

# 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 essi<sup>1</sup>.

## 1. Cornice del progetto

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

<sup>1</sup> 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 competenza<sup>2</sup> 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-

<sup>2</sup> 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 seguenti<sup>3</sup>:

- 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 ricerca<sup>4</sup> 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 qualità<sup>5</sup>;

<sup>3</sup> Cfr. Billero, Nicolás 2017 per una descrizione più dettagliata.

<sup>4</sup> 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.

<sup>5</sup> 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 diffusione<sup>6</sup>; 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.

<sup>6</sup> «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 diversi<sup>7</sup>. 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

<sup>7</sup> «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 lingua<sup>8</sup>.

### 3.2 Creazione degli strumenti necessari alla gestione delle fonti

#### 3.2.1 Condivisione delle cartelle

Per consentire la corretta gestione del flusso di lavoro<sup>9</sup> 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.

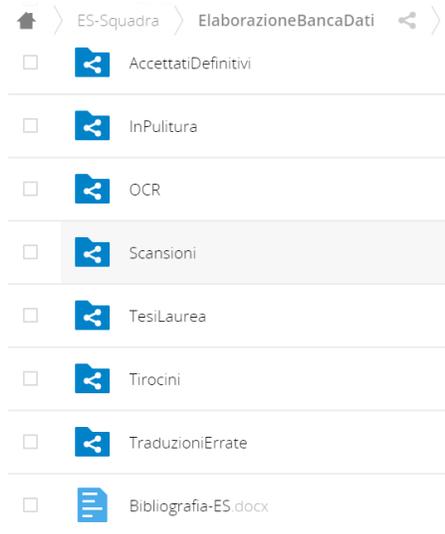


Figura 1. Cartelle Cloud del corpus spagnolo.

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.

<sup>8</sup> 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.

<sup>9</sup> 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 tradotto<sup>10</sup>. 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-

<sup>10</sup> 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 tradotti<sup>11</sup>. 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 testo<sup>12</sup>.

<sup>11</sup> 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.

<sup>12</sup> 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: Broches, 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;



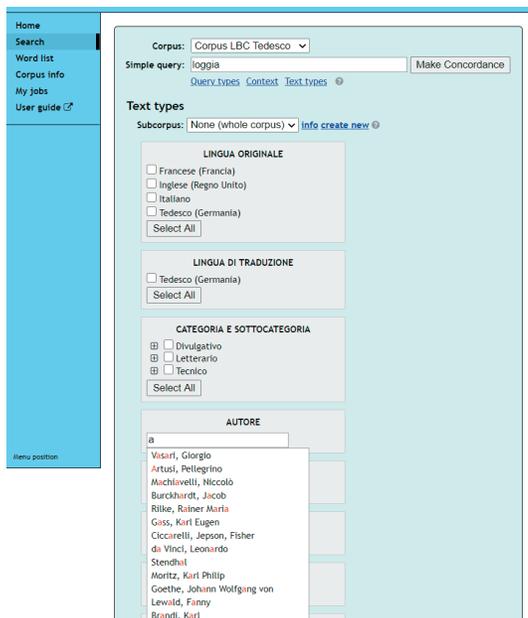


Figura 3. Ricerca con text type su corpus tedesco.

Corpus LBC Francese		General info		Lexicon sizes	
Counts		Corpus description	<a href="#">Document</a>	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	<a href="#">Tagset</a>	<a href="#">Description</a>	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](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.