

Corpora e terminologia artistica: il caso del corpus spagnolo LBC

Riccardo Billero[°], Elena Carpi*

[°]Università degli Studi di Firenze, *Università di Pisa¹

Il lessico relativo al patrimonio culturale italiano è oramai divenuto un ambito di interesse per molte discipline; a partire da tale considerazione, il progetto Lessico multilingue dei Beni Culturali (LBC) si pone l'obiettivo di analizzare e studiare la terminologia del mondo dell'arte in varie lingue, focalizzandosi in particolare sulla Toscana. Nell'ambito di tale lavoro di ricerca sono state realizzate varie banche dati contenenti testi letterari e tecnici con informazioni di rilievo sui beni culturali toscani al fine di diffonderne la conoscenza, mediante la pubblicazione online di corpora nelle lingue coinvolte nel progetto. Tra i risultati del progetto è da notare il corpus "Español LBC", il quale dispone ad oggi di oltre un milione di parole, l'origine delle quali fa sì che il corpus sia caratterizzato da una alta frequenza di terminologia relativa all'arte, costituendo per questo un punto di riferimento per coloro che vogliono occuparsi di lessico spagnolo nell'ambito dei beni culturali. In particolare, lo scopo di questo paper è quello di descrivere il lavoro svolto nell'ambito della progettazione ed implementazione del corpus, e di analizzarne l'utilizzo per lo studio della terminologia spagnola dell'arte. Nell'ambito della terminologia spagnola dell'arte, come esempio delle possibilità di uso del corpus, verranno mostrate alcune tra le più significative unità pluriverbali, assenti nel dizionario utilizzato come riferimento.

Keywords: digital humanities, information technology, lexicography, corpora linguistics, LBC

1. Introduzione

A partire dal 2013, è stata costituita presso il Dipartimento di Lingue, Letterature e Studi Interculturali dell'Università degli Studi di Firenze l'Unità di Ricerca Lessico dei Beni Culturali (LBC) che, attraverso il progetto omonimo, intende promuovere la realizzazione di studi, ricerche e altre attività incentrate sul lessico

¹ I paragrafi 1 e 2 sono redatti dall'autore 1, mentre i paragrafi 3 e 4 dall'autore 2.

di alcune lingue (ivi compreso lo spagnolo) in relazione alla lingua italiana nell'ambito dei Beni culturali della città di Firenze, considerata modello di ogni città d'arte italiana. A tal fine sono stati realizzati vari corpora linguistici, uno per ognuna delle lingue coinvolte nel progetto, che raccolgono una selezione di testi riguardanti il patrimonio culturale fiorentino (Farina, 2016).

2. La costruzione del corpus

Al fine di ottenere una gestione dei contenuti snella ed efficace, si è deciso di utilizzare per la memorizzazione dei testi una soluzione di *storage* su *cloud* ospitato in un apposito server virtuale remoto a disposizione del progetto; in particolare, si è ricorso ad una installazione di NextCloud, un software open source all'altezza dei requisiti progettuali. All'interno del cloud, oltre ai testi di interesse per il progetto, è presente un apposito foglio Excel che consente, tra gli altri scopi:

- la creazione dei nomi dei file testuali presenti nei corpora;
- la memorizzazione dei metadati relativi ai testi.

In particolare, vi sono alcune informazioni che devono essere inserite all'interno del foglio Excel (quali lingua originale, autore, titolo del testo, anno di pubblicazione, ecc...) che contribuiscono alla decisione del nome univoco da assegnare al file contenente il testo² (Billero *et al.* 2017).

Utilizzare un nome di file ottenuto algoritmicamente facilita il lavoro nella fase di consultazione delle concordanze; infatti tale nome è "intelligibile", poiché composto da elementi che consentono di facilitare la corretta individuazione del testo oggetto di interesse, distinguendolo facilmente dagli altri della banca dati.

Come già accennato, il file Excel consente anche di memorizzare i vari metadati relativi al testo; tali informazioni vengono riconciliate con il rispettivo testo, prima del loro inserimento all'interno dei corpora, attraverso un software realizzato ad hoc. In tal modo sono rese possibili ricerche utilizzando filtri in base a varie caratteristiche, quali l'autore, il periodo storico, la tipologia testuale, ecc....

