolo per scopi specifici (EFE) nei settori delle belle arti o del turismo, grazie alle informazioni culturali che contiene e alle possibilità di indagare i possibili significati di parole tecniche. Ad esempio, l'esplorazione del contesto in cui compare il termine *piagnone*, con due occorrenze nel corpus, ci permette di comprendere, grazie alla spiegazione fornita dal Vasari, che si tratta dei seguaci di fra' Girolamo Savonarola (Figura 4):

< .previous , y en lo cual pasó mucho tiempo, y como entre tanto no trabajaba, se produjo gran desorden en su vida. Imprimió muchos otros dibujos, pero de una calidad inferior, pues las planchas estaban mal grabadas, siendo su mejor trabajo el triunfo de la fe de Fray Girolamo Savonarola de Ferrara. Era adherente de la secta de ésta y ello le llevó a abandonar la pintura y, como no tenía renta, se vio envuelto en las más serias dificultades. Pero como permaneció obstinado en su determinación y se convirtió en un **piagnone**, como se les llamaba, abandonó el trabajo, y debido a ello llegó á ser tan pobre en su vejez, que si Lorenzo de Médicis, mientras vivió, no lo hubiera asistido, porque el artista había < .next >

Figura 4. Frammento delle *Vidas* di Vasari (vita di Botticelli), dove appare *piagnone*.

D'altra parte, la ricerca di frequenze e concordanze della parola «diseño» può contribuire ad ampliare il lessico dello studente di spagnolo per fini artistici, poiché la sua analisi nel contesto ci consente di cogliere non solo il significato più esteso del termine, come «1. m. *Traza o delineación* de un edificio o de una figura; 2. m. *Proyecto, plan* que configura algo», ma anche il più ristretto, ovvero «3. m. *Concepción original* de un objeto u obra destinados a la producción en serie» (DLE, 23ª ed., 2014, corsivo nostro), che generalmente viene associato al termine «dibujo». La consultazione comparativa di questi due termini illustra i significati di entrambi, sui quali si può riflettere in aula dal punto di vista linguistico, ad esempio in relazione a collocazioni più o meno ricorrenti («ejecutado con gracia y diseño», «diseño de la cúpula», «diseños de arquitectura», «dibujo anatómico», «dibujo del natural», «dibujo a pluma»). Inoltre, l'ana-

lisi delle concordanze dei testi del corpus riguardanti pittura e architettura può anche contribuire a una riflessione di carattere culturale.

Gli esempi citati si dimostrano utili anche nel campo della traduzione specializzata, poiché il trasferimento del contenuto di un testo specializzato da una lingua-cultura a un'altra «implica [...] el conocimiento de las formas terminológicas específicas en la lengua de partida» (Cabré 1993: 107); quando si parla di traduzione specializzata è infatti necessario non sottovalutare parametri come il genere, il registro e il meta-discorso del testo di partenza (Suau-Jiménez 2010). Come si ricordava, i testi possono essere utilizzati per identificare le opzioni di traduzione più appropriate in base sia alla loro frequenza di uso, alle collocazioni e alla rilevanza in determinati contesti linguistici, sia al testo di destinazione e al destinatario, come nel caso di termini che caratterizzano la tecnica pittorica come «pittura a tempera» (Carpi 2015). Allo stesso modo, il confronto delle traduzioni delle *Vite* contenute nel corpus permette di individuare le soluzioni adottate dai traduttori, in particolare per quanto riguarda i *realia* e i nomi propri di luoghi e persone, e di valutare, ad esempio, se sia possibile tradurre, adattare o mantenere le forme così come sono. Tornando al caso di *piagnone*, termine legato alla situazione politica e religiosa della Firenze rinascimentale⁵, le traduzioni mantengono la forma italiana del termine, conservando la forma originaria e il riferimento culturale e storico, che può essere compreso dal contesto. Questo tipo di ricerca permetterà al traduttore di accedere a tutte le possibili opzioni in fase di documentazione e di traduzione, risolvendo alcune delle carenze dei dizionari bilingui tradizionali. La prospettiva diacronica e i criteri interculturali che stanno alla base del corpus guidano il traduttore nel prendere decisioni terminologiche corrette. Grazie ai vari tipi di testo (generi, temi, carattere divulgativo o specialistico, originali, traduzioni) che compongono il corpus e per l'ampio arco temporale che coprono, ci si aspetta che questa risorsa faciliti l'accesso a termini e informazioni culturali che altrimenti non sarebbero facili da trovare.

3. Sviluppi futuri

Da quanto si è detto, risulta evidente il carattere originale del corpus che presentiamo e la sua grande importanza pratica e teorica per la ricerca sulle lingue specialistiche e sulla traduzione. D'altra parte, il *Corpus LBC Español* si propone come un progetto a lungo termine, la cui utilità risiederà nell'essere aperto e dinamico. Per questa caratteristica, consideriamo la sua pubblicazione il primo segmento di un'opera più ampia, il cui perseguimento includerà l'ampliamento del corpus al fine di massimizzare la coerenza con le altre lingue del progetto, nonché la presenza di alcuni generi sottorappresentati. Il corpus sarà ampliato

⁵ «3. Come termine storico, nome dato a Firenze, negli ultimi anni del sec. 15, ai seguaci di frate G. Savonarola, per il loro continuo sollecitare riforme politiche con l'argomento della collera divina, e per il loro voler imporre un tono di penitenza e di contrizione alla vita quotidiana». Disponibile su: <<http://www.treccani.it/vocabolario/piagnone/>>.

con traduzioni spagnole esistenti di opere francesi, tedesche e inglesi presenti negli altri corpora, e con dizionari e manuali di storia dell'arte, tipologie meno presenti in questa fase. Questo incremento consentirà di stabilire una serie di corpora paralleli, la cui importanza per gli studi di traduzione è ben nota, e che rappresenta uno degli obiettivi per i prossimi anni.

Bibliografia

- Biffi M. 2010, *Arte e critica d'arte, lingua dell'*, Enciclopedia dell'italiano, <<https://bit.ly/31XQyBw>> (08/2019).
- Billero R., Nicolás Martínez M.C. 2017, *Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 4(2): 203-216.
- Billero R., Carpi E. 2018, *Corpora e terminologia artistica: il caso del corpus spagnolo LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 5(1): 85-91.
- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Cabré M.T. 2006, *Lenguajes de especialidad y enseñanza de lenguas*, in Cabré M.T., Gómez de Enterría J. (eds.), *La enseñanza de los lenguajes de especialidad. La simulación global*, Gredos, Madrid: 9-73.
- Calvi M.V., Mapelli G. (eds.) 2011, *La lengua del turismo. Géneros discursivos y terminología*, Peter Lang, Bern.
- Carchia G., D'Angelo P. 2007, *Dizionario di estetica*, Laterza, Roma.
- Carpi E. 2015, *El español de la pintura y los recursos lexicográficos y terminológicos: cómo traducir al español tempera y guazzo*, «Cuadernos AISPI», 6: 111-126.
- 2017, *El lenguaje para fines artísticos: traducciones de tondo al español*, in Curado A. (ed.), *EPiC Series in Language and Linguistics*, 3, *LSP in Multi-disciplinary Contexts of Teaching and Research. Papers from the 16th International AELFE Conference*: 79-84, <<https://doi.org/10.29007/wx3m>>.
- De Mauro T. 1965, *Il linguaggio della critica d'arte*, Vallecchi, Firenze.
- Montijano García, J.M. 2010, *El diseño en los siglos XVI y XVII*, «I+Diseño: revista internacional de investigación, innovación y desarrollo en diseño», 2(2): 39-46.
- Pano Alamán A. 2017, *El léxico del arte y su traducción: un estudio a partir de las 'vidas' de Vasari*, in González Pastor D. M. (ed.), *La traducción de los textos de especialidad: una perspectiva multidisciplinar*, Publicacions Universitat de València, València: 195-221.
- 2019, *El léxico del patrimonio cultural florentino en blogs y reportajes de viaje en línea*, «CLAC. Círculo de Lingüística Aplicada a la Comunicación», 78: 49-62.
- Pérez Hernández M.C. 2002, *Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*, «EliEs», 18, <<http://elies.rediris.es/elies18/>>.
- Rojó G. 2015, *Hispanic corpus linguistics*, in Lacorte M. (ed.), *The Routledge Handbook of Hispanic Applied Linguistics*, Routledge, London: 371-387.
- Suau-Jiménez F. 2010, *La traducción especializada*, Arco Libros, Madrid.
- Tognini-Bonelli E. 2001, *Corpus linguistics at work*, John Benjamins, Amsterdam.

Un itinerario letterario nel corpus LBC in lingua inglese

Ilaria Natali

Abstract (italiano): Questo contributo si propone di illustrare il corpus inglese creato per LBC in un momento importante della sua evoluzione, il raggiungimento di un milione di parole (2018). Ci si sofferma, poi, sul sub-corpus letterario, al momento il più sviluppato; nonostante le opere ivi incluse siano apparentemente diverse tra loro, stabiliscono una complessa rete di connessioni intertestuali che si snoda a più livelli e si estende principalmente attorno alle *Vite* di Vasari.

Parole chiave: poesia dell'Ottocento, età Vittoriana, corpus letterario, viaggio e turismo.

Abstract (inglese): This brief essay discusses the English corpus for the LBC project at an important stage of its development, after it reached one million words (2018). Focus is especially on the literary sub-corpus in English, which is currently the richest: despite the apparent diversity of the works it comprises, most of its texts are connected to one another at various levels, interlinking into a web-like form that extends outward primarily from Vasari's *Lives*.

Keywords: Nineteenth-century poetry, Victorian age, literary corpus, travel and tourism.

1. Premessa: il corpus LBC in lingua inglese

Avvalersi del supporto di *corpora* è irrinunciabile nella compilazione dei nuovi dizionari in formato elettronico, che risultano così «far more exhaustive, authentic, and reliable than printed dictionaries compiled manually with lexical databases obtained from secondary sources» (Dash 2008: 85). Sebbene i primi *corpus-based dictionaries* abbiano fatto la propria comparsa soltanto negli anni Ottanta del ventesimo secolo, la tradizione lessicografica anglofona ha fatto ricorso a citazioni testuali da inserire in fraseologie esemplificative sin dal Settecento: fu Samuel Johnson, autore del notissimo dizionario della lingua inglese del 1755, il primo ad esaminare dati empirici, raccogliendo termini ed espressioni in uso tra il 1560 e il 1660 (Crystal 2007: 214). Oggi, corpora linguistici di notevoli dimensioni quali il *British National Corpus*, *The Bank of English* e il *Cambridge International Corpus* attraggono crescente attenzione per le loro potenzialità sia in ambito teorico, sia negli usi pratici.

Il progetto LBC, che annovera tra i propri obiettivi prioritari la creazione di un dizionario specialistico multilingue dedicato ai beni culturali dell'area fio-

Ilaria Natali, University of Florence, Italy, ilaria.natali@unifi.it, 0000-0003-4484-7994

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Riccardo Billero, Annick Farina, Maria Carlota Nicolás Martínez (edited by), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, © 2020 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5870 (online), ISBN 978-88-5518-253-9 (PDF), DOI 10.36253/978-88-5518-253-9

rentina, poggia le proprie fondamenta proprio sulla linguistica dei corpora, che apre al lessicografo o lessicologo nuove prospettive sulla lingua, oltre a permettere osservazioni più precise su collocazioni, semantica dei lemmi ed espressioni polirematiche. Centrale per il progetto è procedere a raccolte di testi trascritti in formato elettronico che offrano uno sguardo quanto più ampio possibile sul modo in cui i beni culturali fiorentini sono stati «tradotti» in parole (nonché in varie lingue) attraverso i secoli. Il corpus in lingua inglese è stato sviluppato principalmente nel corso del 2018, con l'aggiunta di numerosi nuovi testi e una revisione generale di quelli inseriti precedentemente, operazioni che hanno permesso di raggiungere la soglia indicativa minima per l'estensione dei corpora LBC, stabilita in un milione di parole. Si tratta di un punto di svolta che è opportuno considerare alla stregua di un primo passo in un processo ancora lungo, che richiederà varie fasi di revisione e soprattutto di ampliamento, vista anche l'abbondanza di risorse disponibili in lingua inglese in formato elettronico e cartaceo.

Attualmente composto di testi prodotti a partire dal sedicesimo secolo sino ad oggi, il corpus in lingua inglese è chiaramente un *sample corpus*, ancora di estensione molto limitata, che sarà presto ampliato e aggiornato introducendo nuovi testi selezionati tenendo conto non solo dell'importanza dei contenuti, ma anche di rappresentatività e bilanciamento. È vero che, come nota Francesca Bianchi, i corpora specialistici possono essere di dimensioni più contenute rispetto a quelli generalisti (2012: 34); tuttavia, trovo sempre valida la popolare affermazione spesso attribuita a Bob Mercer e Fred Jelinek, «more data is better data» (cfr. Hans 1992: 110). La rappresentatività è meglio garantita da corpora ampi e multidimensionali, dove è anche possibile incontrare «a reasonable number of hits for infrequent or rare linguistic events» (Bianchi 2012: 33); in altre parole, più informazioni contiene il corpus, più sono precise e dettagliate le osservazioni che se ne possono ricavare.

Se per alcuni studiosi l'estensione dei corpora a vari milioni di parole può compensare una scarsa differenziazione delle varietà di testo rappresentate, per altri l'ampia differenziazione delle varietà è condizione essenziale per poter generalizzare i risultati. La varietà è componente chiave per i corpora LBC, tutti articolati in una medesima struttura di sotto-corpora che, nel caso della lingua inglese, ci si propone di sviluppare in modo quantitativamente bilanciato. I sotto-corpora fanno riferimento a macro-varietà testuali quali, ad esempio, la distinzione primaria tra *fiction* e *non-fiction*; si tratta di una strutturazione non priva di difficoltà, poiché i confini tra categorie diverse si possono fare instabili, come nel caso di *Mornings in Florence* (1875) di John Ruskin, un'anticonvenzionale guida per viaggiatori sulla quale torneremo a breve. Queste difficoltà, a mio parere, possono rivelarsi un vantaggio per il corpus, poiché permettono alle varietà di testi di collocarsi in un *continuum*, sovrapponendosi e integrandosi tra loro.

Il presente contributo non intende soffermarsi sulle fasi tecniche di costruzione del corpus, dal momento che riflessioni, criteri e propositi relativi alla sua progettazione sono stati condivisi all'interno del gruppo di lavoro LBC e sono appannaggio soprattutto dei coordinatori generali del progetto. Vorrei dirigere l'attenzione, invece, sul mio ambito specifico di competenza, la letteratura, per

evidenziare come il sotto-corpus letterario LBC che sta gradualmente prendendo forma possa costituire un terreno fertile per la ricerca. D'altra parte, è solo negli ultimi decenni che l'analisi letteraria ha riconosciuto l'utilità dei corpora quali strumenti atti a rendere «concreti» agli occhi di lettori e studiosi elementi quali «linguistic evidence, style ornamentations, aesthetic choices, manners of expressions or deviations from language norms» (Sciarrino 2019: 144). *Pattern* di uso che solitamente sfuggono ad una lettura «tradizionale» del testo possono essere colti attraverso l'analisi computerizzata, seguendo metodi che presentano il vantaggio (non secondario) di avvicinare studi letterari e linguistici. La ricerca lessicografica su corpus testuale permette di evidenziare correlazioni semantiche, disseminazione, costellazioni lessicali che svelano le linee di tensione annidate nelle pieghe del linguaggio.

Al momento, il sotto-corpus letterario LBC include principalmente testi anglofoni ottocenteschi. Il motivo di questo sbilanciamento temporaneo su un periodo specifico è legato alla reperibilità dei testi stessi e ha innanzi tutto radici storico-culturali: a partire dal 1815, per una combinazione di fattori economici, sociali e politici, fioriscono le opere in lingua inglese che affrontano il tema del viaggio o ne narrano le esperienze. In Italia, sostanziali miglioramenti nei sistemi di trasporto favoriscono l'attrattiva turistica e, con la fine delle guerre napoleoniche, il paese è accessibile non più soltanto agli aristocratici impegnati nel Grand Tour, ma anche ai viaggiatori della media borghesia. Molti visitatori, rientrati in patria, narrano le proprie esperienze italiane, come nota John Pemble, in «Sketches, Notes, Diaries, Gleanings, Glimpses, Impressions, Pictures, Narratives, and Leaves from Journals about Tours, Visits, Wanderings, Residences, Rambles, and Travels» (1987: 7). L'Italia diviene, così, «a major inspiration of Victorian writers» (*Ibid.*), e tra questi scrittori figurano, per ricordare solo alcuni tra i più noti, Robert Browning, Elizabeth Barrett Browning, Charles Dickens, Benjamin Disraeli, George Eliot, George Gissing, George Meredith, Algernon Swinburne, William Thackeray e Anthony Trollope.

Le opere raccolte sino ad oggi nel corpus in inglese mettono in luce quanto centrale per la cultura europea sia il testo identificato come punto di riferimento comune a tutti i gruppi di lavoro LBC, *Le vite de' più eccellenti pittori, scultori, e architettori* di Giorgio Vasari. Pubblicato per la prima volta nel 1550, il libro fu rivisto e riedito con ampliamenti nel 1568 presso il tipografo fiorentino Jacopo Giunti (inclusa nel corpus è la traduzione inglese di quest'ultima edizione, detta «la Giuntina»). *Le Vite*, ben nota raccolta di biografie, offre non solo informazioni dettagliate su opere d'arte e artisti dal tardo medioevo al tempo di Vasari, ma anche opinioni riguardo alla buona pratica dell'arte e, come d'obbligo tra i biografi dell'epoca, qualche gustoso aneddoto. Pur non essendo stato concepito o strutturato attorno alle *Vite*, allo stato attuale il sotto-corpus letterario sembra ruotare a perno proprio attorno a questo testo, poiché a partire da esso rivela una rete di interconnessioni e intersezioni (in parte inedite) tra le opere che lo compongono. È proprio a tale rete inaspettata che vorrei dedicare le pagine seguenti, ricostruendone i nodi e ripercorrendone i fili attraverso un breve itinerario.

2. Un punto di partenza

Mentre la seconda edizione del testo era in preparazione, Vasari discusse le *Vite* con l'amico Vincenzo Maria Borghini, noto storico e filologo fiorentino. In una lettera datata 14 agosto 1564, Borghini commentò:

Il FINE di questa vostra fatica non è di scrivere la vita de' pittori, né di chi furono figl[i]uoli, né quello che e' feciono d'ationj ordinarie; ma solo per le OPERE loro di pittori, scultori, architetti: che altrimenti poco importa a noi saper la vita di Baccio d'Agnolo o del Puntormo. E lo scriver le vite, è solo di principi et huominj che habbino esercitato cose da principi et non di persone basse, ma solo qui havete per fine l'arte et l'opere di lor mano (Vasari 1930: 102).

Pare che Vasari abbia finito col seguire i consigli di Borghini, poiché nella seconda edizione delle *Vite* afferma di voler muovere oltre la visione atomica o frammentata di azioni ed esperienze del singolo, per andare alla ricerca del processo invisibile dell'arte o di una *historia* universale, in cui l'individuo è pensato come un punto d'incontro tra la parte e il tutto. È così che Vasari stabilisce un nuovo modo di guardare all'arte, secondo il quale ciascun artista diviene «impregnato» di storia, e la sua rilevanza nel quadro epocale è data non tanto dal genio o dal talento quanto da una congiunzione di fattori, non ultimo il processo evolutivo che il critico ricostruisce a posteriori.

Non è superfluo ricordare, inoltre, che le *Vite* ha esercitato un'influenza secolare sul modo in cui è stato percepito e descritto il Rinascimento italiano. Questa influenza, come appena menzionato, permea i primi testi inclusi nel sotto-corpus letterario del progetto LBC, tra i quali figurano varie opere di Robert Browning. Egli scoprì le *Vite* nel 1846, nel corso dei sei mesi trascorsi a Pisa con la moglie Elizabeth Barrett: la coppia abitava vicino a Piazza dei Miracoli, in una casa progettata proprio da Vasari. In quel periodo, come testimonia Elizabeth in una lettera scritta il 13 aprile 1853 all'amica Julia Martin, entrambi erano «fond of digging at Vasari» (Browning 2007: 294), amavano scavare negli interstizi delle *Vite*, consultando l'edizione di Gaetano Milanesi pubblicata tra il 1846 e il 1857. Robert usò il testo come una sorta di guida sia per visitare e ammirare il patrimonio artistico toscano, sia per scrivere le sue opere di ambientazione italiana.

Il monologo drammatico *Fra Lippo Lippi* (1855)¹ mette bene in luce come Browning abbia acquisito familiarità con l'opera e la vita del pittore e frate carmelitano Fra Filippo di Tommaso Lippi principalmente attraverso il testo di Vasari, che dell'artista scrive:

¹ Questa e tutte le opere menzionate di seguito sono incluse nel sotto-corpus letterario in lingua inglese del progetto LBC. Le edizioni citate nel presente contributo spesso non sono le stesse di quelle trascritte in formato elettronico per il corpus: qui si sono privilegiate edizioni critiche recenti, provviste di apparato critico aggiornato, mentre nel corpus la scelta delle edizioni è spesso legata a questioni di accessibilità.

Dicesi ch'era tanto venereo che vedendo donne che gli piacessero, se le poteva avere, ogni sua facultà donato le avrebbe: e non potendo per via di mezzi, ritraendole in pittura, con ragionamenti la fiamma del suo amore intiepidiva. Et era tanto perduto dietro a questo appetito, che all'opere prese da lui quando era di questo umore poco o nulla attendeva (Vasari 1971: 331).

Per mostrare questi lati del carattere di Lippi, Vasari prosegue narrando la storia della sua fuga dall'abitazione di Cosimo de' Medici in preda a «furore amoroso» per indulgere «a' suoi piaceri» (*Ibid.*), momento che fa da sfondo a *Fra Lippo Lippi*. Tanto nelle *Vite* quanto nei versi di Browning, Lippi è una figura sfaccettata e complessa, caratterizzata da una passionalità o esuberanza creativa che entra costantemente in conflitto con la disciplina richiesta dalla vita monastica.

Fra Lippo Lippi abbonda di descrizioni di elementi pittorici, allusioni ad opere d'arte e menzioni di luoghi d'interesse fiorentini, con particolare attenzione agli edifici associati alla storia della famiglia Medici. Ad esempio, «the house that caps the corner» (v. 18, Browning 2013: 485) è Palazzo Medici Riccardi, da dove i Medici governavano la città, «St. Laurence» (v. 67, 488) è la chiesa di San Lorenzo, e il «convent» che si trova «over the bridge» (vv. 90-91, 490), ovvero che sorge nell'Oltrarno, è la Chiesa di Santa Maria del Carmine, dove Lippi completò alcuni dei suoi primi affreschi degni di nota. Al verso 73, «Jerome knocking at his poor old breast» (489) sembra riferirsi all'*Adorazione del Bambino con i Santi Gerolamo, Maddalena e Ilarione*, oggi conservato presso la Galleria degli Uffizi, mentre quel «Something» che Lippo promette di dipingere nella Chiesa di S. Ambrogio con «God in the midst, Madonna and her babe» (vv. 346-48, 503-4) potrebbe essere l'*Incoronazione Maringhi*, anch'essa oggi agli Uffizi.

Assieme a le *Vite* di Vasari, una preziosa risorsa per la produzione poetica di Browning fu la stessa città di Firenze, dove lo scrittore visse con Elizabeth Barrett dal 1847 al 1861. È nota la residenza fiorentina dei coniugi, Casa Guidi, un edificio del quindicesimo secolo sito in Piazza San Felice (vicino a Palazzo Pitti), che è oggi un museo. Una volta stabilitosi a Firenze, Browning iniziò a documentarsi avidamente sulla storia artistica e culturale della città, un interesse che trovò massima espressione in *Old Pictures in Florence*. Questa poesia, concepita quasi in imitazione delle *Vite*, tratta di una serie di artisti vissuti tra il dodicesimo e il sedicesimo secolo: Browning menziona Giotto (vv. 15, 17, 24, 133, 233, 287; 2013: 410-34), «The Michaels and Rafaels» (v. 51, 413) e Leonardo da Vinci (v. 64, 414), parla di «Bigordi» (conosciuto anche come Domenico Ghirlandaio, v. 201, 425), «Sandro» (Botticelli, v. 202, *Ibid.*), Alesso Baldovinetti (v. 215, 426) e vari altri. L'attenzione, tuttavia, non è concentrata solo sulle figure più conosciute della storia dell'arte italiana: in *Old Pictures in Florence* trovano spazio anche artisti meno noti quali Niccolò Delli (detto «Dello», v. 64, 414), Nicola Pisano (v. 179, 423), Lorenzo Monaco (v. 208, 425) e Margaritone o Margheritone d'Arezzo (v. 217, 426). Secondo Browning, «A younger [artist] succeeds to an elder brother, / Da Vincis derive in good time from Dello» (v. 64, 414): a gettare le basi per la realizzazione delle grandi opere del sedicesimo secolo sono stati i contributi di personaggi oggi considerati minori, ma non meno significa-

tivi e degni di attenzione rispetto a Leonardo da Vinci. Il poeta offre, quindi, la propria interpretazione della storia dell'arte vasariana, che pare concepire come una frammentazione atta a riprodurre la pluralità del passato in un mosaico di singole figure; Browning cerca di compensare ogni lacuna storica, di mettere a nudo anche la calce che lega i singoli mattoni e che dà forma al tutto.

Secondo Dante Gabriel Rossetti, la competenza che Browning sviluppò nella storia dell'arte italiana durante il suo soggiorno fiorentino fu «superior to that of anyone he had ever met, including Ruskin» (Korg 1983: 100). Il termine del confronto scelto da Rossetti non è casuale: l'opera di Browning e quella di John Ruskin sono strettamente connesse, poiché i due strinsero amicizia e si influenzarono reciprocamente. È naturale che il critico d'arte Ruskin conoscesse molto bene le *Vite* di Vasari; ne aveva persino disapprovato alcuni aspetti ma, dopotutto, disapprovava anche la maggior parte delle opere pubblicate dai propri contemporanei. Di certo, con Browning condivideva una poetica del frammento: difendeva il valore estetico della rovina e, nelle sue indagini della storia architettonica, raramente discuteva degli edifici nella loro interezza (cfr. Chatterjee 2018: 125). Tuttavia, la sua visione del passato artistico pare molto meno democratica di quella di Browning in *Old Pictures in Florence*, e ruota costantemente attorno alla figura di Giotto, per Ruskin uno dei massimi esponenti della pittura naturalistica europea. L'analisi lessicale del testo *Mornings in Florence*, ad esempio, rivela che «Giotto» è la parola piena più frequente, con 143 occorrenze; la seconda parola piena, in ordine decrescente di frequenza, è «work», con sole 95 occorrenze su 43.427 *token* totali.

Mornings in Florence, come preannuncia il titolo, descrive in modo anticonvenzionale sei itinerari mattutini nella città di Firenze, offuscando a tratti i confini tra guida turistica, *fiction* letteraria e diario, ma mantenendo sempre un tono ironico e provocatorio. Nel parlare delle opere del Ghirlandaio in Santa Maria Novella, ad esempio, Ruskin suggerisce ai lettori: «Look at the Madonna, carefully. You will find she is not the least meek – only stupid, – as all the other women in the picture are» (1904: 26). L'originalità della pratica compositiva si riscontra anche negli appelli al lettore, frequenti e piuttosto perentori:

Get this little bit of geography, and architectural fact, well into your mind. There is the little octagon Baptistery in the middle; here, ten minutes' walk east of it, the Franciscan church of the Holy Cross; there, five minutes' walk west of it, the Dominican church of St. Mary (1904: 5-6).

Tanto la visione dell'arte di Vasari (con la mediazione di Browning) quanto quella di Ruskin sono state «a major influence» per altri due testi inclusi nel sotto-corpus letterario LBC, *The Portrait of a Lady* di Henry James (1881; Luckhurst 2009: xiv) e *A Room with a View* di E. M. Forster (1908; Summers 1987: 170-173). In particolare, Forster sembra giocare con il carattere illusorio dell'arte di Giotto così come spiegato da Vasari, il quale sottolinea non solo l'abilità del pittore nel creare effetti ottici, ma anche la sua propensione ad usare questo talento per giocare degli scherzi. In modo simile, nel secondo capitolo di *A Room with a View*, l'*Ascensione di san Giovanni* di Giotto si fa una sorta specchio

deformante che riflette o svela concezione del mondo, vita interiore e rapporto con l'arte di Lucy Honeychurch, Mr. Emerson e suo figlio George, mentre visitano la Cappella Peruzzi nella basilica di Santa Croce a Firenze.

È nuovamente il Giotto di Vasari, o meglio la sua descrizione della torre campanaria di Santa Maria del Fiore, a riecheggiare in *Casa Guidi Windows* (1851) di Elizabeth Barrett Browning. *Le Vite* specifica che la costruzione del campanile si interrompe con la morte di Giotto:

Doveva questo campanile, secondo il modello di Giotto, avere per finimento, sopra quello che si vede, una punta ovvero piramide quadra alta braccia cinquanta, ma per essere cosa tedesca e di maniera vecchia, gli architettori moderni non hanno mai se non consigliato che non si faccia, parendo che stia meglio così (Vasari 1967: 115).

In *Casa Guidi Windows*, Barrett Browning paragona il «Campanile» di Giotto a un «unperplex'd / Fine question Heaven-ward» (vv. 69-70), suggerendo che la sua forma incompleta sia un tributo alle aspirazioni umane e, quindi, conviene con Vasari sul fatto che «the tower is best as it stands, representing the beauty and openness of the future» (Pollock 2016: 168).

La prospettiva di Elizabeth su Firenze è diversa da quella del marito Robert; mentre lui si dedica principalmente alla storia culturale della città, lei si occupa innanzitutto della politica italiana del tempo. Ne è testimonianza proprio *Casa Guidi Windows*, che, pur ricca di riferimenti all'arte e alla letteratura italiana, esprime anche impressioni soggettive della parlante sulla situazione della Toscana, di cui è testimone diretta ma al contempo «distante», poiché osserva tutto dalla finestra della propria abitazione. La particolarità di questa posizione di soglia tra visione cosmica e domestica è affrontata dalla stessa Barrett Browning nell'*Advertisement* che precede la prima edizione del testo:

This poem contains the impressions of the writer upon events in Tuscany of which she was a witness. 'From a window', the critic may demur. She bows to the objection in the very title of her work. No continuous narrative nor exposition of political philosophy is attempted by her. It is a simple story of personal impressions, whose only value is in the intensity with which they were received, as proving her warm affection for a beautiful and unfortunate country, and the sincerity with which they are related, as indicating her own good faith and freedom from partisanship (Barrett Browning 2009: 237)

È l'opposto della «freedom from partisanship», qui dichiarata a mero scopo retorico-formale, a caratterizzare *Casa Guidi Windows*: l'osservatrice delinea i contorni di una nuova Italia per cui parteggia, un'Italia che prende forma in mezzo a chiese, piazze o monumenti, e trova espressione «'Twixt church and palace of a Florence street!» (v. 10, 2009: 238), in una perfetta commistione di passato e presente. Interessante notare che, benché tematicamente improntato su questioni politico-sociali, il testo di Barrett Browning è profondamente intriso di lessico artistico. Ben l'11,5% delle parole impiegate in *Casa Guidi Windows* sono indicate dall'*Oxford English Dictionary* come afferenti al dominio dell'arte, con

un divario non certo significativo rispetto al 13,1% di *Old Pictures in Florence* e al 12,5% di *Mornings in Florence*. In altre parole, pur non dedicandosi espressamente all'osservazione del patrimonio culturale fiorentino, Barrett Browning descrive i volti di donne e bambini, la processione per il Granduca Leopoldo II di Toscana e le tombe degli italiani come se fossero opere d'arte, fissandole in una eternità statica e quasi statuaria.

Simile è la procedura adottata da Barrett Browning in altre poesie, quali *A Tale of Villafranca*, *The Dance*, e *A Child's Grave at Florence*, dove il lessico «artistico» occupa mediamente il 12% dei testi. *A Tale of Villafranca* tratta della reazione fiorentina all'armistizio di Villafranca di Verona, firmato l'11 luglio 1859, che pose fine alla Seconda guerra di indipendenza e ritardò il progetto dell'Unità italiana. Un barlume di speranza riluce nella poesia *The Dance*, in cui il «piazzone» (v. 6) presso «the Cascine» (v. 1, 2009: 277) offre lo sfondo per inediti e inaspettati momenti di tregua, se non di amicizia, tra i fiorentini e gli «invasori» francesi. Immagini di dolore e rassegnazione tornano in modo ancor più marcato in *A Child's Grave at Florence*, i cui versi conclusivi, dove un angelo giunge ad anticipare una morte imminente, rovesciano il significato convenzionale dell'Annunciazione:

Some smiling angel close shall stand
 In old Correggio's fashion,
 And bear a LILY in his hand,
 For death's ANNUNCIATION
 (vv. 129-132; Barrett Browning 1849: 1304).

Come è evidente, i temi portanti dei versi fiorentini di Barrett Browning sono senz'altro vari, ma tutti finiscono per essere inevitabilmente pervasi di riferimenti alla storia dell'arte italiana: l'*Annunciazione* di Correggio (ca. 1524-1525) è un affresco staccato ora conservato nella Galleria nazionale di Parma. Tuttavia, l'angelo del Correggio non porta «a LILY in his hand», come nel tema iconografico dell'Annunciazione tipico del quattordicesimo e quindicesimo secolo: nella seconda parte della strofa di *A Child's Grave at Florence*, sembra che Barrett Browning abbia in mente, piuttosto, un dipinto sul tipo dell'*Annunciazione* di Leonardo da Vinci (ca. 1472-1475), oggi conservato alla Galleria degli Uffizi di Firenze.