Tutti i testi presenti all'interno della raccolta sono stati memorizzati come file Word; in particolare, la scelta di tale programma è dovuta sia alla sua pervasività (e quindi all'essere di uso e conoscenza comune per tutti i componenti le varie squadre) sia alla necessità di mantenere alcuni elementi basilari della formatta-

² Il nome viene stabilito associando (mediante l'uso di una apposita macro Excel) una sigla ai vari metadati.

zione quali grassetto, corsivo, note a piè di pagina, ecc... Inoltre, essendo il formato *.docx* costituito da uno standard XML, è stato possibile realizzare appositi software in grado di rielaborare automaticamente tali file in modo da produrre in output nuovi documenti aventi le caratteristiche desiderate.

Tutti i testi presenti nella banca dati sono stati quindi caricati automaticamente all'interno di SketchEngine, rendendoli così interrogabili attraverso l'utilizzo di questo software specializzato nella gestione di corpora testuali.

3. I contenuti del corpus

La creazione del corpus Español LBC - che supera attualmente un milione di parole - si è sviluppata per fasi successive: durante la prima, sono stati raccolti testi collegati con la città di Firenze, considerata città d'arte per antonomasia; nella seconda, il corpus è stato ampliato con testi riguardanti la Toscana, mentre nella terza l'ambito di interesse si è allargato all'Italia nel suo complesso. Il corpus comprende sia opere originali in spagnolo sia tradotte dalle altre lingue del progetto (concretamente, italiano, francese e inglese), che vanno dal XVI° fino al XX° secolo, scelte per la loro rilevanza culturale e interculturale. Le tipologie testuali (Billero & Nicolás 2017: 209) sono quella tecnica (architettura, arte e storia) e quella letteraria (biografica), in modo da comprendere punti di vista diversificati in relazione con l'arte italiana, e ampliare in tal modo la scelta terminologica.

Tale diversificazione, che è uno dei criteri adottati per la creazione del corpus, giustifica la decisione di includere nella tipologia letterario-biografica opere di autori molto diversi tra loro. Infatti, insieme a testi di famosi scrittori spagnoli, come Alarcón, Blasco Ibañez, Fernández de Moratín, Pardo Bazán e Pérez Galdós e alle opere scritte sull'Italia e Firenze da Chateaubriand, Ruskin e Stendhal, si è deciso di inserire anche i diari di viaggio di autori minori come Galofre, Iribas y Sánchez e Lasa, vere e proprie guide pratiche che offrono uno sguardo più concreto sull'arte italiana rispetto a quello di testi più conosciuti, e che quindi fanno uso di una terminologia diversa.

Per quel che riguarda i testi tecnici, che rappresentano una imprescindibile fonte di termini artistici, si è optato per le traduzioni spagnole delle due edizioni delle *Vite* di Vasari, dei *Dieci libri di Architettura* di Leon Battista Alberti e dei *Quattro libri di architettura* di Palladio, opere che si collocano all'origine della storia dell'arte italiana. Il corpus comprende anche le parti dedicate a pittori italiani ne *Las vidas de los pintores y estatuarios eminentes españoles* di Palomino de Castro e *El Tratado de arquitectura* di Alonso de Vandelvira, la cui importanza

consiste nell'essere stato il veicolo di diffusione della corrente architettonica italianista in Castiglia; inoltre, il *Tratado* è un'opera di estremo interesse, perché contiene non solo termini architettonici, ma anche quelli che caratterizzano l'arte di tagliare le pietre da costruzione. Infine, le traduzioni della *Storia di Firenze* di Macchiavelli e della *Vita* di Cellini apportano al corpus antroponimi e toponimi culturalmente rilevanti per la storia dell'arte italiana.