3. Un punto d'approdo

Il *travelogue* in versi stava ancora riscuotendo un discreto successo di lettori quando Arthur Hugh Clough, nel suo *Recent English Poetry* del 1853, iniziò a mettere in guardia i propri colleghi poeti da un cambiamento nei gusti del pubblico, che sempre più spesso pareva orientarsi verso il romanzo (1964: 144). Ora non giova più, osserva Clough, affidarsi alla tradizione dei «poems after classical models», e un resoconto poetico di viaggio può attirare lettori solo se ha a che fare con gli «ordinary feelings» della vita di tutti i giorni (*Ibid.*) È con lo stesso

spirito che lo scrittore aveva già composto *Amours de Voyage* (1849; pubblicato nel 1858), in cui rivedeva i canoni del diario di viaggio tradizionale. Come Barrett Browning, Clough sceglie di dare un'impronta politica alle proprie opere, anziché soffermarsi sugli splendori dell'arte italiana, relegata a fare da sfondo agli eventi del Risorgimento:

*Is it Florence we follow, or are we to tarry yet longer,
E'en amid clamour of arms, here in the city of old,
Seeking from clamour of arms in the Past and the Arts to be hidden,
Vainly 'mid Arts and the Past seeking one life to forget?*
(vv. 341-344; Clough 2003: 154, corsivo originale)

Il romanzo epistolare in versi *Amours de Voyage* è suddiviso in cinque canti (o capitoli) dedicati alle vicende di un gruppo di viaggiatori inglesi coinvolto nei tumulti politici italiani del 1849. Non c'è dubbio che l'esperienza personale di Clough sia stata d'ispirazione all'opera: anch'egli, nella primavera dello stesso anno, intraprese un tormentato viaggio in Europa. Pare che Giuseppe Mazzini, allora a capo della Repubblica romana e alle prese con i conflitti con la Francia, fosse improvvisamente preso d'assedio anche su un diverso fronte: uno sconosciuto poeta inglese insisteva per ottenere da lui il permesso firmato per visitare i Musei vaticani. Tempo dopo, Clough confessò di essersi sentito in imbarazzo di «bother the Dictator [...] with my trivial English-tourist importunities» (1957: 257). Si tratta forse di un imbarazzo provato a posteriori, maturato quando il poeta ha iniziato a domandarsi se ammirare il patrimonio artistico italiano fosse occupazione consona ad un momento tanto cruciale della storia europea. È da questa domanda che nasce la sua opera più conosciuta.

Clough è morto nel 1861 a Firenze, dove si era ritirato per problemi di salute, ma a Firenze è sempre rimasto uno «straniero», così come nelle altre sue tappe italiane ed esperienze all'estero. La sua opera è lontana dall'appassionato coinvolgimento politico di Barrett Browning, e *Amours de Voyage* sembra tradire un senso di perplessità e incertezza nei confronti dei progetti di unificazione della penisola, piuttosto che genuina partecipazione. In rapporto analogo con l'Italia si è posta anche George Eliot (Mary Ann Evans) che, a differenza dei Browning, non è mai stata «a casa» a Firenze. Vi trascorse solo sei settimane nel 1861, raccogliendo il materiale che poi sarebbe confluito in *Romola* (1863), oggi forse l'opera meno conosciuta dell'autrice. Questo romanzo storico è ambientato a Firenze tra il 1492 e il 1498, due date ricche di significato: la prima coincide con la morte di Lorenzo de' Medici, la seconda con quella di Savonarola. Si tratta di anni segnati da instabilità politica, che poi condussero alle cosiddette Guerre d'Italia (1494-1559); è evidente che Eliot percepì delle similitudini tra la situazione dell'Italia del Quattrocento e quella dell'Ottocento, epoca delle guerre d'indipendenza (1848-1870). *Romola* stabilisce connessioni tra passato e presente proprio attraverso il patrimonio culturale fiorentino sin dalle sue prime pagine:

The Loggia de' Cerchi stood in the heart of old Florence, within a labyrinth of narrow streets behind the Badia, now rarely threaded by the stranger, unless in a

dubious search for a certain severely-simple door-place, bearing this inscription:
 QUI NACQUE IL DIVINO POETA (Eliot 1863: 9).

Il presente della storia è, come abbiamo detto, il 1492; il «now» della narrazione, invece, è verosimilmente il tempo di Eliot, quando la Loggia de' Cerchi ha cessato di essere un'area particolarmente trafficata della città. La connessione inattesa tra il prima e il dopo suggerisce al lettore che, nonostante i cambiamenti, la storia che sta per leggere è rilevante anche per la propria epoca. In effetti, sono vari gli studiosi che leggono il romanzo come un'allegoria non solo della situazione italiana contemporanea ad Eliot, ma soprattutto dell'era vittoriana inglese nel suo complesso (cfr. Morse 1993: 351).

Riprendo qui un filo momentaneamente spezzato del discorso: non susciterà sorpresa, a questo punto, apprendere che anche il romanzo *Romola* possa dirsi profondamente indebitato alle *Vite*. William Sullivan nota che Eliot lesse l'opera di Vasari nel 1861, probabilmente nel corso della propria permanenza a Firenze, e ne trasse spunti importanti per delineare il personaggio di Piero di Cosimo (1972: 391). Secondo Vasari, il Piero storico «teneva una vita da uomo più tosto bestiale che umano», era «capriccioso e di stravagante invenzione» ma «non [...] piacevole per la sua vaghezza» (1976: 61-63). Questa figura dal carattere impossibile, costantemente in preda a manie e paure, dovette affascinare Eliot, che parimenti descrive il «suo» Piero come un uomo «freakish», «whimsical» e «crusty» (1863: 21, 43, 117). Tuttavia, nel romanzo queste caratteristiche vanno a costituire solo l'apparenza esteriore, una patina superficiale cinica e scontrosa del personaggio, dietro la quale si celano grande lealtà e onestà d'animo, assieme ad altri tratti che sono pura invenzione dell'autrice. Eliot si appropria della figura storica di Piero di Cosimo, la rilegge e la trasforma, poiché il suo scopo non è tanto restare fedele ai fatti del passato, quanto strumentalizzarli a fini morali. È a questo proposito che *Romola* trova dei punti di contatto significativi anche con il pensiero di Ruskin: nel Rinascimento italiano quest'ultimo vedeva l'inizio di un declino morale dovuto agli eccessi, ivi inclusi «over-refinement» e «over-luxuriance» (Ruskin 1979: 4). In chiave simile, Eliot enfatizza la corruzione, dissolutezza e amoralità di un periodo e un luogo in cui «lust and obscenity, lying and treachery, oppression and murder, were pleasant, useful, and, when properly managed, not dangerous» e «sublime artists were at hand to paint the holy and the unclean with impartial skill» (1863: 95).

La rete di interconnessioni tra testi del sotto-corpus letterario, di cui si sono discussi qui solo alcuni nodi fondamentali, si estende anche oltre, e si amplierà ulteriormente nell'immediato futuro con l'inserimento di nuovi testi: ad esempio, *Handbook for Travellers in Northern Italy, Part II: Parma, Piacenza, Modena, Florence, Pisa, Lucca and Tuscany* di John Murray, spesso citato (e criticato) da Ruskin, e le opere su Firenze di Bernard Berenson, dalle quali E.M. Forster molto ha tratto per la composizione di *A Room with a View*. Per quanto ancora in fase di sviluppo, il sotto-corpus letterario non è solo utile per l'indagine linguistica, ma anche per quella storica e culturale: racconta soprattutto di quel secolo che ha visto un numero di residenti stranieri a Firenze mai eguagliato prima nella storia della città.

Bibliografia

- Barrett Browning E. 1849, *A Child's Grave at Florence*, «The Athenaeum», 1156: 1304.
 — 2009, *Elizabeth Barrett Browning: Selected Poems*, Broadview Press, Peterborough.
- Bianchi F. 2012, *Culture, Corpora and Semantics. Methodological Issues in Using Elicited Data and Corpus Data for Cultural Comparison*, Università del Salento, Lecce.
- Browning R. 2007, *The Poems of Robert Browning. Volume Three: 1847-1861*, Pearson, Harlow.
 — 2013, *Robert Browning: Selected Poems*, Routledge, London-New York.
- Chatterjee A. 2018, *John Ruskin and the Fabric of Architecture*, Routledge, London-New York.
- Clough A.H. 1957, *Correspondence of Arthur Hugh Clough, vol. I*, Clarendon Press, Oxford.
 — 1964, *Selected Prose Works of Arthur Hugh Clough*, University of Alabama Press, Alabama.
 — 2003, *Arthur Hugh Clough: Selected Poems*, Routledge, New York.
- Crystal D. 2007, *How Language Works*, Avery, New York.
- Dash N.S. 2008, *Corpus Linguistics: An Introduction*, Pearson Longman, New Delhi.
- Eliot G. (Mary Ann Evans) 1863, *Romola*, Harper & Brothers, New York.
- Hans P. 1992, *Computational Analysis and Definitional Structure*, «Lexicographica», 8: 100-129.
- Korg J. 1983, *Browning and Italy*, Ohio University Press, Athens.
- Luckhurst R. 2009, *Introduction*, in James H., *The Portrait of a Lady*, Oxford University Press, Oxford: vii-xxvi.
- Morse D. 1993, *High Victorian Culture*, Palgrave Macmillan, London.
- Pemle J. 1987, *The Mediterranean Passion: Victorians and Edwardians in the South*, Clarendon Press, Oxford.
- Pollock M.S. 2016, *Elizabeth Barrett and Robert Browning: A Creative Partnership*, Routledge, Londra e New York.
- Ruskin J. 1904, *Mornings in Florence: Being Simple Studies of Christian Art for English Travellers*, George Allen, London.
 — 1979, *The Stones of Venice*, vol. 3, Garland Publishing, New York-London.
- Sciarrino C. 2019, *Corpus Stylistics and A Portrait of the Artist as a Young Man: (De-) constructing the Spirit of the Soul*, «Joyce Studies in Italy», 21: 143-158.
- Sullivan W.J. 1972, *Piero di Cosimo and the Higher Primitivism in Romola*, «Nineteenth-Century Fiction», 26, 4: 390-405.
- Summers C.J. 1987, *The Meaningful Ambiguity of Giotto in A Room with a View*, «English Literature in Transition, 1880-1920», 30 (2): 165-176.
- Vasari G. 1930, *Der literarische Nachlass Giorgio Vasaris, Band II*, in Frey K., Frey H.W. (eds.), *Georg Müller*, München.
 — 1967, *Le vite de' più eccellenti pittori, scultori e architettori nelle redazioni del 1550 e 1568*, vol. 2, Sansoni, Firenze.
 — 1971, *Le vite de' più eccellenti pittori, scultori e architettori nelle redazioni del 1550 e 1568*, vol. 3, Sansoni, Firenze.
 — 1976, *Le vite de' più eccellenti pittori, scultori e architettori nelle redazioni del 1550 e 1568*, vol. 4, Sansoni, Firenze.

Il corpus LBC tedesco: costruzione e possibili applicazioni

Sabrina Ballestracci, Claudia Buffagni, Carolina Flinz¹

Abstract (Italiano): In questo contributo si presentano le caratteristiche principali e le possibili applicazioni del Corpus LBC tedesco, creato nell'ambito del progetto *Lessico multilingue dei Beni Culturali* (LBC). Il lavoro si compone di tre paragrafi: il primo paragrafo disamina i generi testuali, gli autori e i titoli presenti nel corpus, illustra i criteri di selezione adottati per la loro scelta e propone una valutazione dell'attuale configurazione del corpus. Successivamente (par. 2) ci si sofferma sui possibili usi del corpus, mettendo in evidenza come esso si presti ad analisi di tipo sia intralinguistico sia interlinguistico, applicabili sia in ambito didattico sia in contesti professionali. Chiudono il lavoro un riassunto e uno sguardo sulle prospettive di sviluppo futuro (par 3).

Parole chiave: linguistica dei corpora, didattica DaF, lessico, terminologia, arte, beni culturali.

Abstract (English): This essay describes the main features of the German LBC corpus, which was compiled within the research project *Lessico multilingue dei Beni Culturali* (LBC) and illustrates its possible applications. The paper is divided into three sections: Section 1 illustrates the German LBC corpus, paying special attention to the selected text genres and authors and their selection criteria; it also proposes an evaluation of the current corpus configuration. Section 2 presents various possible applications of the corpus: The proposed applications are both intralingual and interlingual. Conclusion and outlook (Section 3) conclude the work.

Keywords: Corpus Linguistics, German as Foreign Language Teaching, Lexicon, Terminology, Art, Cultural Heritage.

1. Il corpus tedesco di LBC. Costruzione

La piattaforma del corpus LBC (<<http://corpora.lessicobeniculturali.net>>) è una piattaforma multilingue specializzata composta da corpora monolingui in più lingue (tedesco, inglese, francese, italiano, russo, spagnolo), contenente

¹ Il presente contributo è la traduzione di un più ampio saggio in tedesco dal titolo *Das deutsche LBC-Korpus: Zusammenstellung und Anwendung*, pubblicato in Ballestracci, Buffagni, Flinz 2020. Autrice della traduzione è Sara Congregati. Come il saggio in tedesco, così anche il presente contributo è frutto della concezione comune delle tre autrici, che ne hanno discusso

sia testi originali sia traduzioni (cfr. Farina, Garzaniti 2013; Farina 2016; Billero, Martínez 2017; Billero, 2020; cfr. anche Figura 1):

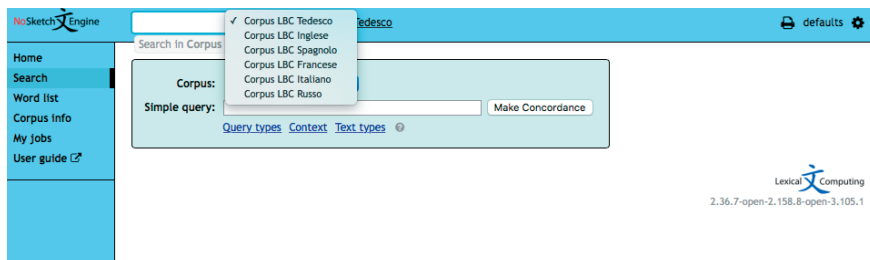


Figura 1. Screenshot dei corpora LBC nell'ambiente NoSketchEngine.

Scopo principale del progetto LBC è la redazione di un dizionario multilingue in rete, che possa anche servire da fonte per ricerche scientifiche (linguistiche, letterarie, culturali ecc.), offrendo al largo pubblico testi per una migliore conoscenza dei Beni culturali italiani. La banca dati, dunque, non solo è funzionale alla creazione di un dizionario da parte dell'unità di ricerca LBC, ma si rivolge anche a utenti esterni, come esperti del settore in storia dell'arte, italianistica ecc., traduttori professionisti, e anche insegnanti di traduzione specialistica.

È a partire da queste premesse che i singoli corpora LBC mirano a illustrare il lessico dei Beni culturali italiani. Per questa ragione si è posta fin dall'inizio particolare attenzione sulla lingua del Rinascimento e dell'arte fiorentina, nonché sulle sue opere e i suoi autori famosi (cfr. paragrafo 1.1 e paragrafo 1.2).

1.1 Generi testuali

Il corpus tedesco di LBC – come gli altri corpora LBC – si compone sia di testi in lingua originale tedesca sia di testi tradotti da altre lingue del progetto (italiano, francese e inglese²). Di questi testi alcuni erano disponibili in versione cartacea nelle nostre biblioteche universitarie o private, altri sono stati scaricati da archivi testuali online. Delle fonti online impiegate fanno parte rinomati archivi testuali come, per esempio, il *Deutsches Textarchiv* (<[in dettaglio le singole parti. Sabrina Ballestracci ha curato in particolare i paragrafi 1.1 e 1.2, Claudia Buffagni il paragrafo 1.3, Carolina Flinz il paragrafo 2. Alla stesura del paragrafo 3 e del cappello del paragrafo 1 hanno collaborato tutte e tre le autrici. Desideriamo qui ringraziare Katharina Müller e Anna Nissen per gli utili spunti di riflessione e i suggerimenti offerti nel corso di uno scambio proficuo.](http://www.deutsche-</p>
</div>
<div data-bbox=)

² Fanno parte dell'intero progetto anche altre lingue straniere, in particolare lo spagnolo e il russo; il corpus tedesco di LBC già caricato online non contiene tuttavia testi tradotti da queste lingue (cfr. anche fig. 6 nel paragrafo 1.2). Un ampliamento del corpus in tal senso sarà uno degli obiettivi delle prossime fasi di lavoro (cfr. paragrafo 1.3 e paragrafo 3).

stextarchiv.de/>), che dispone soprattutto di testi in formato PDF, e il *Projekt Gutenberg* (<<https://www.projekt-gutenberg.org/index.html>>), dove tutti i testi sono presenti in formato HTML. I testi in versione cartacea sono stati scannerizzati e dunque importati in file PDF. Tutti i file PDF – sia quelli attinti da fonti internet, sia quelli di provenienza cartacea – sono stati convertiti in formato Word tramite OCR, i testi del sito web del Progetto Gutenberg sono stati copiati direttamente in un file Word³. Ogni singolo esemplare testuale è stato sottoposto a revisione formale, ovvero linguistica.

Attualmente il corpus tedesco di LBC contiene 763 file Word (in totale 1.183.484 token)⁴, che sulla base del genere testuale di appartenenza possiamo suddividere in tre macrocategorie: divulgativa, tecnico-specialistica e letteraria (cfr. Tabella 1 e Figura 2).

Tabella 1. Ripartizione dei token nelle tre macrocategorie (frequenza assoluta).

Categorie	Frequenza assoluta
Divulgativa	17.450
Tecnico-specialistica	526.516
Letteraria	639.518
Totale	1.183.484

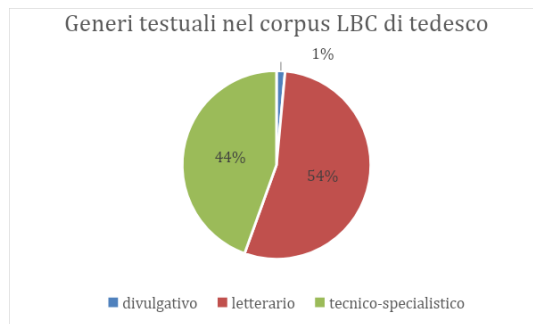


Figura 2. Ripartizione dei token nelle tre macrocategorie (percentuale).

³ Per la realizzazione di questa prima fase della ricerca è stato fondamentale il lavoro di dottorandi e tirocinanti dell'Università di Firenze. Il nostro particolare ringraziamento va qui alla dott.ssa Benedetta Bronzini, alla dott.ssa Sara Congregati, a Valentina Baldi, Domenica Barone, Delia Papandrea, Maria Chiara Susini, Clara Sabatino, Giulia Giannattasio e Lisa Bandinelli.

⁴ Le cifre si riferiscono allo stato del corpus al 31 agosto 2018 (cfr. anche paragrafo 3.1). Da allora il corpus tedesco di LBC è stato fortemente ampliato; i nuovi testi elaborati non sono tuttavia ancora disponibili online. Raggiunta quota 1.000.000 di token, il corpus online non è stato ulteriormente incrementato per permettere di ancorare il lavoro sulle concordanze a una raccolta di testi circoscritta (cfr. Flinz, Buffagni, Ballestracci 2020). Tale decisione è estesa a tutte le lingue del progetto.

Tabella 1 e Figura 2 mostrano che il genere più rappresentato nel corpus tedesco è quello letterario (54%). Ben rappresentato è anche il genere tecnico-specialistico con una percentuale del 45%. I meno rappresentati nel corpus LBC tedesco sono i testi divulgativi con una percentuale dell'1%.

Rientrano nei testi di genere letterario biografie, opere narrative e saggi secondo la ripartizione qui di seguito illustrata (Figura 3):

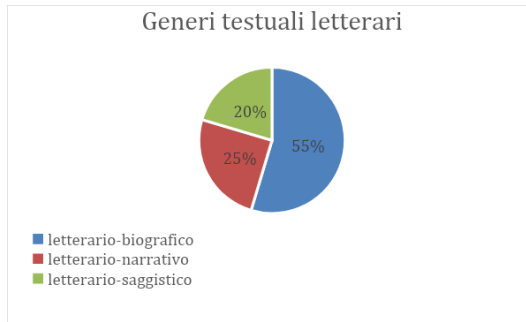


Figura 3. Ripartizione dei testi letterari nelle tre categorie (percentuale).

Come possibile osservare in Figura 3, la categoria più ampia, con una percentuale del 55%, è la biografia, mentre le altre due categorie letterarie – *fiction* e saggistica – presentano una percentuale più bassa, pressoché equivalente (rispettivamente del 25% e del 20%).

Rientrano nei testi biografici sia biografie vere e proprie sia testi più vicini ad altri generi testuali, per esempio diari e resoconti di viaggio. Questo gruppo di testi accoglie in egual misura testi in traduzione – per esempio dall'italiano *Das Leben des Benvenuto Cellini, florentinischen Goldschmieds und Bildhauer. Von ihm selbst geschrieben* (2016)⁵ di Benvenuto Cellini – oppure dal francese – *Reise in Italien (Rome, Naples et Florence en 1817)* (1922) di Stendhal – così come testi originali in lingua tedesca, ad esempio *Das Florenzer Tagebuch* (1984) di Rainer Maria Rilke.

Si annoverano fra i testi di *fiction* novelle, romanzi e altri testi di narrativa letteraria dove Firenze o la Toscana sono al centro o sullo sfondo delle vicende narrate. All'interno di questa categoria figurano soltanto testi di autori tedesco-foni. Ne sono un esempio *Florentiner Novellen* (1984) di Isolde Kurz o *Florentinische Nächte* (1969) di Heinrich Heine.

Anche i testi di saggistica includono soltanto testi in lingua originale tedesca, per esempio *Stadt des Lebens* (1905) di Isolde Kurz, oppure *Geschichte der Renaissance in Italien* (1891) di Jacob Burckhardt.

La categoria dei testi tecnico-specialistici comprende testi di varia provenienza settoriale: arte, enogastronomia e storia (cfr. Figura 4):

⁵ Da qui in poi le indicazioni bibliografiche fanno riferimento alle edizioni dei testi contenuti nel corpus.

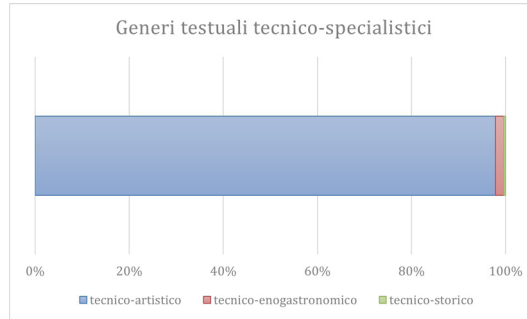


Figura 4. Ripartizione dei testi tecnico-specialistici (percentuale).

Un'alta percentuale è rappresentata dai testi tecnico-specialistici di arte (oltre 500.000 token; 98%). I testi tecnico-specialistici di enogastronomia e di storia sono invece molto poco rappresentati (in totale 11.134 token; 2%).

I testi tecnico-specialistici mostrano una varietà minore rispetto a quelli letterari: si tratta esclusivamente di traduzioni dall'italiano, come per esempio *Der Traktat von der Malerei* (1909) di Leonardo da Vinci, *Kleinere kunsttheoretische Schriften* (1877) di Leon Battista Alberti, *Von der Wissenschaft des Kochens und der Kunst des Genießens* (2000) di Pellegrino Artusi e *Geschichte von Florenz* (1934) di Niccolò Machiavelli.

Per quanto riguarda i testi divulgativi abbiamo al momento un unico esemplare: si tratta di una buona guida turistica, relativamente esaustiva (17.412 token) in lingua originale tedesca, *Florenz. Perfekte Tage in der Toskana – Metropole* (2016) di Ciccarelli, Jepson e Fisher⁶.

La scelta dei testi è stata spesso influenzata dalla reperibilità degli stessi; ciononostante abbiamo tentato di garantire una certa varietà per tutti i generi letterari, privilegiando quelle tipologie testuali particolarmente importanti per le finalità del progetto. Per quanto possibile, sono stati rispettati criteri quali la paternità e l'importanza dell'opera per la cultura e l'arte italiana, come pure la diffusione e la ricezione a livello internazionale. Nel caso di alcuni esemplari l'attribuzione del genere testuale di appartenenza ha presentato non poche difficoltà al gruppo di lavoro. Basti solo citare, a titolo esemplificativo, l'opera di Giorgio Vasari (1966 [1568]), che rappresenta il testo fondante del progetto, ovvero il testo da cui le singole unità di ricerca – divise per lingua – sono partite per l'acquisizione dei dati (edizione tedesca: *Leben der ausgezeichnetsten Maler, Bildhauer und Baumeister von Cimabue bis zum Jahre 1567*, 2008). Questi testi – come altri del corpus – sono caratterizzati da una forte commistione di generi testuali, il cui stile oscilla fra lo specialistico-descrittivo, il biografico-narrativo e il critico-valutativo. Il testo è stato attribuito alla categoria del genere tecni-

⁶ In una prima fase di lavoro sono state prese in esame anche altre guide turistiche in seguito escluse di necessità dal corpus per mancanza di informazioni sulle caratteristiche della lingua impiegata (se per esempio da traduzione o da testo originale).

co-specialistico, da un lato, per il criterio delle peculiarità stilistiche prevalenti, dall'altro, per il criterio dello scopo insito nel progetto stesso. Quest'ultimo criterio ha guidato progressivamente anche la scelta dei singoli autori e delle singole opere presenti nel corpus (cfr. paragrafo 1.2).

1.2 Autori e opere

Nel selezionare gli autori e le opere che costituiscono il corpus LBC abbiamo scelto di privilegiare quelli particolarmente rappresentativi per l'arte e la cultura del Rinascimento. Con questo intento tutti i gruppi di lavoro LBC sono partiti all'inizio del progetto da Giorgio Vasari e dalle sue *Vite* (1550, 1568), cioè da un'opera fondamentale per l'arte e la cultura del Rinascimento, che ha contribuito alla diffusione del mito del Rinascimento italiano in gran parte dei Paesi europei – anche in ambito tedescofono (cfr. Wackenroder 1991 [1797]; Fiorillo 1798; cfr. anche Ballestracci, Buffagni 2016). Questo si è rivelato un vantaggio per l'intero progetto, anche perché l'opera di Giorgio Vasari esiste in tutte le lingue comprese nel progetto (italiano, tedesco, inglese, francese, spagnolo, russo e cinese), cosa che in futuro potrebbe anche consentirci di costruire un corpus parallelo delle *Vite* per ulteriori ricerche (cfr. paragrafo 1.3 e paragrafo 2). Fra le altre opere e gli altri autori italiani che abbiamo scelto per il nostro corpus vi sono le *Opere Volgari* (1849) di Leon Battista Alberti, il *Trattato della pittura* (1651) di Leonardo da Vinci e le *Istorie Fiorentine* (1971) di Niccolò Machiavelli. Si è cercato, per quanto possibile, di ritrovare gli stessi autori in tutte le lingue.

Si ritengono fondamentali per il progetto anche autori non italiani che hanno avuto un ruolo di rilievo nella diffusione della cultura italiana del Rinascimento nei Paesi stranieri: ne fanno parte per esempio testi tradotti da altre lingue come *Wege zur Kunst* (1897) di John Ruskin o testi in lingua originale tedesca come *Die Geschichte der Renaissance in Italien* (1891) di Jacob Burckhardt⁷.

Un altro gruppo importante è rappresentato da autori o letterati famosi che hanno fatto viaggi in Italia e hanno scritto un diario o un resoconto di viaggio. La loro importanza risiede soprattutto nel fatto di aver descritto l'arte e la cultura italiane con occhi stranieri, contribuendo così alla diffusione dei Beni culturali e artistici soprattutto nel XIX secolo, ma anche fino ai giorni nostri, in realtà. Fra gli autori di testi letterari del corpus LBC tedesco si annoverano letterati famosi come, per esempio, Johann Wolfgang Goethe (*Italienische Reise*, 1997), Karl Philipp Moritz (*Reisen eines Deutschen in Italien in den Jahren 1786 bis 1788*, 2015), Fanny Lewald (*Bilderbuch*, 1992), Karl Eugen Gass (*Pisaner Tagebuch. Aufzeichnungen/Briefe*, 1961) e Stendhal (*Reise in Italien (Rom, Naples et Florence en 1817)*, 1922).

⁷ L'opera di Burckhardt nel corpus LBC tedesco non figura ancora in versione integrale a causa delle varie impurità del file Word importato dal formato PDF, che hanno pesato non poco sulla revisione formale e linguistica del testo.

Per tutti gli autori e le loro opere sono stati scelti nelle prime fasi di lavoro solo brani nei quali spiccano Firenze, la Toscana e l'arte del Rinascimento. I criteri inizialmente adottati sono stati in seguito ampliati: da un lato, si è deciso di prendere in considerazione non solo Firenze e la Toscana ma l'Italia tutta, pur mantenendo al centro dell'attenzione l'arte e la cultura; dall'altro, sono stati presi in esame non solo l'arte e la cultura del Rinascimento ma anche l'arte e la cultura di altre epoche storiche. Per questo il corpus da noi impiegato per l'elaborazione delle concordanze (cfr. Buffagni, Flinz, Ballestracci 2020) e qui descritto è l'immagine del corpus in un dato momento, così come rappresentato in Figura 5:

<u>doc.author</u>	<u>document frequency</u>
Cellini, Benvenuto	237
Kurz, Isolde	169
Vasari, Giorgio	126
Moritz, Karl Philip	66
Artusi, Pellegrino	46
Rilke, Rainer Maria	25
Ciccarelli, Jepson, Fisher	21
Brandt, Karl	18
Heine, Heinrich	12
Burckhardt, Jacob	10
Alberti, Leon Battista	7
da Vinci, Leonardo	7
Lewald, Fanny	6
Ruskin, John	4
Stendhal	3
Gass, Karl Eugen	3
Machiavelli, Niccolò	2
Goethe, Johann Wolfgang von	1

Figura 5. Autori nel corpus LBC tedesco.

Come evidenziato in Figura 5 il corpus LBC tedesco comprende in totale 18 autori: di questi, dieci sono tedeschi (Kurz, Moritz, Ciccarelli, Jepson e Fisher, Brandt, Heine, Burckhardt, Gass e Goethe), mentre i restanti otto sono o italiani (Cellini, Vasari, Artusi, Alberti, da Vinci e Machiavelli) o di altre lingue, come Stendhal (francese) e Ruskin (inglese) (cfr. Figura 6).

<u>doc.language</u>	<u>document frequency</u>
Italiano	425
Tedesco (Germania)	331
Inglese (Regno Unito)	4
Francese (Francia)	3

Figura 6. Lingue di partenza rappresentate nel corpus LBC tedesco.

La Figura 6 mostra come dal corpus LBC tedesco restino ancora esclusi testi in traduzione da alcune lingue del progetto (es. spagnolo, russo e cinese); queste

dovranno venir prese in considerazione nelle prossime fasi di lavoro. Allo stato attuale il corpus LBC tedesco offre tuttavia testi che garantiscono una varietà diacronica relativamente ricca, essendo stati scritti o pubblicati fra il 1400 e il 2016 (cfr. Figura 7).

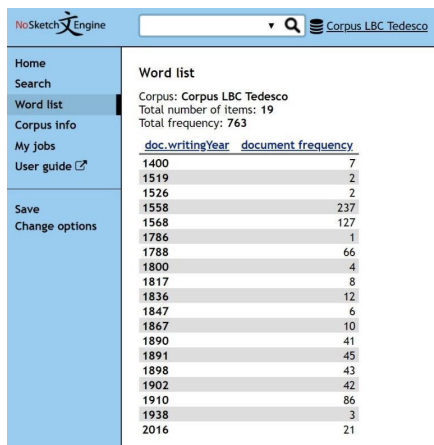


Figura 7. Lingue di partenza rappresentate nel corpus LBC tedesco.

Pur necessitando di alcuni interventi per vari aspetti, il corpus LBC tedesco evidenzia altrettanti aspetti che allo stato attuale non possono che trovare una valutazione positiva (cfr. paragrafo 1.3).

1.3 Lo stato attuale del corpus. Prime conclusioni

Riassumendo, la piattaforma del corpus LBC mira, unitamente alla redazione di un dizionario multilingue in rete, alla reperibilità di testi rilevanti per la conoscenza dei Beni culturali italiani, così da offrire una fonte liberamente accessibile per ricerche di tipo linguistico, letterario e culturale.

Come si evince dal paragrafo 1.1, più della metà dei testi compresi nel corpus appartiene al genere letterario, la quasi totalità dei testi rimanenti è riconducibile invece al genere tecnico-specialistico, e solo una minima parte al settore divulgativo. Tale configurazione potrebbe a prima vista apparire non del tutto equilibrata per la preponderanza dell'ambito letterario. Vanno tuttavia sottolineate innanzitutto le motivazioni anche genetiche dell'attuale configurazione della banca dati: la maggior parte dei testi letterari presenti provengono infatti da archivi online molto prestigiosi e facilmente accessibili; così non è per molti dei testi tecnico-specialistici. È sembrato dunque sensato nella prima fase di lavoro partire dalla raccolta dei testi letterari. La struttura attuale della banca dati che, come evidenziato, registra la situazione del corpus in un dato momento, corrisponde dunque in primo luogo a questo modo di procedere.

La bontà della decisione presa si è palesata anche in un'altra ottica. Dei testi letterari del corpus LBC tedesco va notato innanzitutto che si contraddistinguono per la loro varietà: includono prevalentemente biografie, la restante parte è data invece dalla somma dei generi testuali di *fiction* e saggistica. Le biografie assumono pertanto un'importanza notevole, a queste sono stati infatti aggiunti i testi di viaggio. Come nel caso delle altre lingue rappresentate nella banca dati (es. il francese), le descrizioni e i resoconti di viaggio svolgono sul piano sia quantitativo che qualitativo un ruolo particolare in relazione agli obiettivi del corpus LBC tedesco: quantitativo, poiché (cfr. paragrafo 1.2) costituiscono una parte cospicua del corpus; qualitativo, in quanto riportano la prospettiva di osservatori stranieri di fronte alle opere d'arte italiane, e fiorentine in particolare (si tratta spesso di viaggiatori del XVIII e XIX secolo, come, per esempio, il consigliere segreto Goethe o l'irrequieto soldato Stendhal). Le descrizioni che scrittori e artisti tedeschi, francesi e inglesi hanno dedicato alle opere d'arte italiane sono testimonianze dell'epoca e della vita degli autori; possiedono un elevato valore letterario e hanno contribuito alla ricezione dell'immagine dell'Italia nel mondo. Hanno infine portato alla creazione di un vero e proprio «mito dell'Italia» che avrà ampie ripercussioni sull'arte e la cultura europea ed extraeuropea. Ricordiamo fra questi anche *Wege zur Kunst* (1897, traduzione dall'inglese) di Ruskin. Le altre biografie presenti nel corpus, che comprendono le vite degli artisti del Rinascimento, si presentano da parte loro come «tratteggio di un percorso di vita» e sono ascrivibili a un'epoca più antica (per esempio le *Lebensbeschreibungen* di Vasari e l'autobiografia di Cellini). Opere architettoniche, sculture, dipinti, affreschi, stucchi e ornamenti vengono descritti e valutati rispettivamente secondo criteri indicati dagli autori. Queste opere comprendono «elementi [...] di storiografia e letteratura» (Best 1994, 71) e contengono a tal riguardo preziose informazioni storiche. Entrambe le tipologie di genere biografico possiedono un inestimabile valore storico-culturale, poiché permettono di comprendere lo sviluppo della storia dell'arte, come pure della storiografia dell'arte, fra il XV secolo e il presente.

Nel corpus si trovano anche, classificati come generi letterari – nella sottocategoria dei testi di saggistica – classici moderni della storiografia dell'arte come la *Geschichte der Renaissance in Italien* (1891) di Burckhardt. L'opera di Burckhardt, che ha avuto un'influenza notevole su altri autori presenti nel corpus (es. Isolde Kurz), si distingue per le neoacquisite conoscenze scientifiche dell'epoca e offre un'immagine dell'arte del Rinascimento ancor più ricca di sfaccettature. Sul piano stilistico mostra un lessico variegato, duttile, che accanto ai termini tecnici presenta anche peculiarità estetiche.

Sono poi da menzionare come generi di *fiction* con valore prototipico in ambito letterario i testi narrativi presenti nel corpus, fra cui il racconto *Florentinische Nächte* di Heinrich Heine (1969) e il ciclo di novelle *Florentiner Novellen* (1984) di Isolde Kurz. La città di Firenze con gli usi e costumi della sua popolazione fanno spesso da sfondo alle storie narrate, il cui ampio spettro di varianti si esprime, fra l'altro, nella possibilità di venir ambientate in epoche passate (come nelle novelle di Kurz) o nel presente (come nel racconto di Heine).

Fra i generi testuali tecnico-specialistici, tutti in traduzione, spiccano in particolare i *Kleinere kunsttheoretische Schriften* (1877) di Leon Battista Alberti e *Traktat von der Malerei* (1909) di Leonardo da Vinci: il primo segna una svolta importante nella storiografia dell'arte dell'epoca, avendo introdotto un lessico tecnico innovativo (architettura, scultura e pittura); l'altro rappresenta una riflessione approfondita sull'essenza della pittura che proprio all'epoca vuole affermarsi accanto alla poesia come arte di pari dignità (Gombrich 2000). La *Geschichte von Florenz* (1934) di Niccolò Machiavelli costituisce l'unico testo tecnico di contenuto storico, mentre *Von der Wissenschaft des Kochens und der Kunst des Genießens* (2000) di Pellegrino Artusi è il solo testo dedicato alla gastronomia. Si tratta in entrambi i casi di opere di grande valore culturale. A questo proposito si rivelano rappresentativi per un lungo lasso di tempo (dal 1400 al 1891) anche i generi testuali tecnico-specialistici che occupano quasi la metà del corpus, mentre si annovera fra i testi divulgativi una guida turistica contemporanea in lingua tedesca, *Florenz. Perfekte Tage in der Toskana – Metropole* (2016) di Ciccarelli, Jepson e Fisher.

Come emerge chiaramente da quanto argomentato sopra, la stragrande maggioranza dei testi presenti nel corpus è accomunata da un elevato valore stilistico che contraddistingue la banca dati multilingue e il corpus tedesco. Proprio lo stile alto possiede un valore intrinseco ed è rilevante soprattutto per la storia dell'arte e la storiografia artistica. Non a caso Reiners contrappone allo «stile enciclopedico privo di visione» il «testo creato da un buon scrittore di prosa», per giungere alla conclusione: «Le opere di storia dell'arte contrappongono immagini di uno stile [dello stile vivace, plastico, n.d.a.] all'altro [lo stile enciclopedico, n.d.a.] per metterne in risalto le peculiarità. Anche nel regno della parola la bellezza della grande prosa risalta con maggior chiarezza sullo sfondo dello stile medio⁸» (Reiners 1991, 219 s.).

E quanto più diversi tra loro si presentano i generi, tanto più ricca si rivela la stragrande maggioranza dei testi all'interno del corpus. Da un punto di vista qualitativo il corpus è da considerarsi adeguato agli obiettivi perseguiti. Come annunciato nel paragrafo 1.2, nelle prossime fasi di lavoro la prospettiva verrà ulteriormente ampliata sul versante dello spazio (a tutta l'Italia) e su quello del tempo (all'arte e alla cultura di altre epoche); ci si concentrerà, anche in futuro, sulla qualità stilistica e sul valore storico-culturale dei testi.

Va inoltre aggiunto che, soprattutto in virtù della propria costituzione testuale, il corpus rappresenta una miniera linguistica inesauribile: le prime ricerche effettuate con approcci differenti hanno dimostrato di condurre all'acquisizione di conoscenze fruttuose su numerosi fenomeni linguistici (es. lessicali, sintattici, testuali, traduttivi, cfr. paragrafo 2).⁹

⁸ Con «stile medio» Ludwig Reiners intende il tipico stile enciclopedico che egli critica come astratto e «privo di visione». Agli antipodi troviamo la grande prosa. Secondo l'autore il «buon scrittore di prosa» porta a compimento un lavoro significativo che richiede molta forza d'immaginazione e conoscenza del mondo per riprodurre complessi processi mentali e renderli vivi (Reiners 1991: 219).

⁹ Vanno qui menzionate anche numerose tesi triennali redatte a Firenze e a Siena e incentrate sul testo di Vasari o su altri testi del corpus LBC.

2. Possibilità applicative del corpus

La piattaforma LBC (<<http://corpora.lessicobeniculturali.net/de/>>) mette a disposizione i vari corpora del progetto (tedesco, inglese, francese, italiano, russo, spagnolo) permettendo la loro consultazione tramite *NoSketchEngine* (versione ridotta di *SketchEngine*¹⁰). Gli utenti possono analizzare il corpus¹¹ fruendo di diverse funzionalità, come *Wordlist* con la quale si possono estrarre liste di parole sulla base della frequenza assoluta (parole, lemmi, tags ecc.), *Concordance* con la quale si possono analizzare singole/i parole/lemmi nel loro contesto (*Keywords in Context*) e *Collocations* con la quale si possono identificare possibili collocazioni di un lessema (cfr. Kilgarriff *et al.* 2004; Kilgarriff *et al.* 2014). Si possono poi valutare i risultati ottenuti confrontandoli fra loro in ottica intralinguistica o interlinguistica.

Dato il grande potenziale scientifico e pedagogico dei corpora (cfr. Flinz 2020; Flinz, Katelhön 2019), entrambe le prospettive trovano possibilità applicative in diversi ambiti scientifici e didattici: per la prassi traduttiva dei traduttori professionisti, per la formazione dei futuri traduttori e anche per la didattica del tedesco come lingua straniera (*DaF – Deutsch als Fremdsprache*) in generale¹². L'uso dei corpora LBC si presta inoltre non solo alla ricerca e all'insegnamento del lessico specialistico ma anche all'indagine di fenomeni grammaticali. Illustreremo in seguito alcuni di questi usi.

2.1. Analisi intralinguistiche

Il corpus tedesco risulta composto da 1.183.484 token (cfr. Figura 8). Trattandosi tuttavia di un *corpus monitor* (Lemnitzer, Zinsmeister 2015: 140), è soggetto a integrazioni e ampliamenti costanti.

Corpus LBC Tedesco ©

Corpus LBC Tedesco

Counts	General Info	Lexicon sizes
Tokens 1,183,484	Corpus description Document	word 66,972
Words 1,018,253	Language German	tag 54
Paragraphs 10,935	Encoding UTF-8	lemma 47,026
Documents 763	Compiled 11/12/2019 11:01:29	lc © 62,797
	Tagset Description	lemma_lc © 45,498

Structures and attributes

doc 763 ▾

Figura 8. Informazioni sul corpus LBC tedesco.

¹⁰ Ai tempi iniziali del progetto LBC, *SketchEngine* (<<https://www.sketchengine.eu/>>) era a pagamento, pertanto l'Unità di Ricerca ha optato per la versione ridotta *open source* *NoSketchEngine*. Successivamente grazie ad un progetto europeo *SketchEngine* è diventato gratuito per la comunità scientifica e accademica e rimarrà tale fino al 2022.

¹¹ I corpora vengono lemmatizzati e annotati a livello morfo-sintattico in maniera automatica.

¹² I corpora LBC possono venire impiegati anche con altre finalità, che si estendono per esempio agli studi lessicografici e all'analisi del discorso.

Una possibilità di impiego consiste nel generare liste di frequenza per lemma (Figura 9) che danno informazioni sulla frequenza dei lemmi registrati nel corpus.

Word list

Corpus: Corpus LBC Tedesco
 Total number of Items: 10,324
 Total frequency: 963,888

Page 1 (Go) Next >

lemma	frequency
die	100,366
und	37,338
sein	28,572
in	24,357
eine	22,406
er	21,792
zu	17,789
ich	16,055
von	14,196
mit	10,164
haben	10,113
sich	8,468
sie	8,343
werden	7,985
dies	7,617
nicht	6,994
auf	6,964
so	6,816
als	6,620
dass	6,551
an	6,309
es	6,256
man	5,800
wie	5,308
welche	5,231
ihr	5,181
nach	4,915
alle	4,079
auch	3,840
aus	3,719
andere	3,558
aber	3,549
für	3,524
groß	3,321
noch	3,290
bei	3,251
sehen	3,131
mein	3,058
können	3,050
um	2,855
viele	2,852
durch	2,830
wir	2,746
sehr	2,731
sagen	2,617
nur	2,550
da	2,462

Figura 9. Screenshot della lista dei lemmi estratta.

Da qui si possono per esempio ricavare i sostantivi, i verbi, gli aggettivi più frequenti:

- Sostantivi: *Werk* (2.183), *Zeit* (1.864), *Bild* (1.617), *Florenz* (1.466), *Leben* (1.414), *Kunst* (1.369), *Hand* (1.346), *Kirche* (1.307), *Rom* (1.150), *Haus* (1.145), *Papst* (1.131), *Herr* (1.120) ecc.
- Verbi: *sein* (28.572)¹³, *haben* (10.113), *werden* (7.985), *sehen* (3.131), *können* (3.050), *sagen* (2.617), *lassen* (2.273), *wollen* (2.205), *machen* (1.859), *geben*

¹³ Abbiamo notato che per *sein* si trovano sia le occorrenze dell'ausiliare *sein* sia quelle dell'aggettivo possessivo *sein*, il che può essere dovuto a errori nel *tagging* automatico. Questo pro-

- (1.808), *kommen* (1.800), *sollen* (1.714), *malen* (1.677), *gehen* (1.677), *finden* (1.373), *stehen* (1.307), *arbeiten* (1.274), *müssen* (1.260) ecc.
- Aggettivi: *andere*¹⁴ (3.558), *groß* (3.321), *viele* (2.852), *schön* (2.424), *gut* (2.213), *alt* (1.234) ecc.

Restringendo il campo di osservazione per esempio ai soli sostantivi, si nota subito che a ricorrere più di frequente sono termini tecnici dell'arte come *Werk*, *Bild*, *Kunst* e *Kirche*, e anche nomi di città come *Florenz* e *Rom*. Fra i verbi, come prevedibile, ricorrono più frequentemente gli ausiliari e i modali, ma anche verbi tipici del linguaggio tecnico, come per esempio *malen*. Questo uso può incentivare lo sviluppo di una sensibilità per le collocazioni¹⁵, aspetto di una lingua straniera assai difficile e quindi soggetto a errori (su questo punto cfr. anche il paragrafo 2.2).

Si possono anche effettuare ricerche sull'uso di una parola nel linguaggio tecnico e nella lingua standard, come pure sul suo sviluppo diacronico, specificando nel campo di ricerca rispettivamente tipologia testuale, autore ed epoca storica (Figura 10).

Oltre alle liste di parole, nel corpus si possono ricercare anche parole isolate. Poiché *Kirche* è una parola con un'alta frequenza, è interessante riflettere sui suoi composti. Con l'opzione di ricerca **kirche* possono venir estrapolati tutti i composti della parola:

Peterskirche (51), *Hauptkirche* (18), *Servitenkirche* (14), *Domkirche* (9), *Paulskirche* (7), *Taufkirche* (5), *Andreaskirche* (3), *Weltkirche* (3), *Bettelordenskirche* (2), *Bischofskirche* (2), *Familienkirche* (2), *Franziskanerkirche* (2), *Kathedralkirche* (2), *Klosterkirche* (2), *Porkirche* (2), *Vorgängerkirche* (2), *Augustinerkirche* (1), *Barfüßerkirche* (1), *Benediktinerkirche* (1), *Dorfkirche* (1), *Franziskaner-Kirche* (1), *Gesetzeskirche* (1), *Gewölbekirche* (1), *Hauptporkirche* (1), *Hauskirche* (1), *Kapuzinerkirche* (1), *Karmeliterkirche* (1), *Kartheuserkirche* (1), *Kollegialkirche* (1), *Kuppelkirche* (1), *Madonnenkirche* (1), *Marienkirche* (1), *Markuskirche* (1), *Ordenskirche* (1), *Pfarrkirche* (1), *Unterkirche* (1).

A lezione è possibile riflettere sia sul contenuto (il significato dei composti) sia sulla forma dei composti (grafia con o senza elementi di raccordo), integrando le ricerche eventualmente con dati estratti da altri corpora online. Ciò è possibile anche con apprendenti di livello linguistico diverso.

blema, tuttavia, può anche rivelarsi una risorsa a lezione, laddove si solleciti la sensibilità degli apprendenti verso la polifunzionalità di alcuni lessemi.

¹⁴ La lemmatizzazione automatica di alcune parole (*anders*, *viel* ecc.) avviene erroneamente con il femminile.

¹⁵ Il termine «collocazione» trova varie definizioni in letteratura: una empirica (fra gli altri, Firth 1957; Evert 2009; Belica, Perkuhn 2015), il cui criterio principale è la frequenza, e una teorica (fra gli altri, Hausmann 1984, Burger 2015), che definisce le collocazioni come unità lessicali in relazione sintattica (Hausmann 1984) o come unità lessicali con un significato unitario. La tesi sostenuta in questo contributo è che nel *DaF* entrambe le definizioni possano trovare applicazione. La scelta verrà determinata dagli obiettivi dell'unità di apprendimento.

The screenshot shows a search interface titled "Text types". At the top, there is a "Subcorpus" dropdown menu set to "None (whole corpus)" and a link "info create new". Below this are several filter sections:

- LINGUA ORIGINALE**: Includes checkboxes for "Francese (Francia)", "Inglese (Regno Unito)", "Italiano", and "Tedesco (Germania)", with a "Select All" button.
- LINGUA DI TRADUZIONE**: Includes a checkbox for "Tedesco (Germania)" and a "Select All" button.
- CATEGORIA E SOTTOCATEGORIA**: Includes checkboxes for "Divulgativo", "Letterario", and "Tecnico", with a "Select All" button.
- AUTORE**: A text input field.
- TITOLO**: A text input field.
- FRAMMENTO**: A text input field.
- ANNO DI REDAZIONE**: A text input field.

Figura10. Screenshot del campo di ricerca.

Le concordanze¹⁶ della parola *Kirche* possono inoltre fornire anche informazioni sul suo uso, come per esempio:

- in combinazione con aggettivi in funzione attributiva: *achteckige Kirche, alte Kirche, bischöfliche Kirche, christliche Kirche, große Kirche, kleine Kirche, neue Kirche, römische Kirche, schöne Kirche, untere Kirche*;
- in combinazione con altri sostantivi: *Bild der Kirche, Bronzetür der Kirche, Chor der Kirche, Dach der Kirche, Eingang der Kirche, Fassade der Kirche, Grundriss der Kirche, Hauptaltar der Kirche, Hauptkapelle der Kirche, Haupttür der Kirche, Kapelle der Kirche, Kreuz der Kirche, Mittelschiff der Kirche, Querschiff der Kirche, Schiff der Kirche, Teil der Kirche, Tür der Kirche, Vorderwand der Kirche, Wand der Kirche, ecc.*;
- in combinazione con preposizioni: *in der Kirche, inmitten der Kirche, innerhalb der Kirche, nach der Kirche, neben der Kirche, unterhalb der Kirche, von der Kirche, vor der Kirche, für die Kirche, in die Kirche, ecc.*;
- in combinazione con verbi (filtro a destra): *eine Kirche bauen, eine Kirche malen, eine Kirche vollenden, eine Kirche verfertigen, ecc.*;

¹⁶ I KWICs possono venir ordinati tramite l'opzione *sort Left, sort Right* ecc.. Questa opzione offre la possibilità di richiamare l'attenzione su determinate caratteristiche della lingua analizzata.

- come nome proprio: *Kirche del Carmine, Kirche della Pace, Kirche der Annunziata, Kirche der Madonna, Kirche des Heiligen Andreas, Kirche des S. Agostino, Kirche Ognissanti, Kirche San Giovanni, Kirche St. Andrea, Kirche von S. Agostino, ecc.*

In questo modo è possibile favorire negli apprendenti del tedesco sia una maggiore consapevolezza delle collocazioni di una determinata parola sia una maggior sensibilità per determinati fenomeni grammaticali.

2.2 Analisi interlinguistiche

I corpora compresi nella piattaforma LBC hanno tutti più di un milione di token (Tabella 2):

Tabella 2. Numero dei token per lingua.

	Tedesco	Inglese	Italiano	Francese	Russo	Spagnolo
Token	1.183.484	1.784.990	1.647.595	3.818.747	1.647.595	1.162.849

Per ogni corpus sono presenti informazioni diverse, così da poter ricavare, fra l'altro, il numero dei rispettivi testi per anno di pubblicazione, autore, genere testuale (Figura 11):

<u>doc.printingYear</u>	<u>document frequency</u>
1966	162
1973	125
1970	66
1971	10
1550	2
1910	1

Figura 11. Distribuzione dei documenti nel corpus italiano in base all'anno di pubblicazione.

Il sistema permette anche di costruire subcorpora sulla base della tipologia testuale, così da poter anche effettuare studi interlinguistici orientati su questo aspetto.

Possono venir utilizzate per confronti interlinguistici anche le liste di frequenza estratte dai corpora, che possono essere utili per l'identificazione di possibili equivalenti (cfr. Tabella 3):

Dalla Tabella 3 è possibile infatti identificare i seguenti equivalenti: *cosa/Ding, opera/Werk, mano/Hand, tempo/Zeit, figura/Figur, anno/Jahr* ecc. Il significato dei lessemi rimanenti, per esempio ‚Papst‘, può poi venire eventualmente desunto tramite contesto (Figura 12):

La parola cercata *Papst* può tuttavia venir aperta anche in un'altra finestra al di fuori della rappresentazione KWIC, per esempio come testo (Figura 13):

Tabella 3. Liste di parole basate sulla frequenza (frequenza assoluta): sostantivi nel corpus tedesco e in quello italiano.

Sostantivi (D)	<->	Sostantivi (I)
Werk (2.183)		cosa (4671)
Zeit (1.864)		opera (3.615)
Bild (1.617)		mano (2.143)
Florenz (1.466)		tempo (2.093)
Leben (1.414)		figura (1.731)
Kunst (1.369)		modo (1.693)
Hand (1.346)		anno (1.693)
Kirche (1.307)		uomo (1.510)
Jahr (1.276)		parte (1.510)
Rom (1.150)		città (1.462)
Haus (1.145)		disegno (1.414)
Papst (1.131)		luogo (1.377)
Herr (1.120)		parere (1.319)
Tag (1.118)		Roma (1.309)
Stadt (1.096)		storia (1.304)
Meister (972)		tavola (1.284)
Figur (945)		chiesa (1.278)
Mann (913)		messer (1.216)
Künstler (903)		maniera (1.206)
Ding (899)		casa (1.194)

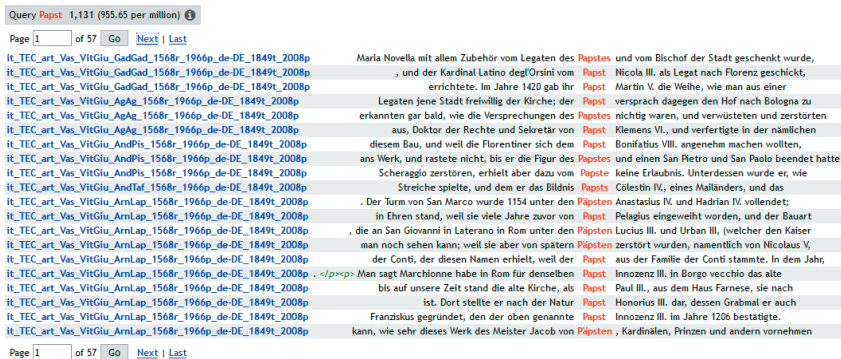


Figura 12. KWICs della parola cercata, Papst'.

Un altro uso particolarmente rilevante per la didattica *DaF* è il confronto interlinguistico delle collocazioni. Prendendo sempre a esempio *Kirche* e in questo caso la traduzione italiana (*chiesa*)¹⁷, possiamo riflettere prima sulle possibili collocazioni delle due parole (per esempio partendo dal tipo AGGETTIVO + SOSTANTIVO) e poi individuare possibili relazioni di equivalenza (Figura 14):

- *Kirche*: *genannte Kirche, bischöfliche Kirche, alte Kirche, christliche Kirche, untere Kirche, römische Kirche* ecc.
- *Chiesa*: *detta Chiesa, nuova chiesa*¹⁸, *chiesa principale, chiesa vecchia, chiesa romana, chiesa piccola* ecc.

< [previous](#) vergrößern. Sie hatten eine bedeutende Summe Geldes zusammen gebracht, und viele in der Stadt versprochen allen Beistand zu leisten. Deshalb ward der Bau der neuen Kirche am Sankt-Lucas-Tag 1278 begonnen, und der Cardinal Latino degli Orsini vom **Papst** Nicola III. als Legat nach Florenz geschickt, legte feierlich den Grundstein. Baumeister jener Kirche waren: der Bruder Giovanni, ein Florentiner, und der Bruder Ristoro da Campi, beide Bekenner eines Ordens, welche auch [next](#) >

Figura 13. Frammento di testo per la parola cercata ‚Papst‘.

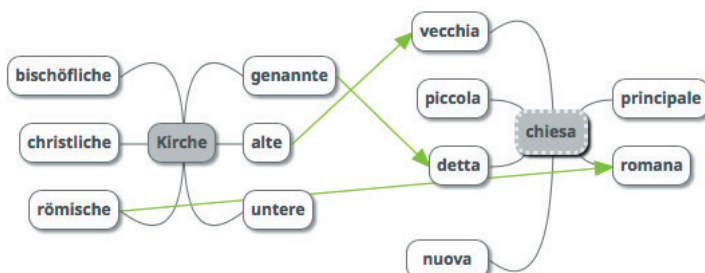


Figura 14. Rapporti di equivalenza fra le collocazioni aggettivali (*Kirche/chiesa*).

I singoli raggruppamenti possono venir considerati anche nei loro contesti e confrontati fra loro (Figura 15):

<p>Query chiesa 1,429 • Positive fitter (excluding KWIC) vecchia 6 (5.10 per million) ⓘ</p> <p>doc#5 - In Florenz, che dipinte nella vecchia chiesa del Tempio, che era dove è oggi la cittadella</p> <p>doc#11 Vigna, dove stanno Frati del Zoccolo, nella chiesa vecchia era in un quadro un Cristo morto, tanto</p> <p>doc#187 quali andarono per terra quando fu rovinata la chiesa vecchia di San Piero per rifar la nuova. <-p>> Fu</p> <p>doc#202 - Il quale, andato a Perugia, fece nella chiesa vecchio di S. Domenico dei Frati Predicatori</p> <p>doc#228 ; il quale per lo interesse che egli aveva nella chiesa vecchio -che la capella e l' altare maggiore era</p> <p>doc#332 deputati, ordinarono che si facesse in mezzo la chiesa vecchia una cappella da Antonio da Sangallo,</p>	<p>Query Kirche 1,307 • Positive fitter (excluding KWIC) alten 13 (10.98 per million) ⓘ</p> <p>doc#8 dorthin kommen, und er arbeitete in der alten Kirche San Domenico, welche den Prädikanten-Mönchen</p> <p>doc#11 Katharina darin darstellte. In der alten Kirche S. Domenico malte er auf einer Wand, wiederum in</p> <p>doc#12 versetzt er in Florenz eine Kapelle in der alten Kirche S. Spirito derselben Stadt, welche beim Brand</p> <p>doc#37 und endlich die Einweihung jener sehr alten Kirche , welche von Papst Paschalis II. vollzogen</p> <p>doc#57 - sagt man, habe er sehr Jung noch in der alten Kirche , die an dem Platz stand, wo nunmehr die alte</p> <p>doc#65 Frati del Zoccolo weihen, sah man in der alten Kirche einen toten Christus auf Herolden von ihm</p> <p>doc#114 fügte hinzu: man müsse es nicht nach der alten Kirche von St. Peter bringen, wo kein Platz mehr dafür</p> <p>doc#137 von Antonio da Sangallo in der Mitte der alten Kirche eine Kapelle zum Teil aus Künemern (spiegel)</p> <p>doc#191 und zu Ravenna der Marmor aus drei alten Kirchen (S. Severo, S. Apollinare in Classe und Galla)</p> <p>doc#193 blieb beharrlich, warf die Hälfte der alten Kirche nieder und legte die Fundamente der neuen (I.B.</p> <p>doc#336 diesen köhnen Gedanken vorlegte. Von der alten Kirche , welche der Kaiser Konstantin auf diesem Fleck</p> <p>doc#336 Hinter grunde oder der Tribune, wo aus der alten Kirche Konstantin, nach des größte Heiligtum,</p> <p>doc#750 II. machte reine Bahn durch Abbruch der alten Kirche - Mit den größten Meilen sollte der schönste</p>
---	---

Figura 15. Collocazioni: *chiesa vecchia* e *alte Kirche*.

¹⁷ È stato scelto a titolo esemplificativo il parametro *range* (-1 to 1). L'opzione (-5 to 5) è imposta per default.

¹⁸ Sono state elencate solo le combinazioni più frequenti. Combinazioni come *chiesa nuova* e *principale chiesa* compaiono anch'esse nel corpus, presentano tuttavia una bassa frequenza.

Come mostrano gli esempi sopra riportati, con l'aiuto degli strumenti messi a disposizione dalla piattaforma LBC gli apprendenti possono fare ricerche all'interno dei vari corpora sviluppando la *corpus literacy* (Mukherjee 2002: 179-180) necessaria per diventare sempre più autonomi.

3. Conclusioni e sviluppi futuri

In questo contributo si sono descritte le caratteristiche più importanti del corpus LBC tedesco unitamente ad alcune delle sue possibilità d'impiego. Tale descrizione, illustrando il corpus tedesco con determinate proprietà in un dato momento, vuole essere l'istantanea di un *work in progress*. Il corpus necessita infatti di ampliamenti testuali continui, qualitativamente elevati, che costituiscano un valido modello linguistico per tutti i gruppi di utenti a cui esso si rivolge: tutte le tipologie testuali presenti verranno arricchite di nuovi esemplari testuali, cui andranno ad aggiungersi altre tipologie testuali; si dovrà spostare l'attenzione su altre epoche storiche di riferimento (non solo i Beni culturali del Rinascimento, ma anche movimenti artistici successivi) e dal punto di vista diatopico si dovranno prendere in considerazione altri territori italiani oltre alla Toscana. Anche il lavoro scientifico e lessicografico è *in progress*, tanto che vecchie e nuove fasi del processo lessicografico vengono a sovrapporsi, come è tipico delle risorse online (Klosa 2016: 29).

Nonostante il carattere *in fieri* del lavoro i dati fin qui esposti dimostrano che già allo stato attuale il corpus tedesco – come gli altri corpora LBC – può venir utilizzato per numerosi scopi di ricerca e anche per attività didattiche. Come illustrato nel paragrafo 1, sono stati impiegati generi e tipologie testuali molto diversi in ottica sia sincronica che diacronica. Questa configurazione offre agli utenti – che si tratti di esperti o di dilettanti – una vasta gamma di applicazioni (cfr. paragrafo 3), come pure riflessioni sulle possibilità traduttive. In sintesi, il corpus tedesco si presta all'analisi di svariati fenomeni lessicali, morfologici, sintattici, morfosintattici e stilistici da prospettive diverse come per esempio:

- in una prospettiva orientata alla tipologia testuale;
- in prospettiva intralinguistica e interlinguistica;
- in prospettiva traduttiva (cfr. fra l'altro il focus sulla preponderanza di italianismi nelle traduzioni raccolte, Farina, Billero 2018);
- in prospettiva contrastiva (italiano-tedesco/tedesco-italiano);
- in prospettiva diacronica e sincronica.

Speriamo in generale con il nostro corpus di aver messo a disposizione di tutti gli interessati che fanno ricerca e lavorano nel campo dei Beni culturali un valido strumento dal quale poter ricavare utili indicazioni sulle proprietà linguistiche delle tipologie testuali inerenti all'arte e alla cultura. A questo proposito il corpus tedesco di LBC è già stato impiegato in concreto come fonte primaria per l'elaborazione del lemmario provvisorio del progettato dizionario LBC (cfr. Farina, Flinz 2020) e per la selezione delle concordanze (cfr. Buffagni, Flinz,

Ballestracci 2020) che fotografano l'attuale stato dell'arte dell'unità di ricerca LBC per il tedesco.

Bibliografia

- Alberti L.B. 1849, *Opere Volgari*, Tipografia Galilei, Firenze.
 — 1877, *Kleinere kunsttheoretische Schriften*, Braumüller, Wien.
- Artusi P. 2000, *Von der Wissenschaft des Kochens und der Kunst des Genießens*, Hahn Verlag, Berlin.
- Ballestracci S., Buffagni C. 2016, *La traduzione dei 'gerundi italiani' in tedesco: un'analisi semantica esemplare*, in Saraçgil A., Vezzosi L. (eds.), *Lingue, letterature e culture migranti*, FUP, Firenze: 11-35.
- Ballestracci S., Buffagni C., Flinz C. 2020, *Das deutsche LBC-Korpus: Zusammenstellung und Anwendung*, in Id., *Das deutsche LBC-Korpus*, FUP, Firenze. In pubblicazione, <<http://corpora.lessicobeniculturali.net/de/>>.
- Belica C., Perkuhn R. 2015, *Feste Wortgruppen/Phraseologie I: Kollokationen und syntagmatische Muster*, in Haß U., Storjohann P. (eds.), *Handbuch Wort und Wortschatz*, de Gruyter, Berlin-Boston: 201-225.
- Best O.F. 1994, *Handbuch literarischer Fachbegriffe. Definitionen und Beispiele*, Fischer, Frankfurt.
- Billero R., Nicolás Martínez M.C. 2017, *Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 4(2): 203-216.
- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Brandi K. 1927, *Die Renaissance in Florenz und Rom. Acht Vorträge von Karl Brandi*, Springer Fachmedien, Wiesbaden.
- Buffagni C., Flinz C., Ballestracci S. 2020, *Das deutsche LBC-Korpus: provisorische Stichwortliste und Konkordanzen*, in Flinz C., Buffagni C., Ballestracci S., Billero R., Farina A., Nicolás Martínez M.C., *Deutsche Lexik der Kunst auf der Basis des Korpus LBC*, FUP, Firenze. In pubblicazione, <<http://lexicon.lessicobeniculturali.net/de>>.
- Burger H. 2015, *Phraseologie. Eine Einführung am Beispiel des Deutschen*, Erich Schmidt Verlag, Berlin.
- Burckhardt J. 1891, *Die Geschichte der Renaissance in Italien*, Ebner & Seubert, Stuttgart.
- Cellini B. 2016, *Das Leben des Benvenuto Cellini, florentinischen Goldschmieds und Bildhauer. Von ihm selbst geschrieben*, Karl-Maria Guth, Berlin.
- Ciccarelli Roming C., Jepson T., Fisher T. 2016, *Florenz. Perfekte Tage in der Toskana – Metropole*, Taschenbuch Verlag, München.
- da Vinci L. 1651, *Trattato della pittura*, Jacques Langlois, Parigi.
 — 1909, *Der Traktat von der Malerei*, Eugen Diederichs, Jena.
Deutsches Textarchiv, <<http://www.deutschestextarchiv.de/>>, (10.2020).
- Evert S. 2009, *Corpora and Collocations*, in Lüdeling A., Kytö A. (eds.), *Corpus Linguistics*, de Gruyter, Berlin-New York: 1212-1248.
- Farina A. 2016, *Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique*, «Publif@rum», 24, <http://www.farum.it/publifarum/ezine_articles.php?art_id=335>.

- Farina A., Billero R. 2018, *Comparaison de corpus de langue «naturelle» et de langue «de traduction»: les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues*, in Lezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 108-116.
- Farina A., Flinz C. 2020, *LBC-Dictionary: a Multilingual Cultural Heritage Dictionary. Data collection and data preparation*, in Gavriilidou Z., Mitsiaki M., Fliatouras A. (eds.), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion – Volume 1*, Ljubljana University Press, Ljubljana: 371-379.
- Farina A., Garzaniti M. 2013, *Un portale per la comunicazione e la divulgazione del patrimonio culturale: progettare un lessico multilingue dei beni culturali on-line*, in Filipovic A., Troiano W. (eds.), *Strategie e Programmazione della Conservazione e Trasmissibilità del Patrimonio Culturale*, Edizioni scientifiche Fidei Signa, Roma: 500-509.
- Fiorillo J. D. 1798, *Die Geschichte der römischen und florentinischen Schule*, Rosenbusch, Göttingen.
- Firth J.R. 1957, *Modes of Meaning*, in *Papers in Linguistics 1934-1951*, Oxford University Press, London: 190-215.
- Flinz C. 2020, *Vergleichbare Spezialkorpora für den Tourismus: eine Chance für den Fachsprachenunterricht*, in Hepp M., Salzmann K. (eds.), *Sprachvergleich in der mehrsprachig orientierten DaF-Didaktik: Theorie und Praxis*, Istituto Italiano di Studi Germanici, Roma: 133-151.
- Flinz C., Buffagni C., Ballestracci S., Billero R., Farina A., Nicolás Martínez M.C. 2020, *Deutsche Lexik der Kunst auf der Basis des Korpus LBC*, FUP, Firenze. In pubblicazione, <<http://lexicon.lessicobeniculturali.net/de>>.
- Flinz C., Katelhön P. 2019, *Korpora nella didattica del tedesco come lingua straniera. Proposte per l'insegnamento del linguaggio specialistico del turismo*, «EL.LE», 8.2: 319-344.
- Gass K. E. 1961, *Pisaner Tagebuch. Aufzeichnungen / Briefe*, Lambert Schneider Verlag, Heidelberg.
- Goethe J.W. 1997, *Italienische Reise*, Deutscher Taschenbuch Verlag, München.
- Gombrich E.H. 2000, *La storia dell'arte raccontata da Ernst H. Gombrich*, Leonardo Arte, Milano.
- Gutenberg-Projekt*, <<https://www.projekt-gutenberg.org/index.html>>, (10.2020).
- Hausmann F.J. 1984, *Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen*, «Praxis des neusprachlichen Unterrichts», 31: 395-406.
- Heine H. 1969, *Florentinische Nächte*, Artemis & Winkler Verlag, München.
- Institut für Deutsche Sprache 2017, *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2017-I (Release vom 08.03.2017)*, Institut für Deutsche Sprache, Mannheim. PID: 10932/00-0373-23CD-C58F-FF01-3.
- Kilgariff A., Rychlý P., Smrz P., Tugwell D. 2004, *The Sketch Engine*, in Geoffrey W., Vessier S. (eds.), *Proceedings of the Eleventh EURALEX International Congress, Lorient, France July 6-10, 2004*, Faculté des Lettres et des Sciences Humaines, Université de Bretagne, Lorient: 99-104.
- Kilgariff A., Baisa V., Bušta J., Jakubiček M., Kovář V., Michelfeit J., Rychlý P., Suchomel V. 2014, *The Sketch Engine: ten years on*, «Lexicography», 1: 7-36.
- Klosa A. 2016, *Der lexikographische Prozess im Projekt elexiko*, in Hildenbrandt V., Klosa A. (eds.), *OPAL Online publizierte Arbeiten zur Linguistik, 1/2016, Lexikographische Prozesse bei Internetwörterbüchern*: 29-38.