3.1 Esempi di uso

In questo paragrafo si presentano alcuni esempi di interrogazione del corpus, in modo da mostrarne le potenzialità e l'interesse per traduttori, studiosi e guide turistiche. I traduttori si trovano a volte ad affrontare testi riguardanti i Beni Culturali senza possedere una specifica competenza, e le *query* del corpus effettuate con il programma *SketchEngine* possono aiutarli a orientarsi tra le traduzioni proposte nei dizionari generalisti. Per esempio, un traduttore che debba rendere in spagnolo il termine *tempera* troverà che TAM considera sinonime le voci *temple* e *témpera*, mentre ZANICHELLI definisce *temple* come tecnica e *témpera* come opera. Interrogando il corpus Español LBC potrà rendersi conto che, a fronte di due occorrenze di *témpera*, ne esistono settantaquattro di *temple*, inteso sia come tecnica, sia come sostanza:

Volvió entonces, famoso y próspero, a Florencia, donde, en San Marcos, realizó un crucifijo de madera, mayor que el natural, *trabajado al temple* sobre campo de oro, que colocó en la zona derecha de la iglesia;

Y haciendo otros descubrimientos de la misma suerte en muchas otras cosas, vio que mezclando los colores con estos aceites, salía un *temple* muy fuerte, que, una vez seco, resultaba impermeable al agua.

Inoltre, potrà verificare che, per indicare un'opera, nel corpus vengono sempre usati i sintagmi *tabla al temple*, o *pintura al temple*.

Inoltre, l'interrogazione effettuata mediante la funzione *Context* di *SketchEngine*, che permette di cercare una parola e le sue co-occorrenze con parole appartenenti a varie categorie grammaticali, mette a disposizione dei traduttori combinazioni lessicali validate dall'uso. La *query* di *capilla* seguita da nome a destra darà come risultato – tra gli altri – *Capilla Brancacci*, *Capilla de los Pazzi*, *Capilla Sixtina*, *Capilla Paulina*, quella della parola *ornamental* seguita da nome a sinistra restituirà *bordes ornamentales*, *carácter ornamental*, *líneas ornamentales* ecc.

Sempre riguardo alla possibilità di individuare le co-occorrenze, è importante sottolineare che, a differenza della maggior parte dei glossari, Español LBC include

anche forme verbali; un traduttore potrà quindi individuare quella più adatta per accompagnare un determinato termine: per esempio, *pechina* appare insieme al verbo *trazar*, e quindi si potranno escludere i sinonimi *delinear* o *dibujar*. *Cruceta*, *planta y tardosa* con il verbo *capialzar*, risultato che permette di escludere *levantar* ecc.

Español LBC può anche essere molto utile agli studiosi che si occupano di traduzione e terminologia, e che intendano studiare la maggiore o minore correttezza delle traduzioni esistenti di testi artistici verso lo spagnolo. Per esempio, le concordanze della parola *retablo* nelle traduzioni delle *Vite* di Vasari consentono di apprezzare l'uso generalizzato di questa parola per descrivere opere d'arte italiane del XV e XVI secolo. In alcuni casi però, queste non corrispondono alla categoria di oggetti artistici definita dalla parola *retablo*³, bensì a quella definita dai termini *díptico* o *políptico*, o semplicemente *tabla*.

In effetti, nella traduzione della vita di Benozzo Gozzoli nell'edizione del 1568 delle *Vite* di Vasari si può leggere: "En su juventud pintó en Florencia el *retablo* para la compañía de San Marcos" (Vasari 1960), mentre il testo italiano recita: "Dipinse in Fiorenza nella sua giovinezza alla Compagnia di S. Marco la *tavola* dello altare" (Vasari [1568] 1966: 292).

In modo analogo, nella vita del Perugino si parla di *retablo*, mentre l'originale parla di *tavola*:

En Asís pintó el *retablo* en la capilla de Santa Catalina, en la iglesia inferior de San Francisco, por encargo del cardenal español Egidio, e hizo otro en San Damiano (Vasari 1960).