- Kurz I. 1905, *Stadt des Lebens*, Hermann Seemann Nachfolger, Leipzig.
 — 1984, *Florentiner Novellen*, Phaidon, Essen.
 LBC-Korpusplattform, <<http://corpora.lessicobeniculturali.net/de/>>, (10.2020).
 Lemnitzer L., Zinsmeister H. 2015, *Korpuslinguistik. Eine Einführung*, Narr, Tübingen.
 Lessico dei Beni Culturali, <<http://corpora.lessicobeniculturali.net/>>, (10.2020).
 Lewald F. 1992, *Bilderbuch*, Urike Helmer Verlag, Frankfurt am Main.
 Machiavelli N. 1934, *Geschichte von Florenz*, Phaidon, Essen.
 — 1971, *Istorie Fiorentine*, Sansoni, Firenze.
 Moritz K.P. 2015, *Reisen eines Deutschen in Italien in den Jahren 1786 bis 1788*, e-artnow, o.O.
 Mukherjee J. 2002, *Korpuslinguistik und Englischunterricht. Eine Einführung*, Peter Lang, Berlin u.a.
 Reiners L. 1991, *Stilkunst. Ein Lehrbuch deutscher Prosa*, Beck, München.
 Rilke R.M. 1984, *Das Florenzer Tagebuch*, Suhrkamp, Frankfurt am Main.
 Ruskin J. 1897, *Wege zur Kunst. Eine Gedankenlese aus den Werken von John Ruskin*, Heitz und Mündel, Straßburg.
 Stendhal 1922, *Reise in Italien (Rome, Naples et Florence)*, Propyläen Verlag, Berlin.
 Vasari G. 1550, *Le vite de' più eccellenti architetti, pittori, et scultori italiani: da Cimabue in sino a tempi nostri*, Torrentino, Firenze. <<http://archiviovasari.beniculturali.it/index.php/1550-la-prima-edizione-de-le-vite/>>.
 — 1966, *Le Vite de' più eccellenti Pittori, Scultori et Architettori*, S.P.E.S/Sansoni, Firenze.
 — 2008, *Leben der ausgezeichnetsten Maler, Bildhauer und Baumeister von Cimabue bis zum Jahre 1567*, Directmedia Publishing, Berlin.
 Wackenroder W.H. 1991, *Herzensergiessungen eines kunstliebenden Klosterbruders*, in Vietta S., Littlejohns R. (eds.), *Sämtliche Werke und Briefe, Historisch-kritische Ausgabe*, Winter Universitätsverlag, Heidelberg.

Il corpus LBC francese¹

Annick Farina, Lorella Sini

Abstract (Italiano): In questo contributo illustriamo le caratteristiche principali ed alcuni esempi di applicazione del Corpus LBC francese, creato nell'ambito del progetto *Lessico multilingue dei Beni Culturali (LBC)*. In primo luogo, presentiamo i generi testuali, gli autori e le opere presenti nel corpus, descriviamo i criteri di selezione adottati per la loro scelta. Successivamente ci soffermiamo sui possibili usi del corpus, mettendo in evidenza come esso si presti ad analisi linguistiche. Chiudiamo il lavoro proponendo una valutazione dell'attuale configurazione del corpus con uno sguardo sulle prospettive di sviluppo futuro.

Parole chiave: linguistica dei corpora, lessico, terminologia, arte, beni culturali.

Abstract (English): This essay describes the main features and possible applications of the French LBC corpus, which has been compiled in the framework of *Lessico multilingue dei Beni Culturali (LBC)*. First, the authors illustrate the main features of the corpus, with a special attention to the selected text genres and authors and selection criteria. Various possible linguistic uses of the corpus illustrating some options of the software are then presented. An evaluation of the current corpus configuration and outlook conclude the work.

Keywords: Corpus Linguistics, Lexicon, Terminology, Art, Cultural Heritage.

Introduzione

Come è già stato accennato nella presentazione generale di questo volume e nelle presentazioni dei corpora delle altre lingue, il corpus LBC francese è stato concepito nell'ambito del progetto interuniversitario *Lessico dei Beni Culturali* con l'obiettivo di creare dizionari monolingue e bilingue collegati tra loro da collegamenti ipertestuali. Questi dizionari, come i corpora che li alimentano, includono diverse lingue: francese, inglese, italiano, spagnolo, russo, tedesco. Nascono per consentire sia a un pubblico specializzato che a un pubblico più ampio di svolgere ricerche, qualitative e quantitative, sul patrimonio artistico delle città di Firenze e della Toscana e, in una seconda fase, del patrimonio artistico italiano. I nostri strumenti – in particolare i corpora – si rivolgono infatti ad autori di guide o a curatori di mostre, commentatori o traduttori, al fine di offrire un supporto nelle loro scelte linguistiche. Il corpus consente infatti di accedere facilmente a una banca dati che può for-

¹ Questo contributo è una versione leggermente modificata e tradotta dal francese dell'introduzione al corpus francese presente sulla pagina del corpus <<http://corpora.lessicobeniculturali.net/fr/>>. Ringraziamo Fernando Funari per la sua traduzione.

nire informazioni affidabili come, ad esempio, equivalenti affidabili di determinati concetti, in particolare quelli relativi a tecniche e materiali artistici, o fornire loro informazioni sulla frequenza di determinate collocazioni o co-occorrenze speciali.

Per il corpus francese abbiamo raccolto, secondo gli stessi criteri delle altre lingue, un certo numero di documenti caratterizzati da una certa variazione tipologica, e redatti in un arco temporale esteso, dai celebri scritti sul Rinascimento fino a tempi moderni e contemporanei. Ciò dovrebbe facilitare, tra le altre cose, gli studi di semantica comparativa inter e intralinguistica sia in diacronia che in sincronia.

Il gruppo francese è stato in un certo senso un pioniere nello sviluppo del progetto, beneficiando dall'inizio di molte risorse, grazie al contributo di molti ricercatori e studenti-tirocinanti, che hanno collaborato a questo progetto sin dal 2013. Questo corpus – che deriva da un approccio «corpus-driven» e non «corpus-based» (Biber 2012) – ha infatti rappresentato un banco di prova per il progetto, che è servito da base di lavoro per le altre lingue, consentendo di evidenziare le difficoltà di gestione delle fonti e di creazione dei metadati. Incorpora alcuni elementi destinati a testare diversi aspetti (principalmente la validità delle nostre categorie lessicografiche), che vorremmo sviluppare nei prossimi anni per tutti i nostri corpora. Il corpus francese contiene quindi campioni di testi diversi da quelli trovati in altre lingue, sia in termini di tipologia testuale (dizionari, corrispondenza di artisti, ad esempio, ma anche siti web destinati ai turisti) che dal punto di vista dell'estensione del nostro principale oggetto – che ha finora guidato le nostre scelte – dal momento che abbiamo iniziato ad incorporare il patrimonio artistico italiano tangibile e intangibile nel patrimonio artistico toscano.

Concepito come un laboratorio didattico (Farina 2016), il progetto ha dato vita a numerose tesi di laurea e stage per studenti. In questo quadro sono stati creati dei subcorpora che sono stati riuniti nel corpus principale o che saranno incorporati nei prossimi mesi e anni². Allo stesso modo, i ricercatori che fanno parte del gruppo hanno lavorato su alcuni autori o temi legati al nostro progetto e per farlo hanno trascritto un certo numero di testi di cui il nostro corpus ha potuto beneficiare. Possiamo citare ad esempio il lavoro di Rosa Cetro (Università di Pisa) su A. Félibien, quello di Nicole Maroger (Università di Firenze) sul Palio di Siena, il progetto di ricerca attualmente sviluppato da Valeria Zotti (Università di Bologna) sui racconti dei viaggiatori francesi e francofoni a Bologna e quello di Dominique Garand (Università del Quebec a Montreal) che sta attualmente collaborando al progetto di raccolta di testi sull'Italia di autori quebecchesi. Quanto alle autrici di questo contributo, hanno raccolto e trascritto la maggior parte del materiale presente oggi nel corpus, passando mano mano in rassegna autori di tutte le epoche³.

² Molti studenti di Valeria Zotti e Annick Farina hanno lavorato ad esempio alle traduzioni di Vasari in francese. Una studentessa seguita da Annick Farina per il francese e Sabrina Ballestracci per il tedesco ha recentemente lavorato al confronto delle traduzioni della corrispondenza di Michelangelo in queste due lingue; le sue trascrizioni saranno presto incorporate. Abbiamo anche diretto diverse tesi sui testi dei viaggiatori francesi sia a Firenze (A. Farina) che a Bologna (V. Zotti).

³ Questi risultati sono serviti come base per la redazione di un libro su Firenze e le sue parole chiave attraverso gli autori che l'hanno descritta (Farina 2019).

Il nostro lavoro deve molto anche alle collaborazioni interuniversitarie franco-italiane come quelle con il *Laboratoire d'Informatique* dell'Università di Tours, il *Laboratoire UMR 5206 Triangle* del CNRS, il gruppo LiCoRN dell'Università della Bretagna-Sud, che hanno guidato la nostra metodologia sia per la costituzione del corpus che per i suoi futuri usi lessicografici.

1. Contenuto del corpus

La ricerca delle fonti è stata effettuata dapprima sulla base di bibliografie esistenti su diari di viaggio in Italia e opere di riferimento, relative alla storia dell'arte italiana e del Rinascimento in particolare. I primi testi integrati nel database erano classici di pubblico dominio e facilmente accessibili online. In particolare, abbiamo tratto una parte sostanziale del nostro corpus dalle risorse disponibili dalla biblioteca digitale Gallica: ad esempio, la traduzione di C. Weiss delle *Vite* di Vasari pubblicata nel 1900, e la biblioteca digitale archive.org, ad esempio l'opera *Des principes de l'architecture, de la sculpture, de la peinture...* di A. Félibien (1676), entrambi ancora in fase di trascrizione in fascicoli di edizione⁴. Abbiamo anche utilizzato opere presenti su Wikisource, che spesso abbiamo contribuito a trascrivere online, come ad esempio *La villa Palmieri* (1843) o *Une Année à Florence* (1841) di A. Dumas.

Questo spiega perché il nostro corpus è principalmente letterario (quasi il 60% del corpus). Contiene infatti sia i racconti di viaggio degli scrittori del Grand Tour e quelli che seguirono le loro orme più tardi fino all'inizio del XX secolo, sia opere di narrativa con Firenze o l'Italia come ambientazione, nonché scritti critici o amatoriali sull'arte o sulla cultura italiana. Sappiamo che una delle mete privilegiate di questi autori, soprattutto durante il periodo romantico, era la città di Firenze e la Toscana, cosa non sempre vera per i viaggiatori europei che spesso preferivano a loro Roma, Venezia o Napoli. Non c'è da stupirsi, quindi, se la maggior parte dei nostri testi sono tratti dalla letteratura di viaggio del XIX e della prima metà del XX secolo. Le numerose traduzioni che compaiono nel nostro corpus, anch'esse redatte nello stesso periodo, confermano il particolare interesse degli studiosi francesi per il patrimonio artistico italiano.

Tale interesse prefigura l'attuale predilezione per il turismo culturale in Italia. Anche se questo rappresenta, per il momento, una porzione minore del nostro corpus, testi più recenti e più contemporanei sono stati incorporati con il progredire della ricerca, in quanto il viaggio in Italia resta un tema molto vivo nella letteratura francese e francofona. Ha suscitato nel tempo e suscita tuttora numerose pubblicazioni che spaziano dai libri di cucina alle guide turistiche, dalle recensioni d'arte ai dizionari di pittori, passando dalla corrispondenza e dagli articoli

⁴ Citiamo in particolare la *Bibliographie italico-française...* (1886), la bibliografia del *Grand Tour* disponibile sul sito della Biblioteca Nazionale di Firenze, quella del sito del *Centre de Recherche sur la littérature des voyages*, il dossier tematico sui viaggi in Italia del sito *Gallica* della Biblioteca Nazionale di Francia e Castiglione Minischetti, Dotoli, Musnik (2002).

di giornale. Di seguito è riportato il dettaglio degli autori che compaiono nel nostro corpus con l'indicazione dei *token*⁵, in ordine decrescente.

Tabella 1. Numero di tokens per ciascun autore presente nel corpus in ordine decrescente.

Viollet Le Duc	853.133	Lafenestre, Georges	24.527	Revel, Adèle	4.588
Fernandez, Dominique	726.437	Perrot, George	20.886	Pommier, Édouard	4.443
SA6	264.073	Allais, Gustave	19.221	Schmitz, Isabelle	4.010
Vasari, Giorgio	252.132	Goncourt, J. et E.	19.194	Dufay, Henri	3.832
Stendhal	218.313	Félibien, André	18.894	Grimaldi, Edouard-F.	3.594
Dumas, Alexandre	168.272	Brosses, Charles de	18.316	Musset, Paul de	3.584
Colet, Louise	135.166	Michel, Émile	17.289	Jaucourt, Chevalier de	3.548
France, Anatole	103.614	Gautier, Théophile	15.326	Bard, Joseph	3.266
Faure, Élie	84.716	La Sizeranne, R. de	15.195	Pasquin, Antoine-C.	3.028
Lang, Jack	73.914	Renan, Ernest	14.695	Powell, Véronique G.	2.612
Bazin, René	73.301	Le Routard	14.463	Favre le Bret, Robert	2.556
Taine, Hippolyte	58.962	Montaigne, Michel de	13.057	Erdan, Andrea	2.470
Auzias, Dominique	55.897	Fréville, Valentin	12.120	Taillasson, Jean-J.	2.106
Cellini, Benvenuto	50.742	Lescure, O.	11.584	Nobecourt, Jacques	1.947
Grandgeorge, Gaston	48.737	Palustre, Léon	10.414	Prieur, Jean	1.909
Klaczko, Julian	44.386	Giono, Jean	10.135	Bertheroy, Jean	1.809
Feuillet, Michel	42.624	Méry, Joseph	9.976	Mallarmé, Camille	1.593
Meyer, Eric	40.029	Staël-Holstein, G.	9.152	Colombari, Colonel F.	1.578
Moran, Robert	34.705	Libri, Guglielmo	8.790	Creuzé de Lesser, A.	1.564
Maurel, André	32.942	Delacroix, Eugène	7.174	Beaugrand, Honoré	1.410
Rosov, Gricha	32.613	Wyzewa, Théodore de	5.299	De la Borie, G.	1.155
De Navenne, Ferdinand	30.664	Mallet, George	5.137	Gaboury, Louise	907
Chateaubriand, François-R.	29.599	Camus, Albert	5.033	Artusi, Pellegrino	388

⁵ In SketchEngine, il software utilizzato per il nostro corpus, i dati statistici sono forniti in numero di “token” e/o “parole”, a seconda dei tipi di ricerca. Un “token” è definito come «the smallest unit that a corpus consists of. A token normally refers to: a word form: going, trees, Mary, twenty-five... / punctuation: comma, dot, question mark, quotes.../ digit: 50,000.../ abbreviations, product names: 3M, i600, XP, FB...anything else between spaces / There are two types of tokens: words and nonwords. Corpora contain more tokens than words. Spaces are not tokens. A text is divided into tokens by a tool called tokenizer which is often specific for each language» <https://www.sketchengine.eu/my_keywords/token/> (consultato nov. 2020).

⁶ Senza autore.

Il corpus francese LBC comprende 3.164.995 parole, corrispondenti a 252 testi (libri, articoli o capitoli di libri) di cui 129 testi in lingua francese originale e 123 testi tradotti dall'italiano al francese⁷ che sono essenzialmente frammenti per consentire la futura creazione di basi parallele, secondo le seguenti proporzioni divise per secolo.

Tabella 2⁸. Distribuzione dei tokens per anno/secolo di pubblicazione.

Secolo	Tokens	Data di redazione	Tokens	Percentuale di tokens ⁹
XVI	13.057	1580	13.057	0,34%
XVII	18.894	1676	18.894	0,55%
XVIII	21.864	1739	18.316	0,63%
		1751	3.548	
		1801	1.564	
		1807	11.258	
		1815	5.137	
		1817	212.493	
		1825	582	
		1827	29.599	
		1829	2.691	
		1830	7.174	
		1832	879	
		1841	4.953	
		1842	3.028	
		1843	122.326	
XIX	1.796.410	1844	3.594	52,17%
		1849	14.695	
		1850	1.212	
		1853	13.808	

⁷ Si tratta delle traduzioni delle *Vite* di Vasari (sono già state incorporate 65 vite d'artista, il lavoro di trascrizione è ancora in corso per le altre vite), di Benvenuto Cellini, *La vita di Benvenuto Cellini* (Libri I e II), e una ricetta di Pellegrino Artusi. Questi testi rappresentano 303.262 parole, ossia circa il 10% del corpus.

⁸ Sono stati presi in considerazione solo testi 'originali'. Con i testi tradotti si pone il problema del momento della stesura del testo originale che non coincide con quello della stesura della traduzione e di conseguenza con le caratteristiche linguistiche della lingua utilizzata, che varia a seconda delle date. Nella consultazione dei dati è possibile filtrare questi testi sia sulla base dell'anno di scrittura dell'originale che dell'anno della sua traduzione.

⁹ Percentuale calcolata sul numero totale di token dei testi non tradotti presenti nel corpus (3.443.357).

		1854	857.977	
		1855	19.194	
		1862	135.166	
		1863	4.588	
		1864	58.962	
		1865	15.326	
XIX	1.796.410	1866	11.584	52,17%
		1867	247	
		1870	5.631	
		1878	20.886	
		1880	44.386	
		1888	141	
		1892	10.414	
		1894	176.915	
		1901	17.289	
		1903	30.664	
		1905	5.299	
		1906	18.896	
		1907	48.737	
		1909	19.221	
		1910	15.195	
		1913	34.751	
XXe	554.153	1923	84.716	16,09%
		1924	1.593	
		1934	1.909	
		1936	5.033	
		1938	2.556	
		1953	10.135	
		1969	1.947	
		1982	221.507	
		1994	34.705	
		2002	249.623	
		2007	2.612	
XXIe	1.038.979	2008	297.931	30,17%
		2012	113.943	
		2013	33.768	

		2015	907	
XXIe	1.038.979	2016	321.722	30,17%
		2017	18.473	

Da questa tabella si evince chiaramente che la lingua più rappresentata nel corpus è la lingua francese del XIX e XXI secolo cui corrisponde una maggiore accessibilità. I testi del XIX secolo sono infatti liberi dal diritto d'autore e i testi contemporanei facilmente accessibili online. È però necessario notare che non tutte le categorie di testi inseriti nel corpus sono distribuite in modo omogeneo sui diversi periodi attestati, come vedremo più avanti (cfr. Fig. 4), aspetto che dovrà essere preso in considerazione durante ricerche specifiche su generi testuali (vedi parte 2.)

I testi sono stati suddivisi in quattro categorie o generi: testi di divulgazione (DIV); dizionari (DIZ); testi tecnici destinati a specialisti (TEC); testi letterari (LET).

Tabella 3. Distribuzione dei documenti in base alle categorie (e sottocategorie) del corpus.

	Percentuale sul totale parole corpus	Sottocategoria	Num. documenti	Num. parole	Percentuale relativa sul totale parole corpus
DIV (divulgazione)	10,92%	BLG (blog)	1	27.029	0,85%
		RIV (riviste e giornali)	8	41.940	1,32%
		GUI (guide turistiche)	4	277.179	8,75%
DIZ (dizionario)	22,26%	MON (monolingue)	7	704.622	22,26%
		ARC (architettura)	13	24.290	0,76%
TEC (tecnico)	8,42%	ART (arti)	67	243.398	7,65%
		GAS (enogastronomia) ¹		321	0,01%
		BIO (biografie)	124	733.487	23,17%
LET (letterario)	58,35%	FIC (fiction)	10	524.493	16,57%
		SAG (saggi)	17	589.228	18,61%

1. La categoria [DIV] comprende testi di divulgazione destinati al grande pubblico, tratti principalmente da documenti recenti, come si può vedere nella Tabella 4). Questi testi sono suddivisi in tre sottocategorie: 1 testo sotto forma di blog partecipativo di appassionato di viaggi concepito come una nomenclatura [DIV_BLG]; 8 testi (XX e XXI) tratti da riviste specializzate

- e dalla stampa [DIV_RIV]; 4 guide turistiche di cui 1 del 1829 e le altre posteriori all'anno 2000 [DIV_GUI].
2. I dizionari ([DIZ]) comprendono 7 testi di cui 2 del XVIII secolo e 5 del XIX secolo (22,26% del corpus generale). Questa categoria è stata creata con l'obiettivo di stabilire una base di lavoro per la costituzione di futuri dizionari monolingue specializzati che comprendano definizioni, citazioni e riferimenti di origine. Come si vede, questa categoria è composta quasi esclusivamente dal *Dictionnaire raisonné de l'architecture française* di Viollet-le-Duc (peraltro non riportato integralmente nel nostro database che comprende 5 volumi su 11). Questo testo fondamentale per la cultura artistica francese, che occupa quasi un quarto dei nostri dati, può apparire sovradimensionato rispetto ad altri testi raccolti e ad altre lingue che non hanno beneficiato di una simile risorsa documentaria. Per questo abbiamo introdotto la possibilità di escludere questo sottocorpus dalle ricerche per evitare di falsare i risultati. Inoltre, 2 articoli dell'*Encyclopédie*: sono stati aggiunti l'articolo «Florence» e l'articolo «École florentine».
 3. Alcuni testi destinati ad un pubblico specializzato nello studio delle arti, sono stati classificati come «tecnici» ([TEC]). Tale categoria è composta da 81 documenti, suddivisi in tre categorie o sottocorpus: testi tecnici specializzati in architettura ([TEC_ARC]); testi relativi ad altre tecniche artistiche, come la scultura o la pittura ([TEC_ART]) ed infine testi riguardanti il campo dell'enogastronomia ([TEC_GAS]), specialità che si presume giocherà un ruolo importante nella costituzione del patrimonio gastronomico italiano. A questo scopo sono state raccolte, tra l'altro, le celebri opere dell'Artusi.
 4. Testi letterari ([LET]): in questa tipologia sono state classificate opere che hanno come ambientazione l'Italia e più in particolare la Toscana e la città di Firenze. Si tratta di 151 documenti che costituiscono più della metà del nostro corpus. Questa categoria è stata suddivisa in tre sub-corpora:
 - a. opere di carattere biografico ([LET_BIO]: 23% del corpus). La maggior parte sono costituite dai resoconti di viaggi in Italia dei grandi scrittori del XIX secolo: dal *Voyage en Italie* di Chateaubriand (1827), a *L'Italie des Voyages* di Ernest Renan (1849), passando da *Une année à Florence* di Dumas (1841). Allo stesso modo, abbiamo integrato la traduzione in francese della vita di Benvenuto Cellini (per il momento, libri I e II, traduzione del 1847 di Léopold Leclanché). Abbiamo già preparato la maggior parte delle lettere di Michelangelo (traduzione di Boyer d'Agen risalente al 1911) ma siamo in attesa di prendere decisioni sulla loro futura classificazione e in particolare sulla possibilità di integrare frammenti di una stessa opera in diverse categorie prima di renderle disponibili nel corpus (cfr. Conclusione). Vorremmo, in un prossimo futuro, aggiungere traduzioni dei taccuini di Leonardo da Vinci, a partire dal suo *Trattato sulla pittura*, e frammenti delle *Istorie fiorentine* di Machiavelli.
 - b. Opere relative alla narrativa ([LET_FIC]: 16,5% del corpus): gran parte di questo sotto-corpus è costituito da opere del XIX secolo tra cui il

famoso racconto di Madame de Staël, *Corinne ou l'Italie* (1807), in particolare *Le séjour à Florence*, dal libro XVIII.

- c. I saggi ([LET_SAG]: 18,6% del corpus) sono costituiti da 17 testi distribuiti in poco più di due secoli, dal Settecento ai giorni nostri. Questi includono, tra gli altri, l'*Histoire de la peinture en Italie* di Stendhal (1817) e alcuni articoli pubblicati nella *Revue des deux mondes* (1832, 1878, 1901, 1905, 1906). Abbiamo anche digitalizzato e poi processato mediante OCR il *Que-sais-je* sull'arte italiana (Michel Feuillet, 2008) ma anche i *Dictionnaire amoureux de l'Italie* di Dominique Fernandez (1997 e 2008). Questa sottocategoria è la più difficile da definire rispetto alle altre perché il patto di lettura potrebbe essere cambiato nel corso dei secoli. Un'opera che nell'Ottocento, infatti, era destinata a un pubblico di specialisti delle arti e delle tecniche artistiche può rivelarsi oggi una testimonianza letteraria lasciata in eredità da illustri critici d'arte dell'epoca.

Tabella 4. Percentuale di documenti in ciascuna categoria divisi per secolo.

	DIV (divulgazione)	DIZ (dizionario)	TEC (tecnico)	LET (letterario)
XVI	0	0	0	0,9%
XVII	0	0	80%	0
XVIII	0	28,6%	0	3,6%
XIX	8,3 %	71,4%	13,3%	61,6%
XX	16,7 %	0	6,7%	28,6%
XXI	75%	0	0	5,3%

2. Esempi di ricerche all'interno del corpus

Grazie alla partecipazione attiva al progetto di un informatico che ha ottenuto contratti di ricerca postdottorale (Riccardo Billero), siamo stati in grado di sviluppare sia un modello di gestione delle fonti che un modulo di ricerca *ad hoc* all'interno del corpus¹⁰. Illustreremo qui alcuni esempi di ricerche all'interno del corpus con diversi filtri e opzioni¹¹.

Nell'ambito di un confronto tra il corpus francese e quello tedesco (Farina, Flinz 2020), abbiamo intrapreso una ricerca sulle diverse denominazioni del monumento principale di Firenze, il Duomo di Santa Maria del Fiore, in entrambe le lingue secondo le tipologie di testi e la loro epoca di redazione.

¹⁰ L'interfaccia di consultazione del corpus è attualmente in italiano ma sarà presto disponibile nella lingua di ogni corpus.

¹¹ Sulla scelta del software NoSketchEngine e sulla gestione dei dati e dei metadati dei nostri corpora, si rimanda agli articoli di Riccardo Billero (Billero, Nicolás Martínez 2017 e Billero 2020).

Per verificare il numero di occorrenze del nome *Sainte-Marie-des-Fleurs* e i contesti in cui si trova, abbiamo dapprima effettuato la ricerca solo nei testi “in lingua originale” e nel sottocorpus di testi appartenenti alla categoria “Letterario” utilizzando il modulo di ricerca disponibile in “Search/Text types” (Figura 1):

Figura 1. Ricerca di concordanze contenenti *Sainte-Marie-des-Fleurs* nei testi in lingua originale della categoria «Letterario».

Line	Snippet	Occurrences
1	n' est pas ou ' une réputation . Elle est bel et bien intégrée dans les mœurs .	SAINTE-MARIE-DES-FLEURS . Notre premier soin , en arrivant à Florence , avait été de déposer aux palais
2	La nouvelle cathédrale commença donc de s' élever , sous la gracieuse invocation de	Sainte-Marie-des-Fleurs , nom qu' elle reçut , ditent les uns , en souvenir du champ de roses sur lequel
3	et sa Tour penché . Arnolfo di Lapo avait jeté sur la place du Dôme les fondemens de	Sainte-Marie-des-Fleurs ; Sienne avait élevé sa cathédrale au clocher rouge et noir , et y avait renfermé
4	s' élevaient en même temps que le Dôme de Sienne , le Campo-Santo de Pise , et le Dôme de	Sainte-Marie-des-Fleurs . Le commencement du XIIIe siècle avait bien vu naître les Hebelungen , et mourir
5	, qui était jour de fête , le cardinal Riario devait entendre la messe dans le Dôme de	Sainte-Marie-des-Fleurs , et comme il avait fait prévenir Laurent et Julien de cette solennité , il était
6	, en comparaison d' une statue de l' un ou d' un tableau de l' autre . En sortant de	Sainte-Marie-des-Fleurs , on se trouva juste en face d' une autre porte . C' est
7	venait de couronner de sa coupole le monument d' Arnolfo di Lapo . Enfin	Sainte-Marie-des-Fleurs était élevée depuis 1400 en esglise métropolitaine par le pape Martin V . lorsque l'
8	d' indulgence pour les créations de l' esprit humain . Il monta en chaire aussi à	Sainte-Marie-des-Fleurs , la cathédrale ornée du campanile de Giotto . Prêches virulents contre la
9	Allort . Pour peu que vous vous promenez , un sort de géomé lune , sur la place de	Sainte-Marie-des-Fleurs , à l' endroit où le Dante aimait à se reposer , et que vous regardiez ses jours de
10	Cellini , et de tous les peintres de l' école dite Florentine ! La cathédrale ,	Sainte-Marie-des-Fleurs , est une immense et hardie création de Brunelleschi ; sa campanille et son
11	que Lucca della Robbia et Donatello avaient exécutés pour la tribune de l' orgue , à	Sainte-Marie-des-Fleurs . Ces bas-reliefs furent enlevés et déposés , vers la fin du XVIIe siècle , à l'

Figura 2. Le undici occorrenze di *Sainte-Marie-des-Fleurs* nelle fonti letterarie del corpus.

Questa ricerca, confrontata con quelle dello stesso tipo nelle altre categorie testuali, poi con le altre forme (*Santa Maria Novella*, *Dôme* e *Duomo*) nelle diverse categorie ha permesso di constatare che, mentre il nome francese del monumento *Sainte-Marie-des-Fleurs* non è mai usato nei nostri testi divulgativi (nelle guide turistiche) né si trova nei testi in traduzione, appare in buona proporzione rispetto ad altri nomi nelle opere letterarie. L'incrocio

di questi dati con la datazione delle occorrenze (visualizzazione delle date di scrittura ottenute modificando le opzioni come mostrato in Figura 3) ha permesso di comprendere meglio la distribuzione della scelta di denominazione del monumento.

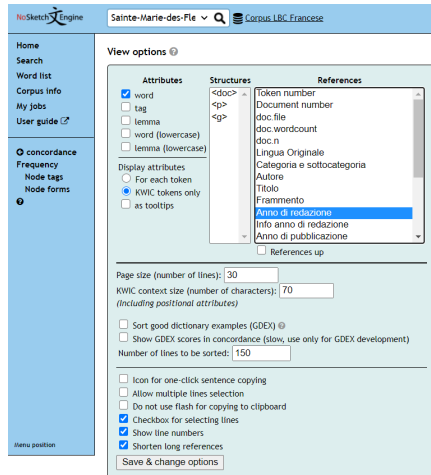


Figura 3. Modifica delle opzioni di visualizzazione dei risultati nel menu “View options”: qui, “anno di redazione”.

Figura 4. Visualizzazione delle occorrenze di Santa Maria del Fiore nella categoria “Letterario” del corpus, classificate per data di redazione (classificazione mediante l’opzione “Sort / References”).

Infatti, il 90% delle occorrenze di *Sainte-Marie-des-Fleurs* risalgono al XIX secolo e quasi l’80% di *Santa Maria del Fiore* sono del XX e XXI secolo. La dif-

ferenza di generi testuali ci è sembrata così un elemento rilevante per spiegare le differenze nella designazione. La designazione in lingua originale italiana si manifesta in particolare più frequentemente nelle guide turistiche, che optano per il mantenimento della forma originale dei nomi propri in generale: tale forma si ritrova anche nei testi in traduzione. Tuttavia, il periodo di redazione dei documenti non è meno importante. La datazione delle occorrenze ci permette quindi di compensare la mancanza di omogeneità nelle proporzioni date a ciascun periodo storico nelle diverse categorie testuali del nostro corpus, che abbiamo già illustrato nella parte 1.

Un'altra opzione estremamente utile per la nostra ricerca lessicografica è quella che ci permette di ottenere informazioni su collocazioni e co-occorrenze sulle diverse parole presenti nel corpus. È possibile effettuare questa ricerca in due modi diversi:

- a. utilizzando l'opzione "Collocation" si ottiene un elenco delle forme presenti a una distanza da definire nel menu di ricerca. Ad esempio, nella Figura 5, vediamo una ricerca delle parole che si trovano a una distanza di una o due parole a destra della parola «arc»:

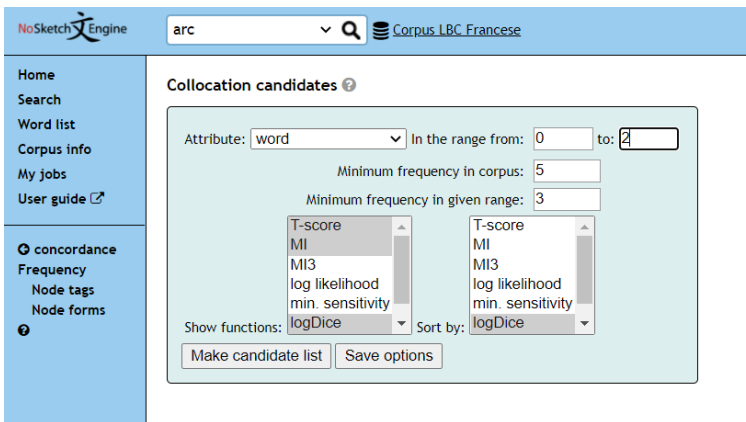


Figura 5. Ricerca di una lista di parole a una distanza di 1 o 2 parole a destra del lemma arc nel menu "Collocation".

Questa ricerca ci permette di ottenere sia collocazioni con struttura «arc + aggettivo o sostantivo», sia quelle con struttura ad «arc + preposizione + nome»: *arc ogive*, *arcs doubleaux*, *arc formeret*, etc. et *arc en ogive*, *arc en tiers-point*, *arc de vouête*, etc.

- b. Facendo una ricerca di concordanza su una parola, tramite l'opzione di visualizzazione "Sort/Left" o "Sort/Right", possiamo inoltre osservare le diverse collocazioni spostandoci tra le diverse pagine dei risultati, come presentato in Figura 7 (sempre nell'ambito di una ricerca di collocazioni con la parola «arc»).

Conclusione: analisi critica e sviluppi futuri

La ricerca di un equilibrio formale e di un bilanciamento dei dati tra così tante lingue diverse costituisce senz'altro una sfida. In questo senso, la scelta dei testi tradotti è stata una decisione collegiale dei gruppi di tutte le lingue. La ricerca delle traduzioni delle opere precedentemente elencate è stata infatti condotta sulla base della bibliografia di riferimento creata in comune. In questo modo sono state selezionate per i testi francesi le traduzioni di Stendhal e Taine in altre lingue, per i testi tedeschi di Burckhardt e Goethe, di Ruskin per l'inglese, ecc. Ma sono soprattutto le numerose traduzioni dell'italiano Giorgio Vasari (*Le Vite de' più eccellenti pittori, scultori, architettori*, 1550) nelle diverse lingue su cui abbiamo tutti lavorato, che permetteranno di costruire un database parallelo essenziale a partire dai nostri diversi corpora (cfr Zotti 2017). Per il francese, si tratta per il momento di 65 vite tra *Les vies des plus excellents peintres, sculpteurs et architectes*, secondo le traduzioni di L. Leclanché e C. Weiss, che potremo anche confrontare con quelle di A. Chastel. Questi testi, che hanno dato luogo a traduzioni e ritraduzioni da parte di specialisti di storia dell'arte nelle diverse lingue, non mancheranno di costituire uno strumento prezioso per la raccolta di equivalenti appartenenti ad un vocabolario specializzato in storia dell'arte. Costituiscono inoltre una ricca documentazione di informazioni sulle strategie traduttive adottate e sulla loro evoluzione nel tempo (cfr. Ballestracci 2018).

L'analisi dei testi nelle traduzioni che abbiamo voluto inserire nel nostro corpus ha inoltre permesso di individuare alcune approssimazioni che ne mettevano in discussione l'attendibilità. Se, infatti, abbiamo considerato i testi tradotti come aventi un valore e un interesse paragonabile ai testi di origine (Frawley 1984), alcuni tipi di testi rilevanti per la nostra ricerca (siti web, guide turistiche) hanno richiesto un controllo di affidabilità linguistica per essere inseriti a pieno titolo nel corpus. Abbiamo quindi effettuato una necessaria e preliminare selezione metodica, un vero e proprio 'controllo di qualità', almeno nella definizione di una prima versione pubblicata del corpus francese. Questa operazione consisteva in un confronto di un primo subcorpus di traduzioni con il corpus in «lingua naturale», che ci ha permesso di evidenziare anomalie nelle statistiche delle occorrenze confrontate e di eliminare alcuni testi (Farina, Billero 2018).

Il fatto che, a differenza di altre lingue, il corpus francese contenga gran parte dei dati sotto forma di dizionari (mentre la categoria [DIZ] è ancora assente nelle altre lingue), non dovrebbe sorprendere dal momento che la presenza riflette una tradizione accademica nazionale di cui abbiamo beneficiato. Abbiamo anche considerato l'introduzione di dizionari bilingue durante la creazione di corpora paralleli, che potrebbero aiutarci nella ricerca di equivalenti traduttivi.

Come abbiamo detto, il nostro obiettivo iniziale era quello di mettere insieme una tipologia testuale estesa sia in sincronia che in diacronia. Questa importante varietà di tipologie che abbraccia diversi secoli ha comportato alcune esitazioni riguardo alla classificazione dei subcorpora. Occorre distinguere tra testi tecnici appartenenti al campo dell'architettura e testi tecnici dell'arte? In particolare, il già citato dizionario di Viollet-le-Duc e quello di Félibien *Des prin-*

cipes de l'architecture, de la sculpture, de la peinture (1676) – quest'ultimo ancora in fase di 'pulizia' – potrebbero piuttosto essere classificati nella sottocategoria [TEC_ART] o [DIZ_TEC] (e non in [DIZ_MON] per la prima e in [TEC_ARC] per la seconda), vale a dire nella categoria dei testi specialistici che rientrano nel campo artistico in generale e/o che hanno una forma di dizionario. Questo senza dubbio ci porterà presto a fondere alcuni subcorpora.

Inoltre, il patto di lettura cambia in base allo sviluppo delle arti e delle tecniche e secondo i mezzi di comunicazione o secondo il medium utilizzato. I testi del Vasari erano originariamente testi destinati a specialisti dell'epoca e la loro diffusione in Europa consentì anche lo sviluppo di un lessico artistico specializzato in tutte le lingue ed in particolare in un francese tecnico che si è costituito sul modello dell'italiano a partire dal XVI secolo, per proseguire nei secoli successivi. Oggi, le *Vite* sono piuttosto considerate come raccolte di aneddoti sugli artisti e le loro opere. Allo stesso modo, le guide turistiche del XIX secolo dovrebbero essere classificate oggi, in base alla loro qualità letteraria, come diari di viaggio.

D'altronde, il limite che ci eravamo imposti di prendere in considerazione come prioritarie la città di Firenze e la Toscana, si sta spostando per andare a comprendere altre città italiane che hanno suscitato l'interesse dei viaggiatori stranieri nel corso dei secoli. Abbiamo così iniziato a introdurre opere, capitoli o frammenti relativi ad altre città, in particolare di Edmond e Jules de Goncourt (*L'Italie d'hier*, 1855) per le città di Venezia, Parma, Modena, Bologna, Verona ecc. Questa espansione ci ha permesso di testare la possibilità di una ricerca per città o regione.

Tutte queste osservazioni illustrano la peculiarità dei corpora che pubblichiamo oggi, che consideriamo un primo importante risultato dopo un percorso di ricerca che il gruppo LBC conduce già da quasi un decennio e che reca i segni di un *work in progress*. È ovvio che aggiunte e adattamenti trasformeranno questi corpora negli anni a venire, data la complessità del nostro lavoro in gruppi differenziati per lingua. La varietà degli usi che ne facciamo porterà a modifiche che saranno testate nelle nostre future ricerche.

Bibliografia

- Ballestracci S. 2018, *Tra norma e variazione: allora, ora e poi in Vite de' più eccellenti pittori scultori e architettori di Giorgio Vasari. Un'analisi in ottica contrastiva con il tedesco*, in Meli M. (ed.), *Le norme stabilite e infrante: saggi italo-tedeschi in prospettiva linguistica, letteraria e interculturale*, FUP, Firenze: 21-40.
- Biber D. 2012, *Corpus-based and Corpus-driven Analyses of Language Variation and Use*, in Heine B., Narrog H. (eds.), *The Oxford Handbook of Linguistic Analysis*, Oxford University Press, Oxford, <<https://doi.org/10.1093/oxfordhb/9780199544004.013.0008>>.
- Bibliothèque nationale de Florence, *Indici*, site sur le Grand Tour, <<https://grandtour.bncf.firenze.sbn.it/indici>> (11/2020).
- Billero R., Nicolás Martínez M.C. 2017, *Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, «CHIMERA Romance Corpora and Linguistic Studies», 4(2): 203-216.

- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Blanc J. 1886, *Bibliographie italico-française universelle, ou catalogue méthodique de tous les imprimés en langue française sur l'Italie ancienne et moderne, depuis l'origine de l'imprimerie (1475-1885)*, Paris.
- Centre de Recherche sur la littérature des voyages, <<http://www.crlv.org/>> (11/2020).
- Farina A. 2016, *Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique*, «Publif@rum», 24, <http://www.farum.it/publifarum/ezine_articles.php?art_id=335>.
- 2019, *Florence en V.O.*, Atlande, Paris.
- Farina A., Billero R. 2018, *Comparaison de corpus de langue «naturelle» et de langue «de traduction»: les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues*, in Izzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 108-116.
- Farina A., Flinz C. 2020, *Analisi linguistica comparativa dei corpora LBC. La visione del patrimonio fiorentino francese e tedesco: l'esempio del Duomo*, in Farina A., Funari F. (eds), *Past in Present / Le passé dans le présent / Il passato nel presente*, FUP, Firenze: 75-98.
- Frawley, W. 1984, *Prolegomenon to a theory of translation*, in William Frawley (ed.), *Translation: Literary, Linguistic and Philosophical Perspectives*, Univ. of Delaware Press, Newark: 159-175.
- Gallica (Bibliothèque nationale de France), *Dossier thématique sur les Voyages en Italie*, <<https://gallica.bnf.fr/dossiers/html/dossiers/VoyagesEnItalie/>> (10/2020).
- Minischetti V.C., Dotoli G., Musnik, R. 2002, *Bibliographie du voyage français en Italie du moyen âge à 1914*, Presses de L'Université de Paris Sorbonne/Schena editore, Paris/Fasano.
- Zotti, V. 2017, *L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle Vite di Vasari*, in Zotti V., Pano Alamán A. (eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*, FUP, Firenze: 105-134.

Fonti del corpus¹²

Testi originali francesi

- Allais G. 1909, *Lamartine en Toscane et les harmonies poétiques et religieuses (d'après la correspondance et les manuscrits)*, Société française d'imprimerie et de librairie, Paris.
- Auzias D., Labourdette J.-P. (2016), *Guide Petit Futé, Carnets de Voyage, Toscane*, Le Petit Futé, Paris.
- Bard J. 1854, *Turin, Gênes, Florence, Rome en 1854*, Timon frères, Vienne: 17-24.
- Bazin R. 1894, *Les Italiens d'aujourd'hui*, Calmann Lévy, Paris.

¹² Si tratta della bibliografia dei testi disponibili al mese di novembre 2020; alcuni sono ancora in fase di codifica prima della pubblicazione e saranno disponibili solo nei prossimi mesi. Nell'anno 2021 verrà inserito un sottocorpus contenente un gran numero di testi su Bologna, preparato da Valeria Zotti.

- Beaugrand H. 1889, *Quinzième lettre*, «Lettres de voyages», Presses de La Patrie, Montréal, 1889: 150-156.
- Bertheroy J. 1913, *La couronne d'épines*, «Le Gaulois», 26 septembre 1913, n. 13131: 3.
- Camus A. 1938, *Le désert*, in id., *Noces*, Gallimard, Paris.
- Colet, Louise 1862, *L'Italie des Italiens*, tome 2, E. Dentu, Paris:
- Chapitre II: 13-26.
 - Chapitre III: 26-36.
 - Chapitre IV: 36-102.
 - Chapitre V: 102-136.
 - Chapitre VI: 136-144
 - Chapitre VII: 144-153.
 - Chapitre VIII: 153-189.
 - Chapitre IX: 190-213
 - Chapitre X: 213-246.
 - Chapitre XI: 246-255.
 - Chapitre XII: 255-267.
 - Chapitre XIII: 268-281.
 - Chapitre XIV: 282-292.
- Colombari C.F. 1854, *Fête du 16 août à Sienne*, «L'Illustration, Journal Universel», n° 602, vol. XXIV, 9 septembre 1854: 1-2.
- de Brosses C. 1858, *Lettres familières écrites d'Italie à quelques amis, en 1739 et 1740, avec une étude littéraire et des notes par Hippolyte Babou*, vol.1, Poulet-Malassis et de Broise, Paris (ed. orig. 1739):
- *Lettre XXIV À M. De Quentin. Mémoire sur Florence*: 184-203.
 - *Lettre XXIII À M. De Blancey. Route de Bologne à Florence*: 204-210.
 - *Lettre XXVI À M. De Blancey. Route de Florence à Livourne*: 210-217.
 - *Lettre XXV À M. De Neuilly. Suite du séjour à Florence*: 229-236.
- de Chateaubriand F.-R. (1833), *Voyage en Italie*, «Œuvres», T. XIII, Paris (ed. orig. 1827), Lefèvre Libraire.
- de Goncourt J. et E. (1894), *L'Italie d'hier*, Charpentier & Fasquelle, Paris (ed. orig. 1855):
- *Domodossola*: 1-3.
 - *Milan*: 5-11.
 - *Brescia*: 13-14.
 - *Vérone*: 15-18.
 - *Venise*: 19-50.
 - *Padoue*: 53.
 - *Mantoue*: 55-56.
 - *Parme*: 57-59.
 - *Modène*: 61-62.
 - *Bologne*: 63-70.
 - *Pistoja*: 71.
 - *Florence*: 73-155.
- de Jaucourt Chevalier 1751-1765a, *Ecole Florentine*, in Diderot D., Le Rond d'Alembert J.-B., *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. 5, Briasson, Paris: 317-318.
- 1751-1765b, *Florence*, in Diderot D., Le Rond d'Alembert J.-B., *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. 6, Briasson, Paris: 876-877.

- de La Borie G. 2013, *Sienna vibre pour sa course de chevaux, le Palio*, <la-croix.com>, (11/08/2013).
- de La Sizeranne R. 1910, *Les Masques et les Visages-Portraits de Florentine, le long de la Seine et de l'Arno I. XVe siècle*, «Revue des Deux Mondes», t. 60, Paris: 160-190.
- de Lesser A.C.F. 1806, “La Toscane”, in *Voyage en Italie et en Sicilie*, Imprimerie de P. Didot l’aîné, Paris (ed. orig. 1801), 1806: 60-65.
- de Montaigne M. 1774, *De Pescia a Viterbe*, in *Journal du voyage de Michel de Montaigne en Italie, par la Suisse et l'Allemagne en 1580 et 1581, avec des notes par M. de Querlon*, Le Jay, Rome/Paris (ed. orig. 1580-1581).
- de Musset P. 1866, *Florence-Pise*, in Id., *Voyage en Italie et en Sicile, cinquième édition revue et corrigée*, Charpentier, Paris (ed. orig. 1843): 282-292.
- de Navenne F. 1903, *Entre le Tibre et l'Arno*, Paris, Plon:
- *Aux sources de l'Arno et du Tibre*: 1-28.
 - *À travers l'Apennin toscan*: 29-116.
- de Wyzewa T. 1905, *À propos d'une nouvelle biographie de Giotto*, «Revue des Deux Mondes», 5e période, tome 29, Paris: 459-468.
- Delacroix E. 1830, *Michel-Ange*, Revue de Paris, vol. XV, 1830, pp. 41-58.
- Dufay H. 1854, *Florence*, in id., *Souvenirs de vacances. Naples, Rome, Florence en 1853*, Dubiez, Senlis (ed. orig. 1853): 52-63.
- Dumas A. 1861, *Une année à Florence*, Michel Lévy, Paris (ed. orig. 1841):
- *Florence*: 168-176.
 - *La Pergola*: 177-186.
 - *Sainte-Marie-des-Fleurs*: 186-220.
 - *Le Palais Riccardi*: 220-235.
 - *Le Palais Vieux*: 235-258.
 - *La place du grand-duc*: 259-278.
- 1865, *La Villa Palmieri*, Michel Lévy, Paris (ed. orig. 1843).
- Erdan A. 1867, *Lettres d'Italie*, «Le Temps», Paris, 20 août 1867, n. 2292: 4.
- Faure É. 1924, *Histoire de l'art III: l'art renaissant (Nouvelle édition, revue et augmentée)*, G. Crès et Cie, Paris (ed. orig. 1923).
- Favre le Bret R. 1938, *Le 'Palio' de Sienna*, «Le Temps», Paris, 1^{er} juillet 1938, n° 28052: 5.
- Félibien A. 1676, *Des principes de l'architecture, de la sculpture, de la peinture, et des autres arts qui en dépendent. Avec un Dictionnaire des termes propres à chacun de ces arts*, Paris, J.-B. Coignard:
- Livre I – Chap. I: 1-8.
 - Livre I – Chap. II: 9.
 - Livre I – Chap. III: 9-13.
 - Livre I – Chap. IV: 14-17.
 - Livre I – Chap. V: 18-21.
 - Livre I – Chap. VI: 22-23.
 - Livre I – Chap. VII: 26-28.
 - Livre I – Chap. VIII: 30-31.
 - Livre I – Chap. IX: 33-35.
 - Livre I – Chap. X: 37-39.
 - Livre I – Chap. XI: 41-43.
 - Livre I – Chap. XII: 44-76.
- Fernandez D. 1982, *Dans la main de l'ange*, Grasset, Paris.
- 1997, *Le voyage d'Italie. Dictionnaire amoureux*, Paris, Plon:
- *François d'Assise*: 158-160.

- *Jardins*: 192-194.
 - *Laurent le Magnifique*: 197-199.
 - *Littérature*: 224-237.
 - *Sienna 1*: 576-582.
 - *Sienna 2*: 576-582.
- 2002, *La course à l'abîme*, Grasset, Fasquelle, Paris.
- 2008, *Dictionnaire amoureux de l'Italie de N à Z*, Plon, Paris.
- Feuillet M. 2009, *L'art italien*, «Que sais-je?», PUF.
- Florence 2017, *Guide du routard en ligne*, <www.routard.com> (consultation juin 2017).
- France A. 1894, *Le lys rouge*, Calmann-Lévy, Paris.
- Fréville V.¹³ 1872, *Vi. a M. et Madame Bar, a Paris*, «Les Apennins et la mer Adriatique», Limoges, Barbou Frères, 1872: 129-151.
- Gaboury L. 2015, *Italie, Le Palio: ils sont fous, ces Siennois!*, 11 avril 2015, <ledevoir.com> (21 février 2017).
- Gautier T. 1865, *Quand on voyage*, Michel Lévy frères, Paris:
 - *Florence*: 177-238.
 - *Venise*: 161-176.
- Germaine de Staël-Holstein A.-L. 1807, *Livre XVIII. Le séjour à Florence*, in Id., *Corinne, ou L'Italie*, Vol. 2, H. Nicolle, Paris: 370-409.
- Geoguide Toscane Ombrie* 2016, Gallimard Loisirs, Paris.
- Giono J. 1953, *Florence*, in id., *Voyage en Italie*, Folio, Gallimard, Paris, 1953.
- Grandgeorge G. 1909, *Toscane et Ombrie*, Librairie Plon, Paris (ed. orig. 1907):
 - Chapitre I: *Pise*: 1-29
 - Chapitre II: *Florence*: 30-177.
 - Chapitre IV: *Florence*: 242-248.
 - Chapitre V: *Sienna*: 249-285.
 - Chapitre VI: *Le retour*: 286-289.
- Grimaldi E.-F. 1850, *Dix jours à Gênes et huit jours à Florence*, E. Thunot, Paris (ed. orig. 1844).
- Klaczko J. 1880, *Causeries florentines: Dante et Michel-Ange. Béatrice et la poésie amoureuse*, E. Plon et Cie, Paris:
 - *Dante et Michel-Ange*: 1-79.
 - *Béatrice et la poésie amoureuse*: 81-155.
- Lafenestre G. 1882, *Chapitre 1. La sculpture italienne aux XIIIe et XIVe siècles*, in *Maitres anciens: études d'histoire et d'art*, H. Loones, Paris (ed. orig. 1870): 1-19.
- 1906, *La Crise de la beauté à Florence au XV^e siècle*, «Revue des Deux Mondes», 5^e période, tome 36: 528-565.
- Lang J., Colin L. 2012, *Michel-Ange*, Fayard, Paris.
- Lescure O. 1906, *Milan, 5 septembre 1866*, in F. Blétit, *Lettres familières sur Milan, Venise, Florence, etc.*, Paris (ed. orig. 1866): 3-29.
- Libri G. 1832, *Toscane – Modène – Parme*, «Revue des Deux Mondes», tome 7: 347-367.
- Mallarmé C. 1924, *Lettre d'Italie. Le 'Palio' siennois*, «Le Gaulois», Paris, 18 juillet 1924: 3-4.
- Mallet G. 1817, *Chapitre III – Florence*, in Paschoud J.J., *Voyage en Italie dans l'année, 1815*, Paris (ed. orig. 1815): 47-67.
- Maurel A. 1913, *Paysages d'Italie*, tome 1, Hachette, Paris:

¹³ Pseudonyme de Alfred Driou.

- I. *La fenêtre entr'ouverte: Volterra*: 1-23.
 - II. *Les fiancés de Sainte Catherine: Sienne*: 24-56.
 - III. *Près du sang: Montepulciano*: 57-75.
 - IV. *La leçon sur la place: Pienza*: 76-89.
 - V. *La fenêtre ouverte: Chiusi*: 90-103.
 - VI. *La fenêtre refermée: Corneto*: 104-129.
- Méry, Joseph 1868, *Les Nuits italiennes: contes nocturnes*, Michel Lévy Frère, Paris (ed. orig. 1853):
- *Livourne: La vallée de l'Arno*: 17-23.
 - *Un dimanche à Florence - La villa Catalani - L'album d'une reine*: 22-34.
 - *Les exilés à Florence*: 35-37.
 - *L'Atelier de Bartolini*: 45-56.
 - *Arrivée à Florence. Sur l'album de Madame Marini*: 127-128.
 - *Florence. A mon ami Adolphe Sturler, peintre d'histoire*: 129-137.
- Meyer E. (ed.) 2012, *La Renaissance italienne*, «Geo Art», Prisma media, Paris.
- Michel É. 1901, *Le Dessin chez Léonard de Vinci*, «Revue des Deux Mondes», 5e période, tome 1: 342-375.
- Moran R. 1994, *Secrets de peintres: apprêts, marouflage, médiums, pigments, glacis, vélatures, dorure, vernis*, Fleurus, Paris.
- Nobecourt J. 1969, *Le Palio, fête de la survivance*, «Le Monde», 8 octobre 1969.
- Palustre L. 1892, *Chapitre III, L'Architecture de la Renaissance*, Librairies-imprimeries réunies: 93-134.
- Perrot G. 1878, *Florence depuis l'annexion au royaume d'Italie*, «Revue des Deux Mondes», 3e période, tome 27: 111-149.
- Pommier É. 2016, *Notes sur la couleur dans la littérature artistique de la renaissance en Italie*, in Pigeaud J. (dir.), *La couleur, les couleurs. XIes Entretiens de La Garenne-Lemot*, Presses Universitaires de Rennes, Rennes: 81-90.
- Powell V.G. 2007, *Vie d'un autre artiste*, in Vasari G., *Vies des artistes* (traduction à partir de L. Leclanché de C. Weiss), Grasset, Les Cahiers Rouges, Paris: 7-13.
- Prieur J. 1934, *La Fête du Palio à Sienne*, «L'Auvergne Littéraire, Artistique & Historique», Clermont-Ferrand, n° 75: 37-40.
- Renan E. 1875, *Italie*, in Id., *Voyages, Italie (1849) – Norvège (1870)*, Éditions Mouton, Paris (ed. orig. 1849): 13-119.
- Revel A. 1863, *Florence*, in Id., *Naples, Rome et Florence, impressions de voyage*, Imprimerie de l'union typographique-éditrice, Turin: 55-69.
- Rosov G. 2013, *Petit dictionnaire des Beaux-Arts et de la Peinture à l'usage des artistes autodidactes, Version 1.0*, Index terminologique (consulté en 2017).
- [sa] 1829, II.I. *Route de Gênes à Lucques par la Rivière du Levant*, in *Nouveau guide du voyageur en Italie*, Sonzogno, Milan, vol. 1: 236-244.
- Schmitz I. 2017a, *Un été en Toscane: Florence, berceau du foot*, <Le Figaro.fr>, (06/06/2017).
- 2017b, *Un été en Toscane: le Palio de Sienne*, <Le Figaro.fr>, (04/06/2017).
- 2017c, *Un été en Toscane: les contrade, un modèle de société*, <Le Figaro.fr>, (05/06/2017).
- Stendhal 1854, *Histoire de la peinture en Italie*, Michel Lévy Frères, Paris (ed. orig. 1817).
- 1854, *Rome, Naples et Florence*, tome 2, Michel Lévy frères, Paris (ed. orig. 1817), (édition revue).
- 1935, *Lettre de Rome sur l'état actuel de la littérature italienne*, in Id., *Courrier anglais*, 4, le Divan, Paris (ed. orig. 1825): 232-257.

- Taillasson J.-J. 1807, *Léonard de Vinci*, in Id., *Observations sur quelques grands peintres*, Duminil-Lesueur, Paris: 1-9.
- Taine H. 1990, *Voyage en Italie II. D'Assise à Florence*, Éd. Complexe, Bruxelles (ed. orig. 1864).
- Viollet Le Duc 1854-1868, *Dictionnaire raisonné de l'architecture française du XIe au XVIe siècle*, Bance éditeur, Paris:
- Tome 1: *Préface V-XX.*
 - Tome 1: *Abaque-aronde.*
 - Tome 2: *Arts-chapiteaux.*
 - Tome 3: *Charnier-console.*
 - Tome 4: *Construction-cyborium.*
 - Tome 5: *Dais-fût.*

Testi tradotti dall'italiano

- Cellini B. 1881, *La vie de Benvenuto Cellini écrite par lui-même* (traduction de Léopold Leclanché), A. Quantin imprimeur-éditeur, Paris (ed. orig. 1558, trad. 1847): 188:
- Livre I, I: 3-4.
 - Livre I, II: 4-6.
 - Livre I, III: 7-10.
 - Livre I, IV: 10-11.
 - Livre I, V: 12-14.
 - Livre I, VI: 14-16.
 - Livre I, VII: 16-18.
 - Livre I, VIII: 18-19.
 - Livre I, IX: 19-23.
 - Livre I, X: 23-25.
 - Livre I, XI: 25-27.
 - Livre I, XII: 27-29.
 - Livre I, XIII: 30-34.
 - Livre I, XIV: 34-36.
 - Livre I, XV: 36-38.
 - Livre I, XVI: 38-40.
 - Livre I, XVII: 40-42.
 - Livre I, XVIII: 42-44.
 - Livre I, XIX: 44-48.
 - Livre I, XX: 48-50.
 - Livre I, XXI: 50-52.
 - Livre I, XXII: 52-53.
 - Livre I, XXIII: 54-56.
 - Livre I, XXIV: 57-61.
 - Livre I, XXV: 61-63.
 - Livre I, XXVI: 63-67.
 - Livre I, XXVII: 67-70.
 - Livre I, XXVIII: 70-72.
 - Livre I, XXIX: 72-76.
 - Livre I, XXX: 76-82.
 - Livre I, XXXI: 82-85.
 - Livre I, XXXII: 85-88.
 - Livre I, XXXIII: 89-94.

- Livre II,XXXIV: 95-99.
- Livre II,XXXV: 99-101.
- Livre II,XXXVI: 101-104.
- Livre II,XXXVII: 104-107.
- Livre II,XXXVIII: 107-110.
- Livre II,XXXIX: 110-112.
- Livre II,XL: 112-116.
- Livre II, XLI: 116-119.
- Livre II, XLII: 119-121.
- Livre II, XLIII: 121-124.
- Livre II, XLIV: 125-128.
- Livre II, XLV: 129-131.
- Livre II, XLVI: 132-135.
- Livre II, XLVII: 135-136.
- Livre II, XLVIII: 137-138.
- Livre II, XLIX: 139-141.
- Livre II, L: 141-142.
- Livre II, LI: 143-145.
- Livre II, LII: 145-149.
- Livre II, LIII: 149-151.
- Livre II, LIV: 151-152.
- Livre II, LV: 152-154.
- Livre II, LVI: 154-157.
- Livre II, LVII: 158-161.

Vasari G. 1839, *Giotto*, in *Vies des peintres, sculpteurs et architectes* (traduction Léopold Leclanché), Jeanron, Philippe-Auguste, Paris (ed. orig. 1568, trad. 1839), 42: 201-225.

— 1900, *Les vies des plus excellents peintres, sculpteurs, et architectes* (traduction de Weiss, Charles), Dorbon-Aine, Paris (ed. orig. 1568, trad. 1900).

- *Alesso Baldovinetti*: 389-391.
- *Andrea dal Castagno di Mugello et Domenico Viniziano*: 402-407.
- *Antonio et Bernardo Rossellino*: 418-421.
- *Antonio Filarete et Simone*: 364-367.
- *Arnolfo di Lapo*: 130-138.
- *Benozzo Gozzoli*: 413-416.
- *Berna*: 224-226.
- *Buonamico Buffalmacco*: 186-190.
- *Cosimo Rosselli*: 437-439.
- *Dello*: 277-279.
- *Desiderio da Settignano*: 422-423.
- *Don Bartolommeo, Abbé de San Clemente et Gherardo*: 439-443.
- *Don Lorenzo Monaco*: 247-249.
- *Donato*: 347-358.
- *Fra Giovanni da Fiesole*: 375-381.
- *Francesco di Giorgio et Lorenzo Vecchietto*: 416-418.
- *Gentile da Fabriano et Vittore Pisanello de Vérone*: 407-410.
- *Gherardo Starnina*: 243-245.
- *Giovanni Cimabue*: 123-129.
- *Giuliano da Maiano*: 367-370.

- *Jacopo della Quercia*: 267-274.
- *Jacopo, Giovanni et Gentile Bellini*: 430-436.
- *Leon-Batista Alberti*: 381-384.
- *Lippo*: 245-246.
- *Lorenzo Costa*: 427-430.
- *Lorenzo di Bicci*: 251-256.
- *Luca della Robbia*: 281-288.
- *Margaritone*: 154-157.
- *Masaccio*: 311-317.
- *Mino da Fiesole*: 424-427.
- *Niccola et Giovanni*: 138-149.
- *Pesello et Francesco Peselli*: 411-412.
- *Piero della Francesca*: 370-374.
- *Pietro Cavallini*: 193-195.
- *Pietro Laurati*: 179-181.
- *Simone et Lippo Memmi*: 195-202.
- *Spinello Aretino*: 235-243.
- *Stefano et Ugolino*: 176-179.
- *Taddeo Bartoli*: 249-251.
- *Vellano de Padoue*: 391-393.
- 2002), *Giotto*, in Id., *Vies d'artistes: édition bilingue français-italien* (traduction de Gérard Luciani), Gallimard, Paris (ed. orig. 1568, trad. 2002): 77-93.
- 2007, *Vies des artistes* (traduction à partir de L. Leclanché de C. Weiss), Grasset, Les Cahiers Rouges, Paris (ed. orig. 1568, trad. 1900).
 - *Andrea Pisano*: 38-42.
 - *Andrea Mantegna*: 161-167.
 - *Antonello da Messina*: 129-133.
 - *Antonio et Piero Pollaiuolo*.
 - *Cimabue*: 15-22.
 - *Domenico Ghirlandaio*: 134-147.
 - *Filippo Brunelleschi*: 75-109.
 - *Fra Filippo Lippi*: 393-401.
 - *Fra Giovanni da Fiesole (Fra Angelico)*: 122-128.
 - *Ghiberti*: 50-66.
 - *Giorgione da Castelfranco*: 194-198.
 - *Giotto*: 23-37.
 - *Giovannantonio da Verzellì*: 122-128.
 - *Jules Romain*: 322-335.
 - *Le Rosso*: 122-128.
 - *Le Titien*: 439-459.
 - *Léonard de Vinci*: 178-193.
 - *Marc Antonio*: 295-321.
 - *Masaccio*: 67-74.
 - *Michel-Ange*: 346-438.
 - *Paolo Uccello*: 43-49.
 - *Pietro Perugino*: 168-177.
 - *Raphaël d'Urbain*: 220-254.
 - *Sandro Botticelli*: 155-160.

Il corpus LBC russo

Valentina Rossi, Marcello Garzaniti, Natalia Zhukova

Abstract (Italiano): In questo contributo si presentano i criteri per la realizzazione, le fasi di sviluppo e le possibili applicazioni del Corpus LBC russo, creato nell'ambito del progetto Lessico multilingue dei Beni Culturali (LBC). Il lavoro si compone di tre parti: la prima illustra il percorso e gli obiettivi del progetto e le fasi della sua realizzazione. La seconda espone i criteri di formazione dei Corpora multilingue LBC e descrive i generi testuali, gli autori e i titoli presenti nel corpus russo, offrendo una prima valutazione dell'attuale configurazione del corpus. La terza presenta le caratteristiche specifiche dei corpora LBC e i possibili usi del corpus russo, mettendo in evidenza come esso si presti ad analisi di tipo diverso in aree di interesse come il turismo, la traduzione specializzata.

Parole chiave: linguistica di corpora, lessico, terminologia dell'arte, terminologia dei beni culturali.

Abstract (English): This essay presents the criteria for the realization, the development phases and the possible applications of the LBC Corpus in Russian, created in the framework of the Lexicon of Cultural Heritage (LBC) project. The work is divided into three sections. Section 1 illustrates path and objectives of the project and the phases of its realization. Section 2 sets out the criteria for the formation of multilingual LBC Corpora and describes the textual genres, authors and titles present in the Russian corpus, offering a first assessment of the current corpus configuration. Section 3 presents the specific characteristics of the LBC corpora and the possible uses of the Russian corpus, highlighting how it lends itself to different types of analysis, in areas of interest such as tourism and specialized translation.

Keywords: corpus linguistics, lexicon, art terminology, cultural heritage terminology.

1. Dalla ricerca lessicografica al corpus LBC russo (e ritorno)¹

Quando decidemmo con A. Farina di dare vita a Firenze all'Unità di ricerca dedicata al lessico dei beni culturali (LBC), perceivamo con chiarezza di dover affrontare un tema fondamentale sia nell'ambito dei nostri studi sia dei nostri insegnamenti di lingua. Questa problematica aveva infatti risvolti sensibili per l'attività di insegnamento, dal momento che molti nostri studenti trovava-

¹ Il primo paragrafo ha per autore Marcello Garzaniti, il secondo Valentina Rossi, il terzo Natalia Zhukova.

no impiego in settori che richiedevano queste competenze, a cominciare dalla professione di 'guida turistica'. Quanti si sono impegnati nel progetto dell'Unità di ricerca LBC, ufficialmente approvato dopo un faticoso iter il 7.10.2013, avevano in primo luogo la percezione che esisteva un iato profondo nella qualità della comunicazione linguistica fra le guide della città di Firenze d'uso corrente (per esempio Bonechi) e gli addetti ai lavori nell'ambito dei beni artistici, a cominciare dagli studiosi di storia dell'arte, che dominavano il lessico specialistico dell'italiano. Era poi sufficiente consultare i siti museali per constatare il monolinguisimo dominante con un inglese spesso elementare, espressione di traduzioni non specialistiche. Erano peraltro davanti agli occhi di tutti i risultati raggiunti dalla lessicografia sia in Italia sia all'estero che da tempo si stava orientando all'elaborazione di banche dati. Questo emerse chiaramente in una delle prime giornate di studio, organizzate dall'Unità di ricerca: "Lessico e Banche dati" (Firenze, Bologna 13.12.2013).