Et in Ascesi dipinse la *tavola* della cappella di Santa Caterina nella chiesa di sotto di San Francesco, per il cardinale Egidio Spagnuolo, e parimente una in San Damiano. (Vasari [1568] 1966)

Infine, anche i professionisti del settore turistico possono trovare nell'interrogazione del corpus un valido strumento per approfondire le loro conoscenze enciclopediche riguardo un monumento o un'opera d'arte. L'uso della funzione *Word Sketch* permette infatti di ampliare la ricerca iniziale, come mostra quella effet-

³ "Mueble de culto que desarrolla una o varias representaciones religiosas figuradas, por lo general pintadas o esculpidas, situado detrás del altar. [...] Durante el Gótico, fue un elemento pequeño, portátil, colocado sobre la mesa de altar para officiar la liturgia sagrada. Pintado o esculpido, adoptó a menudo una disposición tripartita, con hojas laterales que a manera de puertas podían cerrarse sobre la parte central." Tesoros del Patrimonio Cultural de España s.v. *retablo*.

tuata con l'antroponimo *Médecis*: oltre alle 240 occorrenze del cognome è possibile recuperare nomi di opere d'arte, come *Venus de Médecis* o di monumenti, come *Capilla de Médecis*. In modo analogo, il risultato della ricerca del toponimo *Florençia* consente di estrarre nomi di edifici religiosi (Santa Croce, Santo Spirito, Santa Maria del Fiore, Santa Trinita), di luoghi (*plaza San Giovanni*) e di artisti a volte poco conosciuti, come negli esempi che seguono, in cui la ricerca *Florençia* produce informazioni su Rocco Zoppo, Taddeo Gaddi e Simone Memmi:

Otro discípulo fue *Rocco Zoppo*, pintor de *Florençia*, de quien Filippo Salviati posee una Virgen muy hermosa que, en verdad, fue terminada por Pietro.

Si lo hacéis así, al entrar en esta capilla podréis saber que en tiempo de Miguel Ángel *Florençia* entera atribuía estos frescos á Tadeo Gaddi y á *Simón Memmi*. No me he consagrado especialmente al estudio da ninguno de estos dos artistas, y así no puedo deciros nada de positivo sobre ellos ó sobre sus obras.

4. Conclusioni

I risultati finora ottenuti permettono di affermare che il progetto Lessico dei Beni Culturali può costituire un importante punto di riferimento sia per gli esperti linguisti e traduttori sia per gli addetti informatici all'interno dell'ambito dell'Umanistica Digitale.

In particolare, grazie alle varie tipologie testuali che lo compongono e all'arco temporale interessato, il corpus Español LBC permette all'utente di reperire termini e informazioni culturali altrimenti poco accessibili, mentre la prospettiva diacronica così come il criterio interculturale alla base della scelta dei testi permettono al traduttore di operare corrette scelte terminologiche.

References

- 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.
- 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*. 25.
- TAM: Tam, L. 2004. *Grande dizionario di spagnolo: spagnolo-italiano, italiano-spagnolo*. Milano: Hoepli
- Tesauros del Patrimonio Cultural de España. <http://tesauros.mecd.es/tesauros/> (consultato il 1 luglio 2017).

- Vandelvira, A. de 1977. "Introducción". In G. Barbé-Coquelin de Lisle (ed.), *El Tratado de arquitectura de Alonso de Vandelvira*. Albacete: Confederación Española de Cajas de Ahorro (ed. orig. 1591), 3-36.
- Vasari, G. 1966. "Vita di Benozzo". In *Le Vite de' più eccellenti Pittori, Scultori et Architettori*, Firenze: SPES Sansoni, 292-294 (ed. orig. 1568)
- Vasari, G. 1960. *Vida de los más excelentes pintores, escultores y arquitectos*. Barcelona: Éxito (ed. orig. 1568).
- Vasari, G. 2011. *Las vidas de los más excelentes arquitectos, pintores y escultores italianos desde Cimabue a nuestros tiempos* (antología). Madrid: Editorial Tecnos/Alianza (ed. orig. 1550).
- Vasari G. 2012. *Las vidas de los más excelentes arquitectos, pintores y escultores italianos desde Cimabue a nuestros tiempos*. Madrid: Cátedra (ed. orig. 1550).
- ZANICHELLI: Arqués, R. & Padoan, A. 2012. *Il grande dizionario di spagnolo: spagnolo-italiano, italiano-spagnolo*. Bologna: Zanichelli.