Le relazioni avviate con l'Accademia della Crusca hanno contribuito a favorire la percezione dei legami del lessico dei Beni Culturali con il patrimonio della lingua italiana e la tradizione del rinascimento fiorentino, in molteplici ambiti che andavano dall'architettura alla scienza culinaria, dalla letteratura alle competenze artigianali fino a toccare questioni relative a scienze estetiche e sociali, in cui beni materiali e immateriali appaiono profondamente connessi.

In genere nel nostro lavoro si verificava la correttezza del lessico ricorrendo a dizionari generalisti o, se disponibili, a dizionari specialistici, verificando la qualità della comunicazione linguistica nei diversi ambiti dei beni culturali. Per questo è stato importante in una fase iniziale individuare i dizionari disponibili nelle singole lingue per constatare come soprattutto i dizionari universali mostravano una serie di limiti e imprecisioni che rendevano questi strumenti scarsamente utili o persino dannosi per i traduttori e i mediatori linguistici. In ambito russo è sufficiente far riferimento a lemmi che appartenevano all'architettura religiosa per incontrare nei dizionari, eredi dell'epoca sovietica, notevoli difficoltà, conseguenza del lungo periodo di ateismo di Stato. Più in generale si poteva osservare la limitatezza degli esempi forniti che spesso non illustravano la ricchezza semantica di taluni termini e soprattutto determinati usi specifici. Ancor più gravi erano i problemi incontrati nell'uso dei dizionari bilingui.

Una prima verifica delle difficoltà del lavoro con i dizionari è stata svolta attraverso le ricerche condotte da alcuni studenti nelle loro tesi di laurea, con un'analisi delle guide della città di Firenze in lingua russa. Questi lavori sono stati oggetto di discussione e di riflessione nella sezione di russo della nostra Unità di ricerca, e più in generale con i colleghi delle altre lingue². Nell'ambito della lingua

² Offriamo di seguito l'elenco tesi che ho seguito in collaborazione con N.N. Zhukova, e più tardi con V. Rossi su guide in lingua russa o cataloghi di mostre. Fra parentesi è indicato l'anno accademico. Fra le tesi triennali: F. Cassari, *Analisi lessicale dei nomi comuni della Guida Bonechi Florencija. Gorod i ego šedevry* (2013-2014); C. Maccianti, *Forme grafiche e traduzioni nei nomi propri della Guida Florencija. Gorod i vse šedevry* (2013-2014); E. Guerra, *Analisi lessicale della versione russa dell'opuscolo La primavera del Rinascimento*, la

russo il lavoro è stato facilitato dalla possibilità di consultare dizionari classici e contemporanei in rete in alcuni portali disponibili al pubblico, consentendo di confrontare le definizioni offerte dai dizionari monolingue russi con i dizionari monolingue dell'italiano, a cominciare dal prestigioso dizionario Treccani. Questo confronto ha fatto inevitabilmente emergere non solo la limitatezza di questi strumenti, ma soprattutto la complessità dell'eredità culturale della Russia che guarda da una parte al mondo occidentale (con la mediazione francese e tedesca, ma anche italiana), mentre dall'altra rimane ancorata alla sua eredità bizantina che ha informato una parte significativa del lessico dei beni culturali, a cominciare per esempio dal lessico relativo all'iconografia.

Lo studio delle guide ha comportato peraltro l'individuazione dei complessi problemi connessi alla forma dei nomi propri nel processo di traslitterazione dall'alfabeto latino al cirillico (per es. le diverse ortografie di Santa Maria del Fiore) o di adattamento al russo, quando è possibile incontrare traduzioni per intero o parziali del sintagma che designa il nome proprio (per es. le diverse versioni di Porta Romana)³. In questo ambito diventava indispensabile consultare non più soltanto i dizionari, ma anche le enciclopedie, a cominciare dalla tradizionale Grande Enciclopedia sovietica fino a quella più famosa dell'epoca dell'impero russo chiamata con il nome dei suoi curatori, Brokgauz-Efron (1890-1907)⁴. Alle nostre ricerche si sono accompagnate quelle dei nostri laureandi che

scultura e le arti a Firenze 1400-1460 (2013-2014); M. Rinella, *Analisi lessicale e pratiche traduttive dei nomi dal russo all'italiano*. Le guide di Firenze di G. Casetta (2008) e di R. Nesti (2008) (2014-2015); V. Bracali, *La guida turistica Montekatini Terme* in lingua russa. *Analisi del lessico dei beni culturali* (2014-2015); M. Montagnani, *A spasso per un corpus di Firenze. La linguistica dei corpora applicata a due guide fiorentine* (2015-2016); L. Petruzzella, *L'architettura italiana in Russia: inquadramento storico e glossario della terminologia artistica e architettonica* (2016-2017). Fra le tesi magistrali: C. Maccianti, *Analisi del lessico dei beni culturali in un corpus delle guide di Firenze* (2014-2015); V. Barbini, *Comunicazione e divulgazione dei beni culturali medievali. Analisi della traduzione russa dell'articolo Spiritualità bizantina e latina nelle «Madonne» di Giotto* di Timothy Verdon (2014-2015). Connesse a questi lavori per l'approccio metodologico sono le seguenti tesi magistrali: G. Tarquini, *Il libro di cucina di S. A. Tolstaja. Un percorso gastronomico tra elementi autoctoni e alloctoni* (2014-2015); E. Guerra, *Il viaggiatore incantato* di N. S. Leskov. *Analisi della terminologia equestre russa, a confronto con le versioni italiana ed inglese* (2015-2016). La prima tirocinante LBC per la sezione di russo, Maccianti, ha elaborato sulla base dei lavori di tesi precedenti una lista dei nomi propri e comuni con le loro definizioni in italiano e in russo presenti nelle guide e ha messo a punto una bibliografia sulle guide turistiche russe di Firenze.

³ Sulla tematica N.N. Zhukova è intervenuta all'Accademia delle arti del disegno di Firenze con una relazione sul tema *Difficoltà nella trascrizione dei nomi italiani e complessità della traduzione del lessico storico-artistico in russo* (2014).

⁴ Sulla tematica si segnalano la relazione di M. Garzaniti *Guideline proposition for a multilingual cultural heritage dictionary of Florence*, presentata alla X Scuola internazionale di lessicografia "Life beyond Dictionaries" (12-14 settembre 2013; Garzaniti 2015, ma si veda anche il precedente Garzaniti, Farina 2013), e l'intervento di N.N. Zhukova, *Per una nuova guida di Firenze in lingua russa* presentato alla conferenza *La guida turistica come luogo d'incontro tra lessico e immagini dei Beni Culturali* (Pisa-Firenze, 11-12 giugno 2015) (Zhukova 2018).

hanno imparato a percepire le modalità e l'importanza delle trascrizioni e della resa dei nomi propri in russo che spesso, come sappiamo, è mutata nel tempo.

La complessa problematica legata ai dizionari e gli sviluppi più recenti della lessicografia hanno portato alla necessità di costruire delle banche dati sulla cui base si sarebbe potuto costruire un nuovo dizionario. La necessità di attingere alle fonti della formazione del lessico dei beni culturali ha condotto alla ricerca delle sue origini, all'epoca del rinascimento fiorentino e italiano, quando di fatto si è sviluppato il lessico della storia dell'arte con la composizione delle famose *Vite de' più eccellenti pittori, scultori e architettori italiani* di Giorgio Vasari (1550, 1568). In questo ambito è stato importante stabilire dei contatti con la Fondazione Memofonte, che possiede un'importante banca dati degli scritti di Vasari⁵. Le sue *Vite* possono essere considerate un punto di riferimento indispensabile per la formazione del lessico italiano dei beni culturali, che rappresentò per secoli il punto di riferimento per l'arte e l'architettura in Europa⁶.

Ci siamo resi immediatamente conto che le versioni delle *Vite* nelle diverse lingue potevano costituire una base importante per lo studio del lessico dei beni culturali. Nell'ambito del russo N.N. Zhukova, fra i primi membri dell'Unità, si incaricò di svolgere una ricerca che ha consentito di individuare più di una traduzione delle *Vite* a cominciare dal XIX secolo. Queste diverse versioni testimoniano non solo il processo di standardizzazione del lessico dell'arte e dei beni culturali nella lingua russa, ma più in generale danno conto dei diversi approcci alla traduzione che riflettono la complessità dell'eredità culturale della Russia. Grazie anche alla presenza di studenti di madrelingua russa è stato possibile sviluppare dei lavori di ricerca confrontando le diverse traduzioni in lingua russa delle *Vite* di Vasari, lavori che hanno condotto a interessanti risultati⁷ e che speriamo possano essere proseguiti in futuro da G. Siedina (membro dal 2018)⁸.

Fin dall'inizio dunque una delle banche dati fondamentali doveva contenere le versioni delle *Vite* di Vasari, come recita la prima formulazione del progetto dell'Unità di ricerca, in cui si proponeva di procedere «in due diverse direzioni: lo studio delle traduzioni di opere classiche del Rinascimento fiorentino a cominciare dalle traduzioni delle *Vite* del Vasari, come pure la creazione di una banca dati del lessico delle guide di Firenze nelle diverse lingue. In seguito si potrà prevedere di incrementare queste banche dati che serviranno a elaborare

⁵ Cfr. <<https://www.memofonte.it/ricerche/giorgio-vasari/>>.

⁶ Per una dettagliata presentazione alla questione si veda Motolese 2012.

⁷ Facciamo riferimento in particolare alle tesi magistrali di C. D'Ambrosi, A.G. Gabričevskij e la traduzione della *Vita di Giotto* di Vasari (2013-2014) e di E. Novikova, *La vita di Giotto* di Giorgio Vasari nelle traduzioni russe di Ju. N. Verchovskij (1933) e di A. I. Venediktov (1956) (2014-2015). Ai lavori sulle versioni del Vasari si possono aggiungere altri lavori di tesi magistrale che hanno affrontato tematiche importanti per la ricezione del lessico rinascimentale in Russia: V. Alderighi, *Leonardo in Russia: la ricezione della sua figura e del suo lessico* (2014-2015); S. Stefanini, *La civiltà del Rinascimento in Italia* di Jacob Burckhardt in Russia. Analisi lessicale della descrizione dell'uomo (2017-2018).

⁸ Sulla ricezione di Vasari in Russia si veda anche Carbone 2020.

delle schede lessicografiche che potranno essere consultate in un portale dedicato» (7.10.2013). Sulla base dunque di queste banche dati si sarebbe costruito un dizionario che sarebbe stato disponibile a tutti online con l'illustrazione dei lemmi, l'individuazione di una rete semantico-tematica e persino approfondimenti enciclopedici con la collaborazione di storici dell'arte e storici. Per queste ragioni furono condotti degli studi approfonditi per elaborare una scheda modello in cui dovevano confluire i materiali offerti dalle banche dati⁹.

In diversi seminari e convegni è stato presentato il progetto dell'Unità di ricerca, i suoi obiettivi principali e i primi risultati raggiunti. Fra questi si deve ricordare la succitata X Scuola internazionale di lessicografia (Ivanovo, Firenze, 12-14 settembre 2013), organizzata con O. Karpova (Università Ivanovo), a cui hanno partecipato diversi membri dell'Unità di ricerca, i cui atti sono stati pubblicati presso la Cambridge Scholars Publishing¹⁰. Questo evento ha segnato l'inizio della collaborazione con Karpova e la scuola lessicografica di Ivanovo¹¹ che abbiamo conosciuto nel corso dei seminari organizzati dalla Fondazione Romualdo Del Bianco, una fondazione fiorentina che da molti anni svolge un'opera meritoria di relazioni con il mondo accademico russo. Con il sostegno della Fondazione è stato organizzato il seminario internazionale "Heritage Lexicography as Supporting Tool for International Council on Monuments and Sites (ICOMOS)" (Firenze 21-23 luglio 2014), a cui hanno partecipato specialisti di diversa provenienza e di cui sono stati stampati gli atti¹².

Successivamente, nel corso del secondo incontro organizzato presso l'Accademia delle arti e del disegno (Firenze), dedicato al "Progetto per un lessico multilingue dei Beni Culturali. Comunicare il patrimonio artistico e culturale fiorentino" (Firenze, 27 gennaio 2015), insieme a M. Perotto e N.N. Zhukova, abbiamo presentato la relazione: *La traduzione del lessico storico-artistico dall'italiano al russo: complessità e problemi*. Nel frattempo si è concretizzato il progetto di una piattaforma in rete dedicata con la finalizzazione del contratto di edizione con la Firenze University Press per la pubblicazione del Portale digitale "Lessico per i Beni Culturali"¹³.

⁹ Nel corso della ricerca in questo ambito sono stati analizzati una serie di lemmi per testare la struttura della scheda informatica da utilizzare per la creazione del dizionario on-line. In particolare sono state preparate le schede: Cimabue, fiorentino, Giotto, pilastro, pittore, tavola, tempera, tondo. Tutti i membri della sezione, ma anche la tirocinante Maccianti hanno collaborato alla creazione e alla revisione delle medesime offrendo il loro contributo. Nel frattempo alla tabella dei dizionari della lingua russa, M. Perotto ha aggiunto l'indicazione di alcuni fra i più importanti portali dedicati alla tematica.

¹⁰ Karpova, Kartashkova 2015.

¹¹ Cfr. Karpova 2011. In particolare si segnala il progetto relativo a un nuovo dizionario dedicato alla descrizione del ruolo di Firenze nella creazione artistica di scrittori, pittori e altri artisti: *Florence in the Works of World Famous People: Encyclopedic Associative Dictionary for Guides and Tourists* dedicato agli studenti, ma molto interessante sul piano metodologico (Karpova 2009).

¹² Cfr. in particolare il nostro intervento sugli strumenti lessicografici Garzaniti 2014.

¹³ Cfr. a questo proposito il paragrafo successivo, in particolare sulla sezione del portale che riguarda la lingua russa. In quel periodo si era progettata con N.N. Zhukova la presentazio-

Nell'ambito della sezione dell'Unità di ricerca dedicata alla lingua russa si sono sviluppate nuove collaborazioni con i colleghi di istituzioni accademiche in Russia, fra cui si segnala l'Istituto della lingua russa V.V. Vinogradov, diretto allora da A. Moldovan. Questo ha consentito di evidenziare rispetto alle altre sezioni i particolari approcci metodologici sviluppati dalla lessicografia russa, a cominciare da una percezione più viva della diacronia con studi avanzati di lessicografia storica del russo, connessi alle nostre ricerche in ambito filologico e storico-culturale, fino alla riflessione sulle strette relazioni fra lingua e cultura testimoniata in particolare dallo sviluppo della teoria della cosiddetta «visione del mondo» (*kartina mira*)¹⁴.

Nel corso del 2014 la sezione di russo si è arricchita della collaborazione di M. Perotto dell'Università di Bologna, autrice di importanti lavori in ambito socio-linguistico, che ha fatto il suo ingresso nell'Unità nel luglio 2014, collaborando alla redazione delle schede, e che è molto attiva nelle iniziative promosse in collaborazione fra le nostre università. Nell'ottobre 2014 ha iniziato a collaborare con noi F. Rossi, che abbiamo invitato nella nostra Unità per le sue particolari competenze grazie alla sua formazione nell'ambito delle Belle arti e dell'architettura sia in Italia che in Russia, dove per anni è stata docente al prestigioso MArchI (Mosca), mentre ora è collaboratrice del Kunsthistorisches Institut di Firenze. In particolare per la formazione del lessico dei beni culturali in Russia sono importanti i suoi studi sull'eredità di Palladio nell'architettura russa¹⁵.

Particolarmente utile sul piano didattico è stato il volume dedicato ai *Mille anni di architettura italiana in Russia*, a cui ha collaborato F. Rossi, come autrice e traduttrice¹⁶. Dal momento che si tratta di un volume pubblicato nelle due lingue sotto l'egida della nostra ambasciata a Mosca, agli studenti sono stati forniti i singoli capitoli in lingua russa e italiana proponendo di individuare in ciascun articolo il lessico dei beni culturali e di cercarne le definizioni presenti nei dizionari più rilevanti della lingua russa e della lingua italiana¹⁷. Confrontando

ne della parte russa del portale al convegno MAPRIAL (Granada 2015), che però non si è potuto realizzare.

¹⁴ Cfr. per un'introduzione Apresjan 2006. Per un approccio critico cfr. Gebert 2006.

¹⁵ Cfr. Rossi 2010; Rossi 2013.

¹⁶ Švidkovskij, Batalov, Rossi 2013.

¹⁷ Si fa riferimento alle seguenti tesi triennali: S. Ippoliti, *Mosca e l'architettura del Rinascimento di D. Švidkovskij: un'analisi lessicografica* (2017-2018); J. Platia, *Dal primo Novecento agli anni trenta di F. Rossi: parole e concetti della cultura italiana nell'architettura russa e sovietica* (2018-2019); G. Cecchi, *Da Pietro il Grande a F.B. Rastrelli di D. Švidkovskij: parole e concetti di architettura nel Settecento russo* (2018-2019); M. Baldini, *Il tema italiano nell'architettura russa dell'Ottocento e di inizio Novecento di Dimitrij Švidkovskij: analisi del lessico architettonico* (2018-2019); L. Cruciatà, *La creatività rinascimentale nell'architettura russa di A. Batalov: analisi semantica e lessicografica del linguaggio architettonico* (2019-2020); A. Giovannetti, *Il lombardo romanico nell'architettura di Vladimir di D. Švidkovskij. Parole e concetti per descrivere l'arte medievale russa* (2019-2020); A. Guidalotti, *Al tempo di Caterina II di D. Švidkovskij. Analisi della terminologia architettonica* (2019-2020). Fra le tesi magistrali: F. Cassari, *Il Cremlino, una fortezza d'arte italiana e russa* (2015-2016). A queste tesi si devono aggiungere alcuni lavori simili per tema-

le diverse definizioni dei lemmi e con l'aiuto del contesto in cui questi si incontrano, è stato possibile individuare i termini che creano difficoltà nel passaggio da una lingua all'altra, soprattutto per la diversa area semantica. Si prospetta in futuro una ricerca che raccolga i risultati dell'analisi dei singoli capitoli e che potrebbe rappresentare la base di un nuovo dizionario di architettura russo-italiano, italiano-russo.

Grazie alla summenzionata Fondazione Del Bianco, siamo entrati in contatto anche con E.B. Genieva, direttrice della Biblioteca statale delle letterature straniere "M.I. Rudomino" (Mosca), che nel 2013 ha offerto alla nostra Biblioteca umanistica una serie di volumi in lingua russa pubblicati dal Rudomino, fra cui due volumi importanti, intitolati *Viaggio a Firenze*, che raccolgono le testimonianze della letteratura russa sulla nostra città¹⁸. Non meno importanti sono stati i contatti con l'Istituto di filosofia dell'Accademia delle scienze, in particolare con A. Kara-Murza, autore di fortunati saggi e guide dedicate alla presenza russa nella città italiane, fra cui ovviamente spicca Firenze, e O.I. Kusenko, che si occupa delle ricerche degli studiosi russi nell'ambito del Rinascimento italiano. È importante infatti sottolineare che lo studio del lessico dei beni culturali per quanto riguarda la Russia è inevitabilmente intrecciato con la questione più complessa della ricezione dell'umanesimo e del Rinascimento in Russia e con la presenza dell'emigrazione russa in Italia dopo la Rivoluzione. Si tratta di un tema di fondamentale importanza per la storia della cultura russa a cui sono state dedicate una serie di ricerche¹⁹. Nell'ambito dell'attività didattica proprio in questa direzione abbiamo promosso lo studio del lessico fondamentale che caratterizza la versione russa del classico saggio di J. Burckhardt sul Rinascimento²⁰.

A partire dal 2016, tenendo conto delle difficoltà legate alla creazione delle schede lessicali che si basavano principalmente sui dizionari e sulla spinta di C. Nicolás Martínez con la sua pluriennale esperienza nell'ambito della linguistica dei corpora, l'Unità di ricerca si è orientata alla creazione di un proprio corpus del lessico dei beni culturali. Personalmente abbiamo potuto verificare la correttezza e la fruttuosità del metodo proposto, quando abbiamo assegnato a un nostro laureando, diplomato in informatica, il compito di creare un limitato corpus di guide di Firenze in lingua russa adottando il programma SketchEngine per analizzare il lessico in esso contenuto²¹. È stato assolutamente evidente

tica e metodologia: A. Gennaioli, *I termini dell'architettura ecclesiastica russa* (2012-2013); E. Moroni, *Gli architetti italiani in Russia tra fine Quattrocento ed inizi Cinquecento* (2015-2016).

¹⁸ Genieva 2009; Genieva, Fridštejn 2009.

¹⁹ Cfr. Tonini 2012.

²⁰ Cfr. *supra* n.7.

²¹ La succitata tesi di Montagnani ha applicato il software SketchEngine per l'analisi di un piccolo corpus formato da alcune guide turistiche di Firenze in lingua russa (Nesti, Casetta), rendendo disponibile a studenti e colleghi un PowerPoint di istruzione del programma in vista di ulteriori ricerche. Questa metodologia è stata applicata nella succitata tesi magistrale di Maccianti per analizzare il lessico dei beni culturali fiorentini in lingua russa sulla base delle guide Bonechi e Touring Club Italiano in versione russa.

il salto di qualità che l'adozione dei nuovi programmi informatici e della linguistica dei corpora può portare all'analisi del lessico. Del resto da questo punto di vista avevamo ormai sotto gli occhi uno strumento fondamentale per la lessicografia russa, da quando è stato creato il Corpus nazionale della lingua russa (<<https://ruscorpora.ru>>), che sul modello di altri corpora raccoglie oramai una quantità straordinaria di testi che possono essere interrogati sviluppando ricerche sia in ambito lessicale che in ambito grammaticale. Si tratta comunque di un *mare magnum* che fra l'altro comprende anche corpora dei testi medievali e moderni che hanno preceduto la nascita del russo standard. Fra i promotori del Corpus nazionale si annovera anche il succitato Istituto della lingua russa.

La creazione delle banche dati che raccolgono testi storici artistici e letterari su Firenze e Toscana, che dovrà essere la base di partenza per un innovativo Dizionario plurilingue destinato a specialisti di storia dell'arte e lessicografia, ma anche a guide turistiche e traduttori, è stato realizzato sotto la guida di Nicolás Martínez, Farina e R. Billero²². Nel momento in cui è stata decisa la creazione di banche dati delle diverse lingue si è costituito all'interno della nostra sezione un gruppo di lavoro guidato da V. Rossi (che collaborava già dall'inizio del 2016), insieme a M. Perotto, N.N. Zhukova e M. Gasanova Mijat. Il gruppo ha elaborato, anche con l'aiuto di alcuni tirocinanti (Bucelli, Gorbacheva e Lichacheva) una serie di testi che apparentemente possono apparire non omogenei, ma che in realtà riflettono proprio le ricerche e le indagini che sono state condotte negli anni precedenti e che abbiamo voluto esporre nelle pagine precedenti. Si spiega così la centralità delle versioni russe delle *Vite* del Vasari, ma allo stesso tempo la presenza di alcuni dei testi contenuti nel volume *Mille anni di architettura italiana in Russia*, le testimonianze dei viaggiatori russi a Firenze, ma di questo si parlerà più in dettaglio nel paragrafo successivo.

Ci limitiamo solo a sottolineare l'importanza di attingere a banche dati specifiche per lo studio dell'lessico dei beni culturali, come abbiamo potuto constatare quando abbiamo sviluppato la nostra ricerca sul termine *frjaz* che designa gli architetti italiani attivi in Russia fra la fine del Quattrocento e l'inizio del Cinquecento, uno studio presentato al Convegno internazionale LBC – CeSLiC “Comunicare il patrimonio artistico e culturale nella società digitale plurilingue” (Bologna, 13-14 dicembre 2018)²³. La sua presenza nelle fonti presenti nella banca dati LBC russo consente di orientare in maniera molto più precisa la ricerca nell'immenso Corpus nazionale della lingua russa²⁴.

2. I corpora multilingue LBC e il corpus russo

L'Unità di ricerca sul *Lessico multilingue dei Beni Culturali (LBC)*, fra gli obiettivi del progetto, si prefiggeva: la realizzazione di una piattaforma web di rife-

²² Cfr. per una presentazione generale Billero, Nicolás Martínez 2017; Billero, Farina 2018.

²³ Garzaniti 2020.

²⁴ Ivi.

rimento per l'uso e lo studio del lessico dei Beni culturali; la creazione di una banca dati con i corpora delle lingue coinvolte nel progetto; la compilazione di un dizionario multilingue di termini che utilizzi come risorsa la banca dati²⁵. Uno degli obiettivi principali del progetto del gruppo di ricerca LBC infatti «est de constituer des dictionnaires monolingues de huit langues différentes en fonction d'un usage précis relié à un objet particulier: la description (et traduction de descriptions) du patrimoine toscan principalement dans des textes de vulgarisation (guides touristiques, sites de musées, etc.)»²⁶.

Primariamente per questo scopo sono state create delle banche dati testuali nelle diverse lingue del progetto²⁷, che si andranno mano a mano completando nella forma di un *Work in progress* e che possono essere utilizzate per svariati compiti: dalla creazione di nomenclature alla stesura di schede lessicografico/terminologiche monolingui e di schede di traduzione che colleghino tra loro le nomenclature delle diverse lingue (per il momento solo per la traduzione dall'italiano verso le altre otto lingue del progetto)²⁸.

L'obiettivo principale del progetto LBC è colmare la mancanza di strumenti adeguati e di una formazione adeguata per la diffusione delle conoscenze sul patrimonio artistico, promuovendo così lo sviluppo e la realizzazione di studi, ricerche e altre attività incentrate sul lessico delle diverse lingue in relazione alla lingua italiana nell'ambito del patrimonio artistico e culturale, a partire da quello della città di Firenze. In tale prospettiva è prevista a breve anche la pubblicazione dei «lessici essenziali del linguaggio dell'arte» del gruppo di ricerca LBC, che rappresentano un'altra tappa di questo progetto a lungo termine: a partire dai corpora LBC le singole sezioni hanno estratto dal loro corpus un elenco di parole rappresentative della lingua dell'arte italiana, con relativi contesti d'uso (concordanze).

I testi di maggiore rilievo che si è ritenuto dovessero essere inseriti per primi all'interno delle banche dati di tutte le lingue sono le traduzioni di una serie di opere in lingua italiana, tali da poter essere considerate testi “fondatori” per la creazione di una terminologia dell'arte o testi classici dell'arte italiana²⁹. La realizzazione di corpora multilingue costituiti in parte consistente da traduzioni di quei testi “canonici”, e *in primis* delle *Vite* di Vasari, prevede la possibilità di sfruttare l'analisi delle traduzioni come utile risorsa per arrivare a creare le sezioni lessicali bilingui dei futuri Dizionari LBC³⁰.

²⁵ Sulle origini del progetto LBC si veda Garzaniti, Farina 2013: 504 e ss. Per il progetto del *Dizionario* relativamente alla lingua russa si veda Garzaniti 2015: 138 e ss.

²⁶ Farina, Billero 2018: 109. Per maggiori informazioni e una descrizione completa del progetto vedi il link <<http://www.lessicobenculturali.net/contenuti/il-progetto/818>>.

²⁷ Per una descrizione dettagliata della creazione della banca dati si veda Billero 2020.

²⁸ Le lingue del progetto sono: cinese, francese, inglese, tedesco, italiano, portoghese, russo, spagnolo e turco.

²⁹ Billero, Nicolás Martínez 2017: 210.

³⁰ V. Farina, Billero 2018: 108; Zotti 2017: 107.

Una banca dati, quindi, che è composta dai corpora delle diverse lingue coinvolte nel progetto, corpora contenenti testi che illustrano il patrimonio culturale: si va da testi tecnici sulla storia dell'arte a libri divulgativi, come le guide turistiche e i libri di viaggio, che trattano l'arte e la cultura italiana (*ibidem*). I criteri di selezione dei testi sono stati: la rilevanza storico-culturale dell'opera nell'ambito specifico di studio (es. i testi di Vitruvio o Leonardo); la diffusione internazionale di un'opera in rapporto all'ambito di studio (es. le *Vite* di Vasari); il prestigio dato a livello internazionale al patrimonio italiano da parte di un'opera (es. i testi di Stendhal o Ruskin); la specificità dell'argomento in rapporto alla storia dell'arte italiana ed in particolare della Toscana (es. la *Geschichte der Renaissance* di Burckhardt)³¹.

I corpora delle diverse lingue sono perciò eterogenei, composti da testi di diverso tipo, appartenenti ad epoche diverse (racconti di viaggio, guide turistiche, saggi sull'arte italiana, biografie di artisti...), scelti però per un comune interesse per il patrimonio italiano. Il connubio di queste diverse tipologie di testi, anche se può apparire contraddittorio, è in realtà fondamentale per studiare l'attuale stato del lessico dei beni culturali proprio in relazione alla sua formazione nelle origini e alla creazione di uno standard ben riconoscibile.

La banca dati LBC³² consente una consultazione mirata dei diversi corpora e sub-corpora, permettendo di operare confronti, quali, ad esempio, "italiano lingua originale" vs. "russo lingua originale", oppure "russo lingua originale" vs. "russo lingua di traduzione", oppure "russo di registro tecnico" vs. "russo di registro letterario". Queste operazioni di confronto possono aiutare a identificare le approssimazioni o gli errori di traduzione e costruire così informazioni lessicografiche comparative di qualità³³.

Questa banca dati multilingue è stata progettata per avere un ampio uso specialistico, sulla base di tre requisiti fondamentali: «disporre di materiale linguistico per documentare le voci del dizionario multilingue LBC; disporre di materiale linguistico per effettuare studi linguistici, letterali o culturali; disporre di testi con i quali far conoscere al grande pubblico il patrimonio culturale di Firenze e della Toscana»³⁴.

Il metodo scelto per la realizzazione dei corpora, comune a tutte le squadre linguistiche del gruppo LBC, è stato concepito, come ben sintetizzano E. Carpi e A. Pano Alamán, a partire da alcuni aspetti propri della lingua dell'arte come linguaggio settoriale: da un lato, «la genesi e l'evoluzione del linguaggio a fini artistici spiegano perché le parole legate all'arte e alla critica d'arte siano lontane

³¹ Billero, Nicolás Martínez 2017: 208. In particolare, sono state individuate le seguenti categorie e sottocategorie: Divulgativo – Blog, Divulgativo – Guida, Divulgativo – Rivista, Dizionario – Monolingue, Dizionario – Bilingue/plurilingue, Letterario – Biografico, Letterario – Fiction, Letterario – Saggistica, Tecnico – Architettura, Tecnico – Arte, Tecnico – Edilizia, Tecnico – Enogastronomia, Tecnico – Storia.

³² Cfr. <<http://corpora.lessicobenculturali.net/ru/>>.

³³ Farina, Billero 2018: 109.

³⁴ Billero, Nicolás Martínez 2017: 206.

dal costituire un insieme di termini ed espressioni univoche»; dall'altro, «il lessico dell'arte si colloca a metà tra i linguaggi specialistici e il linguaggio comune, e presenta un alto grado di variazione a seconda del tipo di testo in cui si trova»³⁵.

Su questa base, «e tenendo conto della natura eterogenea del lessico artistico e della varietà di tipologie discorsive in cui si manifesta» i gruppi di ricerca del Progetto LBC hanno stabilito due grandi blocchi tematici. In una prima fase sono stati raccolti testi relativi alla città di Firenze quale città d'arte per eccellenza (opere originali o tradotte dall'italiano, dal francese, dal tedesco e dall'inglese, pubblicate tra il XVI e il XX secolo): «questi testi sono stati selezionati per la loro rilevanza culturale e interculturale: si tratta infatti di opere letterario-biografiche e di viaggio, che consentono di comprendere i diversi punti di vista legati all'arte italiana e di ampliare così la scelta terminologica». In una seconda fase è stata raccolta una serie di testi tecnici sull'architettura e l'arte: «in questo caso, la priorità è stata data ai trattati, ai manuali e agli scritti della critica che rappresentano una fonte essenziale di terminologia specialistica; come nel caso precedente, sono state incluse sia opere originali che opere tradotte»³⁶.

Le vite de' più eccellenti pittori, scultori, e architettori di Giorgio Vasari rappresentano la fonte principe dei corpora multilingue in quanto Vasari introduce nel suo trattato termini che uniscono il lessico tecnico e alcuni concetti che anticipano il linguaggio della critica d'arte³⁷.

Nel corpus russo il primo gruppo di testi corrisponde a opere ormai classiche come le *Immagini d'Italia* di Pavel Muratov (*Obrazy Italii*, 1913), il saggio di Nikolaj Berdjaev *Il sentimento dell'Italia* (*Čuvstvo Italii*, 1915), le memorie di viaggio di Nikolaj Anciferov (*Otčizna moej duši: vospominanija o putešestvijach v Italiju*, 1938-1940), memorie legate al viaggio in Italia di personaggi di spicco come la principessa Ekaterina Daškova (*Zapiski knjagini*, 1802), di artisti come Mstislav Dobužinskij (*Vospominanija*, 1922) e Michail Nesterov (*O perežitom*, 1928). Figurano in questo gruppo anche testi poetici di Konstantin Batjuškov, Anton Del'vig, Aleksandr Puškin, Petr Vjazemskij, Aleksandr Blok, Nikolaj Gumilev, Dmitrij Merežkovskij, Anna Achmatova, Iosif Brodskij, ispirati a immagini-simbolo della cultura italiana, nonché impressioni di viaggio affidate alle lettere di artisti e scrittori come Petr Čajkovskij, Valentin Serov, Fedor Dostoevskij, Anton Čechov. Vi figurano infine opere tradotte da autori stranieri, quali il diario di viaggio *Roma, Napoli e Firenze* di Stendhal (*Rim Neapol' Florencija*, 1978 [*Rome, Naples et Florence*, 1817]), le *Notti fiorentine* di Heinrich Heine (*Florentijskie noči*, 1971 [*Florentinische Nächte*, 1836]), le *Mattinate fiorentine* di John Ruskin (*Progulki po Florencii*, 2016 [*Mornings in Florence*, 1881]), *Il diario fiorentino* di Ranier Maria Ril'ke (*Florentijskij dnevnik*, 2001 [*Das Florenzer Tagebuch*, 1898]).

Nel secondo gruppo rientrano sia opere tradotte, come i testi di Leon Battista Alberti (*Desjat' knig o zодčestve*, 1935 [*I dieci libri de l'architettura*, 1500]),

³⁵ Carpi, Pano Alamán 2020.

³⁶ *Ibidem*.

³⁷ *Ibidem*.

Andrea Palladio (*Četyre knigi ob arhitekture*, 1938 [I quattro libri dell'architettura, 1570?]) e Leonardo da Vinci (*Traktat o živopisi*, 2010 [Trattato della pittura, 1500?]), le *Istorie fiorentine* di Machiavelli (1526 [Istorija Florencii, 1973]), l'autobiografia di Cellini (*Žizn' Benvenuto, syna maestra Džovanni Čellini*, 1958 [Vita di Benvenuto Cellini, 1539]), sia saggi di carattere generale sulla storia dell'arte rinascimentale di Heinrich Wölfflin, Ernst Gombrich, Bertrand Zhestaz; sia opere originali in russo: testi di epoca sovietica, come il manuale di Michail Alpatov *L'arte italiana dell'epoca di Dante e Giotto (Ital' janskoe iskusstvo epochi Dante i Džotto*, 1939) e il *Breve dizionario di termini dell'arte figurativa (Kratkij slovar' terminov izobrazitel'nogo iskusstva*, 1965) e manuali più recenti, come quelli di A. Kaptikov e A. Plešivcev, dedicati alla storia dell'arte rinascimentale, ricchi di termini culturalmente rilevanti relativi alla storia dell'arte italiana.

Il *Corpus russo* comprende 248 testi per un totale di più di 1.900.000 parole (<<http://corpora.lessicobeniculturali.net/ru>>).

Ogni corpus LBC può essere utilmente descritto a partire dalla classificazione dei testi che lo compongono, individuando così utili comparazioni, quali ad esempio:

1. “testi di traduzione” (147/248, pari al 59% del corpus)³⁸ vs. “testi originali russi” (101/248, pari al 41%). Le traduzioni dall'italiano ovviamente prevalgono (138/147), corrispondendo al 94% dei testi tradotti. Le *Vite* di Vasari costituiscono l'88% delle traduzioni dall'italiano e corrispondono all'83% del sottocorpus “testi di traduzione”;
2. “testi tecnici o dizionari” (171/248, pari al 69% del corpus)³⁹ vs. “testi letterari o divulgativi” (77/248, pari al 31%)⁴⁰.

All'interno del sottocorpus “testi di traduzione”, la percentuale di “testi tecnici o dizionari” (140/147) è pari al 95%, mentre quella di “testi letterari o divulgativi” (7/147) è pari al 5%. All'interno del sottocorpus “testi originali russi” invece la percentuale di “testi tecnici o dizionari” (31/101) è pari al 31%, mentre quella di “testi letterari o divulgativi” (70/101) è pari al 69%. Nel corpus russo risulta quindi evidente una forte disparità tra la percentuale di testi tecnici presente nel sottocorpus “testi di traduzione” (95%) e la percentuale di testi tecnici presente nel sottocorpus “testi originali russi” (31%). La grande presenza di traduzioni dal Vasari, che sono classificate come “Tecnico – Arte”, incide fortemente su queste percentuali.

Tale disparità sarà oggetto di un'ulteriore analisi nell'*Introduzione al Lessico essenziale dell'arte* elaborato dal gruppo di ricerca di lingua russa del progetto LBC (di prossima pubblicazione).

³⁸ Dei 147 testi di traduzione 138 sono traduzioni dall'italiano (di cui 122 da Vasari), 3 dall'inglese, 3 dal tedesco, 2 dal francese, 1 dal latino.

³⁹ Dei 171 testi tecnici/dizionari 142 sono testi classificabili come “Tecnico – Arte”, 20 come “Tecnico – Architettura”, 8 come “Tecnico – Storia”, 1 come “Dizionario – Monolingue”.

⁴⁰ Dei 77 testi non tecnici 52 sono classificabili come “Letterario – Biografico”, 16 come “Letterario – Fiction”, 8 come “Letterario – Saggistica”, 1 come “Divulgativo – Guida”.

3. Uso del corpus LBC russo

Come già detto, i partecipanti al progetto si sono prefissi un obiettivo ambizioso: creare una risorsa di riferimento multilingue *autorevole* che presenti il vocabolario e le combinazioni fisse relative al patrimonio culturale della città di Firenze; una risorsa che includa sia termini di arte, storia dell'arte, religione, storia, sia i nomi propri, i nomi comuni e i toponimi usati per descrivere quel patrimonio culturale. La realizzazione del portale coinvolge non solo studiosi e docenti universitari, ma anche gli studenti, per i quali rappresenta un'ottima scuola di metodo nell'ambito della lessicologia, della lessicografia e dell'ortografia della lingua russa, nonché una pratica insostituibile nell'ambito della traduzione.

A chi è rivolto il progetto? Innanzitutto ai lavoratori dei musei, alle organizzazioni turistiche e ai traduttori, ovvero agli specialisti che quotidianamente affrontano la necessità di compilare, utilizzare, tradurre guide e cataloghi di mostre. Ma si rivolge anche a una più ampia gamma di utenti interessati sia alle attrazioni turistiche che alla storia dell'arte e della religione. È stato appunto l'orientamento su un *target* così ampio che ha predeterminato la scelta delle fonti per la formazione della banca dati, di cui si è già detto nel precedente paragrafo.

Cosa distingue questo progetto da portali simili già esistenti (*Wikipedia*, enciclopedie, dizionari speciali e bilingue)? La banca dati è limitata al vocabolario relativo al patrimonio culturale di una città, intesa come prototipo di città d'arte, ma essa viene ampliata attirando ulteriore materiale fattuale (antroponimi, nomi di istituzioni e monumenti, toponimi), andando a confluire in un'unica risorsa. Ma la cosa più importante è che ogni unità lessicale acquisisce un polifonismo di suono e di significato, in quanto è accompagnata non solo da una traduzione estesa e raffinata, ma anche da esempi di utilizzo memorizzati nella banca dati.

Non è un caso che il centro di questa iniziativa sia stata Firenze: la città in cui tutta l'arte europea non solo ha preso forma e si è sviluppata fino alla sua massima realizzazione, ma ha trovato anche una comprensione teorica, riflessa nell'opera fondamentale del fiorentino Vasari. Pertanto, è ovvio che la lingua italiana in questo progetto sia il punto di partenza (Figura 1).

LESSICO
BENI CULTURALI

HOME | CORPORA | BANCA DATI | COME CITARE | GLOSSARIO | CORPUS | LINGUAGGIO

Corpus LBC Italiano

Lanini Ludovica (dir.), *Corpus LBC-IT* (2018), Firenze University Press, <<http://corpora.lessicobenculturali.net/>>

Gli altri corpora LBC

- > [Corpus LBC Francese](#)
- > [Corpus LBC Inglese](#)
- > [Corpus LBC Russo](#)
- > [Corpus LBC Spagnolo](#)
- > [Corpus LBC Tedesco](#)

Corpora LBC

Farina, Annick; Nicolás Martínez, Carlota; Billero, Riccardo (2018), *Corpora LBC*, Firenze, Firenze University Press, <<http://corpora.lessicobenculturali.net/>>

1. Farina, Annick; Nicolás Martínez, Carlota, *Presentazione della banca dati LBC*
2. Carpi, Elena; Pano, Ana, *El Corpus LBC español: bases, desarrollo y aplicaciones*
3. Natali, Maria, *The English Corpus for the LBC project*
4. Ballestracci, Sabrina; Buffagni, Claudia; Flinz, Carolina, *Das deutsche LBC-Korpus: Zusammenstellung und Anwendung*
5. Farina, Annick; Sini, Lorella, *Le corpus LBC français*
6. Rossi, Valentina, *Corpus LBC russo: caratteristiche principali*
7. Arezzi, Alessandra, *Il corpus LBC italiano*
8. Billero, Riccardo, *Approach to corpora LBC: technical point of view*
9. Billero, Riccardo; Farina, Annick; Nicolás Martínez, Carlota, *Conclusioni: Dati numerici attuali e bilanciamento dei corpora*

Apri "Corpora.lessicobenculturali.net/rob/na/og/finz_form/corname/it/it" in un nuovo pannello

Figura 1. Pagina iniziale dei Corpora.

Naturalmente, ciascuna di queste lingue ha un suo complesso rapporto con il donatore italiano. Ci concentreremo sulla componente russa del progetto, che ha le sue specificità. Oltre ai problemi legati all'accuratezza della traduzione di alcune unità lessicali, derivanti dalle significative differenze che esistono, ad esempio, tra le tradizioni ecclesiastiche cattolica e ortodossa, differenze che richiedono ulteriori chiarimenti e spiegazioni, ad ogni passo ci si imbatte in problemi di traslitterazione e/o trascrizione di nomi propri, nomi comuni e toponimi dall'alfabeto latino al cirillico.

Daremo solo due esempi di asimmetria semantica a proposito di due parole che si incontrano immediatamente nel momento in cui si fa la conoscenza anche superficiale di una città europea o russa. Stiamo parlando dell'«altare» (*altar'*) e della «cattedrale» (*sobor*).

Se ci affidiamo solo al dizionario bilingue russo-italiano, *altar'* viene tradotto come *altare* e si ritiene che queste unità lessicali descrivano la stessa parte dell'interno della chiesa. Ma questa è una somiglianza 'apparente' e uno specialista capisce subito che si parla di cose diverse. L'altare è presente sia nella chiesa cattolica che in quella ortodossa, ma la loro posizione e funzione sono diverse. L'«altare» (*altar'*) della chiesa ortodossa in realtà corrisponde al *presbiterio* italiano; mentre l'«altare» della chiesa cattolica corrisponde al russo «trono» (*prestol*), alla «tavola per i sacrifici» (*stol dlja žertvoprinošenija, žertvennik*).

La prima cosa che i turisti vedono in qualsiasi città italiana è il Duomo, la cattedrale. In russo, la parola *sobor* si riferisce a una chiesa particolarmente importante in cui si svolgono celebrazioni delle massime autorità ecclesiastiche, ma non possiede uno stato unico, come la cattedrale nella tradizione occidentale, che è legata alla presenza della cattedra episcopale. Quindi, in una grande città come Mosca o San Pietroburgo possono esserci diverse cattedrali. Il traduttore si trova subito di fronte a un problema: lasciare la parola «duomo» nella traslitterazione cirillica o dare una traduzione esplicitiva: «la principale, la più importante cattedrale della città». E poiché non esiste una simile gerarchia lessicale fissa per la parola «duomo» in russo, essa si trasforma in un nome proprio: il Duomo, ovvero la Cattedrale di Santa Maria del Fiore. Molto spesso la parola semplicemente non viene usata, sostituita dalla parola *sobor*⁴¹.

Il corpus in corso di implementazione consente di evidenziare tutti i problemi sopra citati e molti non ancora citati al fine di attirare l'attenzione degli specialisti alla ricerca di soluzioni e per raggiungere la massima unificazione possibile delle strategie di traduzione.

Per illustrare le possibilità del corpus nell'analisi e nella valutazione delle traduzioni, consideriamo un piccolo frammento della *Vita di Filippo Brunelleschi* di Vasari, che descrive il disegno della cupola del Duomo di Firenze e i materiali da utilizzare nella costruzione, in particolare i diversi tipi di pietra (Tabella 1).

⁴¹ Si veda la voce «SANTA-MARIJA-DEL'-F'ORE» in *Bol'shaja Rossijskaja Ėnciklopedija (BRĖ)*, <https://bigenc.ru/fine_art/text/3533826>, (2017), 10/2020.

Tabella 1. Frammento dalla *Vita di Filippo Brunelleschi* di Vasari.

Originale (1568)	Traduzione (1933) ⁴² di A. Gabričevskij	Traduzione (1963) ⁴³ di A. Venediktov e A. Gabričevskij
1. Facciansi poi ventiquattro sproni con le dette volte murati intorno, e sei archi di macigni forti e lunghi, bene sprangati di ferri, i quali sieno stagnati, e sopra detti macigni , catene di ferro, che cinghino la detta volta con loro sproni. <...>	1. Затем следует приступить к постройке этих двадцати четырех ребер вместе с заложёнными между ними сводами, а также шести арок из крепких и длинных известняковых камней , прочно скрепленных процинкованными железными пиронами, а поверх этих каменей наложить железные обручи, которые бы связывали означенный свод с его ребрами. <...>	1. Затем следует приступить к постройке этих двадцати четырех ребер вместе с заложёнными между ними сводами, а также шести арок из крепких и длинных кусков мациньо , прочно скрепленных процинкованными железными пиронами, а поверх этих каменей наложить железные обручи, которые бы связывали означенный свод с его ребрами. <...>
2. Il primo e secondo cerchio da piè, sia rinforzato per tutto, con macigni lunghi per il traverso, sì che l'una volta e l'altra della cupola si posi in sui detti macigni . <...>	2. Первый и второй венец снизу должны быть сплошь завязаны поперечной кладкой из длинных известняковых камней , так чтобы оба свода купола на них покоились. <...>	2. Первый и второй венец снизу должны быть сплошь завязаны поперечной кладкой из длинных известняковых камней так, чтобы оба свода купола на них покоились. <...>
3. Gli sproni murati tutti di macigni e di pietra forte , e similmente le facce della cupola tutte di pietra forte , legate con gli sproni fino all'altezza di braccia ventiquattro, e da indi in su si muri di mattoni, o vero di spugna , secondo che si delibererà per chi l'averà a fare, più leggieri che egli potrà. <...>	3. Ребра должны быть целиком сложены из известняка и прочного песчаника , а также и самые своды — целиком из прочного песчаника , причем и ребра и своды должны быть связаны друг с другом вплоть до вышины двадцати четырех локтей, откуда уже может начаться кладка из кирпича или туфа , в зависимости от решения того, кому это будет поручено, так чтобы это было как можно легче.	3. Ребра должны быть целиком сложены из мациньо и пьетрафорте , а также и самые грани целиком из пьетрафорте , причем и ребра, и своды должны быть связаны друг с другом вплоть до вышины двадцати четырех локтей, откуда уже может начинаться кладка из кирпича или туфа в зависимости от решения того, кому это будет поручено, так, чтобы это было, как можно легче. <...>

⁴² Džiardžio Vazari, *Žizn' Filippo Brunellesko, florentijskogo skul'ptora i arhitekтора*, in Id., *Žizneopisanija naibolee znamenitych živopisce, vajatelej i zodčich*, t. 2, Moskva-Leningrad, Akademija 1933, p. 283 (trad. di A. Gabričevskij); accessibile alla pagina web: <http://www.vostlit.info/Texts/rus17/Vazari_2/framevved1.htm>.

⁴³ Džiardžio Vazari, *Žizn' Filippa Brunellesko florentijskogo skul'ptora i arhitekтора*, in Id., *Žizneopisanija naibolee znamenitych živopisce, vajatelej i zodčich*, t. 2, Moskva-Leningrad, Akademija 1963, (trad. di A. Venediktov e A. Gabričevskij); accessibile alla pagina web: <http://www.vostlit.info/Texts/rus17/Vazari_3/frametext11.htm>.

4. L'acque della cupola terminano in su una ratta di marmo larga un terzo, e getti l'acqua dove di **pietra forte** sarà murato sotto la ratta ...

4. <...> вода с купола должна будет попадать на мраморный желоб, шириной в одну треть локтя, который будет выбрасывать воду туда, где внизу желоб будет сложен из **песчаника**.

4. Вода с купола будет попадать на мраморный желоб шириной в одну треть локтя, который будет выбрасывать воду туда, где внизу желоб будет сложен из **песчаника**.

Come si può vedere dai testi sopra citati, essi abbondano di dettagli tecnici e spiegazioni complesse che richiedono non solo il possesso di una terminologia linguistica appropriata, ma anche la comprensione delle caratteristiche ingegneristiche dell'oggetto in costruzione. Quanto ai nomi delle pietre, in questo brano di Vasari, oltre al marmo, troviamo altri tre tipi di minerali: *macigno*, *pietraforte*, *spugna*. I traduttori hanno dovuto affrontare un compito difficile: trovare un equivalente sufficientemente vicino per un lettore non esperto di mineralogia (1.1933), oppure lasciare il nome italiano, spiegando nel testo che si tratta appunto di una pietra (1.1963).

Gabričevskij, l'autore della traduzione del 1933, traduce costantemente «macigno» con *izvestnjakovyj kamen'* («pietra calcarea») e «pietraforte» con *pesčanik* («arenaria»); ma nella traduzione del 1963, dove egli agisce sia come traduttore che come editore, l'impostazione cambia: «macigno» viene traslitterato come *mačin'o* (1. 1963), oppure tradotto, come nella versione precedente, con *izvestnjakovyj kamen'* (2. 1963). Il lettore ne ricava l'impressione che si tratti di due minerali completamente diversi. Lo stesso approccio ambiguo viene offerto nella versione del 1963 per la traduzione di «pietraforte»: la traslitterazione *p'etraforte* (3. 1963) o la traduzione con *pesčanik* (4. 1963, 3-4. 1933). È difficile trovare una spiegazione logica per questa duplicazione di termini vista l'assenza di spiegazioni nel testo.

Va detto che la traduzione del 1933 è incompleta: in essa manca, oltre alle biografie degli artisti meno famosi, anche la parte introduttiva, la parte più 'tecnica' dell'opera vasariana, dedicata a questioni generali di pittura, scultura e architettura. Nella traduzione iniziata nel 1956 il primo volume⁴⁴ contiene questa parte introduttiva, che presenta una grande quantità di terminologia specifica, ivi inclusi termini tratti dalla mineralogia, nella misura in cui questa parte della scienza della natura era nota al Vasari al momento della stesura della sua opera. A. Venediktov, il traduttore del volume, ha lasciato quasi sempre i nomi italiani dei minerali, dando il prestito in uso o una traslitterazione in russo (Tabella 2).

Se guardiamo al numero di menzioni di «macigno» nel corpus italiano (Figura 2), possiamo vedere che questa parola ricorre 38 volte in Vasari. Nel corpus

⁴⁴ Džiorđio Vazari, *O različnyh kamnjach, služušich arhitektozam dlja ukrašeniij i skul'ptom dlja statuj*, in Id., *Žizneopisanija naibolee znamenitych živopiscev, vajatelej i zodčich*, t. 1, Moskva, Iskusstvo 1956, (trad. di A. Venediktov); accessibile alla pagina web: <http://www.vostlit.info/Texts/rus17/Vazari_2/frameved1.htm>.

Tabella 2. Tabella delle pietre citate nell'Introduzione di Vasari.

Vasari 1568	Вазари 1956 (traduzione di A. Venediktov)
porfido	порфир
serpentino	серпентин
cipollaccio	чиполаччо
mischio	мискио
granito	гранит
(pietra di) paragone	Парагон (Nome comune – пробирный камень)
Marmo (diversi colori)	мрамор (разных цветов)
Specie di marmi:	сорта мрамора:
cipollini	чиполино
saligni	салиньо
campanini	кампанино
mischiati	мискиато
pomice	пемза
Trevertino / travertino	травертин
Piperno / preperigno	Пиперино / препериньо (Nome comune attuale – пеперин)
pietra da calcina	известняк
Pietra serena	пъетрасерена
macigno	мачиньо
Pietra forte	пъетрафорте

russo lo troviamo solo 13 volte (Figura 3), di cui due ripetute nella traduzione del 2011⁴⁵. Negli altri contesti il termine «macigno» è reso, come abbiamo detto, con il russo *izvestnjakovyj kamen'*.

La «pietra arenaria», che non era menzionata nell'*Introduzione*, ricorre nel corpus italiano 6 volte (nei testi su Michelangelo e Brunelleschi), il «calcare» 2 volte e la «pietra calcarea» 4 volte. Descrivendo i diversi tipi di minerali utilizzati nell'edilizia, nell'architettura e nella scultura, Vasari si è affidato alla loro descrizione visiva piuttosto che alla composizione chimica, che oggi viene utilizzata dagli specialisti in mineralogia e petrografia. Tuttavia, la maggior parte dei nomi che egli usa si sono conservati nell'italiano⁴⁶. La sfida per i traduttori

⁴⁵ Questa traduzione è un altro mistero, legato però non all'opera di Vasari, ma solo alla pratica editoriale. Il testo, pubblicato nel 2011 e annunciato come nuova traduzione, è in realtà una riproduzione quasi esatta della traduzione pubblicata dalla casa editrice *Iskusstvo* nel 1956-71. Per quale ragione M. Globačev sia indicato come autore di questa traduzione, resta da indovinare.

⁴⁶ Nel *Vocabolario Treccani* troviamo la seguente definizione di «macigno»: «roccia sedimentaria, arenaria del periodo cretaceo o del paleogene, a grana media o fina, a cemento calcareo o calcareo-argilloso, di colore da grigio a grigio-giallastro o grigio-azzurrognolo, facilmente disgregabile per gelivazione, molto diffusa nell'Appennino, usata come materiale da costruzione e per macine da mulino. Viene detto *pietra forte* se compatto e resistente, *pietra morta* se

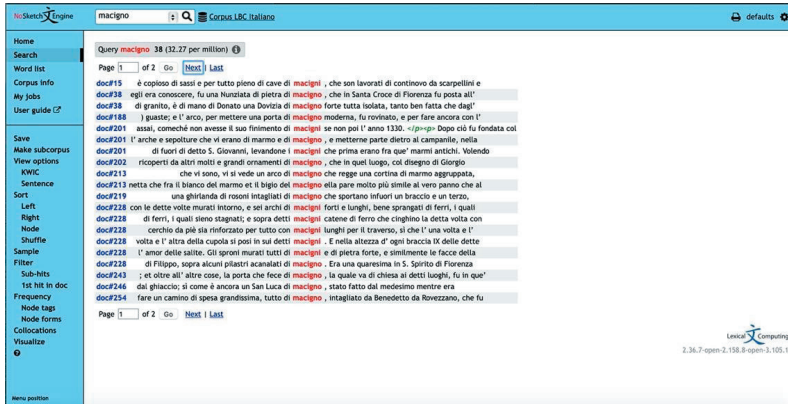


Figura 2. Risultati della ricerca del termine «macigno» nel corpus italiano.

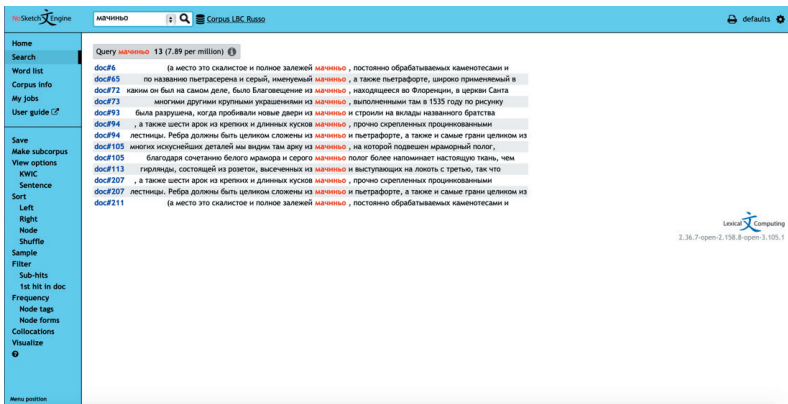


Figura 3. Risultati della ricerca del termine «macigno» nel corpus russo.

era quella di mantenere un equilibrio tra la fedeltà all'originale e l'intelligibilità e l'accessibilità del testo nella lingua di ricezione. In questa situazione, la spiegazione che la "pietra serena" e la "pietra forte" sono tipi di arenaria di una certa tonalità permetterebbe non solo agli specialisti, ma anche agli amatori di comprendere meglio il testo di Vasari. Pertanto, il termine *macin'ò* avrebbe dovuto essere spiegato, almeno nell'*Introduzione sulle pietre*, dicendo che si tratta di una pietra calcarea. Dal nostro punto di vista, è stato un errore usare parallelamente il prestito dall'italiano e il termine russo *izvestnjakovyj kamen'* senza operare una netta distinzione.

friabile, *pietra serena* se di colore azzurrognolo e ben lavorabile» (<<https://www.treccani.it/vocabolario/macigno/>, 11/2020). Questa definizione moderna è pienamente coerente con la traduzione di Gabričevskij 1933, che traduce «macigno» come *izvestnjakovyj kamen'*, e «pietra forte» come *pročnyj pesčanik* (ringrazio la collega M. Perotto per il prezioso suggerimento).

Questo è solo un esempio delle innumerevoli sfide che gli storici dell'arte, gli interpreti, le guide turistiche e gli amanti dell'arte devono affrontare quando si rivolgono a un'opera complessa come quella di Vasari. Ovviamente l'opera di Vasari necessita di una nuova traduzione, che tenga conto delle nuove conoscenze, dei nuovi approcci e delle nuove tecniche di traduzione, e per essa il corpus proposto sarebbe un ausilio essenziale.

Così il progetto, che si era posto l'obiettivo di riportare la lingua italiana al suo legittimo posto di intermediario e donatore nel dialogo tra le culture, ha avuto come logica continuazione l'obiettivo di colmare le lacune, rivedere e chiarire molti fenomeni linguistici nelle versioni delle singole lingue nazionali, incluso il russo. Lavorare con un corpus parallelo ci permetterà di penetrare nell'arte e nella tecnica della traduzione e di seguire sul piano diacronico come sono cambiati gli approcci e i metodi della pratica traduttiva.

Bibliografia

- Apresjan Ju.D. (ed.) 2006, *Jazykovaja kartina mira i sistemnaja leksikografija*, Jazyki slavjanskich kul'tur, Moskva.
- Biffi M. 2010, *Arte e critica d'arte, lingua dell'*, *Enciclopedia dell'italiano*, <<https://bit.ly/31XQyBw>>, (10/2020).
- Billero R., Nicolás Martínez M.C., *Nuove risorser per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC*, in «CHIMERA: Romance Corpora and Linguistic Studies», Vol. 4, No. 2 (2017), pp. 203-216.
- Billero R. 2020, *Cultural Heritage Lexicon: A Case Study*, in Pano Alamán A., Zotti V. (eds.), *The language of art and cultural heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 86-103.
- Billero R., Farina A., Nicolás Martínez M.C. (eds.) 2020, *I Corpora LBC: Informatica Umanistica per il Lessico dei Beni Culturali*, FUP, Firenze.
- Carbone A. 2020, *The reception of Giorgio Vasari's "Vite" in Russia: M.I. Železnov and his "vasarian" approach to the biography of K. Brjullof*, in Afinogenov D.E. (ed.), *"Mare nostrum". Sbornik naučnych statej Instituta Vseobščej Istorii RAN, Institut Vseobščej Istorii RAN, Moskva.*
- Carpi E., Pano Alamán A. 2020, *Il Corpus LBC español: basi, sviluppo e applicazioni*, in Billero R., Farina A., Nicolás Martínez M.C., (eds.), *I Corpora LBC: Informatica Umanistica per il Lessico dei Beni Culturali*, FUP, Firenze: 33-41.
- Farina A. 2015, *Un parcours du combattant: la recherche de traductions de la 'langue de l'art' dans les ressources lexicographiques et terminologiques. L'exemple de tempera et guazzo*, in Carpi E. (ed.), *Prospettive multilingue e interdisciplinari nel discorso specialistico*, Pisa University Press, Pisa: 33-57.
- 2015, *Guideline Proposition for Description and Translation of Proper Names in a Multilingual Cultural Heritage Dictionary of Florence*, in Karpova O.M. (ed.), *Life beyond Dictionaries*, Cambridge Scholars Publishing, Newcastle upon Tyne: 122-132.
- Farina A., Billero R. 2018, *Comparaison de corpus de langue «naturelle» et de langue «de traduction»: les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues* in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 108-116.

- Garzaniti M., Farina A. 2013, *Un portale per la comunicazione e la divulgazione del patrimonio culturale: progettare un lessico multilingue dei beni culturali on-line*, in Filipovic A., Troiano W. (eds.), *Strategie e Programmazione della Conservazione e Trasmissibilità del Patrimonio Culturale*, Fidei Signa, Roma: 500-509.
- Garzaniti M. 2014, *Lexicographical tools in the communication of Italian and Florentine cultural heritage in Russian language*, in *Heritage Lexicography as Supporting tool for International Council on Monuments and Sites (ICOMOS). Proceedings of the International Workshop (Florence, July 21-23, 2014)*, Firenze/Ivanovo: 21-24.
- 2015, *Proposal for a multilingual cultural heritage dictionary: complexities and problems in corresponding terms in Italian and Russian*, in Karpova O.M. (ed.), *Life beyond Dictionaries*, Cambridge Scholars Publishing, Newcastle upon Tyne: 133-143.
- 2020, *Il termino russo friag e le sue radici nelle relazioni culturali e artistiche fra la Russia e l'Italia*, in Pano Alamán A., Zotti V. (eds.), *The language of art and culture heritage: a plurilingual and digital perspective*, Cambridge Scholars Publishing, Newcastle upon Tyne: 104-119.
- Gebert L. 2006, *Immagine linguistica del mondo e carattere nazionale nella lingua. A proposito di alcune recenti pubblicazioni*, «Studi Slavistici», 3 (1): 217-243.
- Genieva E.Ju. (ed.) 2009, *Choždenija vo Florenciju. Florencija i florentijcy v ruskoj kul'ture. Vek XIX*, Rudomino, Moskva.
- Genieva E.Ju., Fridštejn Ju. G. (eds.) 2009, *Choždenija vo Florenciju. Florencija i florentijcy v ruskoj kul'ture iz veka XIX v vek XXI*, Rudomino, Moskva.
- Karpova O.M. 2009, *Florencija v tvorčestve evropejskich pisatelej i chudožnikov: enciklopedičeskij slovar' dlja gidov i turistov: proekt slovarja. Florence in the Works of European Writers and Artists: Encyclopedic Dictionary for Guides and Tourists: Project of a Dictionary*, Ivanovo.
- 2011, *Ivanovskaja leksikografičeskaja škola: tradicii i innovacii*, Ivanovskij Gosudarstvennyj Universitet, Ivanovo.
- Karpova O.M., Kartashkova F.I. (eds.) 2015, *Life Beyond Dictionaries*, Cambridge Scholars Publishing, Newcastle upon Tyne, 2015.
- Lanini L., Nicolás Martínez M.C. 2018, *Verso un dizionario corpus-based del lessico dei beni culturali: procedure di estrazione del lemmario*, in Iezzi D.F., Celardo L., Misuraca M. (eds.), *JADT 2018 – International Conference on Statistical Analysis of Textual Data, Roma, 12-15 giugno 2018*, UniversItalia, Roma: 411-418.
- Motolese M. 2012, *Italiano lingua delle arti. Un'avventura europea (1250-1650)*, Il Mulino, Bologna.
- Rossi F. 2010, *Palladio in Russia. Nikolaj L'vov architetto e intellettuale russo al tramonto del Lumi*, Marsilio, Venezia
- 2013, *Il taccuino italiano di Nikolaj L'vov*, Edizioni della Normale, Pisa.
- Švidkovskij D.O., Batalov A. L., Rossi F. 2013, *Italija-Rossija. Tysjača let architektury*, Torino 2013 (versione it.: *Mille anni di architettura italiana in Russia*).
- Tonini L. (ed.) 2012, *Rinascimento e antirinascimento. Firenze nella cultura russa fra Otto e Novecento*, Olschki, Firenze.
- Zotti V. 2017, *L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle Vite di Vasari*, in Zotti V., Pano Alamán A. (eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*, FUP, Firenze: 105-134.
- Zhukova N.N. 2018, *Per una nuova guida di Firenze in lingua russa*, in Flinz C., Carpi E., Farina A. (eds.), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel*, Vol. I, FUP, Firenze: 171-182.

Fonti del corpus

- Achmatova A. 1998, *Sobranie sočinenij v šesti tomach*, t. 1, Ellis Lak, Moskva (ed. orig. 1936).
- Al'berti L.-B. 1935, *Desjat' knig o zodčestve*, in Id., *Desjat' knig o zodčestve* (trad. di Zubov V.P.), Moskva (ed. orig. 1500), Izd-vo Vsesojuznoj akademii architektury, vol. 1, 11-385.
- Alpatov M. 1939, *Ital' janskoe iskusstvo èpochi Dante i Džotto*, Iskusstvo, Moskva-Leningrad.
- Anciferov N. 2016, *Otčizna moej duši: vospominanija o putešestvijach v Italiju* (a cura di Moskovskaja D.S., Talalaj M.G., Farsetti A.), Staraja Bosmannaja, Moskva (ed. orig. 1938-1940), 37-47.
- Batjuškov K. 1977, *Opyty v stichach i proze*, (a cura di Semenko I.M.), Nauka, Moskva (ed. orig. 1815), 20-28.
- 1964, *Polnoe sobranie stichotvorenij* (a cura di Fridman N.V.), Moskva-Leningrad (ed. orig. 1809).
- Berdjaev N. 1949, *Samopoznanie*, YMCA PRESS, Paris (ed. orig. 1940).
- 1994, *Filosofija tvorčestva, kul'tury i iskusstva*, t. I, Liga, Moskva (ed. orig. 1915), 367-371.
- Blok A. 1962, *Sobranie sočinenij v devjati tomach*, t. 3, t. 8, Goslitizdat, Moskva (ed. orig. 1909).
- Brodskij I. 2017?, *Èlektronnoe sobranie sočinenij. Stichotvorenija i poèmy*, (ed. orig. 1976).
- Čajkovskij P. (1940), *Pis'ma k rodnym*, Muzgiz, Moskva (ed. orig. 1874), 204-205.
- Čechov A. 1976, *Polnoe sobranie sočinenij i pisem*, Nauka, Moskva, (ed. orig. 1891), t. 4.
- 1980, *Polnoe sobranie sočinenij i pisem*, Nauka, Moskva (ed. orig. 1901), t. 9.
- Čellini B. 1958, *Žizn' Benvenuto, syna maèstro Džovanni Čellini, florentinca, napisannaja im samym vo Florencii*, trad. Lozinskij M., GICHL, Moskva (ed. orig. 1539).
- Da Vinči L. 2010, *Traktat o živopisi*, trad. Gubarev A.A., Azbuka, Moskva (ed. orig. 1500?), <Azbuka>.
- Daškova E. 2003, *Zapiski knjagini: Vospominanija. Memuary*, Charvest, Minsk (ed. orig. 1805).
- Del'vig A. 1986, *Sočinenija. Stichi*, Chudožestvennaja literatura, Leningrad (ed. orig. 1819-1820?).
- Dėržavin G. 1986, *Anakreontičeskije pesni*, Nauka, Moskva (ed. orig. 1802), 69.
- Dobužinskij M. 1987, *Vospominanija* (a cura di Čugunov G.I.), Nauka, Moskva (ed. orig. 1922), 258-264.
- Dostoevskaja A. 1971, *Vospominanija* (a cura di Grigorenko V. et al.), Chudožestvennaja literatura, Moskva (ed. orig. 1916).
- Dostoevskij F. 2005, *Zimnie zametki o letnyh vpečatlenijach*, Ripol Klassik, Moskva (ed. orig. 1862).
- 1996, *Sobranie sočinenij v pjatnadcati tomach*, vol. 15, Nauka, Leningrad (ed. orig. 1868).
- Forster È.M. 2014, *Komnata s vidom na Arno*, trad. V.A. Milovidov, Ast, Moskva (ed. orig. 1904).
- Gabričevskij A. 1937, *Al'berti – arhitektor. Florencija*, in Al'berti L.B., *Desjat' knig o zodčestve* (trad. di Zubov V.P.), Izd-vo Vsesojuznoj akademii architektury, Moskva, vol. 2, 201-215.
- Gejne G. 1971, *Stichotvorenija. Poèmy. Proza*, Chudožestvennaja literatura, Moskva (ed. orig. 1836), 688-738.
- Gerzyk E. 1973, *Vospominanija*, YMCA PRESS, Paris (ed. orig. 1943), 117-140.
- Gombrich È. 1998, *Istorija iskusstva*, AST (ed. orig. 1950), 223-245.

- Gumilev N., *Sobranie sočinenij v četyrech tomach* (a cura di Struve G.P. e Filippov B.A.), Victor Kamkin ed., Washington, t. 1, 135-136.
- 1988, *Stichotvorenija i poëmy*, Sovetskij pisatel', Leningrad (ed. orig. 1912).
- 2018, *Elektronnoe sobranie sočinenij*", <<https://gumilev.ru/verses/286/>> (01/2018).
- Gur'eva T. 1965 (a cura di), *Kratkij slovar' terminov izobrazitel'nogo iskusstva*, Sovetskij chudožnik, Moskva (ed. orig. 1957).
- Kaptikov A., *Romanskaja arhitektura Italii*, Tatlin, Ekaterinburg, 29-41.
- Mak'javelli N. 1973, *Istorija Florencii*, trad. Rykova N.Ja., Nauka, Leningrad (ed. orig. 1526).
- Merežkovskij D. 2000, *Dante*, Respublika, Moskva (ed. orig. 1939), 13-20.
- 2000, *Stichotvorenija i poëmy*, Akademičeskij proekt, Sankt-Petersburg, 928.
- 2008, *Voskresšie bogi, ili Leonardo da-Vinči*, AST, Moskva (ed. orig. 1900).
- Muratov P. 1994, *Obrazy Italii*, Respublika, Moskva (ed. orig. 1913).
- Nesterov M. 2006, *O perežitom. 1862-1917. Vospominanija*, Molodaja gvardija, Moskva (ed. orig. 1928).
- Palladio D. 1938, *Četyre knigi ob arhitekture* (trad. Žoltovskij I.V.), Izd. Vsesojuznoj Akademii Arhitektury, Moskva (ed. orig. 1570?).
- Plešivcev A. 2015, *Istorija arhitektury*, MGSU, Moskva, 245-251.
- Puškin A. 1959, *Sobranie sočinenij v 10 tomach*, t. 2, GICH, Moskva, 201-203.
- Ril'ke R.M. 2001, *Florentijskij dnevnik* (trad. di V. Bakusev), Tekst, Moskva (ed. orig. 1898).
- Rossi F. et al. 2016, *Piranezi v Rossii: do i posle*, Antiga Edizioni, Moskva.
- Ruskin J. 2016, *Progulki po Florencii*, trad. Gercyk A., Azbuka-Attikus, Sankt Peterburg (ed. orig. 1881).
- s.a. 2001, *Florencija: Gorod i vse šedevry. Zolotaja kniga (Firenze: tutta la città e i suoi capolavori. Il libro d'oro)*, trad. di Zotova L. e Katkova M., Bonechi, Firenze (ed. orig. 1963).
- Serov V. 1937, *Perepiska 1884-1911*, Iskusstvo, Lenindrad-Moskva (ed. orig. 1887).
- Stendal' F. 1978, *Rim Neapol' Florencija*, trad. Rykova N.A., Pravda (ed. orig. 1817).
- Strachov N. 1990, *Vospominanija o Fedore Michajlovičem Dostojevskom*, in Tjun'kin K. (a cura di), *Dostojevskij v vospominanijach sovremennikov*, vol. 1, Chudožestvennaja literatura, Moskva (ed. orig. 1862?), 439-444.
- Švidkovskij D. (a cura di) 2013, *Italija – Rossija: tysjača let arhitektury*, Umberto Allemandi & C., Torino-Londra-Venezia-New York.
- Vazari D. 1933, *Žizneopisanija naibolee znamenitych živopiscev, vajatelej i zodčich*, (a cura di Dživelegov A. e Ėfros A.), trad. Dživelegov A., Academia, Moskva-Leningrad (ed. orig. 1568).
- 1998, *Žizneopisanija naibolee znamenitych živopiscev, vajatelej i zodčich* (a cura di Dživelegov A., Ėfros A.), Rostov na Donu, Feniks (ed. orig. 1568).
- 2008, *Žizneopisanija naibolee znamenitych živopiscev, vajatelej i zodčich*, trad. A. Gabričevskij e A. Venediktov, Al'fa-Kniga, Moskva (ed. orig. 1568).
- 2011, *Žizneopisanija naibolee znamenitych živopiscev, vajatelej i zodčich*, trad. M. Globačev, Knigovek, Moskva (ed. orig. 1568).
- Vel'flin G. 2004, *Renessans i Barokko*, Azbuka-klassika, Sankt-Peterburg (ed. orig. 1888).
- Vjazemskij P. 1958, *Stichotvorenija*, Leningrad (ed. orig. 1834), 247-248.
- Žestaz B. 2003, *Arhitektura. Renessans. Ot Brunelleski do Palladio*, Ast, Moskva (ed. orig. 1995).

STRUMENTI PER LA DIDATTICA E LA RICERCA

TITOLI PUBBLICATI

1. Brunetto Chiarelli, Renzo Bigazzi, Luca Sineo (a cura di), *Alia: Antropologia di una comunità dell'entroterra siciliano*
2. Vincenzo Cavaliere, Dario Rosini, *Da amministratore a manager. Il dirigente pubblico nella gestione del personale: esperienze a confronto*
3. Carlo Biagini, *Information technology ed automazione del progetto*
4. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza. Medico, antropologo, viaggiatore*
5. Luca Solari, *Topics in Fluvial and Lagoon Morphodynamics*
6. Salvatore Cesario, Chiara Fredianelli, Alessandro Remorini, *Un pacchetto evidence based di tecniche cognitivo-comportamentali sui generis*
7. Marco Masseti, *Uomini e (non solo) topi. Gli animali domestici e la fauna antropocora*
8. Simone Margherini (a cura di), *BIL Bibliografia Informatizzata Leopardiana 1815-1999: manuale d'uso ver. 1.0*
9. Paolo Puma, *Disegno dell'architettura. Appunti per la didattica*
10. Antonio Calvani (a cura di), *Innovazione tecnologica e cambiamento dell'università. Verso l'università virtuale*
11. Leonardo Casini, Enrico Marone, Silvio Menghini, *La riforma della Politica Agricola Comunitaria e la filiera olivicolo-olearia italiana*
12. Salvatore Cesario, *L'ultima a dover morire è la speranza. Tentativi di narrativa autobiografica e di "autobiografia assistita"*
13. Alessandro Bertirotti, *L'uomo, il suono e la musica*
14. Maria Antonietta Rovida, *Palazzi senesi tra '600 e '700. Modelli abitativi e architettura tra tradizione e innovazione*
15. Simone Guercini, Roberto Piovan, *Schemi di negoziato e tecniche di comunicazione per il tessile e abbigliamento*
16. Antonio Calvani, *Technological innovation and change in the university. Moving towards the Virtual University*
17. Paolo Emilio Pecorella, *Tell Barri/Kahat: la campagna del 2000. Relazione preliminare*
18. Marta Chevanne, *Appunti di Patologia Generale. Corso di laurea in Tecniche di Radiologia Medica per Immagini e Radioterapia*
19. Paolo Ventura, *Città e stazione ferroviaria*
20. Nicola Spinosi, *Critica sociale e individuazione*
21. Roberto Ventura (a cura di), *Dalla misurazione dei servizi alla customer satisfaction*
22. Dimitra Babalis (a cura di), *Ecological Design for an Effective Urban Regeneration*
23. Massimo Papini, Debora Tringali (a cura di), *Il pupazzo di garza. L'esperienza della malattia potenzialmente mortale nei bambini e negli adolescenti*
24. Manlio Marchetta, *La progettazione della città portuale. Sperimentazioni didattiche per una nuova Livorno*
25. Fabrizio F.V. Arrigoni, *Note su progetto e metropoli*
26. Leonardo Casini, Enrico Marone, Silvio Menghini, *OCM seminativi: tendenze evolutive e assetto territoriale*
27. Pecorella Paolo Emilio, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2001. Relazione preliminare*
28. Nicola Spinosi, *Wir Kinder. La questione del potere nelle relazioni adulti/bambini*
29. Stefano Cordero di Montezemolo, *I profili finanziari delle società vinicole*
30. Luca Bagnoli, Maurizio Catalano, *Il bilancio sociale degli enti non profit: esperienze toscane*
31. Elena Rotelli, *Il capitolo della cattedrale di Firenze dalle origini al XV secolo*
32. Leonardo Trisciuzzi, Barbara Sandrucci, Tamara Zappaterra, *Il recupero del sé attraverso l'autobiografia*
33. Nicola Spinosi, *Invito alla psicologia sociale*
34. Raffaele Moschillo, *Laboratorio di disegno. Esercitazioni guidate al disegno di arredo*
35. Niccolò Bellanca, *Le emergenze umanitarie complesse. Un'introduzione*
36. Giovanni Allegretti, *Porto Alegre una biografia territoriale. Ricercando la qualità urbana a partire dal patrimonio sociale*
37. Riccardo Passeri, Leonardo Quagliotti, Christian Simoni, *Procedure concorsuali*

- e governo dell'impresa artigiana in Toscana
38. Nicola Spinosi, *Un soffitto viola. Psicoterapia, formazione, autobiografia*
 39. Tommaso Urso, *Una biblioteca in divenire. La biblioteca della Facoltà di Lettere dalla penna all'elaboratore. Seconda edizione rivista e accresciuta*
 40. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2002. Relazione preliminare*
 41. Antonio Pellicanò, *Da Galileo Galilei a Cosimo Noferi: verso una nuova scienza. Un inedito trattato galileiano di architettura nella Firenze del 1650*
 42. Aldo Burresi (a cura di), *Il marketing della moda. Temi emergenti nel tessile-abbigliamento*
 43. Curzio Cipriani, *Appunti di museologia naturalistica*
 44. Fabrizio F.V. Arrigoni, *Incipit. Esercizi di composizione architettonica*
 45. Roberta Gentile, Stefano Mancuso, Silvia Martelli, Simona Rizzitelli, *Il Giardino di Villa Corsini a Mezzomonte. Descrizione dello stato di fatto e proposta di restauro conservativo*
 46. Arnaldo Nesti, Alba Scarpellini (a cura di), *Mondo democristiano, mondo cattolico nel secondo Novecento italiano*
 47. Stefano Alessandri, *Sintesi e discussioni su temi di chimica generale*
 48. Gianni Galeota (a cura di), *Traslocare, riaggregare, rifondare. Il caso della Biblioteca di Scienze Sociali dell'Università di Firenze*
 49. Gianni Cavallina, *Nuove città antichi segni. Tre esperienze didattiche*
 50. Bruno Zanoni, *Tecnologia alimentare 1. La classe delle operazioni unitarie di disidratazione per la conservazione dei prodotti alimentari*
 51. Gianfranco Martiello, *La tutela penale del capitale sociale nelle società per azioni*
 52. Salvatore Cingari (a cura di), *Cultura democratica e istituzioni rappresentative. Due esempi a confronto: Italia e Romania*
 53. Laura Leonardi (a cura di), *Il distretto delle donne*
 54. Cristina Delogu (a cura di), *Tecnologia per il web learning. Realtà e scenari*
 55. Luca Bagnoli (a cura di), *La lettura dei bilanci delle Organizzazioni di Volontariato toscane nel biennio 2004-2005*
 56. Lorenzo Grifone Baglioni (a cura di), *Una generazione che cambia. Civismo, solidarietà e nuove incertezze dei giovani della provincia di Firenze*
 57. Monica Bolognesi, Laura Donati, Gabriella Granatiero, *Acque e territorio. Progetti e regole per la qualità dell'abitare*
 58. Carlo Natali, Daniela Poli (a cura di), *Città e territori da vivere oggi e domani. Il contributo scientifico delle tesi di laurea*
 59. Riccardo Passeri, *Valutazioni imprenditoriali per la successione nell'impresa familiare*
 60. Brunetto Chiarelli, Alberto Simonetta, *Storia dei musei naturalistici fiorentini*
 61. Gianfranco Bettin Lattes, Marco Bontempi (a cura di), *Generazione Erasmus? L'identità europea tra vissuto e istituzioni*
 62. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2003*
 63. Fabrizio F.V. Arrigoni, *Il cervello delle passioni. Dieci tesi di Adolfo Natalini*
 64. Saverio Pisaniello, *Esistenza minima. Stanze, spazi della mente, reliquiario*
 65. Maria Antonietta Rovida (a cura di), *Fonti per la storia dell'architettura, della città, del territorio*
 66. Ornella De Zordo, *Saggi di anglistica e americanistica. Temi e prospettive di ricerca*
 67. Chiara Favilli, Maria Paola Monaco, *Materiali per lo studio del diritto antidiscriminatorio*
 68. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2004*
 69. Emanuela Caldognetto Magno, Federica Cavicchio, *Aspetti emotivi e relazionali nell'e-learning*
 70. Marco Massetti, *Uomini e (non solo) topi (2ª edizione)*
 71. Giovanni Nerli, Marco Pierini, *Costruzione di macchine*
 72. Lorenzo Viviani, *L'Europa dei partiti. Per una sociologia dei partiti politici nel processo di integrazione europea*
 73. Teresa Crespellani, *Terremoto e ricerca. Un percorso scientifico condiviso per la caratterizzazione del comportamento sismico di alcuni depositi italiani*
 74. Fabrizio F.V. Arrigoni, *Cava. Architettura in "ars marmoris"*
 75. Ernesto Tavoletti, *Higher Education and Local Economic Development*
 76. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di*

- Carlo Rosselli (1917-1930)
77. Luca Bagnoli, Massimo Cini (a cura di), *La cooperazione sociale nell'area metropolitana fiorentina. Una lettura dei bilanci d'esercizio delle cooperative sociali di Firenze, Pistoia e Prato nel quadriennio 2004-2007*
 78. Lamberto Ippolito, *La villa del Novecento*
 79. Cosimo Di Bari, *A passo di critica. Il modello di Media Education nell'opera di Umberto Eco*
 80. Leonardo Chiesi (a cura di), *Identità sociale e territorio. Il Montalbano*
 81. Piero Degl'Innocenti, *Cinquant'anni, cento chiese. L'edilizia di culto nelle diocesi di Firenze, Prato e Fiesole (1946-2000)*
 82. Giancarlo Paba, Anna Lisa Pecoriello, Camilla Perrone, Francesca Rispoli, *Partecipazione in Toscana: interpretazioni e racconti*
 83. Alberto Magnaghi, Sara Giacomozzi (a cura di), *Un fiume per il territorio. Indirizzi progettuali per il parco fluviale del Valdarno empoese*
 84. Dino Costantini (a cura di), *Multiculturalismo alla francese?*
 85. Alessandro Viviani (a cura di), *Firms and System Competitiveness in Italy*
 86. Paolo Fabiani, *The Philosophy of the Imagination in Vico and Malebranche*
 87. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli*
 88. David Fanfani (a cura di), *Pianificare tra città e campagna. Scenari, attori e progetti di nuova ruralità per il territorio di Prato*
 89. Massimo Papini (a cura di), *L'ultima cura. I vissuti degli operatori in due reparti di oncologia pediatrica*
 90. Raffaella Cerica, *Cultura Organizzativa e Performance economico-finanziarie*
 91. Alessandra Lorini, Duccio Basosi (a cura di), *Cuba in the World, the World in Cuba*
 92. Marco Goldoni, *La dottrina costituzionale di Sieyès*
 93. Francesca Di Donato, *La scienza e la rete. L'uso pubblico della ragione nell'età del Web*
 94. Serena Vicari Haddock, Marianna D'Ovidio, *Brand-building: the creative city. A critical look at current concepts and practices*
 95. Ornella De Zordo (a cura di), *Saggi di Anglistica e Americanistica. Ricerche in corso*
 96. Massimo Moneglia, Alessandro Panunzi (edited by), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*
 97. Alessandro Panunzi, *La variazione semantica del verbo essere nell'Italiano parlato*
 98. Matteo Gerlini, *Sansone e la Guerra fredda. La capacità nucleare israeliana fra le due superpotenze (1953-1963)*
 99. Luca Raffini, *La democrazia in mutamento: dallo Stato-nazione all'Europa*
 100. Gianfranco Bandini (a cura di), *noi-loro. Storia e attualità della relazione educativa fra adulti e bambini*
 101. Anna Taglioli, *Il mondo degli altri. Territori e orizzonti sociologici del cosmopolitismo*
 102. Gianni Angelucci, Luisa Vierucci (a cura di), *Il diritto internazionale umanitario e la guerra aerea. Scritti scelti*
 103. Giulia Mascagni, *Salute e disuguaglianze in Europa*
 104. Elisabetta Cioni, Alberto Marinelli (a cura di), *Le reti della comunicazione politica. Tra televisioni e social network*
 105. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza e l'Evoluzionismo in Italia*
 106. Andrea Simoncini (a cura di), *La semplificazione in Toscana. La legge n. 40 del 2009*
 107. Claudio Borri, Claudio Mannini (edited by), *Aeroelastic phenomena and pedestrian-structure dynamic interaction on non-conventional bridges and footbridges*
 108. Emiliano Scampolli, *Firenze, archeologia di una città (secoli I a.C. - XIII d.C.)*
 109. Emanuela Cresti, Iørn Korzen (a cura di), *Language, Cognition and Identity. Extensions of the endocentric/exocentric language typology*
 110. Alberto Parola, Maria Ranieri, *Media Education in Action. A Research Study in Six European Countries*
 111. Lorenzo Grifone Baglioni (a cura di), *Scegliere di partecipare. L'impegno dei giovani della provincia di Firenze nelle arene deliberative e nei partiti*
 112. Alfonso Lagi, Ranuccio Nuti, Stefano Taddei, *Raccontaci l'ipertensione. Indagine a distanza in Toscana*
 113. Lorenzo De Sio, *I partiti cambiano, i valori restano? Una ricerca quantitativa e qualitativa sulla cultura politica in Toscana*
 114. Anna Romiti, *Coreografie di stakeholders nel management del turismo sportivo*
 115. Guidi Vannini (a cura di), *Archeologia Pubblica in Toscana: un progetto e una proposta*

116. Lucia Varra (a cura di), *Le case per ferie: valori, funzioni e processi per un servizio differenziato e di qualità*
117. Gianfranco Bandini (a cura di), *Manuali, sussidi e didattica della geografia. Una prospettiva storica*
118. Anna Margherita Jasink, Grazia Tucci e Luca Bombardieri (a cura di), *MUSINT. Le Collezioni archeologiche egee e cipriote in Toscana. Ricerche ed esperienze di museologia interattiva*
119. Ilaria Caloi, *Modernità Minoica. L'Arte Egea e l'Art Nouveau: il Caso di Mariano Fortuny y Madrazo*
120. Heliana Mello, Alessandro Panunzi, Tommaso Raso (edited by), *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*
121. Luciana Lazzeretti, *Cluster creativi per i beni culturali. L'esperienza toscana delle tecnologie per la conservazione e la valorizzazione*
122. Maurizio De Vita (a cura di / edited by), *Città storica e sostenibilità / Historic Cities and Sustainability*
123. Eleonora Berti, *Itinerari culturali del consiglio d'Europa tra ricerca di identità e progetto di paesaggio*
124. Stefano Di Blasi (a cura di), *La ricerca applicata ai vini di qualità*
125. Lorenzo Cini, *Società civile e democrazia radicale*
126. Francesco Ciampi, *La consulenza direzionale: interpretazione scientifica in chiave cognitiva*
127. Lucia Varra (a cura di), *Dal dato diffuso alla conoscenza condivisa. Competitività e sostenibilità di Abetone nel progetto dell'Osservatorio Turistico di Destinazione*
128. Riccardo Roni, *Il lavoro della ragione. Dimensioni del soggetto nella Fenomenologia dello spirito di Hegel*
129. Vanna Boffo (edited by), *A Glance at Work. Educational Perspectives*
130. Raffaele Donvito, *L'innovazione nei servizi: i percorsi di innovazione nel retailing basati sul vertical branding*
131. Dino Costantini, *La democrazia dei moderni. Storia di una crisi*
132. Thomas Casadei, *I diritti sociali. Un percorso filosofico-giuridico*
133. Maurizio De Vita, *Verso il restauro. Temi, tesi, progetti per la conservazione*
134. Laura Leonardi, *La società europea in costruzione. Sfide e tendenze nella sociologia contemporanea*
135. Antonio Capestro, *Oggi la città. Riflessione sui fenomeni di trasformazione urbana*
136. Antonio Capestro, *Progettando città. Riflessioni sul metodo della Progettazione Urbana*
137. Filippo Bussotti, Mohamed Hazem Kalaji, Rosanna Desotgiu, Martina Pollastrini, Tadeusz Łoboda, Karolina Bosa, *Misurare la vitalità delle piante per mezzo della fluorescenza della clorofilla*
138. Francesco Dini, *Differenziali geografici di sviluppo. Una ricostruzione*
139. Maria Antonietta Esposito, *Poggio al vento la prima casa solare in Toscana - Windy hill the first solar house in Tuscany*
140. Maria Ranieri (a cura di), *Risorse educative aperte e sperimentazione didattica. Le proposte del progetto Innovascuola-AMELIS per la condivisione di risorse e lo sviluppo professionale dei docenti*
141. Andrea Runfola, *Apprendimento e reti nei processi di internazionalizzazione del retail. Il caso del tessile-abbigliamento*
142. Vanna Boffo, Sabina Falconi, Tamara Zappaterra (a cura di), *Per una formazione al lavoro. Le sfide della disabilità adulta*
143. Beatrice Töttössy (a cura di), *Fonti di Weltliteratur. Ungheria*
144. Fiorenzo Fantaccini, Ornella De Zordo (a cura di), *Saggi di Anglistica e Americanistica. Percorsi di ricerca*
145. Enzo Catarsi (a cura di), *The Very Hungry Caterpillar in Tuscany*
146. Daria Sarti, *La gestione delle risorse umane nelle imprese della distribuzione commerciale*
147. Raffaele De Gaudio, Iacopo Lanini, *Vivere e morire in Terapia Intensiva. Quotidianità in Bioetica e Medicina Palliativa*
148. Elisabete Figueiredo, Antonio Raschi (a cura di), *Fertile Links? Connections between tourism activities, socioeconomic contexts and local development in European rural areas*
149. Gioacchino Amato, *L'informazione finanziaria price-sensitive*
150. Nicoletta Setola, *Percorsi, flussi e persone nella progettazione ospedaliera. L'analisi configurazionale, teoria e applicazione*
151. Laura Solito e Letizia Materassi, *DIVERSE eppur VICINE. Associazioni e imprese per la responsabilità sociale*

152. Ioana Both, Ayşe Saraçgil e Angela Tarantino (a cura di), *Storia, identità e canoni letterari*
153. Barbara Montecchi, *Luoghi per lavorare, pregare, morire. Edifici e maestranze edili negli interessi delle élites micenee*
154. Carlo Orefice, *Relazioni pedagogiche. Materiali di ricerca e formazione*
155. Riccardo Roni (a cura di), *Le competenze del politico. Persone, ricerca, lavoro, comunicazione*
156. Barbara Sibilio (a cura di), *Linee guida per l'utilizzo della Piattaforma Tecnologica PO.MA. Musei*
157. Fortunato Sorrentino, Maria Chiara Pettenati, *Orizzonti di Conoscenza. Strumenti digitali, metodi e prospettive per l'uomo del terzo millenni*
158. Lucia Felici (a cura di), *Alterità. Esperienze e percorsi nell'Europa moderna*
159. Edoardo Gerlini, *The Heian Court Poetry as World Literature. From the Point of View of Early Italian Poetry*
160. Marco Carini, Andrea Minervini, Giuseppe Morgia, Sergio Serni, Augusto Zaninelli, *Progetto Clic-URO. Clinical Cases in Urology*
161. Sonia Lucarelli (a cura di), *Gender and the European Union*
162. Michela Ceccorulli, *Framing irregular immigration in security terms. The case of Libya*
163. Andrea Bellini, *Il puzzle dei ceti medi*
164. Ambra Collino, Mario Biggeri, Lorenzo Murgia (a cura di), *Processi industriali e parti sociali. Una riflessione sulle imprese italiane in Cina (Jiangsu) e sulle imprese cinesi in Italia (Prato)*
165. Anna Margherita Jasink, Luca Bombardieri (a cura di), *AKROTHINIA. Contributi di giovani ricercatori italiani agli studi egei e ciprioti*
166. Pasquale Perrone Filardi, Stefano Urbinati, Augusto Zaninelli, *Progetto ABC. Achieved Best Cholesterol*
167. Iryna Solodovnik, *Repository Istituzionali, Open Access e strategie Linked Open Data. Per una migliore comunicazione dei prodotti della ricerca scientifica*
168. Andrea Arrighetti, *L'archeologia in architettura*
169. Lorenza Garrino (a cura di), *Strumenti per una medicina del nostro tempo. Medicina narrativa, Metodologia Pedagogia dei Genitori e International Classification of Functioning (ICF)*
170. Ioana Both, Ayşe Saraçgil e Angela Tarantino (a cura di), *Innesti e ibridazione tra spazi culturali*
171. Alberto Gherardini, *Squarci nell'avorio. Le università italiane e l'innovazione tecnologica*
172. Anthony Jensen, Greg Patmore, Ermanno Tortia (a cura di), *Cooperative Enterprises in Australia and Italy. Comparative analysis and theoretical insights*
173. Raffaello Giannini (a cura di), *Il vino nel legno. La valorizzazione della biomassa legnosa dei boschi del Chianti*
174. Gian Franco Gensini, Augusto Zaninelli (a cura di), *Progetto RIARTE. Raccontaci l'Iperensione ARTERiosa*
175. Enzo Manzano, Augusto Zaninelli (a cura di), *Racconti 33. Come migliorare la pratica clinica quotidiana partendo dalla Medicina Narrativa*
176. Patrizia Romei, *Territorio e turismo: un lungo dialogo. Il modello di specializzazione turistica di Montecatini Terme*
177. Enrico Bonari, Giampiero Maracchi (a cura di), *Le biomasse lignocellulosiche*
178. Mastroberti C., *Assoggettamento e passioni nel pensiero politico di Judith Butler*
179. Franca Tani, Annalisa Ilari, *La spirale del gioco. Il gioco d'azzardo da attività ludica a patologia*
180. Angelica Degasperi, *Arte nell'arte. Ceramiche medievali lette attraverso gli occhi dei grandi maestri toscani del Trecento e del Quattrocento*
181. Lucilla Conigliello, Chiara Melani (a cura di), *Esperienze di gestione in una biblioteca accademica: la Biblioteca di scienze sociali dell'Ateneo fiorentino (2004-2015)*
182. Anna Margherita Jasink, Giulia Dionisio (a cura di), *Musint 2. Nuove esperienze di ricerca e didattica nella museologia interattiva*
183. Ayşe Saraçgil, Letizia Vezzosi (a cura di), *Lingue, letterature e culture migranti*
184. Gian Luigi Corinto, Roberto Fratini, *Caccia e territorio. Evoluzione della disciplina normativa in Toscana*
185. Riccardo Bruni, *Dialogare: compendio di logica*
186. Daniele Buratta, *Dialogare: compendio di matematica*
187. Manuela Lima, *Dialogare: compendio di fisica*

188. Filippo Frizzi, *Dialogare: compendio di biologia*
189. Riccardo Peruzzini, *Dialogare: compendio di chimica*
190. Guido Vannini (a cura di), *Florentia. Studi di archeologia: vol. 3*
191. Rachele Raus, Gloria Cappelli, Carolina Flinz (édité par), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel. Vol. II*
192. Lorenzo Corbetta (a cura di), *Hot Topics in pneumologia interventistica*
193. Valeria Zotti, Ana Pano Alamán (a cura di), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*
194. Sabrina Ballestracci, *Teoria e ricerca sull'apprendimento del tedesco L2. Manuale per insegnanti in formazione*
195. Ginevra Cerrina Feroni, Veronica Federico (a cura di), *Società multiculturali e percorsi di integrazione. Francia, Germania, Regno Unito ed Italia a confronto*
196. Anna Margherita Jasink, Judith Weingarten, Silvia Ferrara (edited by), *Non-scribal Communication Media in the Bronze Age Aegean and Surrounding Areas : the semantics of a-literate and proto-literate media (seals, potmarks, mason's marks, seal-impressed pottery, ideograms and logograms, and related systems)*
197. Nicola Antonello Vittiglio, *Il lessico miceneo riferito ai cereali*
198. Rosario D'Auria, *Recall Map. Imparare e Ricordare attraverso Immagini, Colori, Forme e Font*
199. Bruno Bertaccini, *Introduzione alla Statistica Computazionale con R*
200. Lorenzo Corbetta (a cura di), *Hot Topics in Pneumologia Interventistica. Volume 2*
201. Carolina Flinz, Elena Carpi, Annick Farina (édité par), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel. Vol. I*
202. Anna Margherita Jasink, Maria Emanuela Alberti (a cura di), *AKROTHINIA 2. Contributi di giovani ricercatori agli studi egei e ciprioti*
203. Marco Meli (a cura di), *Le norme stabilite e infrante. Saggi italo-tedeschi in prospettiva linguistica, letteraria e interculturale*
204. Lea Campos Boralevi (a cura di), *La costruzione dello Stato moderno*
205. Maria Renza Guelfi, Marco Masoni, Jonida Shtylla, Andreas Robert Formiconi (a cura di), *Peer assessment nell'insegnamento di Informatica del Corso di Laurea in Medicina e Chirurgia dell'Università di Firenze. Una selezione di elaborati di Informatica Biomedica prodotti dagli studenti*
206. Fabio Silari, *Massive Open Online Course. "Un audace esperimento di apprendimento distribuito" nelle università*
207. Raffaele Pavoni, *Gli sguardi degli altri. Filmare il paesaggio urbano come esperienza multi-culturale e multi-identitaria*
208. Luigi Barletti, Giorgio Ottaviani, *Il premio Laboratorio Matematico "Riccardo Ricci" 2014-2016*
209. Josep-E. Baños, Carlo Orefice, Francesca Bianchi, Stefano Costantini, *Good Health, Quality Education, Sustainable Communities, Human Rights. The scientific contribution of Italian UNESCO Chairs and partners to SDGs 2030*
210. Lorenzo Corbetta (a cura di), *Hot Topics in Pneumologia Interventistica. Volume 3*
211. Michele Nucciotti, Chiara Bonacchi, Chiara Molducci (a cura di), *Archeologia Pubblica in Italia*
212. Guido Vannini (a cura di), *Florentia. Studi di archeologia vol. 4*
213. Ioana Both, Angela Tarantino (a cura di), *Cronologia della letteratura rumena moderna (1780-1914) - Cronologia literaturii române moderne (1780-1914)*
214. Mario Mauro, *L'impresa selvicolturale alla luce del decreto legislativo 3 aprile 2018, n. 34 "Testo unico in materia di foreste e filiere forestali"*
215. Guido Carlo Pigliasco, *The Custodians of the Gift. Fairy Beliefs, Holy Doubts and Heritage Paradoxes on a Fijian Island*
216. Francisco Matte Bon, Inmaculada Solís García, *Introducción a la gramática metaoperacional*
217. Annick Farina, Fernando Funari (a cura di), *Il passato nel presente: la lingua dei beni culturali*
218. Riccardo Billero, Annick Farina, María Carlota Nicolás Martínez (a cura di), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*

L'interazione con l'Informatica Umanistica si rivela al giorno d'oggi sempre più fondamentale per gli studi di lessicografia. Questo volume presenta la genesi e la struttura della Banca Dati LBC, strumento informatico di supporto al lavoro sviluppato dell'Unità di ricerca "Lessico multilingue dei Beni Culturali" a cui fa capo l'Università di Firenze, che consente di effettuare ricerche testuali in sei diversi corpora elettronici (francese, inglese, italiano, russo, spagnolo e tedesco). Gli autori di ogni corpus ne illustrano le relative specificità sia per quanto riguarda le scelte delle fonti, sia per le possibilità di uso legate alle attività di lessicografia e di traduzione, svolte all'interno del progetto.

RICCARDO BILLERO è docente di Informatica presso le Università di Bologna, Bolzano e Firenze. Si occupa di Informatica Umanistica come responsabile tecnologico dell'Unità di Ricerca "Lessico dei Beni Culturali".

ANNICK FARINA è professoressa associata di Lingua e Traduzione Francese all'Università di Firenze e presidente del Centro Linguistico di Ateneo. Dirige il gruppo "Lessico dei Beni Culturali" dal 2013.

MARÍA CARLOTA NICOLÁS MARTÍNEZ è ricercatrice di Lingua Spagnola presso l'Università di Firenze. Le sue principali linee di ricerca riguardano l'analisi della lingua orale, la lessicografia e l'applicazione dell'informatica alla ricerca umanistica.

ISSN 2704-6249 (print)
ISSN 2704-5870 (online)
ISBN 978-88-5518-252-2 (print)
ISBN 978-88-5518-253-9 (PDF)
ISBN 978-88-5518-255-3 (XML)
DOI 10.36253/978-88-5518-253-9
www.fupress.com